Plosive voicing acoustics and voice quality in Yerevan Armenian

Scott Seyfarth and Marc Garellek

Abstract

Yerevan Armenian is a variety of Eastern Armenian with a three-way voicing contrast that includes voiced, voiceless unaspirated, and voiceless aspirated stops, but previous work has not converged on a description of how voice quality is involved in the contrast. We demonstrate how voice quality can be assessed in a two-dimensional acoustic space using a spectral tilt measure in conjunction with a measure of spectral noise. Eight speakers produced a list of words with prevocalic word-initial and postvocalic word-final plosives. The results suggest that Yerevan Armenian has breathy-voiced plosives which are produced with closure voicing and a relatively spread glottis that is maintained into a following vowel. These qualitatively differ from some Indic ones in that they do not have an extended interval of voiced aspiration after the closure. For the voiceless unaspirated plosives, most speakers produced acoustically modal voiceless plosives, although two showed evidence for some glottal constriction and tensing. Many acoustic cues contribute to overall reliable discriminability of the three-way contrast in both initial and final position. Nevertheless, closure voicing intensity and aspiration duration together provide a robust separation of the three categories in both positions. We also find that back vowels are fronted after the breathy-voiced plosives, which supports a historical analysis in which early Armenian voiced stops were also breathy, rather than plain voiced.

Contents

1 Introduction

Stop consonants produced at the same place of articulation can be differentiated by a variety of phonetic parameters (Henton, Ladefoged, & Maddieson, 1992), such as phonation during the closure, degree and timing of glottal constriction, duration of the closure and of adjacent vowels (Chen, 1970; Raphael, 1972; Summerfield, 1981), release burst spectrum (Chodroff & Wilson, 2014), and pitch and formants adjacent to the closure (Hanson, 2009; Hombert, Ohala, & Ewan, 1979; Liberman, Delattre, & Cooper, 1958; Ohde, 1984). Because of the many-to-many relationship between articulatory mechanisms and acoustic cues, descriptions of stop contrasts have relied on aggregate acoustic measures such as voice onset time (Lisker & Abramson, 1964) to summarize the dynamics of voicing-related events surrounding the stops (Keating, 1984).

One of the earliest uses of voice onset time was to describe the three-way voicing contrast in Armenian (Adjarian, 1899, cf. Braun, 2013). Armenian historically had a three-way stop contrast that has developed into a range of systems which now include at least two-way and three-way contrasts. The standard description of the modern Armenian languages includes seven different systems derived from Classical Armenian, shown in Table 1 (Gharibian, 1969 cited in Garrett, 1998; Schirru, 2012; Vaux, 1998a; Weitenberg, 2002; see Baronian, 2017 for a reanalysis). In the table, each system is given a schematic representation following standard practice (e.g., D, T, T^h in Standard Eastern Armenian, a Group 6 dialect), though the contrasts occur for labial, dental, and velar plosives, as well as dental and postalveolar affricates. Each system occurs in a diverse group of dialects, but Armenian dialectology is complex beyond the basic Eastern–Western divide (Adjarian, 1899; Vaux, 1998a; Jahukyan, 1972 cited in Baronian, 2017; Weitenberg, 2002).

While the description in Table 1 includes four realizations—voiceless unaspirated T, voiceless aspirated T^h, voiced D, and D^{fi}, which has been called 'voiced aspirated' or 'murmured'—the correct phonetic description of the stops in each series is not entirely clear (Adjarian, 1899; Allen, 1950; Fleming, 2000; Hacopian, 2003; Khachaturian, 1984, 1992; Kortlandt, 1998; Ladefoged & Maddieson, 1996; Pisowicz, 1997, 1998; Schirru, 2012; Vaux, 1998a; Weitenberg, 2002). In this paper, we investigate the acoustics of plosives in the variety of Eastern Armenian spoken in Yerevan, one of the central Group 2 dialects. The Group 2 dialects include both D^{fi} and T realizations, which have each been claimed to involve a range of non-modal voice qualities and other phonetic characteristics. Yerevan Armenian thus serves as a case study for understanding

Table 1: Correspondences for seven modern stop systems in dialect groupings derived from the three-way contrast in Classical Armenian. Each column indicates one group of dialects. There are two literary standard varieties, Standard Western (Group 5) and Standard Eastern (Group 6) Armenian. The reconstruction of the voiced stops in Classical Armenian is disputed; see §5.2.

Classical	1	2	3	4	5	6	7
D	D^{fi}	D^{fi}	D	T	T^h	D	T
T	D	T	D	D	D	T	T
T^h	T^h	T^h	T^h	T^h	T^h	T^h	T^h

the dynamics of voicing and voice quality contrasts in different syllabic positions. We demonstrate how voice quality can be assessed with a combination of two speaker-specific acoustic measures which index glottal constriction (via the difference in amplitude of the first two harmonics, H1–H2) and noise (via cepstral peak prominence, CPP). As well as adding to the phonetic documentation of the Armenian languages, the results point towards acoustic techniques for more accurate descriptions of laryngeal contrasts. More broadly, an exact description of voice quality in Armenian plays an important role in typological questions about the development and cross-linguistic comparison of laryngeal articulations in the Indo-European language family (e.g., Fleming, 2000; Garrett, 1998; Kortlandt, 1985, 1998; Pisowicz, 1997; Schirru, 2012; Vaux, 1998b; Weitenberg, 2002).

1.1   Existing descriptions of Armenian plosives

1.1.1   Voiced plosives

The D$^\text{fi}$ stops are often referred to as 'voiced aspirated' or 'murmured' (Adjarian, 1899, discussed in Pisowicz, 1997; Garrett, 1998; Pisowicz, 1997, 1998; Vaux, 1997; Weitenberg, 2002). In terms of actual phonation during the closure, the Armenian plosives in this category have been described as canonically voiceless in at least word-initial position (Allen, 1950; Khachaturian, 1984; Pisowicz, 1997, 1998, though see Schirru, 2012), though they may have closure voicing in some medial and final nasal clusters (Allen, 1950; Khachaturian, 1984). However, the presence of closure voicing in postvocalic word-final position is disputed (Allen, 1950; contra Pisowicz, 1998; Vaux, 1998a, pp. 16–17, 237; and see also Ladefoged & Maddieson, 1996, pp. 66–67; Hacopian, 2003; Dum-Tragut, 2009, pp. 24–27 on final voicing in the plain-voiced D category). Allen (1950) also suggests that initial plosives may be occasionally voiced, and that in intervocalic position, voicing may carry over from a preceding vowel into the first part of the closure.

At the release, it has been claimed that these plosives have a weakly-voiced murmur (Pisowicz, 1998) or intermittent voicing (Khachaturian, 1992, cited in Vaux, 1997) that begins near the closure offset, or else a noisy voiced release (Allen, 1950; Khachaturian, 1992, cited in Garrett, 1998) or brief aspiration (Khachaturian, 1984). In terms of voice quality, they have been described as murmured, breathy (Garrett, 1998), breathy in initial position (Khachaturian, 1984), and as having slack vocal folds during the first few voicing pulses after the release (Schirru, 2012). Additionally, it has been claimed that the D$^\text{fi}$ stops are associated with lower pitch (Allen, 1950; Benveniste, 1958; Khachaturian, 1992 cited in Garrett, 1998; Schirru, 2012), stronger airflow (Adjarian, 1899; Allen, 1950), or greater intensity (Adjarian, 1899; Khachaturian, 1984, Gamkrelidze & Ivanov, 1995, p. 15).

While the appropriate terminology for these stops has been debated (see discussions in Kortlandt, 1985 and Pisowicz, 1998), much of the variability in describing 'voiced aspirated' and 'murmured' stops may simply reflect different terminological traditions, and possibly an incomplete understanding of laryngeal articulations with their associated acoustics. Languages primarily make contrastive use of up to three broad classes of voice qualities: breathy, modal, and creaky (Garellek, to appear; Gordon & Ladefoged, 2001). These labels are meaningful only in comparison with one another, which is likely why many names for voice qualities exist. Breathy voice (broadly defined) is thus sometimes called 'lax', 'slack', or 'murmured', especially

when the voice quality is not as breathy as some other baseline, which may be based on breathy voice quality in another language, or on another sound category in the same language (Gordon & Ladefoged, 2001; Keating, Esposito, Garellek, Khan, & Kuang, 2011).

Though the $D^{ɦ}$ plosives likely do have breathy voice quality, earlier reports may also be referring to some other voicing dynamic—such as weak or inconsistent voicing perhaps in conjunction with optional aspiration—and it is uncertain how their realization fits into the typology of plosive voicing contrasts more generally. Specifically, the term 'voiced aspirated' has typically been used as part of the four-way plosive contrast in Hindi-Urdu and other Indic languages. The Armenian plosives have been claimed to be both acoustically similar (Garrett, 1998; Vaux, 1997) and dissimilar (Khachaturian, 1984; Pisowicz, 1998) to the Indic ones, which have both voicing during the closure and a release that typically involves an interval of voiced aspiration followed by a more modal vowel target (Henton et al., 1992; Ladefoged & Maddieson, 1996, pp. 57–60).

1.1.2 Voiceless unaspirated plosives

The voiceless T plosives are usually transcribed as /p, t, k/, but there are many reports that they are glottalized in Yerevan Armenian and other varieties, especially those in Iran (Allen, 1950; Baronian, 2017; Dum-Tragut, 2009; Fleming, 2000; Gamkrelidze & Ivanov, 1995; Kortlandt, 1985, 1995, 1998; Ladefoged & Maddieson, 1996; Pisowicz, 1997; Fairbanks & Stevick, 1958 cited in Hacopian, 2003; Job, 1977 cited in Weitenberg, 2002; Kortlandt, 1978 cited in Baronian, 2017). In this context, glottalization has referred to glottal constriction with a pulmonic airstream mechanism (Pisowicz, 1997) as well as to an ejective articulation with a glottalic one (Allen, 1950; Ladefoged & Maddieson, 1996, p. 67; Baronian, 2017; Pisowicz, 1998).

At the same time, it has also been suggested that glottalization may be only weakly perceptible (Pisowicz, 1997) or simply absent in the voiceless plosives (Hacopian, 2003; Macak, 2017), and the reports that these plosives are glottalized have been largely impressionistic, rather than based on instrumental data. Schirru (2012) provides an acoustic analysis of plosive consonants in Yerevan Armenian, and finds that vowels adjacent to the voiceless T series likely have more glottal constriction than those adjacent to the voiced $D^{ɦ}$ series, based on a measure of spectral tilt. However, this measure cannot be used to determine the absolute degree of glottal constriction (see §1.2.1, below). Because the $D^{ɦ}$ series is likely breathy and thus has higher spectral tilt than if it were modal-voiced, the finding that voiceless T has lower spectral tilt is consistent with both a modal (neither glottalized nor breathy) or a glottalized T series. Indeed, Schirru (2012) reports observing fewer than five ejectives with a characteristic double release in a corpus of 225 voiceless T tokens, and Dum-Tragut (2009, p. 18) points out that "normative grammars" typically do not describe these plosives as glottalized.

1.1.3 Voicing realization in word-final position

A third question about the plosive contrast is whether and how it is maintained in final position. Hacopian (2003) reports that for Standard Eastern Armenian (Group 6; reported to have plain voiced stops), the voiced series is always fully voiced in final position in a variety of postvocalic phonological environments, and aspiration duration distinguished the other two series in this position. However, it has also been claimed that for Group 2 varieties like Yerevan Arme-

nian, the Dᶠⁱ series actually has a voiceless closure word-finally (Pisowicz, 1998, contra Allen, 1950), and several researchers have suggested that the major cues associated with Dᶠⁱ are pitch, intensity, or a breathy quality on the following vowel (see §1.1.1 above).

If so, this raises the question as to how Dᶠⁱ is contrasted with T and Tʰ in final position, where there is no following vowel which would be affected by a breathy quality. One possibility is that glottal spreading may occur leading into the closure; another is that different cues, such as vowel duration, distinguish the contrast in final position (Ladefoged & Maddieson, 1996, pp. 66–67). For the voiceless unaspirated T series, Allen (1950) reports that ejectives are especially noticeable in final position, and Ladefoged and Maddieson (1996, p. 67) report that some speakers may have a glottal closure associated with final T stops, which may distinguish the voiced and voiceless unaspirated plosives.

Another possibility is that the three-way voicing contrast might be reduced or absent in final position, which is common cross-linguistically due to the weaker and fewer cues available to plosives in this context (Henton et al., 1992; Keating, Linker, & Huffman, 1983; Steriade, 1997). In various dialects of Armenian, the final voicing contrast is reduced adjacent to nasals, sibilants, and /ɾ/ (Dum-Tragut, 2009; Macak, 2017; Vaux, 1997, 1998a); and even after vowels in final position there may be across-the-board (Vaux, 1997), idiosyncratic (Dum-Tragut, 2009, pp. 24–27; Vaux, 1998a, p. 17), or dialect-specific (Pisowicz, 1997, p. 228; Hacopian, 2003) reduction of the contrast.

1.2 Two acoustic dimensions are necessary to identify voice quality

To assess voice quality using acoustic measures, both spectral tilt and noise measures should be used together. For the following analysis of Yerevan Armenian plosives, we select H1–H2 (the difference between the amplitudes of the first two harmonics of the spectrum) as a representative spectral tilt measure that is known to be correlated with degree of glottal constriction and contact. For a noise measure, we select cepstral peak prominence (CPP; Hillenbrand, Cleveland, & Erickson, 1994), a measure of harmonics-to-noise which is correlated with both aspiration noise and vocal fold irregularity (Blankenship, 2002; Esposito, 2012; Garellek & Keating, 2011; Keating et al., 2011; Misnadin, 2016; Wayland & Jongman, 2003).

1.2.1 Glottal constriction and H1–H2

From an articulatory perspective, differences between breathy, modal, and creaky voice qualities can minimally be described using a one-dimensional model of vocal fold contact (Gordon & Ladefoged, 2001; Ladefoged, 1971, cf. Edmondson & Esling, 2006). Breathy voice occurs when there is relatively less contact, and creaky voice occurs when there is relatively more contact (Gordon & Ladefoged, 2001, cf. other types of creaky voice in Garellek to appear; Keating, Garellek, and Kreiman 2015). While this description serves for phonated sounds, it can also capture distinctions among voiceless ones, such as when the transition into a particular voiceless glottal configuration alters the quality of adjacent voicing.

In the case of stops, the degree of contact during the closure can affect the voice quality of adjacent vowels (or other voiced sounds). For example, voiceless sounds can be made with either minimal or maximal vocal fold contact, as in aspirated [tʰ] or glottalized [t͡ʔ, t'], respectively. Aspirated plosives have a spread-glottis gesture during and after their closure, which

results in a noisy lag (aspiration) between the stop release and onset of voicing (Cooper, 1991; Davidson, 2017; Löfqvist & McGowan, 1992; Löfqvist & Yoshioka, 1984; Munhall & Löfqvist, 1992). Once the vocal folds begin to vibrate, voicing is initially breathier during the transition from a spread-glottis position (Garellek, 2012; Löfqvist & McGowan, 1992). Similarly, for glottalized plosives, the voice quality of adjacent vowels is creakier when the glottal constriction gesture associated with the closure overlaps with adjacent sounds, which makes glottal constriction perceptible near an otherwise-silent closure (Cho, Jun, & Ladefoged, 2002; Gallagher, 2015; Garellek, 2010, 2012; Garellek & Seyfarth, 2016; Seyfarth & Garellek, 2015; Vicenik, 2010).

The acoustic measure H1–H2 is known to correlate with degree of vocal fold contact, such that higher values are associated with breathier voice quality and less vocal fold contact (e.g., Abramson, Tiede, & Luangthongkum, 2015; Berkson, 2013; Bickley, 1982; Blankenship, 2002; Cho et al., 2002; DiCanio, 2009, 2014; Esposito, 2012; Garellek & Keating, 2011; Gordon & Ladefoged, 2001; Khan, 2012; Miller, 2007; Wayland & Jongman, 2003; Yu & Lam, 2014). It has been proposed that H1–H2 reflects differences in the relative duration of the open part of the glottal vibratory cycle (Gordon & Ladefoged, 2001; Klatt & Klatt, 1990; though see also Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Kreiman et al., 2012; Samlan & Story, 2011; Samlan, Story, & Bunton, 2013; Swerts & Veldhuis, 2001; Zhang, 2016, 2017). However, the exact articulatory mechanism is not well-established (cf. Zhang, 2016, 2017), and the relationship between glottal constriction and H1–H2 may not be monotonic (Samlan & Story, 2011; Samlan et al., 2013).

1.2.2   Combining spectral tilt and noise measures

Although H1–H2 correlates with vocal fold contact, voice quality can only be inferred by the relationship among H1–H2 values (Garellek, to appear; Garellek & White, 2015; Simpson, 2012). For example, a voiced sound with higher H1–H2 than another sound might be breathier (when the other sound is breathy or modal) or more modal (if the other sound is at all constricted). There are at least two ways to gain more information about voice quality from a spectral tilt measure like H1–H2. First, it can be compared to a reference sound with a known voice quality. This is useful in some cases—for example, if a sound has lower H1–H2 than a known creaky sound, it must also be creaky—but may be ambiguous, such as if a sound has lower H1–H2 than a known breathy sound (it could be less breathy, or modal, or creaky). Second, a noise measure like CPP can be used in combination with spectral tilt to identify voice quality. Both breathy and creaky voice qualities tend to be noisier than modal voice because of aspiration, in the case of breathiness, or because of irregular voicing, in the case of creakiness (Blankenship, 2002; Garellek, 2012; Gordon & Ladefoged, 2001). For the measure CPP, lower values are associated with more noise (and thus non-modal phonation), while higher values are associated with less noise and modal phonation. Thus, if a particular sound has a lower H1–H2 and lower CPP than a reference sound, it is creakier; if it has a lower H1–H2 and higher CPP, it is more modal (Garellek, to appear).[1]

---

[1]Note that there is still some ambiguity, because breathy and creaky phonation may still be associated with different ranges of low CPP, due to differences in the relative degree of breathiness and creakiness (as used by a particular speaker in a particular language). For example, in a language with breathy, modal, and creaky voice quality, a breathy sound may have relatively higher or lower CPP than a creaky reference sound in the same

1.3 The current study

In the current study, native speakers of Yerevan Armenian produce a large set of words containing the target plosives in word-initial and word-final position. The plosives are elicited in both positions in order to better understand how a three-way contrast can be maintained through voice timing and quality. In particular, voice timing cues must be different in final versus initial position (see Abramson & Whalen, 2017), and any voice quality differences are less likely to be usefully audible during and after a final stop closure, which suggests that voice quality might be used differently in the two positions (e.g., Allen, 1950; Khachaturian, 1984).

