Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops

D. Robert Ladd[1] and Stephan Schmid[2]
University of Edinburgh[1] and *University of Zurich*[2]

To appear in *Journal of Phonetics*, special issue on VOT

Abstract

It is well known that what are commonly called voicing contrasts in many languages are accompanied by effects on the fundamental frequency (F0) of a following vowel: roughly, F0 is higher after 'voiceless' and lower after 'voiced' obstruents. This is true regardless of how the voicing contrast is manifested in differences of voice onset time (VOT). Such effects potentially provide a window on the nature of voicing itself, but our knowledge is based primarily on typical European two-way voicing contrasts. Here we present a detailed study of voice onset time (VOT), closure duration, and obstruent F0 effects in Zurich Swiss German. The native two-way contrast in oral stops (often termed fortis/lenis) is unusual in being signalled not by VOT – both types are unaspirated – but primarily by closure duration. We confirm studies showing that this distinction is indeed based on duration, and we show for the first time that both types are accompanied by F0 effects that are typical of voiceless obstruents in other languages. In addition, Swiss German has a smallish set of words conventionally pronounced with voiceless aspirated stops. We investigate the VOT and F0 effects of these marginally contrastive aspirated stops, showing that they do exhibit long VOT and are accompanied by a different pattern of F0 effects that is much more variable than that found with fortis and lenis stops. Our findings support the view that the phonetic basis of voicing and related distinctions involves complex interactions of timing and articulatory gestures that cannot always be characterised in terms of a simple VOT continuum from 'voiced' to 'voiceless aspirated' or a simple phonological dichotomy between 'aspirating' and 'true voicing' languages.

1. Introduction

This special issue of *Journal of Phonetics* is inspired by the runaway success of the concept of voice onset time (VOT) in the description of voicing contrasts in the world's languages. VOT was first proposed as a phonetic dimension by Lisker and Abramson (1964), who observed that in a sample of eleven languages with different two-way, three-way, and four-way stop contrasts, there were striking regularities in the relative timing of onset of phonation and release of the word-initial stop closure. Specifically, across languages, regardless of the number of contrastive categories, there were three clusters of values for VOT: a substantial (60-100 ms) voicing lead, a very short (0-20 ms) voicing lag, and a rather longer (50-80 ms) lag corresponding to 'aspiration'. On the basis of these findings, Lisker and Abramson argued that the single dimension of VOT (or more precisely, relative laryngeal timing; see Abramson 1977) captured something essential about the phonetic basis of voicing contrasts.

The fact that VOT data from a range of languages seemed to fall into three distinct clusters, which was unexpected, has justifiably influenced thinking in both phonetics and phonology ever since. The most prominent current phonological echo of Lisker and Abramson's work is the theory of 'laryngeal realism' (e.g. Beckman et al. 2013), which proposes a universal feature [voice] and a universal feature [spread glottis]; proponents of this idea argue that languages with two-way stop distinctions are generally either 'true voicing' languages (based on [voice] and typically contrasting fully voiced stops with short-lag voiceless unaspirated ones) or 'aspirating' languages (based on [spread glottis] and typically contrasting short-lag stops with voiceless aspirated ones). However, the apparent separation of clusters of 'unaspirated' and 'aspirated' voiceless stops along the VOT dimension may be to some extent an artifact of their limited sample. The larger study by Cho and Ladefoged (1999) suggests that there is a more or less continuous range of typical VOT values up to about 100 ms, and shows that at least a few languages have values as high as 140 ms; the relevance of this larger linguistic dataset for universalist phonological claims like 'laryngeal realism' has not yet been explored. More generally, despite the unquestioned usefulness of the VOT concept (reviewed at length by Abramson and Whalen 2017), the time seems ripe for a reconsideration of cases that fit awkwardly into the Lisker-Abramson scheme.

Lisker and Abramson's own sample presents certain problems for their reductive approach to the phonetic data, particularly in languages with more than a two-way laryngeal contrast. They acknowledged that the distinction between the 'voiced unaspirated' and 'voiced aspirated' stops of the Indo-Aryan languages (many of which show a four-way laryngeal contrast) requires an additional phonetic dimension, and that the difference between 'tense' and 'lax' stops in Korean (which has three laryngeally distinct stop types) cannot be expressed in terms of VOT alone. It seems at least possible that systems with more than a two-way laryngeal contrast are likely to make use of other phonetic properties besides relative laryngeal timing to ensure the robustness of their lexical contrasts. With this in mind, the purpose of this paper is to report a large laboratory study of the unusual set of laryngeal contrasts in Zurich Swiss German.

Swiss German (*Schwyzertüütsch*) is a rather heterogeneous group of German dialects in the Alemannic dialect continuum that extends from Alsace through southwestern Germany and Switzerland into westernmost Austria. In German-speaking Switzerland, unlike Austria and Germany, the status of Alemannic dialects is affected by the existence of stable diglossia (Ferguson 1959), i.e. the systematic use of distinct 'high' and 'low' language varieties in the same speech community, namely Standard German ('high') and a Swiss German dialect ('low'). While Standard German is used for most written purposes, and as a spoken language in some formal contexts (education, law, much broadcasting), most people – of all social classes – speak a Swiss German dialect in their daily lives, and this is what gives Swiss German a certain sociolinguistic unity despite its dialectal diversity. There are many widely shared phonetic features across the different Swiss German dialects, especially in their consonant systems; although our study is specifically based on speakers from the Zurich area, we believe that the instrumental findings reported here will apply quite generally.

Swiss German is generally described as having a 'fortis/lenis' or 'tense/lax' contrast rather than a voicing contrast. The phonetic sketch of Zurich Swiss German presented by Fleischer and Schmid (2006) takes for granted that the fortis/lenis distinction is pervasive in the obstruent inventory, corresponding to distinctions based on voicing in many other West Germanic varieties. In keeping with a long-standing tradition in the use of the IPA, Fleischer and Schmid transcribe lenis obstruents using the IPA symbols for voiced obstruents together with the devoicing diacritic, distinguishing for example between /p/ and /b̥/ or /t/ and /d̥/. Typical word-medial minimal pairs are /ˈhuːpə/ 'honk (a car horn)' vs. /ˈhuːb̥ə/ 'bonnet, hood of a car' and /ˈlɒtə/ 'lath' vs. /ˈlɒd̥ə/ 'store, shop'. As the examples just given show, the word-medial distinction is found after both long and short vowels, which are phonologically distinct (Schmid 2004). The contrast is also lexically exploited both word-finally and word-initially: cf. /g̊rɒːt/ 'ridge' vs. /g̊rɒːd̥/ 'degree', and /tɒː/ 'done' ~ /d̥ɒː/ 'here'. Phonetic evidence reviewed in the next section makes clear (and our study confirms) that the fortis and lenis stops do not differ in VOT and that both are unaspirated; the contrast is manifested most obviously by greater closure duration.

However, we extend the investigation of the acoustic correlates of this distinction into new phonetic territory in two ways. First, we investigate the patterns of F0 perturbation on the vowel following fortis and lenis stops, looking for differences of the sort that are commonly associated with 'voicing' distinctions in many languages (House and Fairbanks 1953, Kingston and Diehl 1994). Previous work on the Swiss fortis-lenis distinction, discussed in section 2.2, has investigated a number of its articulatory and acoustic aspects, but never, to our knowledge, its effect on F0. Second, in addition to the native Swiss German fortis and lenis stops, we investigate the phonetics of words (mainly but by no means exclusively loanwords) that are normally pronounced with aspirated stops. While the phonological status of the aspirated stops is debatable, the presence of three distinct phonetic categories (lenis, fortis, aspirated) allows us to compare not only fortis with lenis, but fortis (unaspirated) with aspirated as well.

2. Background

2.1. 'Fortis' and 'lenis'

Given the background just sketched, it comes as no surprise that the history of the terms 'fortis' and 'lenis' is closely connected with the study of Swiss German. The terms appear in Sievers (1876), referring to a contrast between two homorganic stops realized through a difference in intensity and duration. Sievers stated that the terms had been suggested by his student Winteler (1876), who had just finished his PhD thesis on a Swiss German dialect, though Braun (1988) notes that they had already been used previously by other scholars. In any case, fortis/lenis (or the ostensibly equivalent pair tense/lax) have often been used since then in dialectological and phonetic studies of Swiss German (e.g. Heusler 1888, Dieth 1950, Willi 1996). Jakobson and Halle (1964), quoting Winteler, explicitly mentioned the 'Swiss German consonantal pattern' as an example of their proposed universal feature [±tense], which was intended to cover duration and height differences between vowels and duration and intensity differences between obstruents.

Beyond Swiss German, the dichotomy fortis vs. lenis has been used in descriptions of numerous languages, referring to various different mechanisms of speech production both at the supraglottal and the glottal level. For instance, in Kohler's (1984) attempt to establish a typology of consonantal contrasts relying on this single universal feature, fortis vs. lenis is conceived of as a 'power feature' which interweaves the dimensions of timing and tension. In this view, fortis and lenis can surface in a variety of different phonetic manifestations. For instance, lenis stops can be 'passively voiced' (as in German) but also 'actively voiced' (as in Italian); fortis stops can appear as geminates (as in Finnish) or can involve tension of the vocal folds (as in Korean). A similarly broad perspective on the tense/lax dichotomy is found in Jessen's (1998) synopsis, which covers both a number of genealogically unrelated languages and a variety of different articulatory and glottal timing mechanisms.

It was this range of possible phonetic correlates that drew Lisker and Abramson's criticism in their original VOT article. As they put it, 'if all these terms are assumed to have phonetic meanings, we must ask whether this change nets us any gain in description, for we have exchanged a phonetic dimension, voicing, which has a clear articulatory and acoustic meaning, for one which is considerably less well defined both articulatorily and acoustically' (1964: 420). More categorically, they claim that 'certainly none of the acoustic features which have been suggested as correlates of a fortis/lenis distinction is demonstrably independent of voicing' (1964: 387). Other phoneticians have simply called for caution, warning against the use of the terms fortis/lenis and tense/lax without a precise specification of the phonetic properties in the language being described (e.g. Catford, 1977: 203; Jaeger, 1983: 188; Ladefoged and Maddieson, 1996: 98-99).

2.2. 'Fortis' and 'lenis' in Swiss German stops

Given the considerable number of empirical studies in recent decades, however, it is not difficult to provide this precise specification for the fortis/lenis distinction in Swiss German. The most obvious phonetic correlates in the acoustic signal are durational; for oral stops, overall duration and/or closure duration are not only the most investigated phonetic correlates of the fortis/lenis contrast in Swiss German, but also the ones that turn out to be most relevant perceptually. Dieth and Brunner (1943: 744) found that the average ratio of overall duration between lenis and fortis intervocalic stops for four speakers from Zurich was 1:2.48 for labials and 1:3.04 for coronals. Similar results have been found by others. The two Zurich speakers measured by Willi (1996: 148) show ratios of overall duration between lenis and fortis ranging from 1:1.6 to 1:2.7. For three speakers from Bern, Ham (2001: 60) found significant differences in mean closure duration of word-medial stops; even word-finally, the ratio between lenis and fortis was 1:1.69. Finally, Kraehenmann (2003: 114-122), in her analysis of four speakers from Thurgau, found a significant difference of mean closure duration in both word-medial and word-final lenis and fortis stops, mean closure durations being 67.3 ms for lenis and 119.2 ms for fortis. The perceptual relevance of closure duration was demonstrated by Willi (1996), who found a quasi-categorical effect of closure duration on the discrimination of the two words /ˈʒ̊ɒtə/ ‘shadow’ ~ /ˈʒ̊ɒd̥ə/ ‘damage’, with responses suddenly shifting towards the fortis category as closure duration exceeded 100 ms.

VOT, by contrast, does not differ between fortis and lenis stops. The first study specifically devoted to VOT in a Swiss German dialect was that of Enstrom and Spörri (1981), based on the speech of five Zurich speakers. They found clear overlap in the VOT of fortis and lenis stops at all three places of articulation and concluded that ‘in Swiss-German, VOT does not constitute the primary feature in differentiating between stop cognate pairs’ (1981: 138). Similarly, Ham (2001: 68-69) found statistically insignificant mean differences of VOT ‘on the order of 2-3 ms’ for fortis and lenis stops produced by his three Bern speakers, while Kraehenmann (2003: 115), in her Thurgau data, found mean VOT values of 27.7 ms for lenis stops and 28.0 ms for fortis. To our knowledge there has never been a perceptual study specifically dealing with Swiss German VOT; on general psychophysical grounds the perceptibility of differences of a millisecond or two is highly implausible, and the practical difficulties of setting up an experiment to detect it would be considerable.