We measure the acoustics of each production, and map the results in a two-dimensional space to determine the appropriate description of voice quality for the three-way contrast. The likely descriptions are schematized in Figure 1. If voiced $D^\text{ɦ}$ involves glottal spreading (breathy voice), it should have similar H1–H2 values as aspirated T^h, which must involve a breathy spread articulation (Cho et al., 2002; Garellek, 2012; Kagaya, 1974; Löfqvist & McGowan, 1992), and both should have higher H1–H2 than voiceless T. If it does not involve glottal spreading (or has only slight breathiness), it should have lower H1–H2 and higher CPP than T^h, indicating a more modal articulation. If voiceless T involves glottal constriction ($[\widehat{tʔ}]$; or as an ejective $[t']$), this would be indexed by lower H1–H2 than both $D^\text{ɦ}$ and T^h, but with a similarly low CPP because of irregular voicing.

Figure 1: Expected ranges of H1–H2 and CPP for the possible realizations of the three plosive series in Armenian.

To characterize the three-way contrast, we evaluate which acoustic variables involved in voicing best separate the each pair of stop categories, and whether all three categories can be reliably discriminated in both initial and final position. We compare the acoustics of the Ar-

language. Therefore, the best reference sound is one that is known to be modal, though this will not be possible for Yerevan Armenian plosives.

menian Dʱ stops with the voiced-aspirated stops in related Indic languages, and finally explore how the effect of voice quality on adjacent vowel formants follows the same pattern as a historical sound change in earlier Armenian.

2 Methods

2.1 Words

We extracted minimal triplets and pairs from three dictionaries (Decours, Ouzounian, Riccioli, & Vidal-Gorene, 2014; Nayiri Institute, 2016; Parker, 2008) and the public domain Electronic Library section of the Eastern Armenian National Corpus (Corpus Technologies, 2009). Triplets and pairs were selected with prevocalic word-initial plosives or postvocalic word-final plosives at labial (/pʰ, p, b/), dental (/t̪ʰ, t̪, d̪/), and velar (/kʰ, k, g/) places of articulation. For example, one such word-initial velar triplet is:

գոռ	կոռ	քոռ
/gɔr/	/kɔr/	/kʰɔr/
'fierce'	'forced labor'	'blind (informal)'

An example word-final velar triplet is:

թագ	թակ	թաք
/t̪ʰɑg/	/t̪ʰɑk/	/t̪ʰɑkʰ/
'crown'	'mallet'	'odd'

Although Yerevan Armenian also has the three-way voicing contrast for affricates at two places of articulation, we did not use minimal affricate sets because voicing and aspiration landmarks are difficult to measure during affricate releases. To further facilitate identification of acoustic landmarks and measurement of voice quality, plosives were limited to prevocalic and postvocalic environments at word edges. Besides this practical consideration, the voicing contrast is also more restricted in the few stop consonant clusters that can occur in medial or final position (see Vaux, 1998a, and §1.1.3).

The EANC Electronic Library includes classical texts which were scanned using optical character recognition (OCR), and thus many of the words extracted from it may not be used in modern spoken Yerevan Armenian, or else contain misspellings or OCR errors. For this reason, a native speaker of Yerevan Armenian verified each word, and excluded minimal sets if any word was not an existing word that was both familiar to her and that she thought many speakers from Yerevan would likely know. She also checked each word's translation, or suggested an alternate translation; function words and proper names were excluded.

This procedure resulted in 155 words containing the target plosives, comprising 14 minimal triplets (including 12 prevocalic word-initial, and 2 postvocalic word-final) and 57 minimal pairs (38 prevocalic word-initial; 19 postvocalic word-final). One word, տափ /t̪ɑpʰ/ 'plain', occurred in both a word-initial triplet and a word-final pair. Table 2 lists the number of minimal sets at each place of articulation, and Appendix A provides a complete list of the words used for the study.

Table 2: Number of minimal sets per position per place of articulation.

		Labial	Dental	Velar
Word-initial	Triplets	2	5	5
	Pairs	7	14	17
Word-final	Triplets	0	1	1
	Pairs	3	5	11

2.2 Speakers

Eight speakers of Eastern Armenian were recruited to record the target words for the study, including six women and two men. In the following discussion and visualizations, speakers are assigned a code based on their gender and age: for example, the code F20 is used for a female speaker, age 20. All eight speakers had grown up in Yerevan, and six had lived there through at least age 17. One had moved to California at age 14, and one had lived in Washington D.C. at ages 5–7 but otherwise lived in Yerevan until age 18. Four speakers were no longer residing primarily in Yerevan when they participated in the study, but all speakers reported that they continue to use Armenian on a daily basis.

In addition to native fluency in Armenian, all speakers reported at least some knowledge of both English and Russian. The mean self-reported fluency for English was 4.2 on at scale ranging from 1 to 5, where 5 indicates native or near-native fluency (reported range 3–5). For Russian, speakers rated their mean fluency at 3.4, with a reported range of 2–5. Some of the speakers had also taken classes or self-study beginning at age 12 or later in French, Japanese, Chinese, Dutch, Spanish, and/or Turkish.

All speakers gave informed consent using protocols approved by the UC San Diego IRB. The first and last two speakers that were recorded (F20, M25, F21) were paid for their participation; the others received a small gift. The first speaker was also paid for additional assistance in selecting the target words, for recording five of the other speakers, and for consulting during the design of the study. Because the speakers were recruited by referral from the first speaker, they are less likely to be representative of the general population of Yerevan Armenian speakers, and any inter-speaker differences should not necessarily be construed as reflecting broader gender or age differences in the population.

2.3 Recording procedure

Carrier sentences Each speaker first read the 155 words (including 111 with target word-initial plosives, 43 with target word-final plosives, and 1 with both) in the carrier sentence ասա «___» բարձր /ɑsɑ ___ bɑɾd͡zɾ/ 'say ___ aloud'. All speakers were recorded in a quiet room using a portable Blue Yeti USB microphone with the Praat software (Boersma & Weenink, 2017), with a 44.1 kHz sampling rate. The first and last two speakers (F20, M25, F21) were recorded by the authors in a sound-attenuated booth at UC San Diego, and the other speakers were recorded by the first speaker in Yerevan.

The carrier sentence was chosen so that word-initial plosives occurred between vowels, which makes identifying the closure in a spectrogram straightforward. However, voice onset

time can be difficult to measure in intervocalic position (cf. Abramson & Whalen, 2017), especially when voicing may carry into the closure from the previous vowel. Additionally, in this carrier sentence, the target word-final plosives occurred between voiced sounds, which is likely to facilitate final voicing, and thus may lead to an inaccurate impression of the voicing contrast in final position. To evaluate the plosive acoustics in an alternative environment, speakers next read a second list containing only the 44 target words with word-final plosives in a second carrier sentence, ասա «___» պարողին /ɑsɑ ___ pɑɾoʁin/ 'say ___ to the dancer'.

Instructions   The first speaker was knowledgeable about the study, and read the items in a random order. The other speakers were naïve to the purpose of the study, and read each list using a pseudorandomized order such that the same voicing category did not occur more than twice in a row in either word-initial or word-final position (regardless of place of articulation). Three of the speakers were recorded with reversed versions of the lists in order to help mitigate fatigue or practice effects on particular words. The two lists were organized into sets of 18 words, and speakers were encouraged to take short breaks between each set. All speakers were asked to pronounce the words as if they were speaking to a friend, to the extent that it was possible to do so. If a word was read disfluently, the speaker was asked to repeat it, and the second recording was used in the analysis.

Prosody   The three words in the carrier sentences were typically produced with rising pitch on the first word, a flat or rising pitch on the second word (most often rising for polysyllabic words), and almost always a fall on the third word. Speaker F21 generally had rising list intonation on the third word instead. Our judgment was that speakers typically had major prosodic breaks before and after each target word, suggestive of an accentual phrase or intermediate phrase. Some sentence productions clearly had a stronger prosodic break before or after the target initial or final plosive, including most of those by speaker M30. These breaks were annotated using the procedure described in §2.4 below.

2.4   Annotation procedure

Each recording was annotated using the waveform and spectrogram editor in Praat. The onset and offset of the closure, release burst, and adjacent vowel were annotated for each target plosive. Additionally, the onset and offset of voicing during the closure were also marked if present. For word-initial plosives, voicing at the beginning of the closure was ignored if it did not last longer than five pulses, since this voicing is most likely carried over from the preceding vowel (Lisker & Abramson, 1964).

Closure   The closure was defined as the portion of silence, or silence with voicing only, preceding the release burst. Because speakers occasionally inserted a pause before the target word, the location of the closure onset was sometimes unclear. If the closure onset could be identified by a visible transient in the waveform, the closure was marked beginning at the transient. If not, the closure interval was always marked as including the full silent portion before the release burst.

Release and aspiration   The release burst included only the transient(s) immediately following the closure, including multiple bursts if present. If a release was fricated without well-defined burst transients, the full fricated portion was included. If aspiration (broadband noise) was distinguishable from burst transients, it was not included in the burst interval. For word-final plosives, if there was aspiration that carried beyond the release burst and which was clearly distinguishable from the burst, the offset of this final aspiration was also marked.

Vowel   For word-initial plosives, the vowel onset was defined as either the release burst offset or the onset of a periodic voicing wave following the release of the closure, whichever occurred later. The landmarks for the vowel offset varied depending on the following sound. For word-final plosives, the vowel was the portion between the previous sound and the closure onset.

Besides these intervals, each token was also annotated for the presence of a strong prosodic break before initial stops or after final stops. Pitch tended to be very similar across words and speakers (see §2.3), and none of the target words were preceded or followed by an intake of breath. Thus, a relatively long silence in the spectrogram was used as an approximation of whether the target word was adjacent to a stronger prosodic break, which indicates that it might be initial or final in a higher-level prosodic domain. For initial stops, a strong prosodic break was annotated if there was either at least 100 milliseconds of silence before the plosive onset transient, or else an apparent closure duration of at least 150 milliseconds in the absence of an onset transient. For final stops, a strong prosodic break was annotated if there was a silence of at least 100 milliseconds between the release and the following stop onset transient. In the absence of a following stop transient, a strong prosodic break was annotated if at least 150 milliseconds elapsed before the following stop release, or else if there was the percept of a pause in the absence of both a final stop release and a following stop transient.

2.4.1   Example waveforms and spectrograms

Figure 2 shows waveforms and spectrograms for two minimal triplets produced by one speaker. The word-initial plosives in the upper row are annotated with the closure between colored lines 1–2 in each of the three waveforms, the release burst between lines 2–3, and the vowel between the last two lines. The aspirated plosive in the upper right also has an additional interval which marks voiceless aspiration after the burst between lines 3–4. The voiced plosive in the top left has voicing throughout the oral closure, but this was not always the case (see §2.5), and we annotated the onset and offset of voicing during the closure separately from the closure interval itself.

In the lower row, the word-final plosives are annotated with the vowel between lines 1–2, the oral closure between the lines 2–3, and the release burst between lines 3–4. The voiceless and aspirated plosives both have two release bursts, which is common for velar plosives, and the aspirated plosive in the lower right has additional aspiration following the two bursts between the lines 4–5 in that waveform diagram. The final voiced plosive in the lower left has either a release that is partially spirantized, or else a short portion of aspiration following the release (see §3.1.1 on similar patterns in initial plosives). As with the word-initial voiced plosive in the upper row, this voiced plosive has voicing throughout the oral closure, but we annotated
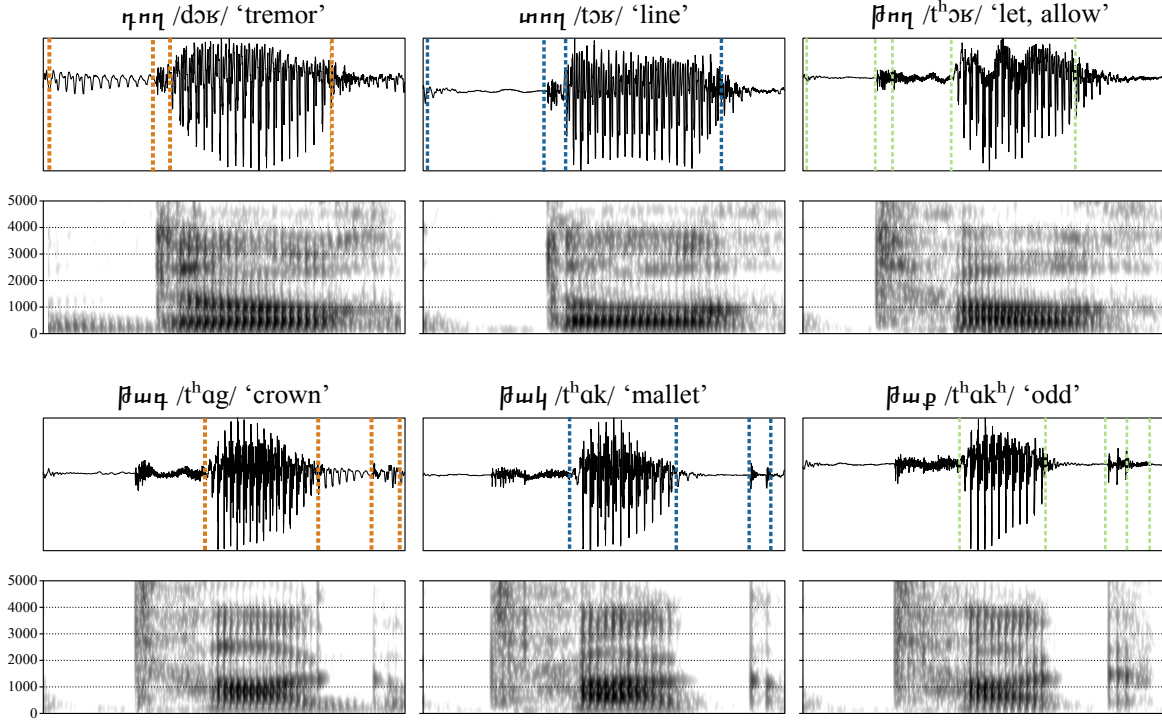
Figure 2: Waveforms and spectrograms for a word-initial minimal triplet (upper row) and a word-final minimal triplet (lower row). Dashed lines show annotation boundaries described in the text.

the onset and offset of voicing during the closure separately for tokens where this was not the case.

2.5 Acoustic measurements

Voice quality   We used VoiceSauce (Shue, Keating, Vicenik, & Yu, 2011) to estimate H1*–H2* and the noise measure CPP over the vowel interval. The asterisks for H1*–H2* indicate that the measure has been corrected for the effects of the estimated formant filter on the harmonics' amplitudes, which facilitates cross-vowel comparisons and provides an approximation of H1–H2 derived from the voice source before vocal tract filtering. All measurement settings were configured to the VoiceSauce defaults (version 1.27). Harmonic amplitudes were estimated at overlapping windows that spanned three pitch periods, with the STRAIGHT algorithm used for pitch tracking (Kawahara, de Cheveigné, & Patterson, 1998). Corrections to harmonic amplitudes were based on Hanson (1997) and Iseli, Shue, and Alwan (2007), with formants measured using the Snack toolkit with default settings (Sjölander, 2004). CPP was calculated over windows comprising five pitch periods. Both measures were smoothed using a moving average over 20 milliseconds.

This procedure produced a series of H1*–H2* and CPP values at 1-millisecond intervals across the full timecourse of each vowel. As summary values, we also calculated the average H1*–H2* and CPP values over a portion of the vowel. For plosives in word-initial position,

we took averages over the first third of the vowel, which is the portion closest to the plosive, excluding the release burst and any portion of voiceless aspiration. For word-final plosives, the summary values were averages over the final third of the vowel, adjacent to the plosive onset.

Voice timing   Based on the annotations, we also measured VOT, defined as the time from closure offset to the onset of voicing (Adjarian, 1899; Lisker & Abramson, 1964). Because voicing during the closure often died before the release (cf. Abramson & Whalen, 2017), it was sometimes difficult to decide whether it was appropriate to mark a plosive as having negative VOT. We used the following rule: if voicing did not stop prior to the closure offset, or if at least half of the closure was voiced, it was measured as having negative VOT (though see Davidson, 2016, 2017). In syllable-final position, the equivalent to VOT is voice offset time, which we measured as the time between the closure offset and the offset of voicing (VOFT; Abramson & Whalen, 2017). However, we note that VOFT has not been consistently defined in the literature (cf. Singh, Keshet, Gencaga, & Raj, 2016).

Voicing strength and aspiration   Because of the challenges in measuring VOT, we also used VoiceSauce to identify voicing epochs (peak excitation of pulses) and to measure their strength-of-excitation (SoE; Mittal, Yegnanarayana, & Bhaskararao, 2014; Murty & Yegnanarayana, 2008) during the closure. These measurements occur at each epoch, with 1-millisecond resolution. Because SoE is the peak excitation strength of the harmonic component of the signal, SoE thus serves to measure the intensity of the voicing. As a summary value, we also calculated average SoE over each closure interval.

Finally, we measured the durations of the closure and vowel, the duration of voicing during the closure, and the duration of aspiration. For word-initial plosives, aspiration duration was defined as the time between the closure offset (i.e., the onset of the first burst) and the vowel onset (i.e., the onset of voicing; see discussion in §3.3). For word-final plosives, it was defined as the time between the closure offset and either the offset of the burst or the offset of any aspiration that followed the burst.

2.6   Amount of data

Number of tokens   In total, there were 1600 tokens included in the study, including 896 with word-initial plosives (112 words per speaker; totaling 264 voiced tokens, 400 voiceless, 232 aspirated) and 704 with word-final plosives (44 words per speaker, each recorded in two carriers; totaling 176 voiced tokens, 304 voiceless, 224 aspirated). Of the 1600 tokens, 32 word-final plosives (2%) were unreleased, making it impossible to annotate the closure interval given the following plosive, and therefore do not include any measurements relating to the closure or release.

Exclusions   Since accurate H1*–H2* (and f0) measurements depend on accurate pitch tracking, we used estimated the pitch tracks in two passes, using the following procedure. In the first pass, we used VoiceSauce to estimate f0 for all vowel tokens, while allowing it to search for f0 within wide limits (the default of 40–500 Hz). We then visually inspected per-speaker histograms of this set of estimated f0 values, including all windows in each pitch track. Based on this

inspection, we revised the lower and upper pitch limits for each individual speaker. The limits for each speaker were chosen so that they would eliminate outliers which fell outside that speaker's apparent normal f0 distribution, but otherwise include the full empirical tails of the distribution. These speaker-specific pitch limits are given in Appendix B.

In the second pass, we used VoiceSauce to re-estimate all acoustic measurements while constraining the f0 estimates to be within the speaker-specific limits. We then inspected all pitch tracks where the estimate in any measurement window was more than five semitones different from the estimate in the preceding window. The majority of these tokens had an obviously-mistracked pitch doubling or halving which persisted for only a few windows. We manually excluded the mistracked windows from each token, so that the H1*–H2* summary values would not include incorrect pitch-doubled or halved estimates. The portion of each track that did not appear to be mistracked was not excluded, nor were any measurements which do not depend on f0 (i.e., all other measurements). Additionally, we manually excluded all measurements from seven tokens (4 voiceless unaspirated, 3 voiced) produced by speaker F21 with substantial phrasal creak, which made pitch measurements unreliable.