As for articulation, different techniques have been used to measure air pressure and articulatory contacts in fortis and lenis obstruents, and there is clear evidence of phonetic differences that cannot be related to VOT but which seem to involve something that might reasonably be called force of articulation. In the first experimental study on Swiss German consonants, Dieth and Brunner (1943) measured amplitude curves on kymographs produced via a rubber tube. Though they found considerable differences between their subjects (six Zurich-area speakers), they reported generally higher intraoral pressure for the fortis consonants (sometimes even twice the pressure of the lenis counterparts). More recently, Schmid, Studer and Dellwo (2011) investigated nine female speakers of Zurich German and found that fortis stops are characterized by both higher intraoral air pressure during closure and higher velocity of the airstream at release. Dieth and Brunner (1943: 747) also produced kymographs reflecting the articulatory pressure of the tongue for coronal consonants, again showing higher values for fortis than for lenis. Finally, Kraehenmann and Lahiri’s (2008) electropalatographic study of four Thurgau speakers showed that closure duration (more specifically, ‘duration of maximum contact’) is significantly longer in fortis stops even where it cannot be directly perceived or measured acoustically, namely in absolute initial position. However, it should be noted that Kraehenmann and Lahiri interpret their results not in terms of fortis and lenis, but as evidence for a phonological distinction between singleton and geminate stops; we return to this issue briefly in the final section of the paper.

2.3. Aspirated stops in Swiss German

The investigation of Swiss German fortis and lenis stops is interestingly complicated by a further feature of the language, namely the presence in some contexts of aspirated stops. As Fleischer and Schmid (2006: 244) put it: ‘aspiration is lexically determined and typical of borrowed items, such as [pʰɒkχ] (parcel) and [tʰeː] (tea)’. This phenomenon was already described by Winteler (1876: 56), who recognized that aspirated stops mainly occurred in loanwords from Standard German (such as *Pack* and *Tee* cited by Fleischer and Schmid) and in ordinary personal names like *Paul* [pʰaʊ̯l]. Nowadays, we can add not only a series of loanwords from English, such as *Team* [tʰiːm] and *Party* [ˈpʰɒrti], but also a number of foreign place names such as *Poole* [ˈpʰoːlə] ‘Poland’ and *Thailand* [tʰæi̯lɒnd̥], as well as the names of the letters <p> and <t>, both when used on their own and in abbreviations such as the names of the political parties *FDP* and *SP*.

It is by no means clear how to treat these aspirated stops in the phonology of Swiss German. On the one hand, we are clearly not dealing with a pervasive three-way laryngeal contrast such as we find in Korean (e.g. Lee 1999; Cho, Jun and Ladefoged 2002) or Thai (e.g. Tingsabadh and Abramson 1999). It is possible to cite a few potential minimal pairs like /puːr/ ‘farmer’ ~ /pʰuːr/ ‘pure’ or /tæi̯l/ ‘part, component’ ~ /tʰæi̯l/ ‘thing, gadget’, but in some important sense the aspirated stops remain outside the core phonological system. On the other hand, it would be equally misleading to ignore the phenomenon as merely a matter of a few foreign words, comparable to the use in English of a front rounded vowel in *Debussy* or a velar fricative in *Bach*. As just noted, the list of words regularly pronounced with aspirated stops includes ordinary given names (e.g. *Paul, Peter, Pia, Tina*) and everyday objects (including not only *Pack* and *Tee* just cited but also *Theek*, the typical satchel or backpack carried to school by many Swiss children). There is evidence (Schifferle 2010) that the use of aspirated stops is on the rise among younger generations, which inevitably means that some words are variably pronounced, but (as we shall see) the lexical distribution of aspirated and fortis (unaspirated) stops is broadly consistent across speakers, and a post-recording questionnaire showed that speakers are generally aware of which words they pronounce which way.

The phonological status of the distinction between aspirated and unaspirated stops is certainly problematic. We might invoke a notion such as ‘quasi-phoneme’ (Kiparsky 2014) or ‘marginal contrast’ (e.g. Hall 2013; Renwick and Ladd 2016) to describe the aspirated stops. It is also conceivable that the current situation – or perhaps any situation of marginal contrast – reflects the slow lexical diffusion of a sound change (e.g. Wang 1969; Labov 1981, 1994); if so, then we might prdict that at some point in the future all fortis stops will be pronounced with aspiration and the phonology will revert to a typical European two-way pattern. These issues lie well beyond the scope of the present paper, though we provide further detail in Appendix 4. However, the existence of ***phonetically*** distinct three-way sets like [b̥] / [p] / [pʰ] is of clear interest for investigating the relation between VOT and other phonetic dimensions, and that is the focus of our investigation of the aspirated stops in this paper.

2.4. Obstruent F0 effects (CF0)

It has been known for decades (House and Fairbanks 1953, Lehiste and Peterson 1961) that voicing contrasts in many languages are accompanied by differing effects on the fundamental frequency (F0) of a following vowel. The difference can be summarised roughly by saying that F0 is higher after voiceless consonants than after voiced. This difference is found irrespective of how the contrast is mapped onto phonetic differences of VOT (Sonderegger et al. 2017), and thus irrespective of the supposed distinction between 'true voicing' and 'aspirating' languages: it appears in studies of languages that contrast voiceless unaspirated stops with fully-voiced ('prevoiced') stops (e.g. Dutch: Löfqvist et al. 1989; French and Italian: Kirby and Ladd 2016) and in languages that contrast voiceless aspirated stops with stops that are often realised as voiceless unaspirated (e.g. English: Hanson 2009; Cantonese: Francis et al. 2006). Following Kingston (2007), we refer to such effects collectively as CF0.

In much of the research on CF0, the main goal has been to determine what causes these differences. In particular, a long-standing issue is whether CF0 reflects the automatic effect of laryngeal articulatory gestures (e.g. Kohler 1982; Löfqvist et al. 1989; Hanson 2009), or involves the deliberate 'enhancement' of voicing contrasts (Kingston and Diehl 1994). It has also been suggested (Connell 2002) that these two types of explanation do not actually form a meaningful dichotomy. The fact that the effects occur regardless of the way VOT is used in the 'voicing' distinction seems broadly consistent with an enhancement account; on the other hand, the apparent universality of CF0 (most convincingly demonstrated in the study by Sonderegger et al. 2017) suggests that at least part of the explanation lies in the automatic physical consequences of the laryngeal activity involved in producing the voicing distinctions.

Until fairly recently, part of the reason for the different viewpoints on these issues has been a lack of clarity about how to characterise CF0 effects. Some investigators have focused on the *contour shape* immediately after the release of the obstruent, noting that F0 often falls after voiceless obstruents and rises after voiced ones, and that this may be the most perceptually relevant feature (e.g. Whalen et al. 1990). Others (e.g. House and Fairbanks 1953) have focused on *F0 level,* observing that, in otherwise identical contexts, F0 after voiced obstruents is typically lower than F0 after voiceless obstruents. This point of view then gives rise to the question of whether F0 is being lowered by voicing or raised by voicelessness. We emphasise that this is a meaningful question, not just a matter of point of view. The course of F0 is linguistically significant in itself, either as lexical/grammatical tone or as intonation, which means that CF0 effects are best seen as deviations from an underlying or idealised F0 pattern. Lehiste and Peterson (1961) clearly saw this, and some later researchers (e.g. Kohler 1985; Silverman 1986) have also recognised the methodological importance of considering interactions with intonation in studying CF0. From this viewpoint, the drop in F0 often seen after voiceless obstruents can be interpreted as the consequence of returning from locally raised F0 to the linguistically specified pitch level.

In our opinion, these questions were effectively settled with the publication of Hanson (2009), a thorough study of CF0 in American English that carefully controlled intonational effects on F0 and, importantly, studied the F0 trajectories that accompany nasals (in syllables like /mɑm/) as well as those that accompany voiced and voiceless obstruents (in syllables like /bɑm/ and /pɑm/). By taking the F0 pattern that accompanies nasals as indicative of the unperturbed (i.e. intonationally intended) F0 contour, Hanson established clearly that F0 is locally raised after voiceless stops but is largely unaffected (i.e. is indistinguishable from the nasal reference contour) after voiced stops. This local raising of F0 is what causes the 'falling F0' reported by others. Unfortunately – probably because of the influence of Kingston & Diehl 1994, who assumed that the enhancement of voicing contrasts was a matter of lowered F0 with voicing – Hanson's finding is still underappreciated, and it is still not uncommon in papers on the phonology of voicing (e.g. Vietti et al. 2018: 80) to find it stated, *as fact*, that voicing lowers F0. Note, though, that in some sense Hanson's finding does not affect Kingston and Diehl's fundamental point, which was that phonemic contrasts based primarily on one phonetic property can be enhanced by the systematic deployment of other phonetic properties, and that such 'phonetic knowledge' is part of phonological competence. It is irrelevant to this general idea whether CF0 effects are a matter of lowering for voicing or raising for voicelessness.

What is particularly interesting in the present context is that Hanson also found raised F0 after the voiceless unaspirated stops that occur in English syllable-initial /s/+stop clusters (in syllables like /spɑm/). This finding calls into question any purely VOT-based definition of 'voicing', because by such a definition, both the /p/ of /spɑm/ and the /b/ of /bɑm/ in Hanson's American English speakers are phonetically 'voiceless unaspirated' or short-lag VOT stops (IPA [p]). Yet they exhibit clearly different effects on F0. One possible explanation for this difference is that the single VOT-based category 'voiceless unaspirated' actually lumps together at least two distinct phonetic types; if this is so, then the challenge for phonetic research is to determine how the distinct types differ in physical detail. Hanson herself seems to favour an explanation of the sort proposed by Halle and Stevens (1971) and Löfqvist et al. (1989), namely that the phonologically voiceless obstruents (including both aspirated [p^h] in /pɑm/ and unaspirated [p] in /spɑm/) are characterised by a voicing inhibition gesture involving glottal stiffness during the obstruent closure and that the physical effects of this gesture take some time to decay after phonation is resumed. The phonologically voiced (but phonetically voiceless unaspirated) [p] in /bɑm/, by contrast, is not accompanied by a voicing inhibition gesture and therefore has no perturbing effect on F0 when phonation resumes. That is, though the [p] in /bɑm/ and the [p] in /spɑm/ have similar VOT, they are *not phonetically identical*; our phonetic descriptive framework for discussing voicing must be enriched to be able to take such differences into account.

Herein lies the relevance of CF0 research for this special issue of the journal: any elaborated descriptive framework that helps explain CF0 will also provide insight into the nature of voicing itself. Specifically, the Swiss German fortis/lenis distinction provides us with a new opportunity to observe the CF0 behaviour of phonologically distinct categories of stops that do not differ in VOT (to see, for example, if lenis and fortis stops differ in a similar way to voiced and voiceless stops in other studies), and to compare CF0 effects in aspirated stops with the two phonologically distinct sets of voiceless unaspirated ones. Our primary contribution thus consists of empirical data from a typologically unusual system of laryngeal distinctions, which should serve to clarify the contribution of VOT to the supposed fortis/lenis contrast and more generally to further our understanding of voicing-related contrasts and their F0 effects.

2.5. Interim summary: Research questions

Our research questions dealt with both durational and F0 correlates of the 'laryngeal' contrasts in Zurich Swiss German stops. Concerning duration, we investigated the following questions:

- Do the fortis and lenis stops of Swiss differ in VOT? In closure duration? Our expectation here was of course that we would replicate the findings of the previous studies reviewed above: no difference in VOT, and substantial differences of closure duration.
- Do we find evidence for the reported distinction, in words beginning with orthographic <p> and <t>, between those pronounced with aspirated stops and those pronounced with unaspirated (fortis) stops? If so, is the difference manifested in VOT, in closure duration, or both? Our expectation was that we would find VOT values in the typical range of aspirated stops in other Germanic languages (roughly 50-80 ms); we had no basis for a prediction about closure duration.

Concerning CF0, we wanted to answer the following:

- Do lenis and fortis stops exhibit differences in CF0 effects of the sort that have been found for 'voiced' and 'voiceless' stops in a wide variety of other languages (i.e. higher pitch after fortis than after lenis)? A small pilot experiment suggested that such a difference was indeed present.
- Do aspirated stops exhibit their own pattern of CF0 effects, or do they pattern with fortis (or with both fortis and lenis)? Here we had no basis for a specific prediction, but we assumed that they would be comparable to effects found with voiceless stops in other languages.
- Do the CF0 effects of fortis and lenis stops, whatever they turn out to be, provide any evidence relevant to theories of CF0 effects in general? For example, a clear difference between fortis and lenis might provide support for the idea that CF0 effects serve to enhance contrasts; on the other hand, if fortis and lenis show similar CF0 effects (consistent with the similar VOT), this might support a more biomechanical explanation.

Through these specific questions, we aimed to explore the relevance of the typologically unusual Swiss German obstruent system for the notion of VOT itself.

3. Method

3.1. Materials

The data reported here are based on laboratory speech – recordings of carefully designed test sentences read aloud – produced by twenty native speakers of Zurich Swiss German. We modelled our speech materials on those used in Kirby and Ladd's (2016) study of CF0 in Italian and French: we embedded test words in pragmatically natural alternative questions of the general form 'Do you prefer A or B?', 'Is her name A or B?', 'Do they live in A or B?', and so on. This ensures that the intonational context of the test words is highly uniform from sentence to sentence and from speaker to speaker.

We created 80 test sentences, half with labial consonants in the test word and half with coronal consonants (we avoided velar stops, as they are generally affricated in word-initial position in Swiss German). The test word was always in the A position, ensuring that it had a prominent rising pitch accent. For example, the test word *Baan* 'train' was embedded in the sentence *Gönd er mit de Baan oder mit em Auto?* 'Are you going by train or by car?'. A complete list of test sentences is given in Appendix 1. There were four groups of 20 test words, with each of four different test consonant types: lenis stop (/b̥ d̥/), fortis unaspirated stop (/p t/), aspirated stop (/p^h t^h/), and nasal (/m n/). The test consonant was always the first phoneme of the test word. In addition to the 80 test sentences, there were 20 warm-up sentences of the same general form, which were not analysed.

Our choice of test words was constrained in some cases by the near absence of everyday items meeting the phonetic criteria. For etymological reasons, words beginning with orthographic <p> – both aspirated and unaspirated – were especially difficult to find. To expand the pool of potential test words, we therefore used both monosyllabic and disyllabic words, though this inevitably has some effect on the shape of the intonational rise. All disyllabic test words had lexical stress on the first syllable. Insofar as possible, we avoided using test words with a high vowel in the first syllable, in order to minimise the effects of vowel intrinsic F0 (Whalen and Levitt 1995), but here too the rarity of words with initial orthographic <p> means that our set of test words is not perfectly balanced. Note also that, because the use of aspirated stops is at least somewhat variable, there was no guarantee that speakers would produce exactly twenty aspirated stops and twenty fortis unaspirated ones; we return to this point in sec. 3.4.

In just over two-thirds of the test sentences (54 out of 80), the test word was a noun immediately preceded by a syntactically close unstressed function word (usually an article, a preposition, or a preposition-article combination, but occasionally also a possessive determiner or a numeral). Although this means that many of the sentences share a common syntactic and prosodic plan, it also means that the segment immediately preceding the test consonant cannot be very well controlled: for example, the final segment of the article is determined by gender and case and may be any of /m/, /n/, /s/, or /ə/. In most of the remaining 26 test sentences, the test word was a bare noun (including proper names), adjective, infinitive, or adverb, preceded by a syntactically more distant unstressed syllable (e.g. *Isch er taub oder blind?* 'Is he deaf or blind?'; *Häsch lieber Tee oder Kafi?* 'Do you prefer tea or coffee?' [lit. 'have-you rather tea or coffee?']; *Ässe mer daa oder uf em Balkon?* 'Shall we eat here or on the balcony?' [lit. 'eat we here or on the balcony?']). Here, too, the primary criteria were pragmatic naturalness and prosodic similarity, and consequently the phonetic segment immediately preceding the test consonant varied.

3.2. Speakers

The speakers were twenty female students at the University of Zurich, ranging in age from 19 to 34 (mean 23.9). The reasons for recording only female speakers were both practical and methodological: first, it was easiest to reach students in the second author's department, a substantial majority of whom are women; second, restricting ourselves to female speakers meant that we did not need to make provision for male-female differences of pitch range in our statistical analyses. Since all the speakers were students of languages, linguistics, and/or phonetics, all but one spoke two or more other languages fluently; twelve had experience of living abroad for periods ranging from 3 months to 2 years; and nine of them had at least one parent whose native language was not any variety of Swiss German. However, all had grown up in the city of Zurich or in nearby towns or villages in the Canton of Zurich, and all used Zurich German in their daily lives. All were judged by the second author (a middle-aged native speaker of Zurich German who grew up near Wetzikon) to be fluent native speakers. The speakers signed informed consent forms in accordance with the University of Zurich research ethics guidelines, and were paid a small fee for their participation.

3.3. Recording procedure

Recordings were made direct to disc in a professional recording studio in the University of Zurich Phonetics Laboratory (sample rate 48 kHz; 24-bit encoding). The sentences were presented one at a time on a computer screen in the recording booth, but before the recording session the speakers had an opportunity to look over a full printed list of the sentences. We used the Dieth orthography (Dieth 1986) for the written form of the sentences: there is no standard Swiss German spelling, but there are a number of widely shared conventions, and it is quite normal to see written Swiss German in some contexts (advertising, text messaging, etc.). None of the speakers had any trouble with the reading task, and the overwhelming majority of the productions were fluent and natural-sounding.

The test sentences were presented in the same pseudo-random order to all speakers. The twenty warm-up sentences were presented first, in order to give the speakers an opportunity to fall into a natural rhythm and intonation for producing alternative questions. As reported by Kirby and Ladd (2016), this method of eliciting a controlled intonation contour proved very successful without resorting to explicit metalinguistic coaching of the sort used by Hanson (2009). The recording session lasted approximately 15 minutes.

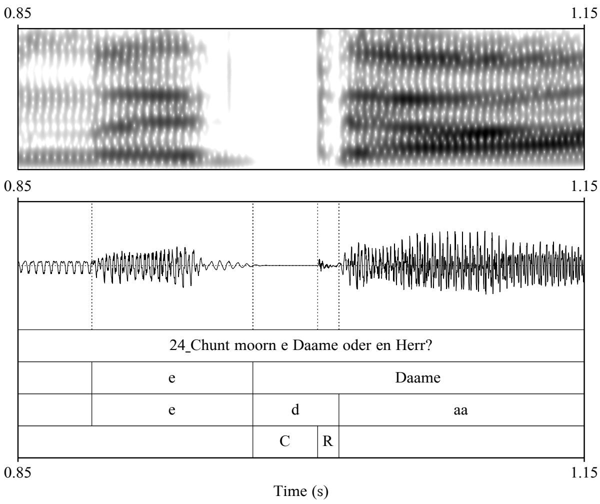
3.4. Acoustic and auditory analysis procedure

Acoustic analysis was carried out using Praat (Boersma and Weenink 2016). The long files for each speaker's whole recording session were broken up into separate sentence files, each with a separate TextGrid. F0 analysis was based on Praat's default analysis settings; because we were concerned only with the F0 contour on an intonationally-prominent stressed vowel we had very few problems with irregular phonation, spurious values, and so on. For the duration measures, segmentation was carried out by hand by the two authors. We marked only those landmarks relevant for the duration and F0 analyses reported below: the beginning of the word preceding the test word, the beginning of the segment preceding the test consonant, the beginning of the closure phase of the test consonant, the release of the test consonant, the onset of voicing (not applicable for nasal test consonants), the end of the sonorant portion of the test word's stressed syllable (e.g. /a:r/ in *Baart* 'beard'), and the end of the test word. Some of these (e.g. end of test word) were not used for precise duration measurements but served only to define intervals over which F0 was computed, and were in any case often difficult to locate with precision, especially in test words ending in unstressed vowels.

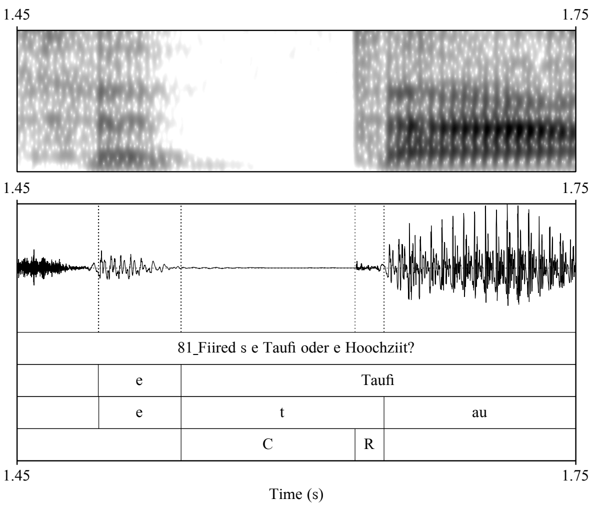
We carefully defined the three landmarks on which the measures of closure duration and VOT are based; after some preliminary exploration, in order to increase inter-labeller reliability (see Appendix 2.i), we decided to locate them on the basis of the waveform rather than using a combination of spectrogram and waveform. Specifically, except in a few specific contexts (see Appendix 2.ii), we marked the onset of closure at the end of any visibly periodic residual voicing rather than attempting to identify the moment of closure from the spectrogram or any change in amplitude. This can be seen in Panels A and C of Figure 1. In the same way, we marked the onset of voicing at the beginning of clear periodicity even if the waveform was also still quite noisy or breathy, as can be seen in Panel C of Figure 1.

This decision to prioritise the waveform in our segmentation criteria undoubtedly has the consequence that our measures of closure duration (in the case of all the oral stops) and VOT (in the case of the aspirated stops) are slightly shorter than they might have been given other criteria. Especially when the test consonant was preceded by a sonorant, there was some degree of residual voicing in many of the stops, corresponding to the ‘bleed’ type of voicing implementation reported for English by Davidson (2016). Nevertheless, we emphasise that in all but a few dozen cases residual voicing in oral stop closures of all three types is followed by an interval of complete voicelessness. Perhaps a more important problem with our procedure is that it may slightly exaggerate the difference in closure duration between lenis and fortis stops: informal inspection of the corpus carried out after the segmentation work was complete suggests that fortis stops may actually be somewhat less susceptible to ‘bleed’ than the other two types. There is a great deal of individual and contextually-conditioned variability, and we doubt that there is a systematic difference in ‘voice offset time’ that might serve as a correlate of the Swiss German fortis/lenis contrast, as has been suggested for Itunyoso Trique by Di Canio (2012), but this is a potentially important question for future experimental study. The question cannot be investigated more rigorously on the basis of our data because our materials do not control the segmental context immediately preceding the stops.

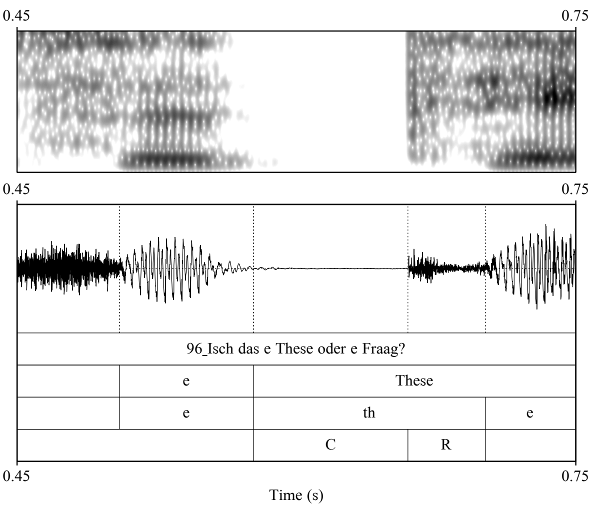
Finally, recall that there is some variation regarding the aspiration of words beginning with orthographic <p> and <t>. For all 40 test words beginning with <p> or <t>, regardless of how they were originally intended in the design of our materials, we judged auditorily whether they had been produced as ‘fortis’ (i.e. unaspirated) or ‘aspirated’, and treated the items in subsequent analysis on the basis of this perceptual classification. That is, even though the materials were designed (based on the intuitions of the second author) to have equal numbers of fortis and aspirated consonants, most speakers aspirated one or more words that were intended as fortis in the original design. The mean number of aspirated words per speaker was 22.4 (median 21). Further detail is given in Appendix 4.



Panel A, lenis



Panel B, fortis



Panel C, aspirated

Figure 1. Illustrations of our segmentation criteria. The three panels show 300 ms. excerpts, plotted on the same horizontal scale so that segment durations can be informally compared by eye across panels. 'C' and 'R' stand for closure and release. For further explanation see text.

3.5. Data reduction and statistical analysis

The segmented TextGrids were processed by means of a Praat script written for us by Dieter Studer-Joho. The script computed the durations of the intervals of interest (in particular, closure duration and VOT) and extracted a number of F0 values. There were no missing duration values; the analyses are all based on 80 items per speaker.

As for F0, the values of most interest for our purposes are those that immediately follow the test consonant. The script used the Praat function *get_F0_at_time* to compute these values at 10 ms intervals over the first 150 ms of the test vowel, beginning with the labelled onset of the test vowel. The first value (0 ms) was normally either missing or spurious, and the data are plotted and analysed beginning at 10 ms. On the basis of these values, as will be seen in the next section, we plotted pitch contours for the four categories of consonant (aspirated, fortis, lenis, and nasal) on a single set of axes to visualise the perturbation effects of the different consonant types. The nasal contours – the reference value for evaluating the other three contours – show a brief period (roughly 30 ms) of approximately level low pitch, followed by a steady rise. This is what we would expect from descriptions of Swiss German and other Southern German varieties, in which rises are often described as L*+H, i.e. low followed by a rise (e.g. Fleischer and Schmid (2006) for Zurich Swiss German, Fitzpatrick-Cole (1999) for Bern Swiss German, Atterer and Ladd (2004) for Bavarian Standard German). The contours that follow the oral stops generally begin higher and fall, often steeply, before the steady rise; on average the rise begins about 45-50 ms into the vowel with all three stop types. Consequently, for all contours we took the F0 minimum (*minF0*) that preceded the rise – corresponding in some sense to the L* pitch target – as the theoretically motivated comparison point for statistical analysis. In most cases identifying this point was completely straightforward; additional details are given in Appendix 2.iii.