In addition, the corrections to H1*–H2* depends on accurate estimation of the formant filter. We visually inspected two-dimensional distributions of all F1–F2 estimates, for each vowel type for each speaker. We excluded outlying F1 and F2 values which fell outside ranges that were chosen based on visual inspection, which are listed in Appendix B. As before, the portion of any formant track which was not outside these limits was not excluded, nor were any measurements which do not depend on formant estimates. In total, measurements which depend on pitch tracking were fully excluded for 29 tokens (1.8%), and measurements which depend on formant tracking were fully excluded for 48 tokens (3.0%). All acoustic measurements reported in the following sections are derived from these second pass estimates only, with exclusions.

Prosody   103 initial plosives (11.5% of all initial plosives) and 123 final plosives (17.5% of all final plosives) were determined to be adjacent to strong prosodic breaks. In particular, speaker M30 had strong prosodic breaks for the majority of the target words (51.7% of initial and 61.6% of final plosives produced by M30), and across speakers, half of all tokens with strong prosodic breaks (50.4%) were produced by speaker M30. Since this may affect certain acoustic measurements, we discuss how the presence of a stronger prosodic break affected each measurement in the corresponding subsections below. Because the current study was not designed to explore these effects, this discussion should be taken only as suggestive for future work.

2.7   Analysis procedure

In the following analysis, we first characterize the data descriptively, focusing on each of four components of plosive realization: voice quality (H1–H2 and CPP), voice timing (VOT and VOFT), voicing intensity (including the presence and duration of voiceless aspiration), and the durations of the closure and adjacent vowel. This section (§3) provides information about the expected values and variability in each acoustic measure for Yerevan Armenian plosives, and discusses some broader techniques and problems that arise when using each measure.

The next section (§4.1) summarizes and explores our findings on voice quality: we show that the voiced series is classically breathy, while the voiceless unaspirated series is modal for six speakers, but may have tense voice quality for two speakers. The last section of the analysis

(§4.2) addresses two questions about phonetic and phonological contrast: which individual variables most robustly differentiate the three plosive categories? Can all three categories be statistically distinguished from one another in both syllabic positions at a rate that is usefully above chance?

Throughout the analysis, we primarily use descriptive rather than inferential statistics. The acoustic measurements are drawn from three categorically different sounds, and given a relatively large number of tokens, we expect to find that most acoustic variables will have significantly different distributions between the three stop categories. However, small but significant differences between categories do not imply that an acoustic variable is useful in discriminating the contrast, nor that it reflects distinct articulatory processes. We instead draw conclusions mainly based on the magnitude and variability of acoustic differences between the plosive categories, as well as the relationship of multiple variables in Yerevan Armenian, and supplement these descriptions with a classification model in §4.2.2.

3 Acoustics of the voicing contrast

3.1 Voice quality

3.1.1 Word-initial plosives

Voiced plosives   The top left panel of Figure 3 illustrates the differences in H1*–H2* by position and by voicing series. H1*–H2* indexes glottal constriction, and lower values are associated with greater constriction. In word-initial position, H1*–H2* tends to be higher overall for the voiced plosives than the voiceless unaspirated ones. This relationship is compatible with two interpretations. First, the voiced series could be breathy, involving a spread-glottal configuration, while the voiceless unaspirated series could be modal or constricted. Alternatively, the voiced series could be modal and the voiceless unaspirated series could be constricted. However, for word-initial position, H1*–H2* has similar values for voiceless aspirated and voiced plosives. Because aspirated plosives by definition must involve glottal spreading, this implies that the voiced plosives have a similar degree of glottal spreading, and thus involve a relatively breathy voice quality.

The left panel of Figure 4 shows the timecourse of H1*–H2* during the vowel for each of the three voicing categories following word-initial plosives, beginning with the onset of voicing.[2] The voiceless aspirated series begins with high H1*–H2*, indexing a spread glottis that begins during the closure (Kagaya, 1974; Kagaya & Hirose, 1975), but it rapidly drops. At the vowel

---

[2]The curves in each panel of Figures 4 and 11 were modeled with a linear mixed-effects regression (Bates, Mächler, Bolker, & Walker, 2015). The data were the acoustic measurements taken at 1-millisecond intervals within each token (see §2.5) over the first half of the vowel following word-initial plosives, and over the second half of the vowel preceding word-final plosives. Predictors were a natural cubic spline function for measurement time (with time scaled within each vowel token so that 0 is the onset of each vowel, 0.5 is the vowel midpoint, and 1 is the offset) with two internal knots, which interacted with voicing category (voiced, voiceless, or aspirated). Models also included group-level intercepts for speaker, minimal-pair (or triplet), and token; as well as group-level slopes for voicing for each speaker, and a group-level cubic spline function for each minimal-pair. The models fit to individual speakers in Figure 11 do not include group-level predictors for speaker. The standard errors in each panel do not take into account group-level effects.
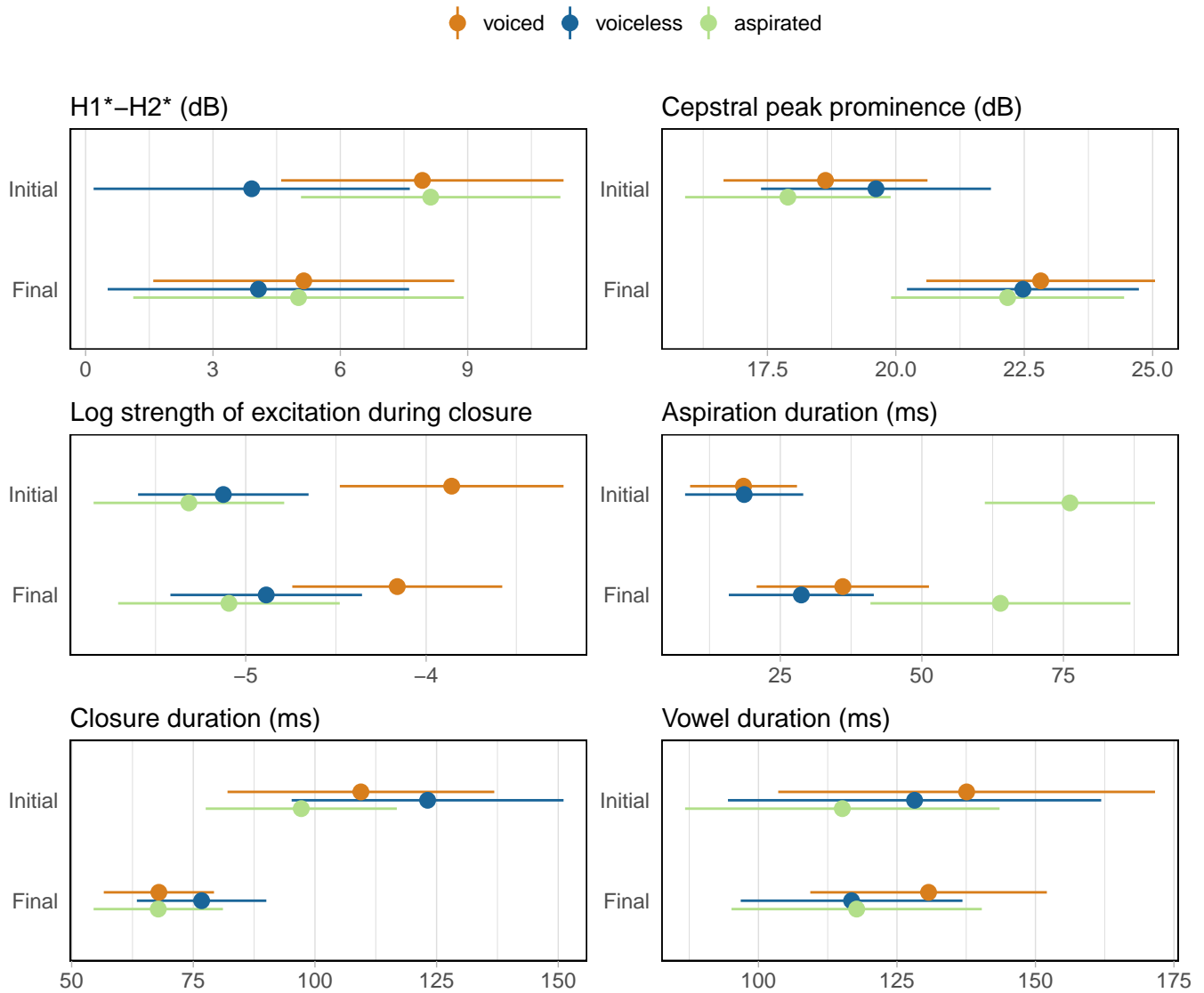
Figure 3: Observed mean values of six acoustic variables, divided by position and voicing series. Points show the mean for each group of tokens; lines show one standard deviation above and below the mean. All variables were mean-centered within-speaker before standard deviations were calculated. Strength of excitation is a proportion from 0 to 1, shown here after natural-log-transformation.

midpoint (0.5 on the x-axis), it is similar to the voiceless unaspirated series, likely indicating similar modal voice quality at that point. This can be compared to the voiced series: the voiced series also begins with high H1*–H2*, but it has a somewhat less rapid drop that does not reach the same level at the vowel midpoint.

   Overall, the voiced series might thus best be characterized as having a breathy voice quality which likely begins during the closure and extends into the vowel. Visual inspection of the waveforms suggested that some word-initial voiced plosives had a portion of voiceless aspira-
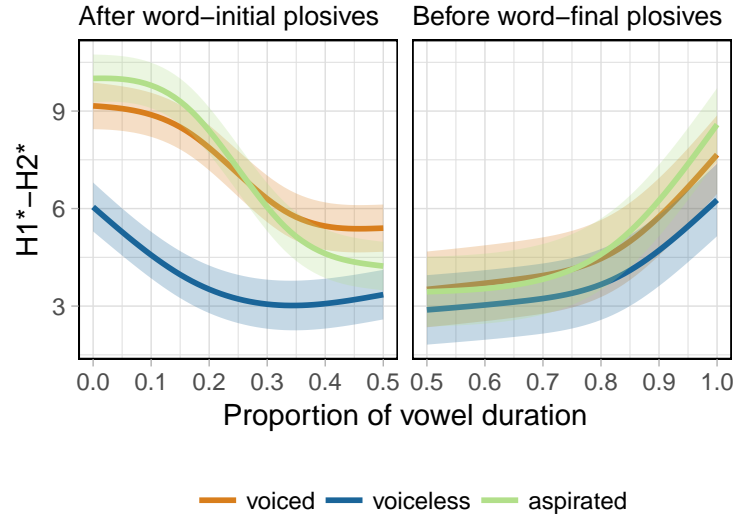
Figure 4: H1*–H2* (y-axis) over time during the vowel (x-axis) for the three voicing categories, estimated by cubic spline regression (see footnote 2 for model details). Lines show estimated means, and shaded areas show one standard error above and below the estimated means.

tion, but this was primarily restricted to /g/, and therefore might also be velar spirantization (which is common in many languages), or else a noisy, fricated release. However, most tokens did not have a distinct portion of voiceless aspiration following the voiced closure.

Voiceless unaspirated plosives   For the voiceless series, the acoustics suggest that there is more constriction than for the breathy-voiced and voiceless aspirated series: in Figure 3, speakers have overall lower H1*–H2* for the voiceless unaspirated series than the other two series. However, because H1*–H2* is a relative measure, and both of the other two series have less constriction, the voiceless unaspirated series could have either modal or creaky voice quality (see Figure 1).

To assess whether the voiceless unaspirated series involves glottal constriction, it is necessary to use two acoustic variables in combination. CPP serves as a measure of noise in the signal, and thus distinguishes modal phonation (higher CPP) from creaky and breathy phonation (lower CPP). Figure 5 shows the bivariate distribution of H1*–H2* and CPP in the vowel immediately following the word-initial plosives. In a two-dimensional acoustic space with H1*–H2* on the x-axis and CPP on the y-axis, the breathiest part of the space is on the right (where H1*–H2* is highest) at the bottom (where CPP is lowest).

In Figure 5, the relative ordering along the x-axis (H1*–H2*) is similar for all speakers, consistent with the overall means shown in Figure 3. For CPP (on the y-axis), there is variation among speakers. Speakers F18, F19, F20, F21, F22, and F49 have numerically the highest CPP values for the voiceless series, which suggests that their voiceless plosives are accompanied by more modal voicing than either the breathy-voiced or voiceless aspirated series. This is reflected to some extent in the H1*–H2* tracks in Figure 4 (left panel). If the voiceless series were glottalized and followed by a modal vowel target, we would expect to see the following: first, a

lower H1*–H2* near the vowel onset, reflecting a constricted glottis; and second, a subsequent rise in H1*–H2* as the vowel transitions from the constricted release of a glottalized plosive towards the more modal vowel target (e.g., Cho et al., 2002). Instead, H1*–H2* falls from the vowel onset and levels off as it reaches the vowel midpoint, where it is roughly similar to the other two series (see Berkson, 2013, Figure 31 for a similar H1*–H2* trajectory for modally-voiced plosives in Marathi). In addition, we did not identify any glottalized onsets (defined as the presence of irregular pitch periods in the waveform) during manual inspection of the voiceless plosive waveforms.[3]
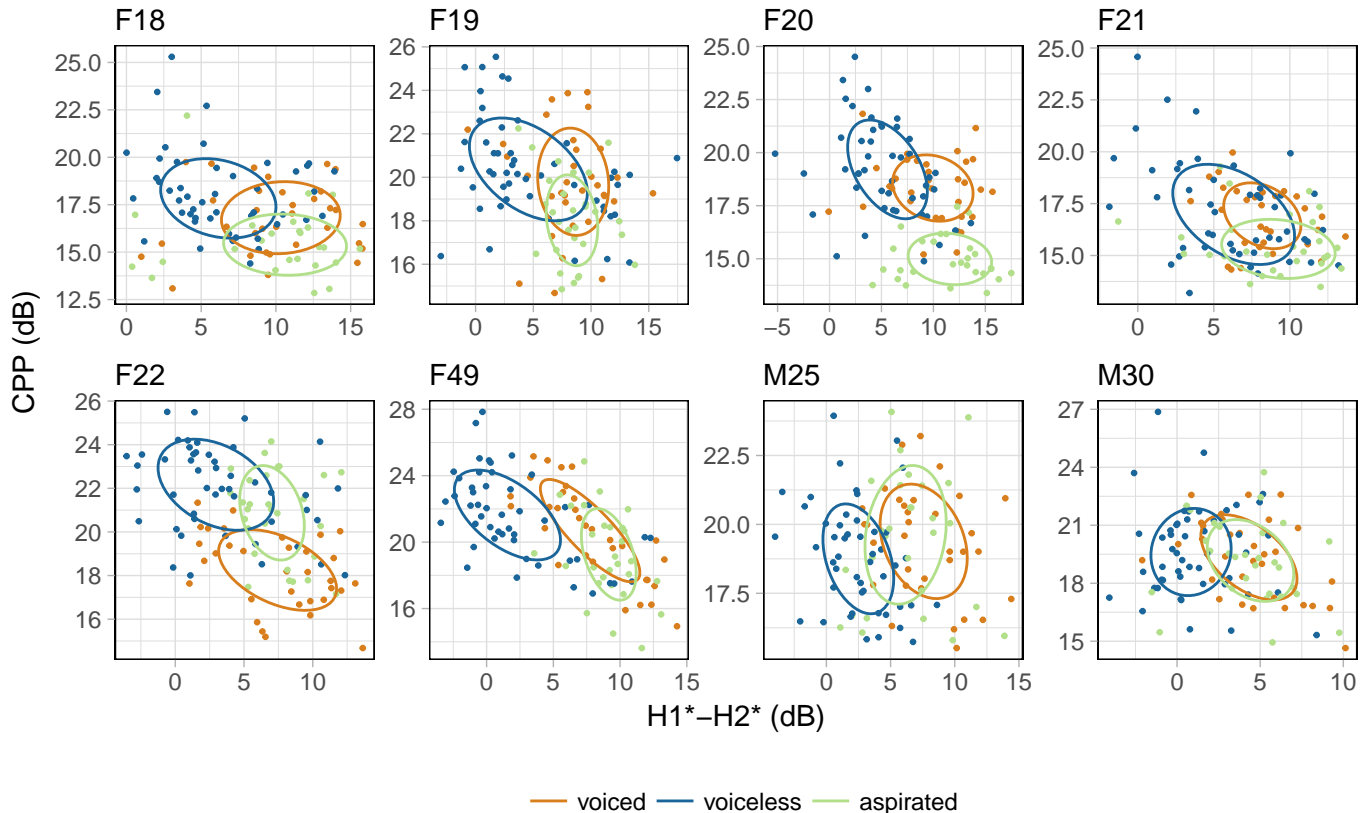

Figure 5: Summary H1*–H2* (x-axis) versus CPP (y-axis) measurements for word-initial plosives. Measurements were calculated by averaging over the first third of the following vowel. Ellipses are drawn around the center 50% of points for each category.

On the other hand, speakers M25 and M30 have similar CPP values for all three series. This means that vowels following the voiceless plosive series are as noisy as the ones following the breathy-voiced or voiceless aspirated series, but with lower H1*–H2*. It is thus possible

---

[3]The exception was a small number of creaky tokens which were excluded for speaker F21 (see §2.6). However, for these few tokens, creak generally extended across the entire carrier sentence, and there were about the same number of voiced plosives with creak as voiceless ones. This creak is therefore unlikely to be a component of speaker F21's voiceless plosive realization. Note that the differences between the voicing categories for both H1*–H2* and CPP tend to be much smaller for speaker F21, which is probably due to this speaker's overall more frequent use of somewhat-creaky phonation.

that these speakers produce the voiceless plosives with vocal fold constriction, which results in irregular creaky voicing on the following vowel. Creaky voicing is associated with lower values of both H1*–H2* (due to the increased vocal fold constriction) and CPP (due to the irregular voicing). We return to this question in §4.1.2.

3.1.2 Word-final plosives

Overall means for H1*–H2* and CPP adjacent to word-final plosives are shown in the top row of Figure 3. For all individual speakers, the three voicing categories have almost complete overlap on both measures in final position. There were no differences in CPP between the three categories, either overall or for individual speakers. There was a small difference in overall H1*–H2* (about 1 dB, shown in the top left panel of Figure 3), such that vowels preceding the voiced and aspirated series had higher H1*–H2* (indicating greater glottal spreading) than vowels preceding the voiceless series. This effect is about one-quarter of the difference in initial position, and it is unlikely to meaningfully separate the categories (see §4.2).

The right-hand panel of Figure 4 shows the timecourse of H1*–H2* beginning at the vowel midpoint and leading into the word-final plosives, for each voicing series. All three trajectories show a rise into the plosive closure. While there is a small difference in H1*–H2*, it begins at the vowel midpoint and is constant throughout, which suggests that it is less likely to be due to a glottal constriction or spreading gesture leading into the plosive closure.