We analysed the influence of the consonant type on *minF0* in two ways: first, on *minF0* itself, expressed in terms of absolute Hz values, and second, on the size of the initial pitch drop (*F0drop*), i.e. the interval in Hz between the first F0 measurement (*F0@10ms*) and *minF0*. This second measure was intended to compensate for the existence of overall range differences between speakers and random differences of overall level from one utterance to the next; however, as we shall see, the two measures point to essentially the same conclusion. In the few cases where a value for *F0@10ms* was missing, we used the difference between *F0@20ms* and *minF0* as the value of *F0drop*; if the values at both 10 and 20 ms were missing, or if *minF0* was missing (usually due to creaky voice), we treated the value of *F0drop* as missing. Out of 1600 items, there were 17 missing values for *minF0* and 20 for *F0drop*.

In all the statistical analyses we evaluated the effect of consonant type on the dependent variable of interest with linear mixed effects models, using the `lme4` package (Bates et al., 2015) in R, with speaker and item as random factors. For the duration analyses (closure duration and VOT), we also included consonant place (labial or coronal) as a fixed effect, together with its interaction with consonant type. For the CF0 analyses (*minF0* and *F0drop*) we included fixed effects of consonant place and vowel height (high, mid, low), together with all two-way interactions. In addition, all models included by-speaker random slopes for both consonant place and vowel height.

For post-hoc pairwise comparisons, we reported differences in estimated marginal means computed using the `emmeans` package (Lenth, 2018). In all cases, fractional degrees of freedom were computed using the Kenward-Roger approximation, and *p*-values were adjusted using the Tukey method for comparing a family of the relevant number of estimates (3 or 4) when averaging over the effects of place or height, respectively.

4. Results and Discussion

4.1. Closure duration and VOT

4.1.1 Effects of consonant type: A summary of mean closure duration and VOT based on all 1600 test tokens in the corpus is presented graphically in Figure 2. These data confirm the key findings of earlier work summarised in sections 2.2 and 2.3. First, they confirm that the distinction between fortis and lenis is signalled by a substantial difference in closure duration and not by VOT. Second, they show that there is a clear difference in the distribution of VOT values between fortis and aspirated stops; this represents the first instrumental validation of impressionistic reports that some words beginning with orthographic <p> and <t> are aspirated. Closure duration for aspirated stops is on average somewhat shorter than for fortis (unaspirated) stops.

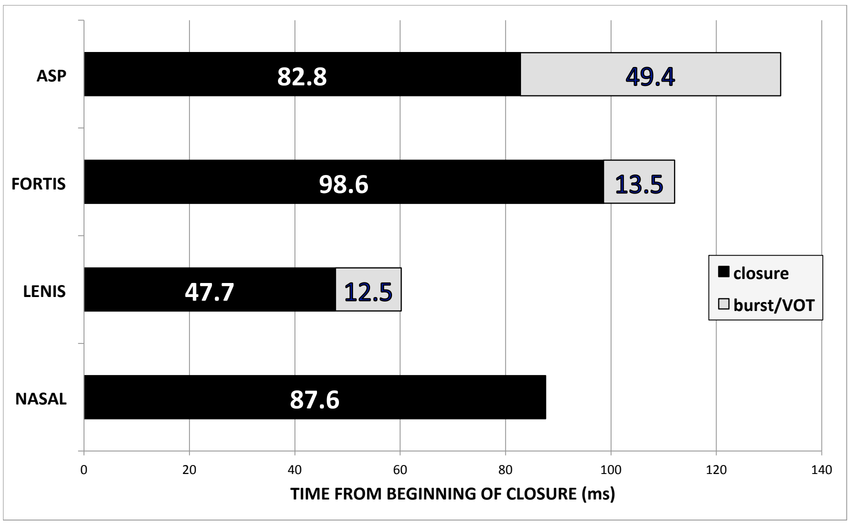


Figure 2. Mean closure and VOT data for all four consonant types and all 20 speakers.

Figure 3 gives an idea of the variability behind the data in Figure 2, showing a scatterplot of closure duration and VOT for all tokens of the three oral stop types. This shows clearly that, despite the marginally contrastive nature of the aspirated category and its somewhat variable lexical distribution, for purposes of investigating the phonetic basis of voicing distinctions, Swiss German uses three phonetically distinct types of stops. It also shows clearly that the three types are not distinct along a single phonetic dimension.

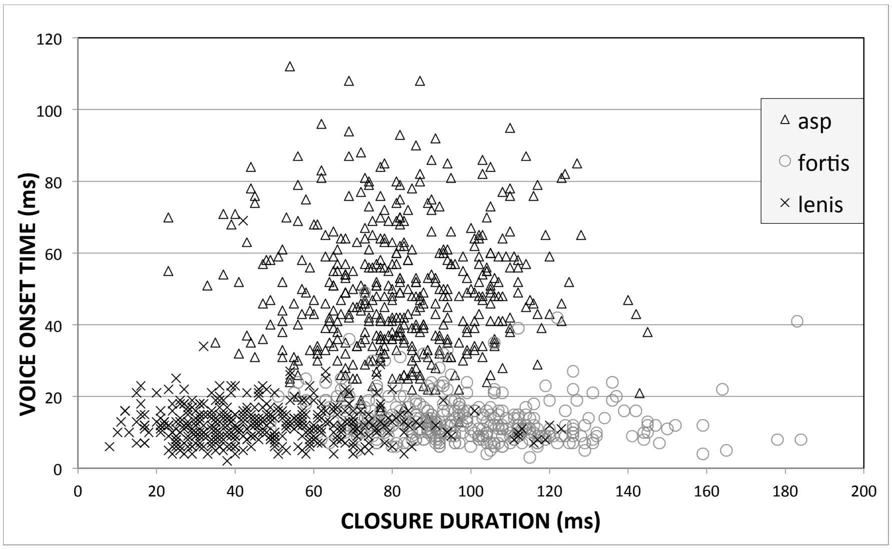


Figure 3. Scatterplot of all tokens of the three oral stop types as a function of closure duration and VOT.

Our statistical model for closure duration included fixed effects for consonant type, consonant place, and their interaction, along with random intercepts for speaker and word and correlated by-speaker slopes for consonant type and consonant place. Our model for VOT was identical except that the data for nasals were not analysed. Full details of the models are given in Appendix 3. Table 1 shows the results of post-hoc pairwise differences between the estimated marginal means of the four consonant types for closure duration. Table 2 shows pairwise differences for VOT within the three oral stop types. As expected on the basis of past studies (e.g. Cho and Ladefoged 1999), there was a small but significant effect of consonant place on VOT, with the labial stops showing shorter VOT than the coronal. This difference was statistically significant only with the aspirated stops (p = .009; see Appendix 3.ii).

contrast	estimate	SE	df	t.ratio	p.value
nasal - lenis	39.81	3.98	60.83	9.99	>.0001
nasal - fortis	-11.33	4.21	56.25	-2.69	0.045
nasal - asp	6.42	4.22	52.35	1.52	0.432
lenis - fortis	-51.14	4.08	59.43	-12.54	>.0001
lenis - asp	-33.39	3.70	63.89	-9.02	>.0001
fortis - asp	17.75	3.13	83.21	5.67	>.0001

Table 1. Pairwise comparisons of closure duration among the four consonant types; the column 'estimate' shows the difference in estimated marginal means (in ms) between two types. Results are averaged over the levels of consonant place. The lenis stops show a significantly shorter closure duration than all the other three types, and fortis significantly longer.

contrast	estimate	SE	df	t.ratio	p.value
lenis - fortis	-1.98	1.57	59.69	-1.26	0.423
lenis - asp	-36.62	2.70	33.49	-13.58	>.0001
fortis - asp	-34.65	2.59	28.51	-13.37	>.0001

Table 2. Pairwise comparisons of estimated marginal means (in ms) for VOT among the three oral stop types lenis, fortis, and aspirated. Results are averaged over the levels of consonant place. Lenis and fortis stops show significantly shorter VOT than aspirated stops, but do not differ significantly from each other.

4.1.2 Notes on variability: An important source of the variability seen in Figure 3 is cross-speaker differences in speaking rate – specifically, 'articulation rate', which eliminates pause durations from the rate calculation and is now generally considered to be the most useful general characterisation of speaking rate (Lee and Doherty 2017). Mean articulation rate calculated over the first ten test sentences for all 20 speakers ranged from 3.95 to 5.78 syllables per second. Because our statistical model treated speaker as a random factor, and because individual variability is not under investigation here, we do not report a full analysis of the link between articulation rate and individual variability in the durational data, but simply note informally that across the twenty speakers there is a correlation between mean articulation rate and mean closure duration ($r = -.65$), i.e. the faster the articulation rate, the shorter the closure durations. These predictable effects of consonant place and articulation rate in our data suggest that the overall results are reliable and representative.

speaker	mean closure	mean VOT	VOT/closure
ZH17	100	40	0.39
ZH12	87	37	0.43
ZH06	76	34	0.45
ZH08	87	40	0.46
ZH11	89	47	0.53
ZH13	73	39	0.53
ZH19	91	48	0.53
ZH03	98	53	0.54
ZH07	66	36	0.54
ZH14	80	43	0.54
ZH20	89	51	0.57
ZH10	88	56	0.64
ZH05	80	52	0.65
ZH09	107	71	0.66
ZH16	88	58	0.66
ZH04	70	50	0.71
ZH18	64	51	0.79
ZH15	66	54	0.82
ZH01	71	59	0.83
ZH02	84	70	0.84

Table 3. Speaker means (in ms) of closure duration and VOT in aspirated stops, showing the range of variation in the ratio of the two values.

In this connection, however, we should mention an area of individual variability that does not seem to be explainable in terms of articulation rate. Individuals vary quite considerably in the proportional duration of VOT in the aspirated stops. That is, when expressed relative to the closure duration, VOT in aspirated stops is extremely variable between speakers, beyond any effect of articulation rate. The data are given in Table 3. We have no explanation for this variability, but we speculate that it may be related to a point of variability with CF0 in the aspirated stops. We return to this point in section 4.2.3, following the presentation of the CF0 data.

4.2. Obstruent F0 effects

4.2.1. Oral stops vs. nasals: Mean F0 contours accompanying the four different categories of consonant, based on the values for all twenty speakers, are plotted in Figure 4. Individual plots for each speaker separately are provided in Appendix 5 and full details of the statistical models are given in Appendix 3. Relative to the nasal reference contour, it can be seen that the three oral stop contours exhibit clear deviations at the beginning of the vowel: all begin higher than the nasal contour and then fall before beginning to rise. Expressed as *F0drop* in Hz, these differences from the nasal contour are all statistically significant, as can be seen in Table 4. There are also clear differences of overall level, expressed as *minF0*, but only the fortis and aspirated contours differ significantly from the nasal contour (Table 5). Presumably after the lenis stops the F0 reverts to near the intonational target level manifested in the nasal contour, whereas the fortis and aspirated contours remain higher.

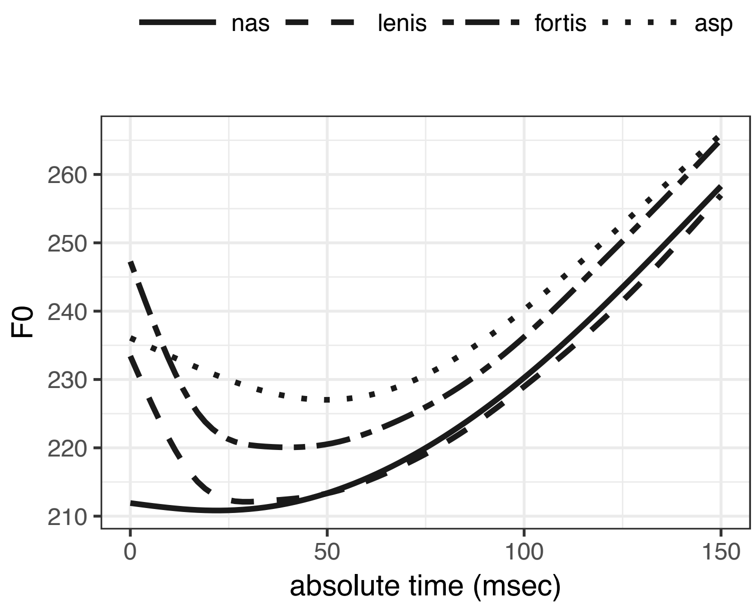


Figure 4. Summary pitch contour data based on all 20 speakers.

In Figure 4, both the lenis and the fortis stops show a qualitatively similar abrupt drop from the beginning of the vowel to the *minF0*. In the aspirated stops, *F0drop* appears somewhat less abrupt; as we shall see, this is due primarily to the existence of considerable individual variation in the aspirated stops. In the following subsection we focus first on the fortis and lenis contours.

C.type	mean	SE	contrast	est.	SE	df	t.ratio	p.value
nasal	1.26	1.11	nasal –lenis	-5.77	1.48	64.56	-3.91	0.001
lenis	7.03	1.26	nasal – fortis	-9.24	1.55	67.59	-5.95	<.0001
fortis	10.50	1.11	nasal – asp.	-10.77	2.82	28.10	-3.81	0.004
asp.	12.03	2.62	lenis – fortis	-3.47	1.52	69.33	-2.28	0.112
			lenis – asp.	-5.00	2.85	27.92	-1.76	0.315
			fortis – asp.	-1.53	2.77	24.23	-0.55	0.945

Table 4A Table 4B

Table 4. F0drop in Hz. Estimated marginal means (4A) and pairwise comparisons between consonant types (4B).

C.type	mean	SE	contrast	est.	SE	df	t.ratio	p.value
nasal	213.65	5.39	nasal –lenis	-3.99	3.57	66.67	-1.12	0.680
lenis	217.64	5.61	nasal – fortis	-11.66	3.44	82.30	-3.39	0.006
fortis	225.31	5.32	nasal – asp.	-14.89	3.58	83.64	-4.16	<.0001
asp.	228.54	5.45	lenis – fortis	-7.67	3.33	77.54	-2.30	0.106
			lenis – asp.	-10.90	3.50	83.19	-3.11	0.013
			fortis – asp.	-3.23	2.30	149.52	-1.40	0.499

Table 5A Table 5B

Table 5. minF0 in Hz. Estimated marginal means (5A) and pairwise comparisons between consonant types (5B).

4.2.2. Fortis vs. lenis: The abrupt initial drop seen in the fortis and lenis contours is similar to what is reported for contours that follow 'voiceless' stops in a number of other studies, including Chen 2011 (esp. Figure 3, Panels 2A and 2B), Hanson 2009 (esp. Figure 7, male subjects), and Kirby and Ladd 2016 (esp. Figure 6, Italian data). This seems consistent with the observation that both fortis and lenis stops are normally characterised by a period of complete voicelessness during the stop closure. On a strict interpretation of our statistical results, this would be our central finding. The differences between the fortis and lenis contours – in both *F0drop* and *minF0* – are not statistically significant (p = .112 and .106, respectively). We might therefore conclude that both fortis and lenis count as voiceless and exhibit typical voiceless CF0 effects, but that there is no difference between them.

However, this conclusion may be premature. Impressionistically, the beginning of the fortis contour is higher overall than the beginning of the lenis contour, and some of the pairwise comparisons indirectly support this comparison. As just noted, fortis *minF0* is significantly higher than nasal *minF0,* whereas lenis and nasal *minF0* do not differ significantly. That is, the patterns of differences between the means shown in Tables 4 and 5 are approximately what we would expect if fortis and lenis differed according to the apparently universal pattern of 'voiceless' and 'voiced' CF0 found by Sonderegger et al. in their cross-language survey. Inspection of individual plots (Appendix 5) suggests that a significant aspect of the overall results is that some speakers (e.g. ZH15, ZH18) show a clear difference of overall level – with fortis higher than lenis – and others (e.g. ZH02, ZH11) do not; none of the speakers shows a clear difference in the opposite direction.

On the other hand, certain aspects of the data seem to favour the more conservative interpretation of the statistical results. In particular, there is an interaction with vowel height that is plausibly the source of the impressionistically higher level in the fortis contours. Recall that we were unable to find enough test words with low vowels to fill out the design, and used high vowels in a few cases where necessary. Unfortunately, three of the fortis test words (*Tuume* 'thumb', *Puur* 'farmer', and *Puuder* 'powder') had high vowels, as against only one of each of the other consonant types (lenis *diich* 'you [accusative]', nasal *nüüni* 'nine o'clock', and aspirated *Team*). As can be seen in Figure 5, the effect of vowel height on *F0drop* and *minF0*, though not unanticipated (Whalen and Levitt 1995, Whalen et al. 1998), was unexpectedly large and may almost entirely account for the numerical difference in overall level. For now, then, it remains unclear whether we are justified in talking of an overall level difference between fortis and lenis CF0 to match Sonderegger et al.'s general pattern.

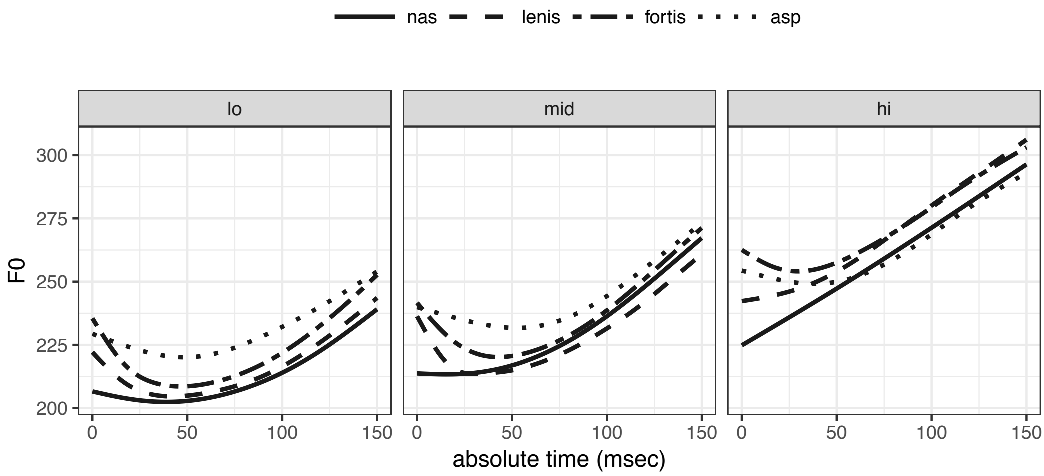


Figure 5. Effect of vowel height on CF0. NB: The high vowel plots are based on only six test words and may not be as representative as those for the mid and low vowels. For more detail see text.

<u>4.2.3. Aspirated stops:</u> For the contours that follow aspirated stops, the individual plots shows a good deal more inter-speaker variability than for the other three contours. We have no clear explanation for this variability, but given that the aspiration contrast is phonologically marginal and lexically variable, it may not be surprising that its phonetic realisation is also somewhat inconsistent. The variability can be roughly described in terms of three typical patterns, which can be seen from inspection of individual plots in Appendix 5. Specifically, some speakers (e.g. ZH06, ZH11, ZH19) show a sharp drop similar in shape to the fortis and lenis contours, and with an overall level even higher than after fortis; some (e.g. ZH02, ZH08, ZH12) show a much more gentle drop, generally starting at a level somewhat lower than the beginning of the drop that follows fortis stops and reaching a minimum somewhat higher than the minimum of the fortis contours; finally, four speakers (ZH01, ZH04, ZH16, ZH18) show first a <u>*rise*</u> in F0 from at or below the level of same speaker's fortis contour, then levelling out or falling again slightly before the beginning of the intonational rise.

However, it may be more appropriate to treat the variability not in terms of three distinct patterns but as a continuum. If we quantify the rising pattern of speakers like ZH04 and ZH18 simply as a negative value of *F0drop* (Appendix 2.iii), then the variation can be straightforwardly ordered from the largest drop to the largest rise, as shown in Table 6. This way of looking at things suggests a link between the variable CF0 and the variable ratio of VOT to closure duration reported in section 4.1.2 (Table 3). Our data, as we have already pointed out, are not well suited to the investigation of individual variation, but a simple analysis of the speaker means in Tables 3 and 6 shows a correlation of $r = -.40$, i.e. steeper falls in F0 are associated with proportionally shorter VOT values. This whole question certainly seems worthy of further research.

speaker	F0drop
ZH05	29.5
ZH19	29.4
ZH11	27.5
ZH20	22.3
ZH06	21.2
ZH09	15.4
ZH15	15.0
ZH07	14.8
ZH17	14.4
ZH03	14.1
ZH14	12.7
ZH02	10.3
ZH13	9.4
ZH12	8.5
ZH10	8.0
ZH08	6.6
ZH01	4.7
ZH16	-0.7
ZH04	-1.6
ZH18	-16.3

Table 6. Initial F0 drop (in Hz) in aspirated stops, showing the range of variation, which may be related to the variation in VOT shown in Table 3. Negative values here indicate an initial F0 rise. For more detail see text.

5. General Discussion

5.1. Summary

Our study makes clear that past auditory and instrumental phonetic descriptions of Swiss German are broadly accurate. We have confirmed that there is a clear distinction between what have often been termed fortis and lenis stops, which correspond to voiceless and voiced stops in many other West Germanic varieties, and that the distinction is not based on VOT but on other properties, notably closure duration, and possibly on CF0 effects. In addition, we have confirmed reports that a number of words beginning with stops spelled with <p> or <t> are quite consistently pronounced with aspiration, and we have shown that the acoustic details of the aspiration are rather variable, both in the proportional duration of the aspiration relative to the closure and in the patterns of CF0 effects found with aspirated stops. In the following sections we discuss some of the implications of these findings.

5.2. Obstruent effects on F0

First, our results appear to shed new light on the question of why CF0 effects occur at all. The Swiss German CF0 patterns are unlike those in English (Hanson 2009) or Italian and French (Kirby and Ladd 2016) in one crucial respect, namely the fact that the rapid initial drop in F0 applies to both lenis and fortis stops. In English, French and Italian, the beginning of the 'voiceless' contours diverges significantly from the nasal reference contour while the beginning of the 'voiced' contours does not. In Swiss German, by contrast, the initial drop in F0 in the lenis and fortis contours means that neither of them matches the nasal reference contour. We interpret this rapid initial drop in F0 as evidence for the idea, put forth in various forms by Halle and Stevens (1971), Löfqvist et al. (1989), and Hanson (2009), that the main source of the basic CF0 effect is the biomechanical consequence of a gesture to inhibit voicing, the effects of which take a short while to decay once phonation begins. This is consistent with the fact that all the fortis stops and almost all the lenis stops have a period of complete voicelessness. We suggest that both lenis and fortis stops are typically accompanied by some such voicing inhibition gesture.

We also believe that the interaction with vowel height discussed in section 4.2.2 and illustrated in Figure 5 potentially provides interesting evidence that the sources of CF0 effects are essentially biomechanical. In Figure 5 it appears that, in high vowel contours, pitch begins to rise immediately following nasal and lenis consonants, and *F0drop* is substantially reduced following fortis and aspirated stops. This suggests that the sources of vowel height differences in F0 (which Kingston 2007 refers to as VF0) may interact biomechanically with the sources of CF0 effects: after a lenis stop, for example, the relaxation of a voicing inhibition gesture that normally yields an abrupt drop in F0 may be overridden by the articulatory configurations that result in higher F0 with high vowels. This seems a potentially fruitful area for further research.

5.3. Aspiration

Our findings are also relevant to the assumption that aspiration is merely delayed voice onset – as Lisker and Abramson (1971: 774) put it, 'that voiceless aspiration is essentially no more than the consequence of delay in the resumption of the voicing position by the larynx'. This line of thinking appears to be widespread. Nevertheless, there are at least two reasons why this may not be a good idea. First, more detailed studies of VOT like Cho and Ladefoged's major comparative study of 18 languages (1999) suggest that the cross-linguistic clustering of VOT values reported by Lisker and Abramson (1964) is at best an over-simplification and possibly merely an artefact of an inadequate sample. More importantly, the possibility that aspirated stops might involve not only voicing lag but also additional glottal gestures during the period of aspiration could go some way to explaining contradictory findings in the literature about CF0 effects involving aspirated stops. Because of methodological differences, comparison across studies of CF0 in different languages is difficult (and in any case full discussion of this topic is well beyond the scope of our paper), but it appears that there may be consistent differences between languages that resemble the differences we find between speakers in our data.

Specifically, in some languages with three-way stop systems (for example, Korean: e.g. Silva 2006, Bang et al. 2018; or Madurese: Misnadin et al. 2015) aspirated stops tend to be followed by overall higher F0 and (often) an initial drop – comparable to the pattern shown by speaker ZH06 or ZH11 – while in others (for example, Shanghai Wu Chinese: Chen 2011) aspirated stops are followed by a rise from a low F0 level – as seen in the data from speaker ZH04 or ZH18. That is, just as 'unaspirated' stops can apparently be produced with or without a voicing inhibition gesture (as suggested by Hanson's 2009 results), there may be different articulatory strategies to create aspiration. Instrumental measurements of speech production would be necessary to determine whether different CF0 patterns reflect consistent articulatory differences; if so, this would provide another instance in which VOT alone is insufficient to describe the phonetic basis of potential 'laryngeal' distinctions in the world's languages.

5.4. The phonology of fortis and lenis

Finally, the Swiss German results would appear to show that Lisker and Abramson's rejection of fortis/lenis or tense/lax as a meaningful phonological dichotomy 'demonstrably independent of voicing' was premature, and that researchers like Kohler who continue to assume that fortis/lenis is independent of VOT have evidence on their side. However, because the Swiss German fortis/lenis distinction can also be analysed as a distinction between geminate and singleton obstruents, such a conclusion merits further discussion. In this final section, we briefly consider Kraehenmann's (2001) proposal – succinctly summarised in her title – that there are 'geminates all over the word'.

The most obvious potential objection to the geminate analysis is that it is typologically rare to find word-final and especially word-initial geminates. The most obvious rebuttal to such an objection is that a featural contrast based on closure duration with no VOT difference, especially given the additional presence of words that consistently have aspirated stops, is also typologically unusual. One way or another, that is, Swiss German is typologically odd. The question is 'odd in what way?'