During manual inspection of the plosive waveforms, we identified a small number of ejectives with a characteristic double release, but this might be attributed to hyperarticulation; ejectives are also attested in English phrase-final stops (Ladefoged, 2006, p. 135; Gordeeva & Scobbie, 2013). Otherwise, there was no clear effect of voicing category on voice quality measures in the final third of the vowel adjacent to word-final plosives, and no convincing evidence for non-pulmonic articulation in this environment.

3.2 Voice timing

3.2.1 Voice onset time for word-initial plosives

Figure 6 shows histograms of voice onset times for the eight speakers. All speakers have three modes which generally correspond to the three voicing categories, though the correspondence is not perfect. Six of the eight speakers produced several tokens in the voiced series with zero or short-lag VOT, while speakers F18 and M25 had no such tokens (mean = 4.9 tokens per speaker, comprising 14.8% of all voiced tokens). Additionally, five of the eight speakers had 1–3 tokens in the voiceless unaspirated or aspirated series with lead VOT, using the criteria in §2.4 (2.4% of all voiceless unaspirated and aspirated tokens), though there was only one such token that was adjacent to a strong prosodic break. For the voiced and voiceless unaspirated series, VOT was not substantially different for tokens that were produced after a strong versus weak prosodic break (on average, both were < 2 ms smaller next to a strong prosodic break). The aspirated series had somewhat longer VOT following a strong prosodic break (difference of means $= 19$ ms, $\sigma = 27$ ms).

One notable feature of the distributions is the large gap between the voiced and voiceless series (also observed by Lisker & Abramson, 1964, p. 407). There is a gap because negative VOTs
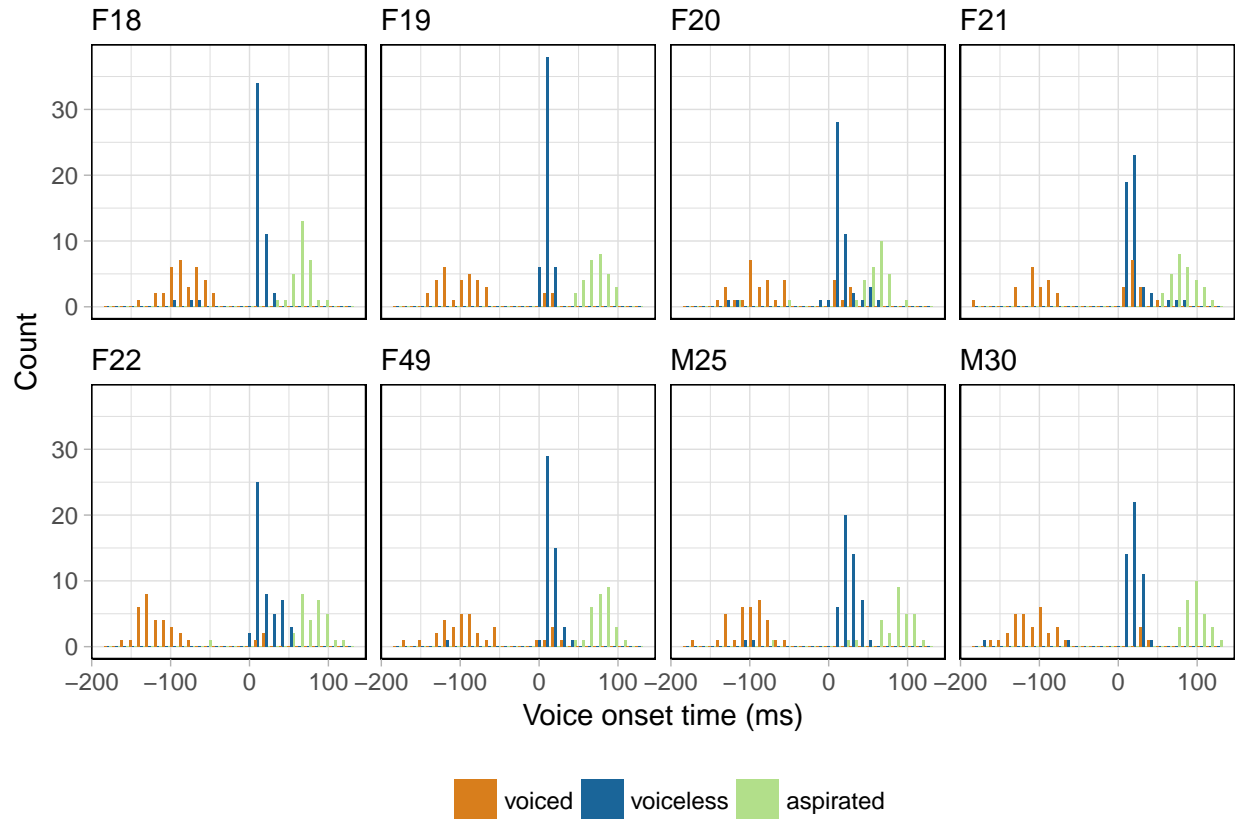
Figure 6: Histograms of voice onset times for each speaker, colored by category of the initial plosive.

are usually equal to exactly the closure duration, and there are no short ($<$ 20 ms) closures. There were 896 tokens with word-initial plosives, of which 331 had some amount of voicing during the closure. 136 of these tokens (41.1%) had voicing for the full duration of the closure. Of the remaining tokens, there were only 31 (9.4%) in which voicing began after the closure onset, which may even be an overestimate, since the annotated closure onset was probably later than the actual closure onset in some cases (see §2.4). Because any closure voicing usually begins at the closure onset, negative VOT is thus almost always equal to the closure duration. The voiced histograms, then, actually show the distribution of closure durations, and the voiceless and aspirated histograms show the distributions of short-lag and long-lag VOTs.

3.2.2 Voice offset time for word-final plosives

Figure 7 shows histograms of voice offset times for the eight speakers. Shaded density estimates are included in the plots, due to the wide range of observed values for all three voicing series. The expectation is that VOFT should be closer to zero for voiced tokens, and more negative for voiceless ones. However, because it does not capture aspiration after the closure, it is not likely to distinguish stop voicing contrasts in languages with final aspiration. Indeed, while VOFT was greater for voiced tokens ($\mu = -21$ ms, $\sigma = 22$ ms) compared to voiceless ($\mu = -46$ ms,

$\sigma = 29$ ms) and aspirated ones ($\mu = -44$ ms, $\sigma = 27$ ms), there was little difference between voiceless and aspirated VOFT. Unlike initial VOT, there was substantial variability and overlap in VOFT among the three series. Voice offset time was also about 10 ms more negative (i.e., earlier voice offset) for all three series when the final plosive preceded a strong prosodic break.

Figure 7: Histograms of voice offset times with overlaid density estimates for each speaker, colored by category of the final plosive.

3.3 Voicing strength and aspiration

3.3.1 Word-initial plosives

Voicing during stop closures typically becomes weaker over time due to increasing supraglottal pressure. When voicing ceases during the stop closure, it is not clear whether it should be measured as negative VOT (Abramson & Whalen, 2017) or whether it belongs to some other descriptive category of laryngeal timing (Davidson, 2016). Of the 254 word-initial tokens in the voiced series with some amount of visible voicing, voicing ceased before the closure release in 99 (39.0%). For 28 such tokens (11.0%), voicing ceased 50 milliseconds or more before the closure offset.

Although voicing ceases for many tokens, closure strength-of-excitation (SoE) robustly distinguishes the voiced series from the other two (see Figure 3, second row left). In conjunction

Figure 8: Summary log strength-of-excitation (x-axis) versus aspiration duration (y-axis) measurements for word-initial plosives. Measurements were calculated by averaging over the first third of the following vowel. Ellipses are drawn around the center 50% of points for each category. The shape of the points indicates whether or not each token followed a strong prosodic break.
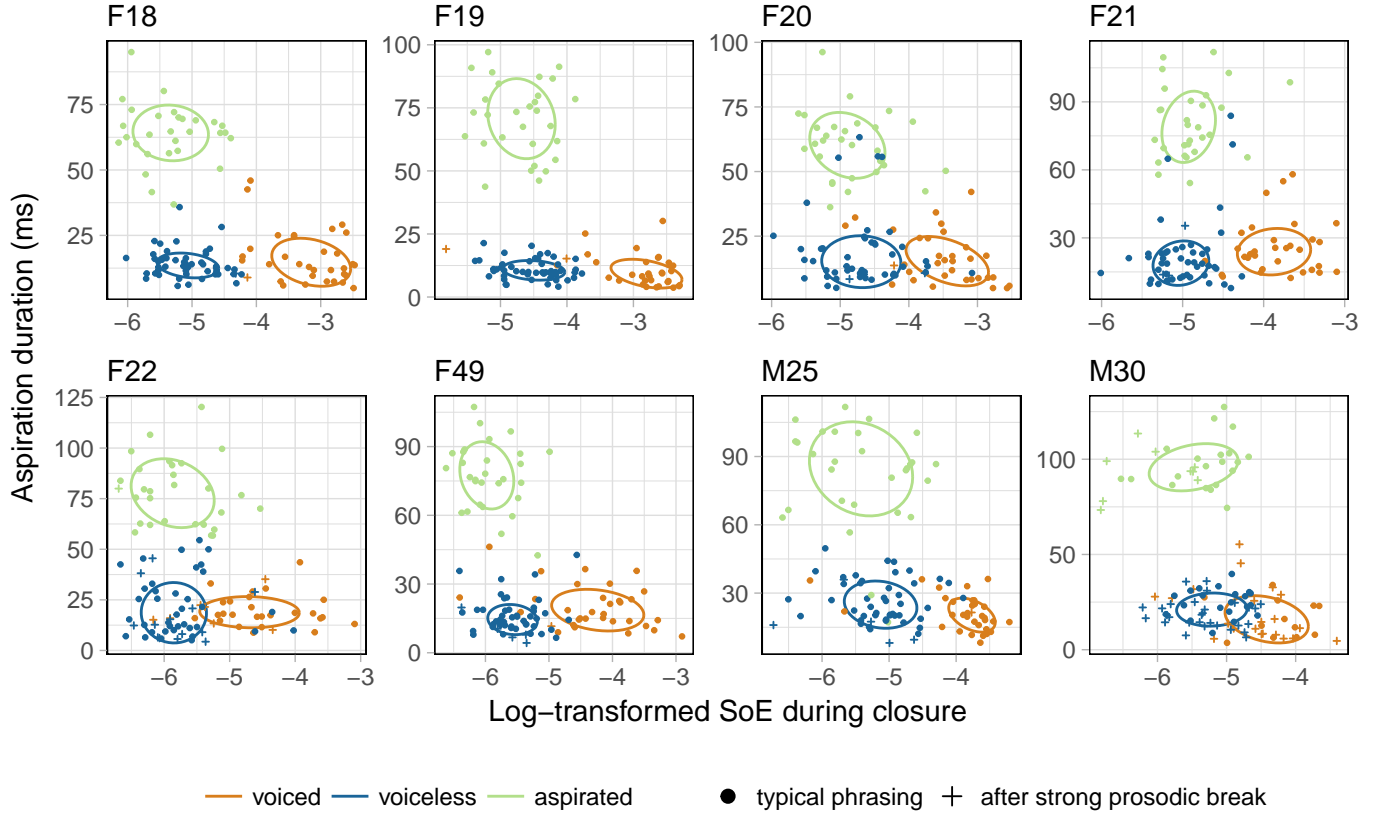
with the duration of aspiration (Figure 3, second row right), this is sufficient to separate the three series. Figure 8 shows the averaged SoE during closure on the x-axis plotted against aspiration duration on the y-axis. While the voiced series occasionally has a portion of voiceless aspiration (e.g., speaker F22; cf. §3.1.1), the aspirated series has reliably longer aspiration than the other two series, and the voiced series has reliably stronger voicing than the other two series.[4] For all speakers, these two cues in combination provide a clean separation between the series.

---

[4]Plosives following a strong prosodic break had lower mean SoE than those after weaker breaks, but this was true for all three categories (difference in means for voiced = -1.01; voiceless unaspirated = -0.32; voiceless aspirated = -0.84), and voiced stops had stronger voicing than the other two categories regardless of the break. This is reflected in the data for speaker M30 (e.g., in Figure 8), who produced the majority of tokens with a strong prosodic break (see §2.4). Aspiration duration for the voiced and voiceless unaspirated series was not affected by following a strong prosodic break (both differences in means < 3 ms), while the aspirated series had somewhat longer aspiration (17 ms) after a strong prosodic break. See Figure 3 for standard deviations.

3.3.2 Word-final plosives

Figure 9 shows log-transformed strength-of-excitation (on the x-axis) and aspiration duration (on the y-axis). As with word-initial plosives, these two dimensions provide good separation of the word-final voicing contrast (see the second row of Figure 3). The contrast is somewhat diminished for speaker F21 in word-final position. All three categories had lower SoE before a strong prosodic break, but even for these, the voiced plosives had greater SoE than the other two categories in both carrier sentences (smallest difference $= 0.47$ before voiceless sounds with a typical weak prosodic break; compare with differences in Figure 3).

Speakers F20 and F22 (and occasionally other speakers) have longer portions of aspiration for the final voiced plosives, which fall between the voiceless series and the aspirated series in aspiration duration. Additionally, the voiced and aspirated series both had greater aspiration before a strong prosodic break (difference in mean aspiration for voiced $= 23$ ms; aspirated $= 24$ ms), while the voiceless unaspirated series did not (difference $= 3$ ms; see Figure 3 for standard deviations).

Figure 9: Summary log strength-of-excitation (x-axis) versus aspiration duration (y-axis) measurements for word-final plosives. Measurements were calculated by averaging over the final third of the preceding vowel. Ellipses are drawn around the center 50% of points for each category. The shape of the points indicates whether or not each token was followed by a voiceless /p/, a voiced /b/, or a strong prosodic break.

While initial voiced plosives tended not to have substantial aspiration, in contrast with the final voiced plosives, this difference might partially be due to differences in how they were annotated. In prevocalic initial position, aspiration was considered to be the duration of voiceless aspiration only, including the release burst. Breathy voicing (sometimes called voiced aspiration; see §5.1 for more discussion) was not included in the measurement of aspiration duration, though it can be seen clearly at the vowel onset as higher H1*–H2* in Figure 4. In postvocalic final position, aspiration was considered to be the full duration of noise after the release preceding the following plosive consonant in the carrier sentence. This potentially includes both voiced and voiceless aspiration in final position, but on visual inspection of all tokens, we found that there were few word-final plosives which had voicing during the release phase. The few tokens which appeared to have voiced releases all occurred in the first carrier sentence, in which the following plosive was voiced /b/, and so it may be that this voicing is primarily due to the surrounding voiced consonants.

3.4 Closure and vowel duration

Besides voice quality, excitation strength, aspiration duration, and voice onset time, we also examined closure and vowel duration as cues to the voicing contrast, shown in the third row of Figure 3. The left two panels of Figure 10 shows mean closure durations, divided by stop place, and the right two panels show mean vowel durations, divided by vowel quality. For both closure and vowel durations, all speakers have the same pattern as the overall summaries shown in Figure 10.

For both word-initial and word-final plosives, closure duration is slightly longer in the voiceless series (initial: $\mu = 123$ms, final: $\mu = 77$ms) compared to the voiced (initial: $\mu = 109$ms, final: $\mu = 68$ms) and aspirated (initial: $\mu = 97$ms, final: $\mu = 68$ms) series.[5] Vowel durations follow typical cross-linguistic patterns: vowels tend to be longer adjacent to voiced plosives relative to voiceless ones; and word-initial voiceless aspirated plosives are followed by shorter vowels, which is likely due to the exclusion of all voiceless aspiration from vowel duration.

---

[5]The aspirated stops were also produced less often with an apparent strong prosodic break, at a rate of 4.5% in initial position and 10.3% in final position; as compared to 14.8% and 18.9% for initial and final voiced stops and 17.3% and 32.1% for initial and final voiceless stops ($p < 0.001$ by Fisher's exact test with token counts). Because the criteria for annotating a strong prosodic break were based on silence duration, the presence of an apparent strong prosodic break is confounded with closure duration. It is likely that strong prosodic breaks were over-annotated for voiceless unaspirated stops and under-annotated for voiceless aspirated ones. Alternatively, it is possible that strong prosodic breaks were not produced equally often across the three categories, and closure durations are more similar across the categories than shown here.
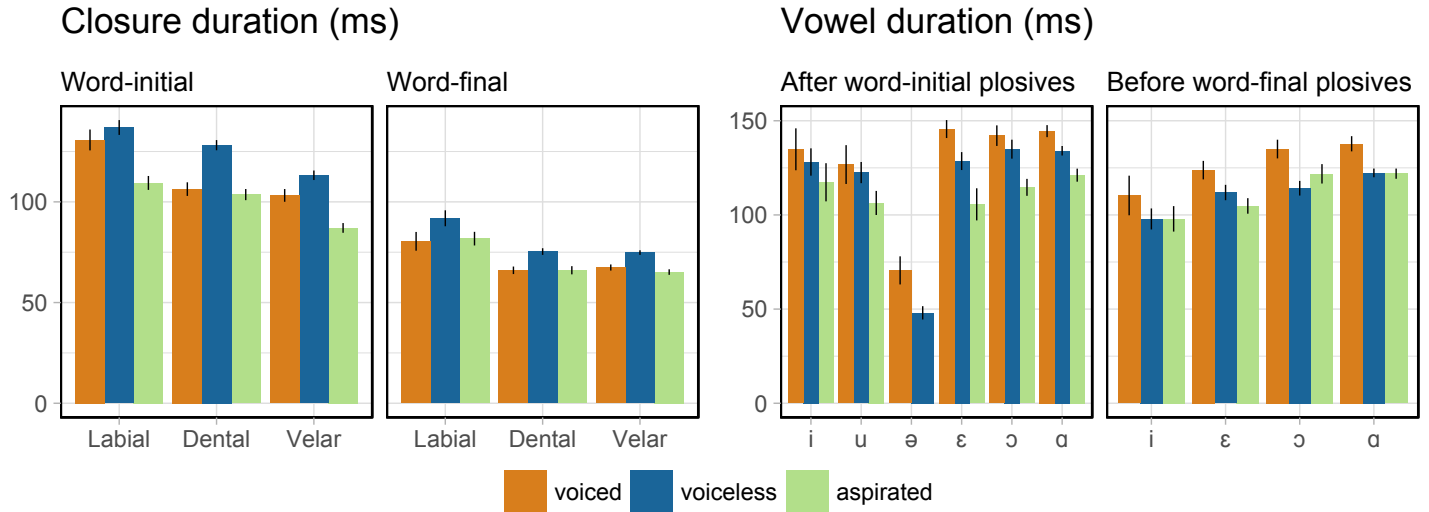
Figure 10: Mean closure and vowel durations, with closures divided by stop place and vowels divided by vowel quality. Lines show one standard error above and below the means.

4 Analysis of findings

4.1 Voice quality

4.1.1 Voice quality for voiced plosives

We found good evidence that the D^{fi} series is classically breathy-voiced in word-initial position, with relatively strong closure voicing as well as a spread-glottis configuration that likely begins during the closure and extends through the transition into the following vowel. In final position, there was little evidence for voice quality distinctions, at least as measured in the vowel preceding the closure. However, we also observed that there tended to be somewhat longer aspiration for D^{fi} in final position, intermediate between T and T^h. This indicates that D^{fi} may nevertheless involve glottal spreading in final position. If the glottal spreading gesture is timed to begin during the final closure rather than leading into it, it would be unlikely to manifest in the acoustics of the preceding vowel.

4.1.2 Voice quality for voiceless unaspirated plosives

In §3.1.1, we observed that two speakers had H1*–H2* and CPP values that were consistent with glottalization for the voiceless unaspirated series. For these speakers, H1*–H2* was lower for the T plosives compared to the other two series, but CPP was similar across all three series. If the T plosives were modal, CPP should be higher than the breathy D^{fi} or T^h series, as was the case for the other six speakers. This result lends limited support to previous claims that the voiceless unaspirated series in Yerevan Armenian may involve glottal constriction, though only for some speakers.

We did not find evidence that this series involved an ejective articulation. However, there is instead evidence that the T series involves tense voice, a specific subtype of creaky voicing
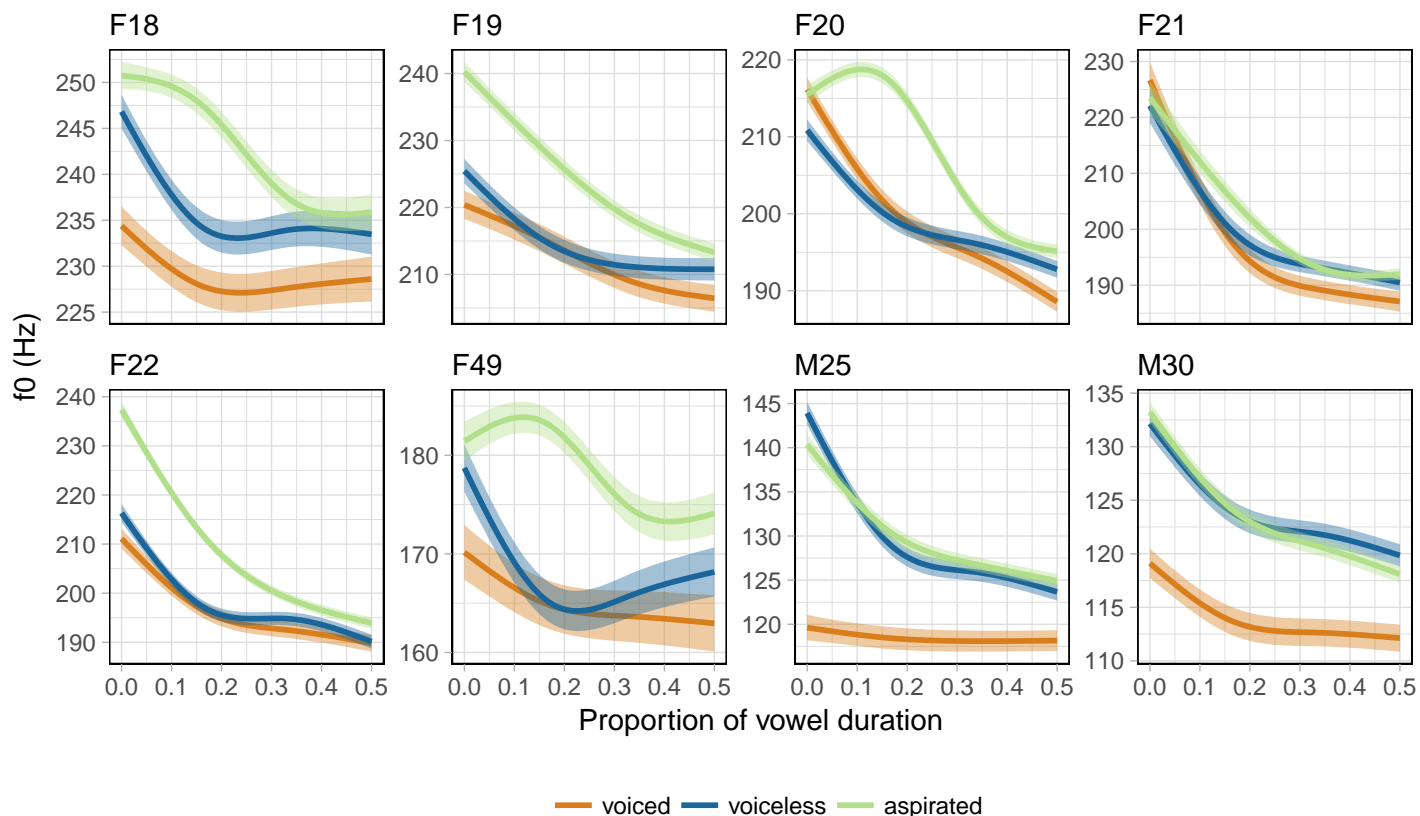
Figure 11: f0 (y-axis) over time during the vowel (x-axis) for the three voicing categories estimated by cubic spline regression for each speaker (see footnote 2 for model details). Lines show estimated means, and shaded areas show one standard error above and below the estimated means.

(see also discussion in Schirru, 2012). In one taxonomy of creaky voice qualities (Garellek, to appear; Keating et al., 2015), 'creaky' voice is a superset of several distinct articulations which are perceived as sharing a creaky quality, or that can be used to implement a phonological contrast. These articulations include prototypical creaky voice, which is low-pitched, constricted, and irregular; but also voice qualities that do not have all three characteristics. In particular, tense voice is characterized by its increased vocal fold constriction (like prototypical creaky voice) but also higher f0.

Figure 11 shows f0 tracks in vowels following word-initial plosives. Most speakers have relatively lower f0 following voiceless unaspirated plosives. However, the two speakers with evidence for glottalized plosives (M25 and M30) also produce that series with higher f0 near the vowel onset following word-initial plosives. Higher f0 onsets are more typically associated with glottal spreading. For instance, vowels following aspirated plosives typically have higher f0 than those following unaspirated plosives due to the aerodynamics associated with the greater airflow produced during aspiration (Hombert et al., 1979).[6] In contrast, voiceless unaspirated

---

[6]The breathy-voiced plosives, which we argue involve glottal spreading, instead have lower f0 in Figure 11. However, this is likely the consequence of aerodynamics or laryngeal adjustments associated with voicing and/or

plosives are expected to raise f0 on following vowels only if they are accompanied by increased vocal fold tension, which would stiffen the folds and cause them to vibrate faster (Hombert et al., 1979; Kirby & Ladd, 2016; Löfqvist, Baer, McGarr, & Story, 1989).

For six of the eight speakers, the voiceless unaspirated plosives have similar f0 onsets as the breathy-voiced ones. This implies that these are not produced with tense voice, but instead have modal voicing, which is supported by the H1*–H2* and CPP measurements. However, for the two speakers whose voiceless unaspirated plosives are followed by irregular creaky voicing (as characterized by lower values of both H1*–H2* and CPP), the voiceless unaspirated plosives are also followed by higher f0 onsets, similar to those found for the aspirated series (see Figure 11). The combination of irregular creaky voicing and higher f0 at vowel onsets lends support to the interpretation that, for these two speakers, the voiceless unaspirated series involve tense voice.

4.2 Discriminability of the voicing contrast

4.2.1 Distance between voicing categories along single acoustic dimensions

To evaluate how different the three voicing categories are with respect to each cue, we calculated the standardized distances (Cohen's d) between each pair of categories along each acoustic dimension, for each individual speaker. These distances are calculated for each acoustic variable by taking the absolute difference of the mean values between two categories, and then dividing it by the pooled standard deviation of that variable. For example, to calculate how well H1*–H2* separates voiced plosives from aspirated plosives for speaker F18, we took the absolute difference between the mean H1*–H2* for the voiced plosives and the mean for the aspirated plosives produced by speaker F18. This value was then divided by the pooled standard deviation of speaker F18's H1*–H2* values (i.e., calculated for each of the three categories separately, and then pooled together) to produce a standardized distance between voiced and aspirated plosives for H1*–H2*, as produced by F18. Because the distances are standardized in this way, they can be compared across different variables.

---

breathy phonation during a stop closure (Hombert et al., 1979; Honda, Hirai, Masaki, & Shimada, 1999, though see Kirby & Ladd, 2016), and breathy voice is often accompanied by lower pitch in languages with mixed tone-phonation systems (Brunelle, 2012; Brunelle & Kirby, 2016; Gordon & Ladefoged, 2001; Hombert et al., 1979).

Figure 12: Standardized distances (Cohen's d) between voicing categories on eight acoustic dimensions, for plosives in word-initial position. Distances are represented by the lengths of the connecting lines (on the x-axis) between each pair of dots; longer lines indicate larger standardized distances between two categories. Distances are unitless and can be compared across different panels.
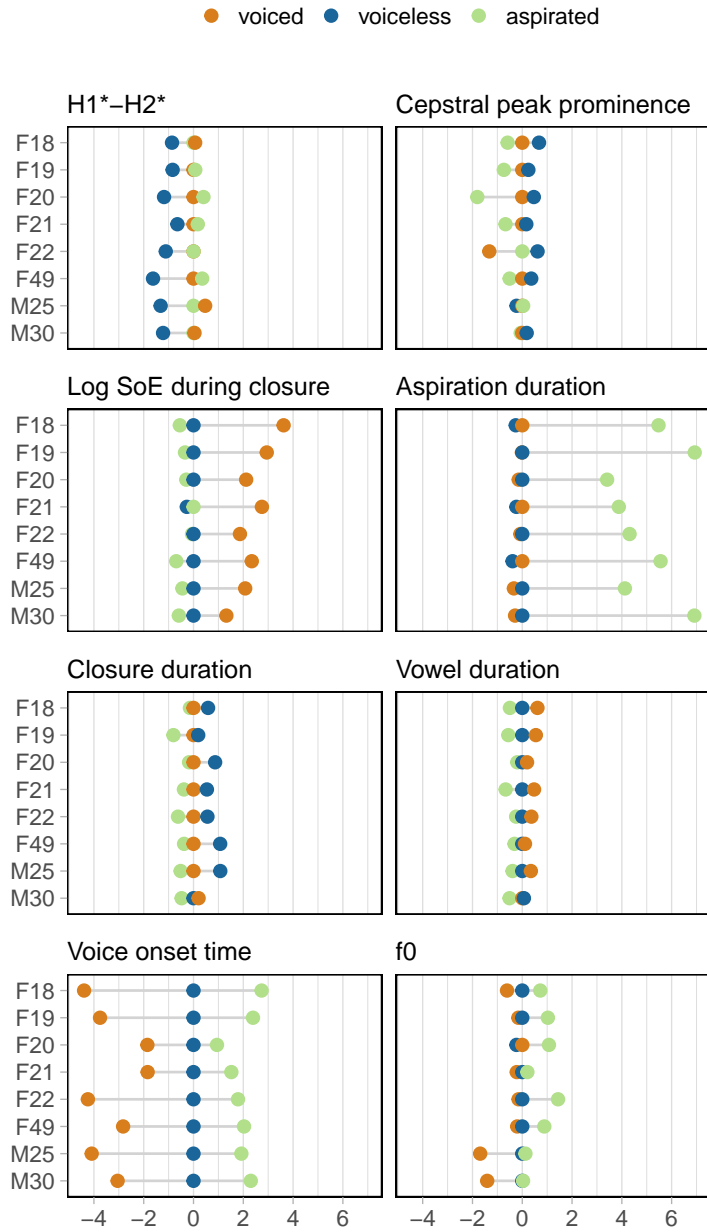
Figure 13: Standardized distances (Cohen's d) between voicing categories on eight acoustic dimensions, for plosives in word-final position. Distances are represented by the lengths of the connecting lines (on the x-axis) between each pair of dots; longer lines indicate larger standardized distances between two categories. Distances are unitless and can be compared across different panels.
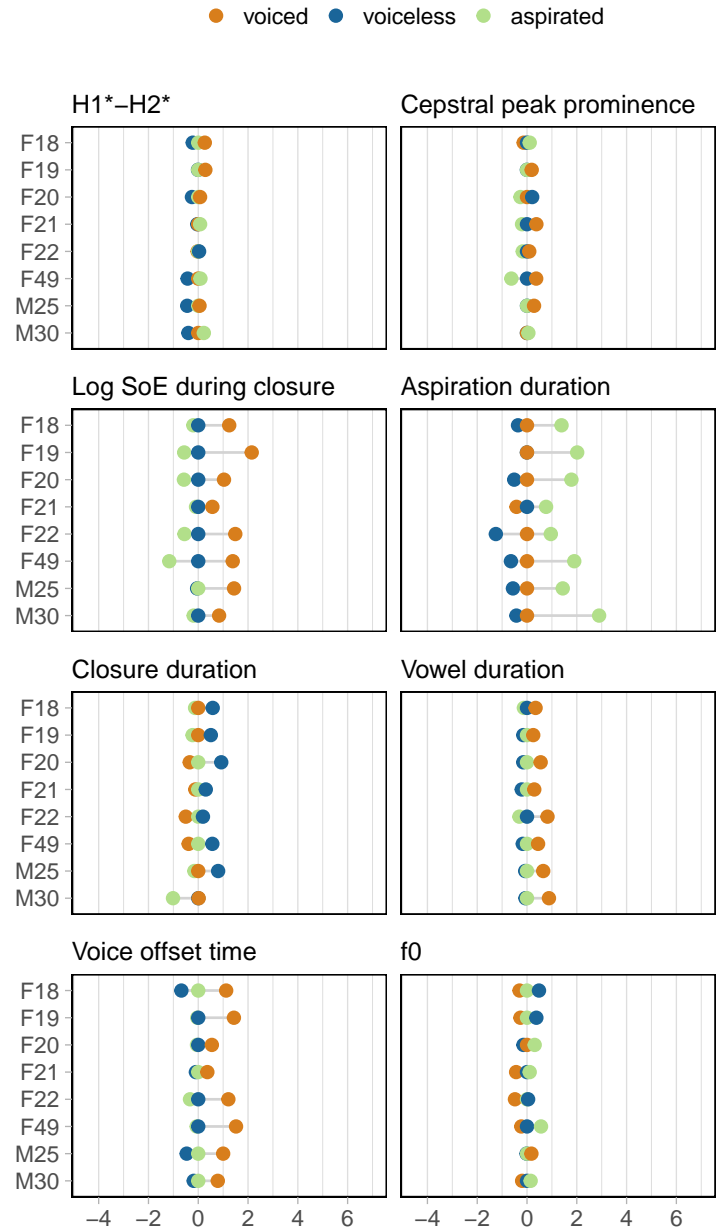
Figure 12 shows the distances between the three categories in word-initial position for each acoustic variable for each speaker. For example, for speaker F18, voice onset time (lower-left panel) separates the voiced and voiceless categories with $d = 4.4$, which is shown as the distance between the orange (left) and blue (center) dots in the first row of that panel. The distance between the blue (center) and green (right) dots shows that the distance between the voiceless and aspirated categories is somewhat smaller ($d = 2.7$) for this speaker. However, the voiced and aspirated categories are very well-separated on the dimension of voice onset time, as can be seen from the total distance between the orange and green dots ($d = 7.1$).

Across speakers, the lower-left panel shows that speaker F18 uses voice onset time to separate the three categories to the greatest extent, while speaker F20 uses voice onset time the least (see also Figure 7). Nevertheless, voice onset time provides good separation of the three voicing categories for all eight speakers. Since the distances in this figure are all on the same scale, the distances in the VOT panel can be compared directly to the distances for the other acoustic variables. Compared to VOT, vowel duration does not distinguish the categories especially well in prevocalic word-initial position (largest $d = 1.1$, for speaker F21 between voiced and aspirated); though we note that different vowels have different intrinsic vowel durations, which was not necessarily balanced across plosive voicing categories in our study (see Figure 10).

Does voice onset time provide the best separation between categories? In addition to voice onset time, aspiration duration and strength-of-excitation during the plosive closure together provide good separation between the categories. Across speakers, aspiration duration separates the aspirated plosives from each of the other two categories with average $d = 5.4$; and strength-of-excitation during closure separates the voiced plosives from each of the other two categories with average $d = 2.6$. By comparison, VOT separates the aspirated plosives from the others with average $d = 3.6$, and the voiced plosives from the others with average $d = 4.2$. Thus, VOT provides roughly the same separation overall as the combination of aspiration and strength-of-excitation.

There is also some variation between speakers. For example, speaker F20 does not use aspiration duration to distinguish aspirated plosives as strongly as the other speakers, but produces relatively much noisier aspirated plosives, as measured by CPP. As discussed in §4.1.2, pitch is used in different ways by different speakers. Five speakers (F18, F19, F20, F22, F49) have relatively higher pitch only for the aspirated plosives, while two speakers (M25, M30) have higher pitch for both voiceless series. For these two speakers M25 and M30, in fact, the voiced category is roughly as distinct from the other two categories on pitch as it is on strength-of-excitation. For the last speaker F21, the three categories have similar pitch.

Figure 13 shows the distances between categories for plosives in postvocalic word-final position. In this position, it can be seen that the three categories are not separated by voice quality, as measured by H1*–H2* and CPP (see also Figure 3), at least when measured during the vowel. The categories are overall much more similar in final position on most of the acoustic variables, except for vowel duration, which provides slightly more separation in final position. The voicing categories are still reasonably well separated by strength-of-excitation and aspiration duration. In addition, voice offset time separates the voiced plosives from the other categories almost as well as SoE.

Although the distances in Figure 13 are averaged across both carrier sentences—one with a following /p/, and one with following /b/—they are generally similar in both phrases. Speak-

ers F20, F21, and M30 do not voice the word-final voiced plosives as strongly before /p/ as before /b/ (see also Figure 9), and speaker M25 has more similar voice offset times for all three categories before /p/, but the voiced plosives are still overall distinct from the others in both contexts (see §4.2.2 below).

4.2.2 Multivariable discriminability

Although Figures 12–13 show that the three voicing categories are more separated along some acoustic dimensions than others, they do not show how well a plosive's voicing category can be identified on the basis of the overall acoustic contrast. This question is of particular interest for word-final postvocalic plosives, as it has been suggested that the voiced–voiceless contrast may be neutralized in this context (see §1.1). Further, they do not show which variables provide independent information about voicing category. For example, although Figures 3 and 12 show that initial aspirated plosives have longer aspiration, breathier following vowels, and higher pitch (for most speakers) than voiceless unaspirated ones, it is likely that these are caused by the same articulatory mechanism (cf. Gordon & Ladefoged, 2001; Hombert et al., 1979; Klatt & Klatt, 1990). If they are highly correlated, it may be the case that not all three variables provide unique information to a potential listener about a plosive's voicing category. To help answer questions about the robustness of the contrast and the unique contribution of each variable to discriminability, we fit a series of multivariable classification models.