Superficially, the geminate analysis has phonetic transparency in its favour. The most conspicuous and apparently most salient perceptual correlate of the distinction is closure duration, which maps very simply on to a contrast between single and geminate segments and seems comparable to what we find phonetically in languages with uncontroversial consonant gemination like Italian or Finnish. But duration data are actually difficult to relate directly to phonological status, in any language, and many durational phenomena can be interpreted in different ways. In part, this is because segment duration is affected by many interacting factors, such as stress and position in word (usefully catalogued by Turk et al. 2006); in part, it is because phonological gemination can be reflected in many ways other than phonetic duration. For example, consider D'Imperio and Rosenthall's discussion of vowel duration in Italian (1999), which attributes the uncontroversial phonetic fact that stressed vowels are longer than unstressed vowels to a ‘phonological’ constraint based on syllable weight in penultimate syllables and to ‘phonetic’ effects in other positions. Or again, consider the general morphophonological process of ‘consonant gradation’ in Finnish: in some cases this is a matter of alternations based on duration (often analysed as gemination) and in others involves differences in voicing and manner of articulation (Karlsson 2013: 28-37). Many languages, in short, have distinctions involving interactions between segment duration and other phonetic properties that are hard to fit unambiguously into a phonological analysis. In many cases these are precisely the distinctions for which the oppositions fortis/lenis or tense/lax have been proposed.

Both the fortis/lenis analysis and the geminate analysis of the Swiss German stops agree on the existence of a phonological distinction whose primary phonetic basis is closure duration. Whether this distinction is analysed phonologically in terms of features (e.g. fortis/lenis) or positions in structure (e.g. geminate/singleton) depends on a variety of considerations, very few of which are based exclusively on the phonetic properties of the segments in question. In the end, that is, we may need to acknowledge that the difference between the geminate analysis and the fortis/lenis analysis of Swiss German obstruents is in many respects an argument about terminology. What seems clear from our study, however, is that this distinction – whatever we call it – really is ‘demonstrably independent of voicing’.

Acknowledgements

We thank Dieter Studer-Joho and James Kirby for much practical assistance with data processing and statistical analysis; James Kirby and Pavel Iosad for discussion; and audiences in Edinburgh and Zurich and at PaPE 2017 in Cologne for comments and criticisms. We also thank Marion Brändle for carrying out a pilot experiment in a course project in autumn 2014, when the first author was a visiting professor in Zurich, which suggested that the F0 difference between fortis and lenis was worth rigorous investigation. Finally, we thank the editor-in-chief and three anonymous referees for extensive comments that led to substantial improvements in the exposition.

Appendix 1: Test sentences

The number before each sentence indicates its position in the order of presentation. The test word in each sentence is underlined.

Ia. Lenis, labial

28. Chunt si vo Bèèrn oder chunt si vo Züri? *'Does she come from Bern or Zurich?'*
41. Faart er mit em Boot oder mit em Zuug? *'Is he travelling by boat or train?'*
46. Schafft er uf em Bou oder i de Fabrik? *'Does he work in construction or in a factory?'*
51. Häsch lieber Boone oder Èrpsli? *'Do you prefer beans or peas?'*
54. Isch das en Bèèr oder en Wolf? *'Is that a bear or a wolf?'*
69. Spilsch mit em Bääbi oder mit em Chaschperli? *'Are you playing with the doll or the puppet?'*
71. Hät er en Baart oder en Schnauz? *'Does he have a beard or a moustache?'*
82. Häsch lieber Beeri oder Truube? *'Do you prefer berries or grapes?'*
84. Hebsch en am Bäi oder am Arm? *'Are you holding his leg or his arm?'*
87. Gönd er mit de Baan oder mit em Auto? *'Are you going by train or by car?'*

Ib. Lenis, coronal

22. Findsch das doof oder findsch es cool? *'Do you think that's silly or cool?'*
24. Chunt moorn e Daame oder en Herr? *'Is a lady coming tomorrow, or a man?'*
26. Woned s im ene Doorf oder in ere Schtadt? *'Do they live in a village or a town?'*
39. Häisst si Doris oder Marlies? *'Is her name Doris or Marlies?'*
64. Lueged mer dèè Film oder en andere? *'Shall we watch this film or another one?'*
66. Chaufsch s Bier i de Doose oder i de Fläsche? *'Do you buy beer in cans or bottles?'*
70. Ässe mer daa oder uf em Balkon? *'Shall we eat here or on the balcony?'*
72. Isch das für diich oder für miich? *'Is this for you or for me?'*
76. Zäiged s de Film deet oder daa? *'Are they showing the film there or here?'*
100. Isch s Probleem de Darm oder de Mage? *'Is the problem the intestine or the stomach?'*

IIa. Fortis, labial

32. Nimmt er e Päitsche oder en Schtock? *'Is he taking a whip or a walking stick?'*
40. Isch das en Panda oder en Bruunbèèr? *'Is that a panda or a brown bear?'*
45. Isch das us Pelz oder us Schtoff? *'Is that made of fur or fabric?*
56. Sind di bäide es Pèèrli oder Fründ? *'Are they a couple or just friends?'*
75. Isch er en Puur oder en Arbäiter? *'Is he a farmer or a workman?'*
80. Wont deet es Paar oder e Familie? *'Is it a couple or a family living there?*
89. Hät s deet es Poort oder e Muur? *'Is there an embankment there or a wall?'*
91. Hät er en Porsche oder en Mercedes? *'Does he have a Porsche or a Mercedes?'*
93. Hät si en Page-Schnitt oder langi Haar? *'Does she have a pageboy haircut or long hair?'*
95. Bruuchsch es Puuder oder e Grèème? *'Do you need a powder or a cream?'*

IIb. Fortis, coronal

31. Gönd er zwäi Tääg oder zwäi Wuche? *'Are you going for two days or two weeks?'*
36. Gaat s en Taag oder e Schtund? *'Is she gone for a day or an hour?'*
37. Isch dèè Chrueg us Toon oder us Lèèm? *'Is that jug made of potter's clay or earthenware?'*
42. Macht si en Täigg oder e Soosse? *'Is she making a batter or a sauce?'*
44. Isch er taub oder blind? *'Is he deaf or blind?'*
47. Isch er toot oder läbig? *'Is he dead or alive?'*
53. Isch s Doorf im Taal oder uf em Bèèrg? *'Is the village in the valley or on the mountain?'*

57. Machsch es mit em Tuume oder mit dem Zeigfinger? *'Do you use your thumb or your forefinger?'*
59. Isch das diini Tante oder diini Cousine? *'Is that your aunt or your cousin?'*
81. Fiired s e Taufi oder e Hoochziit? *'Are they celebrating a baptism or a wedding?'*

IIIa. Aspirated, labial
23. Schickt er es Pack oder en Brief? *'Is he sending a package or a letter?'*
25. Schpile mer Poker oder Bridge? *'Shall we play poker or bridge?'*
30. Häisst din Vater Paul oder Fredi? *'Is your father called Paul or Fredi?'*
34. Hät s deet en Park oder en Platz? *'Is there a park there or a square?'*
43. Hät si Paanik oder Schtress? *'Is she suffering from panic or stress?'*
52. Isch si vo Poole oder vo Russland? *'Is she from Poland or Russia?'*
77. Gaat si an e Party oder is Kino? *'Is she going to a party or the cinema?'*
78. Hät er Pause oder Fiiraabig? *'Is he on break or is he finished for the day?'*
86. Schpilt er Pauke oder Trumle? *'Does he play kettledrum or snare drum?'*
97. Häisst diin Fründ Peter oder Hans? *'Is your friend's name Peter or Hans?'*

IIIb. Aspirated, coronal
27. Hät si en Theek oder e Mappe? *'Does she have a knapsack or a briefcase?'*
50. Gaat s um de Takt oder um de Rhythmus? *'Is it about the beat or the rhythm?'*
55. Schafft er im ene Team oder aläige? *'Does he work in a team or on his own?'*
62. Häsch lieber Tee oder Kafi? *'Do you prefer tea or coffee?'*
68. Isch das Timing oder Glück gsii? *'Was it timing or good luck?'*
73. Tüend s t Schtraass teere oder pflaschtere? *'Are they paving the street with tarmac or cobbles?'*
74. Häisst diin Cousin Theo oder Otto? *'Is your cousin's name Theo or Otto?'*
94. Isch es uf em ene Tape oder uf ere Schallplatte? *'Is it on a tape or a vinyl record?'*
96. Isch das e These oder e Fraag? *'Is that a thesis or a question?'*
99. Schtaat das i de Thora oder im Koran? *'Is that in the Torah or the Koran?'*

IVa. Nasal, labial
21. Tuet si lieber maale oder zäichne? *'Does she prefer painting or drawing?'*
38. Interessiersch di für Moode oder für Kunscht? *'Are you interested in fashion or art?'*
61. Choschtet säb mee oder weniger? *'Does that cost more or less?'*
63. Isch t Hoochziit im Mäi oder im Juni? *'Is the wedding in May or June?'*
65. Hät s am Bode Moos oder Laub? *'Is the ground covered with moss or leaves?'*
67. Isch das en Maa oder e Frau? *'Is that a man or a woman?'*
79. Isch das e Maanig oder e Rächnig? *'Is that a reminder or a bill?'*
90. Gönd er as Meer oder i t Bèèrg? *'Are you going to the seaside or the mountains?'*
92. Flüüged s uf de Moond oder uf de Mars? *'Are they flying to the moon or Mars?'*
98. Chömed er am Mèèntig oder am Ziischtig? *'Are you coming Monday or Tuesday?'*

IVb. Nasal, coronal
29. Isch s Reschtorand nööch oder wiit? *'Is the restaurant near or far?'*
33. Säit me dèm e Naat oder e Fuege? *'Would you call that a seam or a joint?'*
35. Bruuchsch e Naadle oder e Gufe? *'Do you need a needle or a pin?'*
48. Hät si näi oder jaa gsäit? *'Did she say no or yes?'*
49. Isch das Buech nöi oder alt? *'Is the book new or old?'*
58. Isch es en Nèrv oder en Muskel? *'Is it a nerve or a muscle?'*
60. Zaled si mit ere Noote oder mit de Chaarte? *'Are they paying with a banknote or a card?'*
83. Chunt er am Nüüni oder am Zäni? *'Is he coming at nine or at ten?'*
85. Tuet si lieber nèèje oder lisme? *'Does she prefer sewing or knitting?'*
88. Gaat s um s nèè oder ums gèè? *'Is it a matter of taking or giving?'*

Appendix 2: Further methodological details on acoustic measurements

i. Reliability: Except for two speakers whose recordings were segmented entirely by the first author, each author segmented half the utterances of each speaker; the speakers' recordings were divided alternately so that neither author's work was concentrated on the first half or the second half of the materials. As a reliability check, both authors independently segmented all 80 sentences from one speaker (ZH07), and the duration measures (closure duration and VOT) resulting from the two segmentations were compared. Differences were expressed by subtracting DRL's measures from SS's; a positive value thus means that SS's was larger and a negative value means that DRL's was larger. For closure duration (measured on all 80 test sentences), the mean difference was +.0375 ms, the median was -1ms, and the differences ranged from -14 to +15 ms. Only five tokens showed a difference greater than ±10 ms. For VOT (measured on the 60 test sentences involving oral stops) the mean difference was -.433 ms, the median was 0, and there was only one outlier with a difference greater than ±10 ms. There were no obvious patterns by consonant type or place of articulation.

ii. Exceptional segmentation criteria: There were two exceptions to the waveform-based segmentation principles sketched in section 3.4. First, in placing the onset-of-closure boundary, we had to make an arbitrary decision in the case of nasal test consonants preceded by a nasal (e.g. *im Mäi* 'in May', *am Nüüni* 'at nine o'clock'). On the basis of rough duration comparisons between such nasal-nasal sequences and nasals preceded by vowels, we placed the boundary for the beginning of the nasal test consonant one-third of the way into the nasal-nasal closure interval. Second, in a small number of cases of lenis stops preceded by a nasal (e.g. *am Bäi* 'on the leg') or (in a few speakers) by a vowel (e.g. *e Daame* 'a lady'), there was residual voicing throughout the closure. Rather than record a closure duration of zero in such cases, we relied on a combination of the waveform and the spectrogram to estimate the beginning of the closure. An illustration is given in Figure A. For purposes of computing VOT, we did not treat these cases as involving a long period of prevoicing; instead, we equate VOT with the 'R' interval corresponding to the release burst.