Model procedure In the following sections, multinomial logistic regressions are fit to predict voicing category (voiced, voiceless unaspirated, or voiceless aspirated) as a function of the combined set of acoustic predictors. The models were fit by maximum-likelihood using the nnet R package (R Core Team, 2017; Venables & Ripley, 2001). A multinomial logistic regression with a three-way categorical outcome can be written as two binomial logistic regression equations. Each of the two equations models the relative log-odds of the reference category compared to one of the other categories, using a set of predictor variables. Here, the reference category is voiceless unaspirated, and one equation compares it to the voiced series, and the other equation compares it to the aspirated series. To predict the most likely voicing category for a new observation, the relative log-odds for the two comparisons (voiceless versus voiced; and voiceless versus aspirated) are first calculated. Then, these two odds ratios are converted into absolute probabilities for the three voicing categories which sum to one.

The predictors in the following regression models were H1*–H2*, CPP, log-transformed SoE, f0, aspiration duration, closure duration, voice onset/offset time (as appropriate), vowel duration, plus all interactions with word position. The interactions with word position mean that the model can effectively fit different parameter coefficients for initial and final position. The H1*–H2*, CPP, SoE, and f0 measurements used in the model were the summary values averaged over the first third of the nucleus vowel for initial plosives, and over the final third for final plosives (see §2.5). Because the measures vary between speakers, all measures were centered within-speaker so that the mean value of each variable was zero for each speaker.

Classification accuracy The full dataset used for the following models included 1527 of 1600 tokens overall (see §2.6), including 426 voiced, 658 voiceless, and 443 aspirated tokens. The ability of the model to classify each plosive token in our data as voiced, voiceless, or aspirated

was evaluated using the following procedure. For each token, we predicted its voicing category using a model fit to a reduced dataset. The token to be classified was first held out from the full dataset. This was done so that the model which would be used to predict that token's voicing was not fit to that token's own acoustic measurements. Then, a reduced dataset was sampled so that it had an equal number of voiced, voiceless, and aspirated tokens (i.e., 426 tokens each, or 425 each if the held-out token was voiced). This was done to ensure that the higher proportion of voiceless tokens in the full dataset did not result in misleadingly higher accuracy in classifying voiceless tokens. A multinomial regression was then fit to the reduced dataset, and used to predict the most likely voicing category of the held-out token, following the procedure described above. This process was repeated to generate a prediction for each token in the full dataset.

We then calculated the proportion of predictions that were correct, across the full dataset. The model was generally accurate at discriminating the voicing contrast in both initial and final position, with an overall accuracy of 86%. Table 3 shows the percentage of correct model predictions for each category. Performance was lowest for the voiced series, which was frequently confused with voiceless unaspirated tokens in final position before /p/, though these were still categorized correctly a majority of the time. Before /b/, voiceless unaspirated tokens were most often miscategorized as voiced. No other kinds of errors were especially common. For four words with final plosives (ҍηҍq /jɛʁɛg/, ᕍһq /t͡ʃig/, ᕍһᕄ /t͡ʃik/, ᕍһᴜn /t͡ʃit/) and one word with an initial plosive (ᕄһᴘᴘ /kiɾkʰ/), categorization accuracy was 50% or below, but the model categorized at least half of the tokens correctly for every other word type.

Table 3: Percentage of plosives that were correctly categorized, with marginal averages.

	D^fi	T	T^h	
Word-initial	90%	93	97	93
Word-final before /p/	61	76	85	75
Word-final before /b/	80	73	84	78
	82	85	91	86

Predictor evaluation   To evaluate which acoustic variables improved the model, we fit a model with the same set of predictors to the full dataset. We then fit a series of reduced models using the same procedure, which each omitted one predictor (including its interaction with word position). Next, we calculated the Bayesian information criterion (BIC) for the full model, and for each reduced model. BIC is a measure of the inverse likelihood of the data under the model, with a penalty for the number of parameters in the model. A lower BIC indicates a better model. If a reduced model which omits a predictor has a lower BIC than a model which contains that predictor, that suggests that that predictor does not improve likelihood enough to justify its inclusion in a parsimonious model.[7] As the model does not involve any perception

---

[7]There is no significance test for BIC, though see Wagenmakers (2007). However, testing for lowered BIC is generally much more conservative than a likelihood ratio test with $\alpha = 0.05$. Appendix C shows BIC values as well as likelihood ratio test statistics. A significant likelihood ratio test indicates that a model is significantly improved by the inclusion of a predictor.

data, the results should not be interpreted as indications about which acoustic cues are used by listeners (though see McMurray & Jongman, 2011; Toscano & McMurray, 2010).

Of the eight predictors in the multinomial regression, BIC was lowered only when either CPP or f0 were omitted, suggesting that only these two did not provide a unique contribution to the model of voicing category. The omission of f0 and CPP can be explained by the between-speaker differences in these variables in word-initial position, which we have argued is due to qualitatively different voice-quality patterns between speakers (see §3.1.1 and §4.1.2). Note that this does not necessarily mean that CPP and f0 are not useful cues to listeners who encounter such variation, only that they would need to accommodate it in their mental model of Armenian stop voicing, which we did not attempt to do in this analysis (see Kleinschmidt & Jaeger, 2015, for a review). A full table of BIC values is provided in Appendix C. The predictors which lowered BIC by the most, and thus improved the model the most, were (in order): aspiration duration, SoE during the closure, voice onset/offset time, $H1^*$–$H2^*$, vowel duration, and closure duration.

Next, we evaluated whether it was necessary to fit different parameter values for each acoustic predictor in word-initial compared to word-final position. To do so, we fit a series of reduced models which each omitted the interaction parameter between only one acoustic predictor and word position, and compared each of them with the full model by BIC. If a model had higher BIC when it omits the interaction between a given predictor and word position, that suggests that the values of that variable are associated with different voicing categories in initial versus final position.

Of the eight interaction parameters, BIC was higher when the interactions between word position with $H1^*$–$H2^*$ or with voice onset/offset time were omitted. This suggests that of all of the predictors, only $H1^*$–$H2^*$ and voice onset/offset time have different relationships with voicing category in initial versus final position. For voice onset/offset time, this is certainly because the voiced plosives typically have negative voice onset time in initial position but greater voice offset time in final position. For $H1^*$–$H2^*$, we observed in §3.1.1 (e.g., Figure 4) that none of the speakers seemed to have differences in glottal constriction leading into the three series of word-final plosives, although they did so following word-initial plosives. A full table of BIC values is provided in Appendix C. It is noteworthy that closure SoE is both the second most useful predictor in the model overall, and that its association with plosive category varies the least between positions. This points to the robustness of acoustic voicing as a distinguishing cue in both word-initial and word-final positions.

5   General discussion

We have argued that the Yerevan Armenian voiced plosives have breathy voicing which begins with the closure and extends beyond the release. In initial position, this can be measured acoustically from an index of glottal spreading in the following vowel; in final position, this manifests as short post-aspiration that is typically voiceless. The voiceless unaspirated plosives were modal for most speakers, but likely tense (a subtype of creaky) for at least two speakers, as measured through relatively increased noise (lower CPP) and raised f0. Although voice quality is difficult to perceive or to measure directly during an oral stop closure—and impossible dur-

ing a voiceless one—future instrumental work might explore the timing and degree of glottal constriction via articulatory glottography during the closure interval.

The discriminability analyses showed that the three stop categories can be identified at a rate usefully above chance levels in both prevocalic word-initial and postvocalic word-final position. Many different variables contribute to identification, but the three-way contrast was also well separated in both positions by a combination of only voicing strength and aspiration duration. VOT provides a similar degree of separation in initial position, and voice quality and f0 of the following vowel are also particularly distinctive in this position.

How do the breathy-voiced Armenian stops fit within the phonological typology and history of the broader language family? Their similarity to the breathy-voiced or voiced-aspirated stops found in some related Indic languages is unclear (Garrett, 1998; Khachaturian, 1984; Pisowicz, 1998; Vaux, 1997), and it is disputed whether breathy-voicing was present in earlier Armenian—and thus inherited from Indo-European—or whether it represents a recent dialectal innovation (see e.g. Baronian, 2017; Garrett, 1998; Kortlandt, 1985; Vaux, 1998b). Garrett (1998) has proposed that the acoustic effect of breathy voicing on adjacent vowels might have been a phonetic precursor for a vocalic sound change in earlier Armenian, which would provide evidence for early breathy-voicing. Here, we discuss how breathy-voiced plosives in Yerevan Armenian differ from those in Gujarati and perhaps some other Indic languages, and we present acoustic evidence which supports the proposal that breathy voicing plausibly conditioned this historical sound change.

5.1   Comparison with Indic breathy-voiced plosives

The voiced plosives in the Group 1–2 Armenian dialects have been compared to the voiced aspirated or breathy-voiced plosives that occur in many Indic languages (Garrett, 1998; Khachaturian, 1984; Pisowicz, 1998; Schirru, 2012; Vaux, 1997). In terms of acoustics, H1*–H2* after prevocalic breathy-voiced plosives follows a similar pattern in Yerevan Armenian as in three Indic languages (compare Figure 4 of this paper with Marathi in Figure 31 of Berkson, 2013; Gujarati in Figure 1 of Esposito & Khan, 2012; and Hindi in Figure 6.1 of Dutta, 2007). In each language, H1*–H2* is at its maximum during the first third of the vowel, and then descends over time through the vowel midpoint, indicating a spread glottis that is moving towards a relatively more modal configuration. By the vowel midpoint, H1*–H2* is still slightly higher after breathy/aspirated plosives than after the other plosives in all four languages. However, at the vowel midpoint, the numerical difference between the voicing categories is smaller in Yerevan Armenian (about 2 dB; see Figure 4) than in the other languages (about 4-5 dB; see previous references). Although this may be affected by recording context and inter-speaker differences, it may also indicate that the glottal spreading gesture in Armenian breathy-voiced plosives is less extreme, shorter, or begins earlier than in the homologous Indic sounds.

The Indic breathy-voiced plosives are typically described as having a longer portion of very noisy voiced aspiration after the release of the closure (Ladefoged & Maddieson, 1996, p. 58; Berkson, 2012). Although we found that prevocalic Yerevan Armenian voiced plosives have a relatively spread glottis after the closure, they appear to be less noisy, with well-defined formant structure visible immediately after the closure. This can be seen in the upper panels of Figure 2: although there is a short noisy interval after the word-initial voiced plosive closure, it is about the same duration as the release burst after the voiceless unaspirated one. We can

compare these Yerevan Armenian breathy-voiced plosives with their counterparts in the Indic language Gujarati based on the recordings in the freely-available Production and Perception of Linguistic Voice Quality project at UCLA.[8] This database has recordings from Gujarati (Esposito & Khan, 2012; Khan, 2012) and several other languages, including three Gujarati words with word-initial /bɦ, ɖɦ/ that were repeated several times each by nine speakers, with 131 tokens of the relevant Gujarati plosives in total.
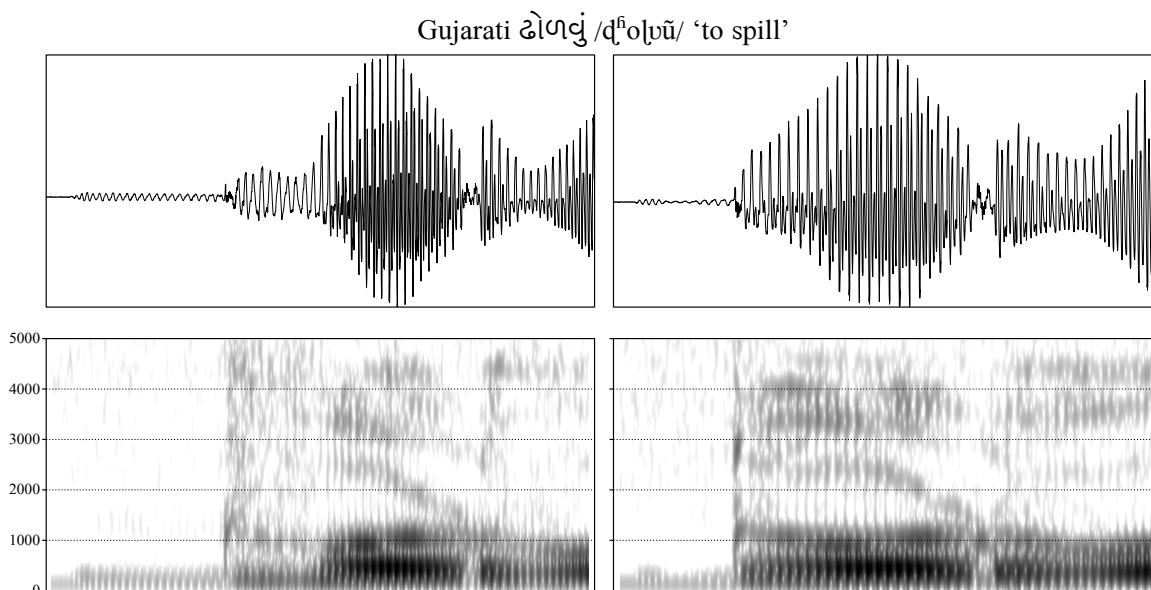
Gujarati ઢોળવું /ɖɦoɭʋ̃ũ/ 'to spill'

Figure 14: Waveforms and spectrograms for two tokens of the Gujarati word /ɖɦoɭʋ̃ũ/ produced by the same speaker.

The Gujarati plosives /bɦ, ɖɦ/ involved a wide range of acoustic variation during the release phase. Figure 14 shows two tokens of the Gujarati word ઢોળવું /ɖɦoɭʋ̃ũ/ 'to spill' produced by the same speaker. In the left panel, the initial stop closure is followed by a long interval of very noisy frication (about 80 milliseconds). This interval is voiced, as can be seen by the periodicity in the waveform and the voice bar in the spectrogram, but the formants are poorly defined or missing (see discussion in Berkson, 2012; Davis, 1994; Mikuteit & Reetz, 2007). We would therefore characterize this interval as voiced aspiration, rather than as part of a breathy-voiced vowel. This interval is followed by a vowel with the expected formant structure. In contrast, the initial stop closure in the right panel is almost immediately followed by a strongly-voiced vowel, with no such interval of voiced aspiration. Nearly all of the Yerevan Armenian plosives were similar to this token, as in Figure 2, in that the release was closely followed by a vowel with well-defined formants that typically began immediately and almost always within 30 milliseconds.[9] The acoustic analysis of Armenian showed that these vowels are relatively noisy; however, because of their clear formant structure, we would characterize these kinds of

---

[8]Available online at http://www.phonetics.ucla.edu/voiceproject/voice.html.

[9]As we noted in §3.1.1, many tokens of Yerevan Armenian /g/ were noisier for a longer duration, but we attribute this noise to velar spirantization rather than aspiration.

tokens in Gujarati and in Armenian as being breathy-voiced during the closure and part of the following vowel, rather than as having post-closure voiced aspiration.

Although a subset of the Gujarati breathy-voiced plosives are thus similar to the Yerevan Armenian ones, it can be seen that such sounds are not a homogeneous category in terms of laryngeal timing, even within a language. While the transcriptions /dʱ/ and /d̤/ have both been used for Armenian breathy-voiced plosives (dʱ: Gamkrelidze & Ivanov, 1995; Vaux, 1998a; d̤: Macak, 2017, p. 1048; and also other symbols in Allen, 1950; Kortlandt, 1998; Pisowicz, 1997, 1998), in principle they differ in that [dʱ] is a more appropriate label for a Gujarati token like the one in the left panel of Figure 14, which has a voiced-aspirated release with a weaker vowel formant structure (i.e., [ɦ]). On the other hand, [d̤] is more appropriate for a token such as in the right panel of Figure 14 in Gujarati or the upper-left Armenian plosive in Figure 2, which probably has breathy voicing during the closure followed immediately by a vowel with very clear formant structure.

We observed that all eight Yerevan Armenian speakers produced breathy-voiced plosives that were roughly similar to the one in Figure 2, with none that had an extended interval of voiced aspiration like the Gujarati token in the left panel of Figure 14. On the other hand, the nine Gujarati speakers showed substantial variation both within- and between-speakers in terms of whether a given breathy-voiced plosive was more similar to the left or right panel in Figure 14. Gujarati, then, differs from Yerevan Armenian in that its breathy-voiced plosives can be implemented as a wider range of phonetic values, including [dʱ, d̤] and in some cases [tʰ], while Yerevan Armenian permits only a narrower set of possibilities, comprising mainly [d̤]. For future work, /b̤, d̤, g̈/ might thus be appropriate transcriptions for the Armenian breathy-voiced plosives (as in Macak, 2017, though this may vary by dialect; see Khachaturian, 1984, p. 61).

From an articulatory perspective, this could be implemented as differences in the magnitude or in the phasing of glottal spreading. For example, Gujarati might differ from Yerevan Armenian in that Gujarati speakers allow more variation in glottal width during the production of breathy-voiced sounds. Tokens produced with greater glottal width might have an interval of aspiration because of the relatively longer transition needed to reach the modal glottal state (cf. Kagaya, 1974, pp. 172–173). Alternatively, the difference might be in phasing: Yerevan Armenian breathy-voiced plosives have the glottal spreading gesture aligned in phase with the closure, resulting in an acoustically breathy plosive without an interval of voiced aspiration, whereas their Gujarati homologues might have the glottal spreading gesture variably aligned in or out of phase with the closure (see Kagaya & Hirose, 1975, on out-of-phase alignment in Hindi), allowing an interval of voiced aspiration. Future work could quantify the range of variation allowed in voiced aspirated plosives within different languages, and to evaluate whether this variation can be ascribed to the magnitude or time-alignment of glottal spreading gestures.

5.2 Reconstruction of voiced plosives in Armenian and Indo-European

5.2.1 Challenges for reconstructing plain-voiced plosives in early Armenian

The phonetic qualities of Armenian voiced stops are significant for the reconstruction of the historical Armenian languages, and of Indo-European more generally. Classical Armenian is

usually reconstructed with plain voiced stops, based on the geography of attested plain and breathy-voiced stops in the modern Armenian languages (Vaux, 1998a, Ch. 7; Kortlandt, 1985; though see Garrett, 1998). However, the corresponding stops in the standard reconstruction of Proto-Indo-European are breathy-voiced stops. If Classical Armenian had plain voiced stops, they would have had to become plain in the earlier Proto-Armenian, and then later revert to being breathy again in only the Group 1–2 Armenian dialects (which include Yerevan Armenian). This change is shown below, using the dental plosive series as an example.

Proto-IE		Proto-Armenian		Mod. Arm. (Groups 1–2)
$*d^{ɦ}$	>	$*d$	>	$d^{ɦ}$
$*d$	>	$*t$	—	t
$*t^{h}$	—	$*t^{h}$	—	t^{h}

The loss and subsequent restoration of breathy-voice has been seen as implausible (Garrett, 1998). One proposal which avoids this scenario involves a version of the glottalic theory of Proto-Indo-European (see Hopper, 1973), in which Indo-European originally had plain voiced stops and glottalized voiceless ones. Under this theory (as discussed in Garrett, 1998), plain voiced stops were preserved from Proto-Indo-European in Proto-Armenian and later Classical Armenian. Breathy-voice is then a later innovation in the Group 1–2 Armenian dialects. The revised correspondences under this proposal are shown below.

Proto-IE		Proto-Armenian		Mod. Arm. (Groups 1–2)
$*d$	—	$*d$	>	$d^{ɦ}$
$*t'$	>	$*t$	—	t
$*t^{h}$	—	$*t^{h}$	—	t^{h}

By reconstructing Proto-Indo-European as having plain voiced stops, the glottalic theory arguably avoids the loss and restoration of breathy voice in Armenian (see Garrett, 1998), as well as avoiding a reconstruction in which the Armenian and Germanic language families underwent parallel consonant shifts after they diverged (Hopper, 1973). However, the glottalic theory has also reconstructed the Proto-Indo-European voiced stops as murmured (Hopper, 1973) or allophonically breathy (Gamkrelidze & Ivanov, 1995, pp. 37–38; Gamkrelidze, 2010), and thus does not clearly reconcile the need to reconstruct a temporary loss of breathy voice in Armenian.

5.2.2 Evidence for breathy-voiced plosives in early Armenian

An alternative analysis is that Classical Armenian had breathy-voiced stops preserved from Proto-Indo-European, which were lost in most modern varieties (Benveniste, 1958; Garrett, 1998; Macak, 2017; see also Gamkrelidze & Ivanov, 1995, pp. 37–38 and Baronian, 2017, p. 15). This analysis avoids the early loss and subsequent restoration of these stops, it is compatible with both the standard and glottalic reconstructions of Proto-Indo-European, and it eliminates the need to reconstruct parallel consonant shifts in Armenian and Germanic, even within the standard reconstruction (Garrett, 1998). This analysis is shown below, assuming the standard reconstruction of Proto-Indo-European.

Proto-IE		Proto-Armenian		Mod. Arm. (Groups 1–2)
*dʱ	—	*dʱ	—	dʱ
*d	>	*t	—	t
*tʰ	—	*tʰ	—	tʰ

This revised reconstruction of breathy-voiced stops in Proto-Armenian is supported by the distribution of Adjarian's Law (Garrett, 1998). Adjarian's Law is a sound change which occurred during the transition between Classical and some modern Armenian dialects, but not including the Group 1–2 dialects (Adjarian, 1901, cited in Vaux, 1998a; Vaux, 1998a, Ch. 7; and see Vaux, 1992). It describes the fronting of /ɑ/ and in some dialects /ɔ, u/ following what were voiced stops in Classical Armenian. Fronting also occurred following post-Classical /ɦ/ (Weitenberg, 1986, cited in Garrett, 1998), as well as perhaps other voiced consonants, although its application following other consonants is less definitively attested (see Vaux, 1992; Garrett, 1998, footnote 5).

To explain the conditioning environments for Adjarian's Law, Garrett (1998) argues that the most plausible phonetic precursor would have been the voice quality of breathy-voiced stops and /ɦ/. Breathy voice quality is claimed to be associated with higher F2 (i.e., vowel fronting). This would explain why the Group 1–2 dialects, which preserve breathy-voiced stops, did not undergo Adjarian's Law (Garrett, 1998): in these dialects, the glottal spreading associated with breathy voice was retained rather than converted into vowel fronting (see Kirby, 2010; Ohala, 1993). Moreover, the reconstruction of breathy-voiced stops in early Armenian would crucially simplify the analysis of consonant shifts in Armenian (Garrett, 1998).

At the same time, it has not been shown that breathiness is actually associated with vowel fronting, especially for the vowels /ɑ, ɔ, u/ which underwent Adjarian's Law. Garrett (1998) observes that English vowels tend to be more fronted after /h/ in a report by Lehiste (1964, p. 148; as reproduced in Garrett, 1998, p. 17). Those data, however, show fronting most strongly in the front vowels—to which Adjarian's Law did not apply—and /ɔ, u/ actually show backing after /h/ that is numerically greater than the fronting of /ɑ/. Moreover, voiceless glottal spreading in /h/ is likely to have different phonetic consequences than glottal spreading which occurs with voicing (e.g., see footnote 6), although Garrett (1998) does point out that English /h/ is often realized as voiced /ɦ/. Besides English, Kuang (2011) reports that the lax (slightly breathy) vowels in Yi are fronted relative to their tense counterparts, but only for the mid vowels.

The data from the speakers in this study, however, provide clear evidence that Armenian breathiness is associated with vowel fronting in Yerevan Armenian, and even that this is limited to the vowels to which Adjarian's Law applied in other Armenian dialects. Figure 15 shows the average F1 and F2 values in Yerevan Armenian vowels following word-initial plosives of all three voicing categories. For all eight speakers, /ɑ/ is fronted when it occurs after breathy-voiced plosives relative to when it occurs after voiceless unaspirated plosives. The vowels /ɔ, u/ also tend to be fronted, though to a lesser extent, with two exceptions (of 16) in the overall means (/u/ for speakers F18 and F21). There is no general pattern of F2 raising for the front vowels, and Adjarian's Law did not apply to the front vowels.

This phonetic evidence provides strong support for Garrett (1998)'s proposal that breathiness may have conditioned Adjarian's Law. One alternative proposal is that voicing, rather than breathiness, was the conditioning environment for Adjarian's Law (Vaux, 1992). How-
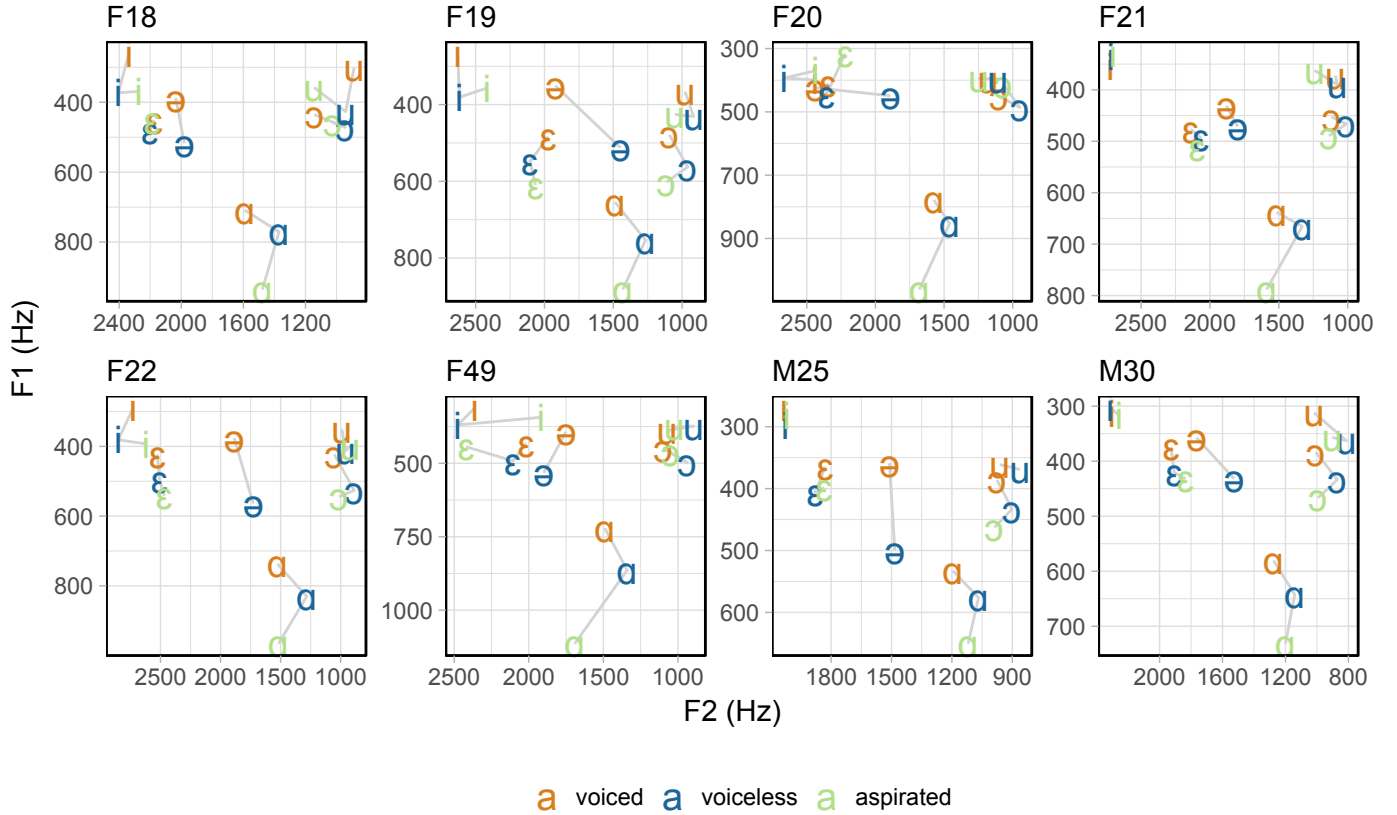
Figure 15: Mean formant frequencies vowels following word-initial plosives, for each speaker, vowel, and voicing category. Measurements were calculated by averaging over the first third of the following vowel, as described in §2.5. Note that there were only four words with /ə/, so its mean values are more variable across speakers than the other vowels.

ever, Figure 15 shows that back vowels after the aspirated plosives also tend to be fronted to a similar extent relative to the voiceless unaspirated ones. This further supports the claim that it was breathiness, and not voicing, which conditioned Adjarian's Law. If so, this would imply that the early Armenian stops were in fact originally breathy-voiced: since Adjarian's Law did not occur in dialects with modern breathy voice, it must have occurred prior (or simultaneous) to the loss of breathy voice in the other dialects.

5.3   Summary and conclusions

We found that the Yerevan Armenian voiced plosives are breathy-voiced, although they differ from at least Gujarati and perhaps other Indic languages in that they are qualitatively less variable and do not have an extended interval of noisy voiced aspiration. Nevertheless, in word-initial prevocalic position, the following vowel is reliably breathy adjacent to the stop closure. Word-initial voiced plosives were also associated with measurable fronting of the back vowels, among other acoustic cues, which supports the proposal that breathiness may have conditioned a historical back-vowel fronting process in other Armenian dialects. In word-final postvocalic

position, the preceding vowel is longer before voiced plosives, though not measurably breathy. The voiced plosives in this position are characterized by phonation during closure and by aspiration that is relatively short, variable, and usually voiceless, which is consistent with glottal spreading that begins during the closure. There was variation between speakers for the voiceless unaspirated plosives, such that two of the eight had acoustic measurements consistent with tense voice (but not ejectives) in word-initial position, while the other six showed no acoustic evidence of glottal constriction. In both syllable positions, these plosives were more consistently voiceless, unaspirated, and had slightly longer closures than the other two series.

The three-way voicing contrast could be reliably discriminated far above chance in both syllable positions on the basis of a set of acoustic variables, although discriminability was somewhat worse in word-final position, especially when the following sound did not facilitate phonation for the voiced series. While the three-way contrast was well separated by a combination of voicing strength and aspiration duration (or by VOT in initial position; though VOT could not always be straightforwardly measured), nearly all of the acoustic predictors improved the fit of a classification model, pointing towards the need for a multidimensional understanding of the contrast. Future work might examine especially the articulatory basis for the qualitatively different acoustic patterns in spread-glottis plosives across languages, and in particular their effects on the vowel formants which may have conditioned sound change in Indo-European.

Acknowledgments

References

Abramson, A. S., Tiede, M. K., & Luangthongkum, T. (2015). Voice register in Mon: acoustics and electroglottography. Phonetica, 72, 237–256.

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. doi: 10.1016/j.wocn.2017.05.002

Adjarian, H. (1899). Les explosives de l'ancien arménien: étudies dans les dialectes modernes. La Parole. Revue internationale de rhinologie, otologie et laryngologie, 119–127.

Adjarian, H. (1901). Lautlehre des Van-Dialekts. In Zeitschrift für Armenische Philologie 1 (pp. 74–87, 121–138). Marburg.

Allen, W. (1950). Notes on the phonetics of an Eastern Armenian speaker. Transactions of the Philological Society, 180–206.

Baronian, L. (2017). Two problems in Armenian phonology. Language and Linguistics Compass, 11(8), e12247. doi: 10.1111/lnc3.12247

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. doi: 10.18637/jss.v067.i01

Benveniste, É. (1958). Sur la phonétique et la syntaxe de l'arménien classique. Bulletin de la Société de Linguistique de Paris, 54, 46-68.

Berkson, K. H. (2012). Capturing Breathy Voice: Durational Measures of Oral Stops in Marathi. Kansas Working Papers in Linguistics. doi: 10.17161/KWPL.1808.10495

Berkson, K. H. (2013). Phonation types in Marathi: An acoustic investigation (Unpublished doctoral dissertation). University of Kansas.

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. MIT Speech Communication Working Papers, 1, 71-81.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. Journal of Phonetics, 30, 163-191.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.28. Retrieved from http://www.praat.org

Braun, A. (2013). An early case of "VOT". In Proceedings of interspeech (pp. 119–122).

Brunelle, M. (2012). Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. Journal of the Acoustical Society of America, 131, 3088–3102.

Brunelle, M., & Kirby, J. (2016). Tone and phonation in Southeast Asian languages. Language and Linguistics Compass, 10, 191–207.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. Phonetica, 22, 129–159.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. Journal of Phonetics, 30, 193-228.

Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. The Journal of the Acoustical Society of America, 136(5), 2762–2772. doi: 10.1121/1.4896470

Cooper, A. M. (1991). An articulatory account of aspiration in English (Unpublished doctoral dissertation). Yale University.

Corpus Technologies. (2009). EANC: Eastern Armenian National Corpus. Retrieved from http://www.eanc.net

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. Journal of Phonetics, 54, 35–50. doi: 10.1016/j.wocn.2015.09.003

Davidson, L. (2017). Phonation and laryngeal specification in American English voiceless obstruents. Journal of the International Phonetic Association.

Davis, K. (1994). Stop voicing in Hindi. Journal of Phonetics, 22(2), 177-194.

Decours, A., Ouzounian, A., Riccioli, T., & Vidal-Gorene, C. (Eds.). (2014). Project Calfa.fr, dictionary of Classical Armenian [online]. Paris. Retrieved from http://calfa.fr

DiCanio, C. (2009). The phonetics of register in Takhian Thong Chong. Journal of the International Phonetic Association, 39, 162-188.

DiCanio, C. (2014). Cue weight in the perception of Trique glottal consonants. Journal of the Acoustical Society of America, 119, 3059–3071.

Dum-Tragut, J. (2009). Modern Eastern Armenian (No. 14). John Benjamins Publishing Company.

Dutta, I. (2007). Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency, and spectral tilt (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. Phonology, 23, 157-191.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong phonation. Journal of Phonetics, 40, 466–476.

Esposito, C. M., & Khan, S. (2012). Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong. Journal of the International Phonetic Association, 42, 123-143.

Fairbanks, G. H., & Stevick, E. W. (1958). Spoken East Armenian. American Council of Learned Societies.

Fleming, H. (2000). Glottalization in Eastern Armenian. Journal of Indo-European Studies, 28(1-2), 155–196.

Gallagher, G. (2015). Natural classes in cooccurrence constraints. Lingua, 166, 80-98.

Gamkrelidze, T. V. (2010). In Defense of Ejectives for Proto-Indo-European (A Response to the Critique of the "Glottalic Theory"). Bulletin of the Georgian National Academy of Sciences, 4(1), 168–178.

Gamkrelidze, T. V., & Ivanov, V. V. (1995). Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture (W. Winter, Ed. & J. Nichols, Trans.). Berlin, New York: Mouton de Gruyter. doi: 10.1515/9783110815030

Garellek, M. (2010). The acoustics of coarticulated non-modal phonation. UCLA Working Papers in Phonetics, 108, 66-112.

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. Journal of Phonetics, 40, 152-161.

Garellek, M. (to appear). The phonetics of voice. In W. Katz & P. Assmann (Eds.), Routledge handbook of phonetics. Routledge.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. Journal of the International Phonetic Association, 41, 185-205.

Garellek, M., & Seyfarth, S. (2016). Acoustic differences between English /t/ glottalization and phrasal creak. In Proceedings of interspeech 2016 (p. 1054-1058). San Francisco.

Garellek, M., & White, J. (2015). Phonetics of Tongan stress. Journal of the International Phonetic Association, 45, 13-34.

Garrett, A. (1998). Adjarian's law, the glottalic theory, and the position of Armenian. In Proceedings of the 24th Annual Meeting of the Berkeley Linguistics Society: Special Session on Indo-European Subgrouping and Internal Relations (pp. 12–23). Berkeley, CA.

Gharibian, A. (1969). A propos de la première mutation des consonnes occlusives dans l'arménien. In Studia classica et orientalia Antonino Pagliaro oblata (Vol. 2, pp. 161–166). Roma.

Gordeeva, O. B., & Scobbie, J. M. (2013). A phonetically versatile contrast: Pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality. Journal of the International Phonetic Association, 43, 249-271.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. Journal of Phonetics, 29, 383-406.

Hacopian, N. (2003). A three-way VOT contrast in final position: data from Armenian. Journal of the International Phonetic Association, 33(1), 51–79.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. Journal of the Acoustical Society of America, 101, 466-481.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. The Journal of the Acoustical Society of America, 125(1), 425–441. doi: 10.1121/1.3021306

Henton, C., Ladefoged, P., & Maddieson, I. (1992). Stops in the world's languages. Phonetica, 49, 65–101.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy voice quality. Journal of Speech and Hearing Research, 37, 769-778.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. Journal of Speech and Hearing Research, 38, 1212-1223.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. Language, 55, 37-58.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. Language and Speech, 42, 401-411.

Hopper, P. (1973). Glottalized and murmured occlusives in Indo-European. Glossa, 7(2), 140-166.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. Journal of the Acoustical Society of America, 121, 2283–2295.

Jahukyan, G. (1972). Hay barbaṙagitut'yan neratsut'yun: (vichakagrakan barbaṙagitut'yun). Yerevan: Haykakan SSH GA Hratarakch'ut'yun.

Job, D. M. (1977). Probleme eines typologischen Vergleichs iberokaukasischer und indogermanischer Phonemsysteme in Kaukasus. Frankfurt.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. Journal of Phonetics, 2, 161-180.

Kagaya, R., & Hirose, H. (1975). Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 9, 27-46.

Kawahara, H., de Cheveigné, A., & Patterson, R. (1998). An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. In ICSLP-1998.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. Language, 60, 286-319.

Keating, P., Esposito, C., Garellek, M., Khan, S., & Kuang, J. (2011). Phonation contrasts across languages. In Proceedings of the international congress of phonetic sciences (p. 1046-1049). Hong Kong.

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In Proceedings of the 18th international congress of phonetic sciences. Glasgow.

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

Khachaturian, A. (1984). The nature of voiced aspirated stops and affricates in Armenian dialects. Annual of Armenian Linguistics, 4, 57–62.

Khachaturian, A. (1992). Voiced aspirated consonants in the Noy Bayazet dialect of Armenian. In J. C. Greppin (Ed.), Proceedings of the Fourth International Conference on Armenian Linguistics (pp. 115–128). Delmar, NY: Caravan Books.