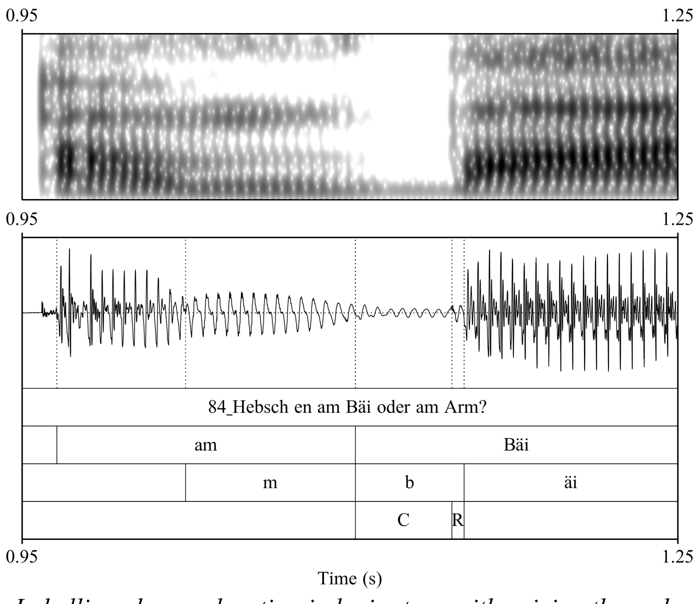


Figure A. Labelling closure duration in lenis stops with voicing throughout.

iii. Identification of *minF0* and calculation of *F0drop*: In the great majority of the items, the series of F0 values for the first 150 ms of the stressed vowel of the test word showed an initial drop followed by a steep rise, as shown in all four mean contours in Figure 4. In these cases *minF0* was defined simply as the lowest F0 value in the series, and *F0drop* as the difference between the first F0 value and *minF0.* This definition of *minF0* was also used in cases that showed only a steady rise from the first F0 value (as illustrated by the mean nasal contour in the high vowel panel of Figure 5), in which case the value of *F0drop* was zero. In both of these cases, *minF0* occurs at approximately the beginning of the intonational pitch rise and corresponds to what might be analysed as the L* of a L*+H pitch accent. However, as discussed in section 4.2.3 and illustrated in the individual plots in Appendix 5, for some speakers the contours with aspirated stops began with an initial rise and then levelled off or dropped again before the beginning of the intonational rise in pitch that covered the rest of the test word. In these cases we defined *minF0* so as to pick out the beginning of the intonational rise: in any contour where the F0 series during the first 100 ms showed an initial rise followed by a fall – however slight the rise and fall – *minF0* was taken to be the lowest F0 value *after* the initial rise. Since *F0drop* was still defined as the difference between the first F0 value and *minF0,* this frequently resulted in negative values (i.e. overall rises) for *F0drop* in contours that showed this pattern.

Appendix 3. Statistical models

In all analyses, consonant type ('C.type.actual'), vowel height ('V.height'), and place of articulation ('C.place') are treatment coded, with reference levels nasal, low, and coronal, respectively. Except where noted, results are averaged over levels of C.place. Fractional degrees of freedom were computed using the Kenward-Roger method, and p-values were adjusted using the Tukey method for comparing a family of estimates.

i. ***F0drop*** and ***minF0***

Fixed effect coefficient estimates, F0drop

term	estimate	Std. error	t value
(Intercept)	4.54	1.38	3.29
C.type.actuallenis	5.15	1.79	2.87
C.type.actualfortis	10.03	1.80	5.58
C.type.actualasp	7.20	3.07	2.34
V.heightmid	-2.17	1.50	-1.44
V.heighthi	-4.54	3.09	-1.47
C.placelab	-1.76	1.47	-1.20
V.heightmid:C.placelab	0.33	1.32	0.25
V.heighthi:C.placelab	-1.25	3.10	-0.40
C.type.actuallenis:V.heightmid	1.44	1.77	0.81
C.type.actualfortis:V.heightmid	0.09	1.88	0.05
C.type.actualasp:V.heightmid	0.73	1.75	0.42
C.type.actuallenis:V.heighthi	1.30	4.09	0.32
C.type.actualfortis:V.heighthi	-1.64	3.80	-0.43
C.type.actualasp:V.heighthi	3.91	3.85	1.02
C.type.actuallenis:C.placelab	-0.53	1.71	-0.31
C.type.actualfortis:C.placelab	-0.67	1.83	-0.36
C.type.actualasp:C.placelab	4.02	1.72	2.33

Pairwise comparisons, F0drop

contrast	V.height	estimate	SE	df	t.ratio	p.value
nas - lenis	lo	-4.88	1.51	66.17	-3.24	0.010
nas - fortis	lo	-9.69	1.61	64.18	-6.01	< 0.0001
nas - asp	lo	-9.21	2.85	28.27	-3.22	0.016
lenis - fortis	lo	-4.81	1.56	68.27	-3.09	0.015
lenis - asp	lo	-4.32	2.87	28.26	-1.51	0.446
fortis - asp	lo	0.49	2.83	26.57	0.17	0.998
nas - lenis	mid	-6.32	1.19	57.00	-5.29	< 0.0001
nas - fortis	mid	-9.79	1.59	64.68	-6.17	< 0.0001
nas - asp	mid	-9.94	2.75	24.79	-3.61	0.007
lenis - fortis	mid	-3.47	1.53	69.18	-2.27	0.115
lenis - asp	mid	-3.62	2.76	24.73	-1.31	0.567
fortis - asp	mid	-0.15	2.83	26.78	-0.05	1.000
nas - lenis	hi	-6.19	3.85	64.08	-1.61	0.381
nas - fortis	hi	-8.06	3.70	78.52	-2.18	0.138
nas - asp	hi	-13.12	4.39	79.61	-2.99	0.019
lenis - fortis	hi	-1.87	3.68	77.08	-0.51	0.957
lenis - asp	hi	-6.93	4.39	79.40	-1.58	0.397
fortis - asp	hi	-5.06	3.76	71.45	-1.34	0.538

Fixed effect coefficient estimates, minF0

term	estimate	Std. error	t value
(Intercept)	201.18	5.42	37.10
C.type.actuallenis	4.72	4.36	1.08
C.type.actualfortis	5.89	4.00	1.47
C.type.actualasp	15.40	4.47	3.44
V.heightmid	9.53	3.70	2.57
V.heighthi	26.47	7.55	3.51
C.placelab	-1.32	3.70	-0.36
V.heightmid:C.placelab	2.01	3.28	0.61
V.heighthi:C.placelab	4.80	7.45	0.64
C.type.actuallenis:V.heightmid	-1.60	4.47	-0.36
C.type.actualfortis:V.heightmid	4.04	4.33	0.93
C.type.actualasp:V.heightmid	0.48	4.20	0.11
C.type.actuallenis:V.heighthi	6.28	10.31	0.61
C.type.actualfortis:V.heighthi	13.78	9.25	1.49
C.type.actualasp:V.heighthi	2.37	9.31	0.25
C.type.actuallenis:C.placelab	-4.58	4.32	-1.06
C.type.actualfortis:C.placelab	-0.34	4.27	-0.08
C.type.actualasp:C.placelab	-2.93	4.16	-0.70

Pairwise comparisons, minF0

contrast	V.height	estimate	SE	df	t.ratio	p.value
nas - lenis	lo	-2.43	3.61	66.74	-0.67	0.907
nas - fortis	lo	-5.72	3.37	86.98	-1.70	0.332
nas - asp	lo	-13.94	3.59	87.30	-3.88	0.001
lenis - fortis	lo	-3.29	3.26	82.49	-1.01	0.745
lenis - asp	lo	-11.51	3.52	87.26	-3.27	0.008
fortis - asp	lo	-8.22	2.66	184.95	-3.09	0.012
nas - lenis	mid	-0.83	2.78	67.65	-0.30	0.991
nas - fortis	mid	-9.76	3.18	99.39	-3.07	0.015
nas - asp	mid	-14.42	3.09	80.12	-4.66	< 0.0001
lenis - fortis	mid	-8.93	3.07	95.89	-2.91	0.023
lenis - asp	mid	-13.59	3.01	82.09	-4.52	< 0.0001
fortis - asp	mid	-4.65	2.52	214.82	-1.85	0.253
nas - lenis	hi	-8.71	9.63	64.08	-0.90	0.803
nas - fortis	hi	-19.50	8.77	73.32	-2.22	0.126
nas - asp	hi	-16.31	8.84	75.37	-1.84	0.261
lenis - fortis	hi	-10.80	8.73	72.03	-1.24	0.606
lenis - asp	hi	-7.60	8.82	74.50	-0.86	0.824
fortis - asp	hi	3.19	5.05	494.26	0.63	0.922

ii. *closure duration* and *VOT*

Fixed effect coefficient estimates, closure duration

term	estimate	Std. error	t value
(Intercept)	86.47	5.27	16.42
C.type.actuallenis	-35.52	5.30	-6.71
C.type.actualfortis	12.97	5.41	2.40
C.type.actualasp	-8.33	5.50	-1.51
C.placelab	2.16	5.13	0.42
C.type.actuallenis:C.placelab	-8.57	6.98	-1.23
C.type.actualfortis:C.placelab	-3.27	6.86	-0.48
C.type.actualasp:C.placelab	3.83	6.61	0.58

Pairwise comparisons, closure duration.

contrast	estimate	SE	df	t.ratio	p.value
nas - lenis	39.81	3.98	60.83	9.99	< 0.0001
nas - fortis	-11.33	4.21	56.25	-2.69	0.045
nas - asp	6.42	4.22	52.35	1.52	0.432
lenis - fortis	-51.14	4.08	59.43	-12.54	< 0.0001
lenis - asp	-33.39	3.70	63.89	-9.02	< 0.0001
fortis - asp	17.75	3.13	83.21	5.67	< 0.0001

Fixed effect coefficient estimates, VOT

term	estimate	Std. error	t value
(Intercept)	14.14	1.71	8.29
C.type.actualfortis	-0.51	2.22	-0.23
C.type.actualasp	37.82	3.15	12.01
C.placelab	-3.38	2.40	-1.40
C.type.actualfortis:C.placelab	4.98	3.13	1.59
C.type.actualasp:C.placelab	-2.38	3.03	-0.79

Pairwise comparisons, VOT.

contrast	estimate	SE	df	t.ratio	p.value
lenis - fortis	-1.98	1.57	59.69	-1.26	0.423
lenis - asp	-36.62	2.70	33.49	-13.58	< 0.0001
fortis - asp	-34.65	2.59	28.51	-13.37	< 0.0001

Between-place pairwise comparisons, VOT.

contrast	C.type.actual	estimate	SE	df	t.ratio	p.value
cor - lab	lenis	3.38	2.40	62.55	1.40	0.165
cor - lab	fortis	-1.60	2.28	91.57	-0.70	0.484
cor - lab	asp	5.76	2.14	79.43	2.69	0.009

Appendix 4: Lexical distribution of aspirated stops

As discussed in section 2.3, the phonological status of aspirated stops in present-day Swiss German is puzzling. It is tempting to treat their marginally contrastive status as transitional, i.e. as reflecting some sort of overall change in progress. There is indeed plenty of evidence that change of some sort is underway, which we summarise here.

Although aspirated stops have been attested in Swiss German since the second half of the 19th century and are described in much of the dialectological literature of the 20th century (for an overview see Schifferle 2010), they continue to have characteristics of an innovation that is still spreading. For example, this is an obvious interpretation of the areal distribution of the variants of the word *Theek* 'school backpack' documented in the Swiss German Linguistic Atlas (SDS): [th] dominates in the North and in the Midlands, whereas [t] appears more frequently in the South-East and in the Alpine regions which are known for their more conservative dialects (cf. Schifferle 2010: 51). More generally, intergenerational variability in the use of aspirated stops is documented in Wolfensberger's (1967: 83) survey of 72 informants from the village of Stäfa in the Canton of Zurich, where the traditional unaspirated form of the given name *Peter* prevails among the older speakers but is less frequent among the middle-aged and almost absent among the younger generation. There is also variability in the speech of individual speakers: Schifferle (2010: 43) reports observing both aspirated and unaspirated pronunciations of the word *Tämperatuure* 'temperatures' spoken within the space of a few minutes by the same TV weather forecast presenter. All of this suggests that we are dealing with the lexical diffusion of a sound change (Wang 1969, Labov 2007).

Our own data, as we pointed out in section 3.4, show that most of our 20 speakers used aspirated stops in some words that the second author – a generation older than our speakers – had intended as unaspirated (fortis) test words. Only one speaker agreed exactly with the second author's original categorisation; the remaining 19 speakers all aspirated *Panda* 'panda'. *Päitsche* 'whip' was aspirated in seven cases, *Porsche* 'Porsche' in five, *Page-Schnitt* 'pageboy haircut' in four, *Poort* 'embankment' and *Puuder* 'powder' in three, and *Pelz* 'fur' in two. (It should be noted that some of the speakers were not entirely sure of the meaning of *Poort* or *Page-Schnitt.*) In addition, there was one aspirated production each of *Toon* 'potter's clay', *Taal* 'valley', *Paar* 'couple', and *Pèèrli* 'couple [diminutive]', all by the same speaker. There were only two cases of the opposite divergence from the original categorisation, i.e. unaspirated (fortis) pronunciation of a test word intended as aspirated: one instance of the name *Theo* and one of *teere* 'pave'. This asymmetry is consistent with the conclusion that the use of aspiration is spreading to new lexical items. We may also note that the more variable lexical items are in general less frequent (e.g. *Päitsche, Poort, Page-Schnitt*) and/or more obviously foreign (e.g. *Panda, Porsche*) than the words that were pronounced without aspiration by all speakers (e.g. *Puur* 'farmer', *Tante* 'aunt', *Taag* 'day'). The fact that the words beginning with orthographic <p> are more variable than those beginning with orthographic <t> is probably due to the very fact that most of them are either rare or foreign or both.

Because our speakers produced only one token of each test sentence, we cannot investigate variability within speakers, but we can make one potentially relevant observation about variation between speakers. By chance, our group of experimental participants consisted of nearly equal-sized groups of speakers with two Swiss German parents ($n = 11$) and speakers with at least one non-Swiss German parent ($n = 9$). Speakers with two Swiss German parents generally aspirated fewer words than those with only one or none: the mean number of aspirated stops in the former group was 21.4 and in the latter group 23.6 ($t = 2.27$, $df = 18$, $p < .05$ two-tailed). This apparent influence of an individual's parents' linguistic background echoes findings in sociolinguistic studies of sound change (e.g. Payne 1980, Labov 2007).

Nevertheless, the case of aspiration in Swiss German stops remains unusual. For those who believe in lexical diffusion as the main mechanism of sound change, an obvious problem is that the change seems to affect primarily low-frequency words. For those who assume that regular sound change largely follows Neogrammarian principles and that 'lexical diffusion is the result of the abrupt substitution of one phoneme for another in words that contain that phoneme' (Labov 2010: 260), the problem is precisely that the phonemic status of the aspirated stops is unclear. For either position, the fact that the change has apparently been going on for at least a century and a half – considerably longer than a human lifespan – means that speakers necessarily acquire a phonological system that is obviously in flux. We therefore believe that careful sociophonetic study of the apparent spread of aspirated stops in Swiss German has the potential to provide considerable insight into the mechanisms of lexical diffusion and perhaps into the nature of phonemic contrast itself.

Appendix 5: Individual F0 plots

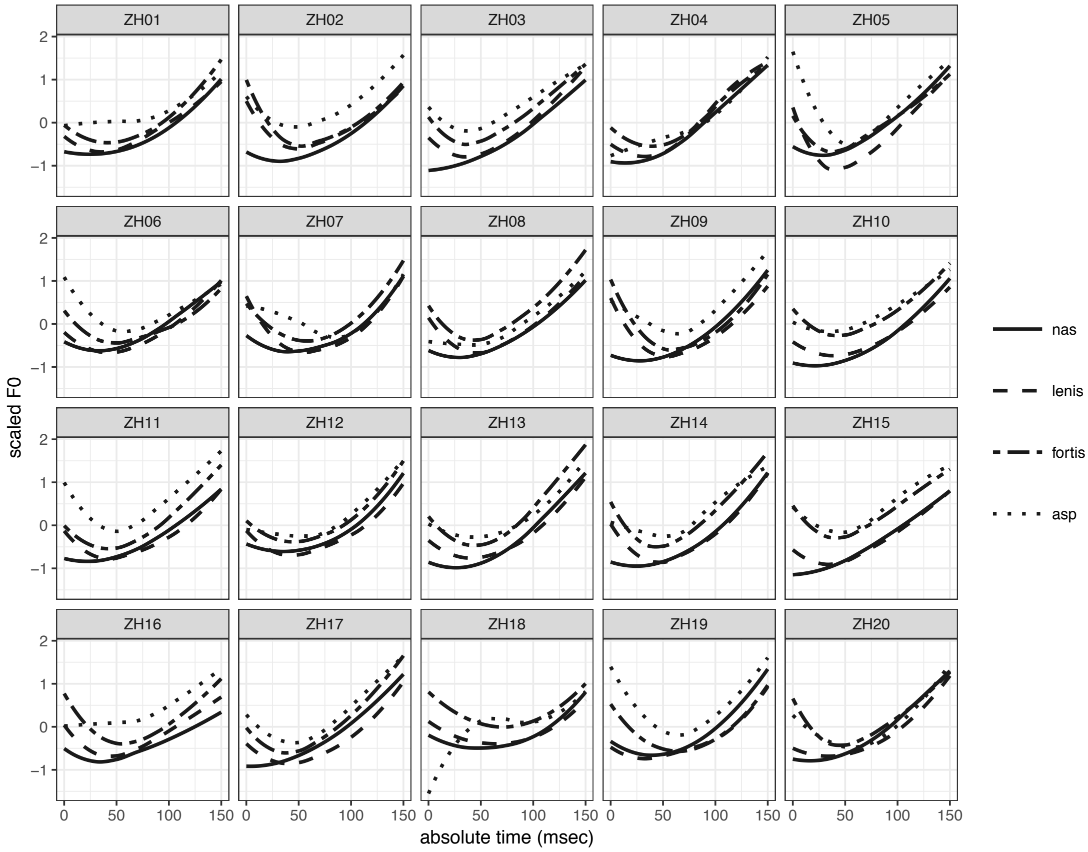


Figure B. Individual F0 plots for all twenty speakers. To normalise for individual differences of overall pitch range, the y-axes show F0 expressed in terms of individual z-scores.

References

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica*, 34, 295-303.
Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics,* 63, 75–86.
Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of ‘segmental anchoring’ of F0: Evidence from German. *Journal of Phonetics,* 32, 177–197.
Bang, H., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, 66, 120-144.
Bates, D. M., Mächler, M., Bolker, B., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67. https://doi.org/10.18637/jss.v067.i01.
Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs true voice languages. *Journal of Linguistics,* 49, 259–284.
Boersma, P., & Weeninck, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.09, retrieved 1 January 2016 from http://www.praat.org/
Braun, A. (1988). *Zum Merkmal “Fortis/Lenis”. Phonologische Untersuchungen und instrumentalphonetische Untersuchungen an einem mittelhessischen Dialekt.* Stuttgart: Steiner.
Catford, J. C. (1977). *Fundamental problems in phonetics.* Bloomington: Indiana University Press.
Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics,* 39, 612–625.
Cho, T., Jun, S.-A. and Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30, 193-228.
Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27, 207–229.
Connell, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics,* 30, 101–129.
Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54, 35-50
DiCanio, C. (2012). The phonetics of fortis and lenis consonants in Itunyoso Trique. *International Journal of American Linguistics*, 78, 239–272.
Dieth, E. (1950). *Vademekum der Phonetik.* Bern: Francke.
Dieth, E. (1986). *Schwyzertütschi Dialäktschrift: Dieth-Schreibung* (2nd edition, revised by Christian Schmid-Cadalbert). Aarau: Verlag Sauerländer.
Dieth, E., & Brunner, R. (1943). Die Konsonanten und Geminaten des Schweizerdeutschen experimental untersucht. *Romanica Helvetica*, 20, 757–762.
D’Imperio, M. & Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology*, 16, 1–28.
Enstrom, D., & Spörri-Bütler, S. (1981). A voice onset time analysis of initial Swiss German stops. *Folia Phoniatrica*, 33, 137–150.
Ferguson, C. (1959). Diglossia. *Word*, 15, 325–340.
Fitzpatrick-Cole, J. (1999). The alpine intonation of Bern Swiss German. *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 941–944). Berkeley: University of California.
Fleischer, J., & Schmid, S. (2006). Zurich German. *Journal of the International Phonetic Association,* 36, 243–253.
Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *Journal of the Acoustical Society of America,* 120, 2884–2895.
Hall, K.C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review,* 30, 1–14.
Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT Quarterly Progress Report*, 101, 198–212.
Ham, W. (2001). *Phonetic and phonological aspects of geminate timing.* London: Routledge.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America,* 125, 425–441.
Heusler, A. (1888). *Der alemannische Consonantismus in der Mundart von Baselstadt.* Strassburg: Trübner.
Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language,* 55, 37–58.
House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustic characteristics of vowels. *Journal of the Acoustical Society of America,* 25, 105–113.
Jaeger, J. J. (1983). The fortis/lenis question: evidence from Zapotec and Jawoñ. *Journal of Phonetics,* 11, 177–189.
Jakobson, R. & Halle, M. (1964). Tenseness and laxness. In Abercrombie, D. et al. (Eds.), *In honour of Daniel Jones* (pp. 96–101). London: Longmans.
Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: Benjamins.
Karlsson, F. (2013). *Finnish: An essential grammar*. London: Routledge.
Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language,* 70, 419–454.
Kingston, J. (2007). Segmental influences on F0: Automatic or controlled? In C. Gussenhoven, & T. Riad (Eds.). *Tones and Tunes. Volume 2: Experimental Studies in Word and Sentence Prosody* (pp. 171–201). Berlin: Mouton de Gruyter.
Kiparsky, Paul. (2014). New perspectives in historical linguistics. In C. Bowern & B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics* (pp. 64–102). London: Routledge.
Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *Journal of the Acoustical Society of America,* 140, 2400–2411.
Kohler, K. J. (1982). F0 in the production of fortis and lenis plosives. *Phonetica,* 39, 199–218.
Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica,* 41, 150–174.
Kohler, K. J. (1985). F0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America,* 78, 21–32.
Kohler, K. J. (2012). From phonetics to phonology and back again. *Phonetica,* 69, 254–273.
Kraehenmann, A. (2001). Swiss German stops: geminates all over the word. *Phonology, 18*, 109–145.
Kraehenmann, A. (2003). *Quantity and prosodic asymmetries in Alemannic: synchronic and diachronic perspectives.* Berlin: Mouton de Gruyter.
Kraehenmann, A., & Lahiri, A. (2008). Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *Journal of the Acoustical Society of America,*123, 4446–4455.
Labov, W. (1981). Resolving the Neogrammarian controversy. *Language,* 57, 267–309.
Labov, W. (1994). *Principles of language change*. *Volume 1: Internal factors*. Oxford: Blackwell.
Labov, W. (2007). Transmission and diffusion. *Language,* 83, 344-387.
Labov, W. (2010). *Principles of language change*. *Volume3: Cognitive and cultural factors*. Oxford: Blackwell.
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford: Blackwell.
Lee, A., & Doherty, R. (2017). Speaking rate and articulation rate of native speakers of Irish English. *Speech, Language and Hearing*, 20, 206–211.
Lee, H. B. (1999). Korean. In *Handbook of the International Phonetic Association* (pp. 120-123). Cambridge: Cambridge University Press.
Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America,* 33, 419–425.
Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.1.2. https://CRAN.R-project.org/package=emmeans
Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word,* 20, 384–422.
Lisker, L., &. Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language*, 47, 767–785.

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.

Misnadin, Remijsen, B., & Kirby, J. (2015). Temporal and spectral properties of Madurese stops. *Proceedings of the 18th International Congress of Phonetic Sciences.* Paper 0789. Glasgow, University of Glasgow.

Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In W. Labov (ed.), *Locating language in time and space* (pp. 143-178), New York: Academic Press.

R Core Development Team. (2016). *R: A language and environment for statistical computing*, version 3.3.2. Retrieved from http://www.R-project.org/.

Renwick, M. E. L., & Ladd, D. R. (2016). Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology,* 7, 1–29.

Schifferle, H. (2010). Zunehmende Behauchung. Aspirierte Plosive im modernen Schweizer-deutsch. In H. Christen et al. (Eds.), *Alemannische Dialektologie: Wege in die Zukunft* (pp. 43–55). Stuttgart: Steiner.

Schmid, S. (2004). Zur Vokalquantität in der Mundart der Stadt Zürich. *Linguistik online*, 20(3): 93–116.

Schmid, S., Studer, D., & Dellwo, V. (2011). Aerodynamische Korrelate von fortis- und lenis-Verschlusslauten im Zürichdeutschen. Paper presented at the *17th Arbeitstagung zur alemannischen Dialektologie*. Strasbourg, October 2011.

SDS: *Sprachatlas der deutschen Schweiz*. Bern (I-VI)/Basel: Francke (VII-VIII), 1962-2003.

Sievers, E. (1876). *Grundzüge der Lautphysiologie zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen.* Leipzig: Breitkopf und Härtel.

Silva, D. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology,* 23, 287–308.

Silverman, K. E. A. (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica,* 43, 76–91.

Sonderegger, M., McAuliffe, M., & Bang, H.-Y. (2017). Segmental influences on F0: A large-scale study of cross-linguistic and interspeaker variability. Paper presented at *Phonetics and Phonology in Europe (PaPE)*. Cologne, June 2017.

Tingsabadh, M. R. K., & Abramson, A. S. (1999). Thai. In *Handbook of the International Phonetic Association* (pp. 147-150). Cambridge: Cambridge University Press.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff et al. (eds.), *Methods in Empirical Prosody Research* (pp. 1-27), Berlin: de Gruyter.

Vietti, A., Alber, B., & Vogt, B. (2018). Initial laryngeal neutralisation in Tyrolean. *Phonology,* 35, 79–114.

Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language,* 45, 9–25.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1998). Cricothyroid activity in high and low vowels: Exploring the automaticy of intrinsic F0. *Journal of Phonetics,* 27, 125–142.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics,* 23, 349–366.

Whalen D. H., Abramson, A. S., Lisker L., & Mody M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica,* 4, 36–49.

Willi, U. (1996). *Die segmentale Dauer als phonetischer Parameter von 'fortis' und 'lenis' bei Plosiven im Zürichdeutschen.* Stuttgart: Steiner.

Winteler, J. (1876). *Die Kerenzer Mundart des Kantons Glarus in ihren Grundzügen dargestellt*. Leipzig: Winter.

Wolfensberger, H. (1967). *Mundartwandel im 20. Jahrhundert: dargestellt an Ausschnitten aus dem Sprachleben der Gemeinde Stäfa*. Frauenfeld: Huber.