Khan, S. (2012). The phonetics of contrastive phonation in Gujarati. Journal of Phonetics, 40, 780-795.

Kirby, J. (2010). Cue Selection and Category Restructuring in Sound Change (Unpublished doctoral dissertation). University of Chicago.

Kirby, J., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. The Journal of the Acoustical Society of America, 140(4), 2400–2411.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 87, 820-857.

Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. Psychological Review.

Kortlandt, F. (1978). Notes on Armenian historical phonology II: The second consonant shift. Studia Caucasica, 4, 9–16.

Kortlandt, F. (1985). Proto-Indo-European glottalic stops: The comparative evidence. Folia Linguistica Historica, 6(2), 183-201. doi: 10.1515/flih.1985.6.2.183

Kortlandt, F. (1995). General Linguistics and Indo-European Reconstruction. Rask, 2, 91-109.

Kortlandt, F. (1998). Armenian glottalization revisited. Annual of Armenian Linguistics, 19, 11–14.

Kreiman, J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J., & Alwan, A. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. Journal of the Acoustical Society of America, 132, 2625–2632.

Kuang, J. (2011). Production and Perception of the Phonation Contrast in Yi (Unpublished master's thesis). UCLA.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics. Chicago: University of Chicago.

Ladefoged, P. (2006). A Course in Phonetics (5th ed.). Boston: Thomson Wadsworth.

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. Oxford, OX, UK ; Cambridge, Mass., USA: Blackwell Publishers.

Lehiste, I. (1964). Acoustical characteristics of selected English consonants (No. 34). Bloomington: Indiana University Press.

Liberman, A., Delattre, P., & Cooper, F. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech, 1, 153–167.

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. WORD, 20(3), 384–422. doi: 10.1080/00437956.1964.11659830

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. Journal of the Acoustical Society of America, 85, 1314–1321.

Löfqvist, A., & McGowan, R. S. (1992). Influence of consonantal envelope on voice source aerodynamics. Journal of Phonetics, 20, 93-110.

Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. Speech Communication, 3, 279-289.

Macak, M. (2017). The phonology of Classical Armenian. In J. Klein, B. D. Joseph, & M. Fritz (Eds.), Handbook of Comparative and Historical Indo-European Linguistics (Vol. 3). Berlin, Boston: De Gruyter Mouton. doi: 10.1515/9783110523874-016

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review, 118(2), 219–246. doi: 10.1037/a0022325

Mikuteit, S., & Reetz, H. (2007). Caught in the ACT: The Timing of Aspiration and Voicing in East Bengali. Language and Speech, 50(2), 247-277. doi: 10.1177/00238309070500020401

Miller, A. L. (2007). Guttural vowels and guttural co-articulation in Ju|'hoansi. Journal of Phonetics, 35, 56-84.

Misnadin. (2016). The phonetics and phonology of the three-way laryngeal contrast in Madurese (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh.

Mittal, V. K., Yegnanarayana, B., & Bhaskararao, P. (2014). Study of the effects of vocal tract constriction on glottal vibration. The Journal of the Acoustical Society of America, 136(4), 1932–1941.

Munhall, K., & Löfqvist, A. (1992). Gestural aggregation in speech-laryngeal gestures. Journal of Phonetics, 20, 111-126.

Murty, K. S. R., & Yegnanarayana, B. (2008, November). Epoch Extraction From Speech Signals. IEEE Transactions on Audio, Speech, and Language Processing, 16(8), 1602–1613. doi: 10.1109/TASL.2008.2004526

Nayiri Institute. (2016). Nayiri. Retrieved from http://nayiri.com

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), Historical Linguistics: Problems and Perspectives (p. 237-278). London: Longman.

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop voicing. Journal of the Acoustical Society of America, 76(1), 224–230.

Parker, P. M. (2008). Webster's Armenian–English Thesaurus Dictionary. San Diego, CA: ICON Group International.

Pisowicz, A. (1997). Consonant shifts in Armenian dialects during the Post-Classical period revisited. In N. Awde (Ed.), Armenian Perspectives: 10th Anniversary Conference of the Association Internationale des études arméniennes (pp. 215–230). School of Oriental and African Studies, London: Curzon Press.

Pisowicz, A. (1998). What did Hratchia Adjarian mean by "voiced aspirates" in Armenian? Annual of Armenian Linguistics, 19, 43–55.

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. Journal of the Acoustical Society of America, 51(4), 1296–1303.

Samlan, R. A., & Story, B. H. (2011). Relation of structural and vibratory kinematics of the vocal folds to two acoustic measures of breathy voice based on computational modeling. Journal of Speech, Language, and Hearing Research, 54, 1267-1283.

Samlan, R. A., Story, B. H., & Bunton, K. (2013). Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. Journal of Speech, Language, and Hearing Research, 56, 1209-1223.

Schirru, G. (2012). Laryngeal features of Armenian dialects. In B. N. Whitehead, T. Olander, B. A. Olsen, & J. E. Rasmussen (Eds.), The sound of Indo-European: Phonetics, phonemics, and morphophonemics (pp. 435–457). Museum Tusculanum Press.

Seyfarth, S., & Garellek, M. (2015). Coda glottalization in American English. In Proceedings of the 18th international congress of phonetic sciences. Glasgow.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. In Proceedings of ICPhS XVII (pp. 1846–1849).

Simpson, A. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. Journal of Phonetics, 40, 477–490.

Singh, R., Keshet, J., Gencaga, D., & Raj, B. (2016). The relationship of voice onset time and voice offset time to physical age. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 5390–5394). IEEE.

Sjölander, K. (2004). The Snack Sound Toolkit. Retrieved from http://www.speech.kth.se/snack/

Steriade, D. (1997). Phonetics in phonology: the case of laryngeal neutralization. UCLA.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 7, 1074–1095.

Swerts, M., & Veldhuis, R. (2001). The effect of speech melody on voice quality. Speech Communication, 33, 297-303.

Toscano, J. C., & McMurray, B. (2010, April). Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. Cognitive Science, 34(3), 434–464. doi: 10.1111/j.1551-6709.2009.01077.x

Vaux, B. (1992). Adjarian's Law and consonantal ATR in Armenian. In J. C. Greppin (Ed.), Proceedings of the Fourth International Conference of Armenian Linguistics (p. 271-293). New York: Caravan.

Vaux, B. (1997). The phonology of voiced aspirates in the Armenian dialect of New Julfa. In N. Awde (Ed.), Armenian Perspectives. Richmond: Curzon Press.

Vaux, B. (1998a). The Phonology of Armenian. New York: Oxford University Press.

Vaux, B. (1998b). Recent Armenological research of Indo-European relevance. UCLA Friends and Alumni of Indo-European Studies Newsletter, 6.

Venables, W. N., & Ripley, B. (2001). Modern Applied Statistics with S (Fourth ed.). New York: Vienna. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vicenik, C. (2010). An acoustic study of Georgian stop consonants. Journal of the International Phonetic Association, 40, 59-92.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review, 14(5), 779–804.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. Journal of Phonetics, 31, 181-201.

Weitenberg, J. (1986). Additional h-, initial y-, and Indo-European *y- in Armenian. In M. Leroy (Ed.), Place de l'arménien dans les langues indo-europeénnes: Colloque Bruxelles, Palais des académies, 21 mars 1985. Louvain: Peeters.

Weitenberg, J. (2002). Aspects of Armenian dialectology. In J. Berns & J. van Marle (Eds.), Present-day Dialectology: problems and findings (pp. 141–157). Berlin/New York: Mouton de Gruyter.

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. Journal of the Acoustical Society of America, 136, 1320-1333.

Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. Journal of the Acoustical Society of America, 139, 1493–1507.

Zhang, Z. (2017). Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model. Journal of the Acoustical Society of America, 142, 2311–2321.

A List of words

Table 4: Table of minimal sets containing word-initial plosives.

	Voiced	Voiceless	Aspirated
Labial	բոկ /b̥ɔk/ 'barefoot (informal)'	պոկ /pɔk/ 'reed (music)'	փոկ /pʰɔk/ 'band, strap'
	բուտ /b̥ut/ 'nourishment'	պուտ /put/ 'spot, a bit (informal)'	փուտ /pʰut/ 'rotten'
	բահ /b̥ɑh/ 'spade, hoe'	պահ /pɑh/ 'moment, guard'	
	բաղ /b̥ɑʁ/ 'garden (informal)'	պաղ /pɑʁ/ 'cold'	
	բան /b̥ɑn/ 'thing, affair'	պան /pɑn/ 'round loaf, guardian'	
	բարուրում /b̥aɾuɾum/ 'swaddling clothes, cradle'	պարուրում /paɾuɾum/ 'wrapping, enclosing'	
		պայտ /pɑjt/ 'horse shoe'	փայտ /pʰɑjt/ 'wood'
		պոկեր /pɔkɛɾ/ 'to tear apart'	փոկեր /pʰɔkɛɾ/ 'seal (animal)'
		պող /pɔʁ/ 'fire'	փող /pʰɔʁ/ 'tube'
Dental	դարը /d̥ɑɾə/ 'pungent'	տարը /tɑɾə/ 'letter'	թարը /tʰɑɾə/ 'perch'
	դասը /d̥ɑsə/ 'lesson'	տասը /tɑsə/ 'ten'	թասը /tʰɑsə/ 'cup'
	դափ /d̥ɑpʰ/ 'tambour, drum'	տափ /tɑpʰ/ 'plain'	թափ /tʰɑpʰ/ 'power, wave'
	դեղի /d̥ɛʁi/ 'medicine'	տեղի /tɛʁi/ 'place'	թեղի /tʰɛʁi/ 'elm tree'
	դող /d̥ɔʁ/ 'tremor'	տող /tɔʁ/ 'line'	թող /tʰɔʁ/ 'to let, allow'
	դաժանագին /d̥aʒanaɟin/ 'cruel, harsh'	տաժանագին /taʒanaɟin/ 'rigorous, arduous'	
	դասական /d̥asakan/ 'classical'	տասական /tasakan/ 'decimal'	
	դատ /d̥at/ 'trial, litigation'	տատ /tat/ 'grandmother'	
	դարան /d̥aɾan/ 'closet'	տարան /taɾan/ 'to take away'	
	դարի /d̥aɾi/ 'century'	տարի /taɾi/ 'a year'	
	դեղնի /d̥ɛʁni/ 'yellow'	տեղնի /tɛʁni/ 'suitable, appropriate'	
	դեր /d̥ɛɾ/ 'role'	տեր /tɛɾ/ 'owner'	
	դում /d̥um/ 'flu'	տում /tum/ 'ginger bread tree'	
		տաղ /taʁ/ 'song'	թաղ /tʰaʁ/ 'neighborhood (informal)'
		տան /tan/ 'home, house'	թան /tʰan/ 'skimmed milk'
		տանկ /tank/ 'tank'	թանկ /tʰank/ 'expensive'
		տիզ /tiz/ 'tick'	թիզ /tʰiz/ 'span, hand'
		տուշ /tuʃ/ 'Indian ink (makeup)'	թուշ /tʰuʃ/ 'cheek (informal)'
		տուփ /tupʰ/ 'box'	թութ /tʰupʰ/ 'shrub'
Velar	գարի /ɟaɾi/ 'barley'	կարի /kaɾi/ 'stitch'	քարի /kʰaɾi/ 'stone'
	գերել /ɟɛɾɛl/ 'to attract, captivate'	կերել /kɛɾɛl/ 'ate'	քերել /kʰɛɾɛl/ 'to scratch, scrape'
	գող /ɟɔʁ/ 'gun'	կող /kɔʁ/ 'side'	քող /kʰɔʁ/ 'screen'
	գոչող /ɟɔt͡ʃʰɔʁ/ 'crier'	կոչող /kɔt͡ʃʰɔʁ/ 'caller'	քոչող /kʰɔt͡ʃʰɔʁ/ 'nomad'
	գոռ /ɟɔr/ 'fierce'	կոռ /kɔr/ 'forced or unpleasant work'	քոռ /kʰɔr/ 'blind (informal)'
	գահ /ɟah/ 'throne, crown'	կահ /kah/ 'furniture'	
	գետ /ɟɛt/ 'river'	կետ /kɛt/ 'point'	
	գեր /ɟɛɾ/ 'fat'	կեր /kɛɾ/ 'to eat'	
	գին /ɟin/ 'price'	կին /kin/ 'woman'	
	գիրք /ɟiɾkʰ/ 'book'	կիրք /kiɾkʰ/ 'passion'	
	գորով /ɟɔɾɔv/ 'tender, emotion'	կորով /kɔɾɔv/ 'vehemence'	
	գուռ /ɟur/ 'bathtub, puddle'	կուռ /kur/ 'compact, solid'	
	գտան /ɟətan/ 'to find'	կտան /kətan/ 'to give'	
	գրել /ɟəɾɛl/ 'to write'	կրել /kəɾɛl/ 'to wear'	
		կաղել /kaʁɛl/ 'limp'	քաղել /kʰaʁɛl/ 'to pick, to harvest'
		կանոն /kanɔn/ 'rule, regulation'	քանոն /kʰanɔn/ 'ruler, guide'
		կաշի /kaʃi/ 'leather'	քաշ /kʰaʃi/ 'weight'
		կով /kɔv/ 'cow'	քով /kʰɔv/ 'side'
		կոռ /kɔɾ/ 'curved'	քոռ /kʰɔɾ/ 'itch'
		կորել /kɔɾɛl/ 'to get lost'	քորել /kʰɔɾɛl/ 'to scratch'
		կուղ /kuʁ/ 'fold'	քուղ /kʰuʁ/ 'thread'
		կույր /kujɾ/ 'blind'	քույր /kʰujɾ/ 'sister'

Table 5: Table of minimal sets containing word-final plosives.

	Voiced	Voiceless	Aspirated
Labial	շտաբ /ʃtɑb̪/ 'headquarters'	շտապ /ʃtɑp/ 'urgent'		
		կապ /kɑp/ 'knot'	կափ /kɑpʰ/ 'knocker'
		տապ /tɑp/ 'sultry'	տափ /tɑpʰ/ 'plain'
Dental	կոդ /kɔd̪/ 'code'	կոտ /kɔt/ 'dry measures'	կոթ /kɔtʰ/ 'handle'
	հոդ /hɔd̪/ 'joint'	հոտ /hɔt/ 'smell'	
	յոդ /jɔd̪/ 'iodine'		յոթ /jɔtʰ/ 'seven'
		անոտ /anɔt/ 'smth that doesn't have a leg'	անոթ /anɔtʰ/ 'vessel'
		ճիտ /t͡ʃit/ 'neck'	ճիթ /t͡ʃitʰ/ 'bunch of grapes'
		մատ /mɑt/ 'finger'	մաթ /mɑtʰ/ 'molasses'
Velar	թագ /tʰɑɡ̆/ 'crown'	թակ /tʰɑk/ 'mallet'	թաք /tʰɑkʰ/ 'odd'
	եղեգ /jɛʁɛɡ̆/ 'cane'		եղեք /jɛʁɛkʰ/ 'be'
	ծագ /t͡sɑɡ̆/ 'apex'	ծակ /t͡sɑk/ 'hole'	
	ճիգ /t͡ʃiɡ̆/ 'effort, endeavour'	ճիկ /t͡ʃik/ 'cry, scream'	
	մեգ /mɛɡ̆/ 'mist'	մեկ /mɛk/ 'one'	
	սագ /sɑɡ̆/ 'goose'	սակ /sɑk/ 'price'	
	սեգ /sɛɡ̆/ 'majestic'	սեկ /sɛk/ 'leather'	
		բակ /b̪ɑk/ 'park'	բաք /b̪ɑkʰ/ 'vessel'
		երեկ /jɛɾɛk/ 'yesterday'	երեք /jɛɾɛkʰ/ 'three'
		հասակ /hɑsɑk/ 'height'	հասաք /hɑsɑkʰ/ 'to get somewhere, to come, to reach'
		տակ /tɑk/ 'bottom'	տաք /tɑkʰ/ 'hot'
		տեսակ /tɛsɑk/ 'type'	տեսաք /tɛsɑkʰ/ 'you saw'

Slashes indicate a transcription based on the orthography. Some words are inflected or otherwise morphologically complex, which is not indicated in the English glosses. The word-initial minimal sets do not include voiced/aspirated minimal pairs because the word-initial sets were originally collected for another experimental protocol which did not require voiced/aspirated pairs.

B   Frequency limits for pitch measurements and formant exclusions

Table 6: Speaker-specific VoiceSauce settings for pitch analysis floor and ceiling

Speaker	Lower bound (Hz)	Upper bound (Hz)
F18, F20, F22	125	300
F19, F49	115	300
F21	100	300
M25, M30	60	175

Table 7: Formant value exclusions. Formant frequency values and H1*–H2* values were excluded if formant frequencies were measured in these ranges.

Speaker gender	Vowel quality	Excluded F1 (Hz)	Excluded F2 (Hz)
All	i	> 500	< 1750
All	u	> 500	> 1750
All	ɛ	> 700	< 1250
Female	ɔ	> 700	> 1750
Male	ɔ	> 700	> 1500
All	ə	> 700	—
Female	ɑ	> 1200	> 2000
Male	ɑ	—	> 1500

C   Model comparisons

Table 8: Model comparison statistics for models with omitted predictors. Positive ΔBIC (left column) and/or $p < 0.05$ by likelihood ratio test (right column) suggests that an acoustic predictor sufficiently improves model fit to justify its inclusion in a parsimonious model of the voicing categories.

Omitted predictor	ΔBIC	BIC	ΔDev	$p(\chi^2(4))$
Aspiration duration	492.8	1831.9	522.15	< 0.001
Log SoE during closure	134.1	1473.2	163.45	< 0.001
Voice onset/offset time	60.3	1399.4	89.65	< 0.001
H1*–H2*	27.2	1366.3	56.56	< 0.001
Vowel duration	23.3	1362.4	52.67	< 0.001
Closure duration	15.7	1354.8	45.03	< 0.001
Cepstral peak prominence	-16.2	1322.9	13.16	0.010
f0	-16.9	1322.2	12.41	0.015

Table 9: Model comparison statistics for models with omitted interactions. Positive ΔBIC (left column) and/or $p < 0.05$ by likelihood ratio test (right column) suggests that the different ranges of a predictor are associated with the voicing categories in word-initial versus word-final position.

Omitted interaction	ΔBIC	BIC	ΔDev	$p(\chi^2(2))$
Voice onset/offset time $\times$ Position	38.8	1377.9	53.46	< 0.001
H1*–H2* $\times$ Position	1.2	1340.3	15.82	< 0.001
Aspiration duration $\times$ Position	-0.3	1338.8	14.36	0.001
Closure duration $\times$ Position	-3.8	1335.3	10.91	0.004
Vowel duration $\times$ Position	-5.0	1334.1	9.68	0.008
Cepstral peak prominence $\times$ Position	-6.9	1332.1	7.72	0.021
f0 $\times$ Position	-11.1	1327.9	3.52	0.172
Log SoE during closure $\times$ Position	-12.3	1326.7	2.32	0.313