The cross-linguistic distribution of vowel and consonant intrinsic F0 effects

Connie Ting, Meghan Clayards, Morgan Sonderegger, and Michael McAuliffe

McGill University

connie.ting@mail.mcgill.ca meghan.clayards@mcgill.ca
morgan.sonderegger@mcgill.ca michael.mcauliffe@mail.mcgill.ca

v4, July 2024

Abstract

Vowels vary systematically in F0 based on intrinsic properties of the vowel (e.g., height: VF0) and preceding obstruent (e.g., voicing: CF0). These patterns have been attested in most languages studied, raising the possibility that they stem from physiological sources. However, previous results show a range of variability in maximum effect size and duration. One explanation is that variation could be learned and increased to enhance phonological contrasts or suppressed for functional reasons (e.g. for tone languages). Alternatively, differences could be due to physiological factors such as laryngeal contrast type. We map out the distribution of intrinsic F0 effects across 20 languages, using large corpora of read speech, operationalizing CF0 as the difference between stop series which most closely approximate phonologically "voiced" and "voiceless" stops. We find both VF0 and CF0 effects are present and in the same direction in all languages examined, but languages vary greatly in effect size. While some of this variability may be due to phonological properties of the language (e.g. tone), and some variability in CF0 may be due to the diverse phonetic realizations of "voicing" across languages, much of this variability remains to be explained. We find that the CF0 effect is consistently at least as large as the VF0 effect and is more variable across languages, suggesting a possible explanation for the tendency for CF0 effects to lead to sound change much more often than VF0 effects. While our results on variability in CF0 effects rely on the validity of 'lumping' together diverse phonetic realizations, our results on robustness of CF0 and VF0 and on variability in VF0 hold regardless. These results motivate further investigation to deepen our understanding of intrinsic F0 effects, their cross-linguistic distribution, and their role as precursors to sound change.

---

1. Introduction

The production of speech sounds requires the complex coordination of multiple articulators along with the airstream. The result is a rich acoustic signal with each phonological contrast having measurable acoustic signatures (or cues) in many dimensions (e.g. F0, formants, etc). Traditionally, some of these cues have been seen as playing a primary role in signalling the contrast and others a secondary or supporting role. In an influential formulation, Wang and Fillmore (1961) distinguish between two types of secondary cues: *extrinsic* cues "which reflect the speech habits of a particular community" and *intrinsic* cues "which reflect the structure of the speech mechanism in general." For any given contrast, such as obstruent voicing or vowel height, there can be debate about whether secondary cues vary for intrinsic or extrinsic reasons, i.e. because they must, or because the speaker chooses to. Furthermore, how these secondary cues vary may differ by speech community. This is particularly relevant in the context of sound change, where the sounds of a community change over time, for which a primary source is thought to be phonologization (Hyman 1976, 2013), where an intrinsic phonetic pattern within a language becomes an extrinsic cue to a phonological contrast. This paper aims to shed light on these issues by investigating one such cue, so called *intrinsic F0 effects*, from a broad perspective, using large scale analysis of many speakers and languages and encompassing both vowel and consonant driven intrinsic F0 effects.

It has been widely reported that vowels vary systematically in fundamental frequency (F0) based on intrinsic properties of (i) the vowel itself and (ii) the preceding obstruent. For vowels, the relevant intrinsic property is *height*, such that high vowels tend to be produced with higher F0 compared to low vowels (e.g. Connell, 2002; Gonzales, 2009; Van Hoof and Verhoeven, 2011; Whalen and Levitt, 1995). For preceding obstruents, the relevant intrinsic property is arguably *voicing*: F0 at vowel onset tends to be higher when the preceding obstruent is phonologically voiceless than when it is phonologically voiced (e.g. House and Fairbanks, 1953; Lehiste and Peterson, 1961). This generalization has been observed in languages with different phonetic implementations of its laryngeal contrasts, though the source of the effect and the relationship between the phonological and phonetic implementation of laryngeal contrasts remains a topic of ongoing debate. Section 2.2 provides a more detailed discussion of laryngeal contrasts as it relates to intrinsic F0 effects and clarifies how the terms 'voiced' and 'voiceless' are used in this paper.

Following Kingston (2007), we will refer to these intrinsic F0 perturbations as vowel intrinsic F0 (VF0) and consonant intrinsic F0 (CF0) effects.[1] Figure 1 illustrates VF0 and CF0 patterns based on empirical data from two languages in our sample (German for VF0; Portuguese for CF0). In Figure 1a, there is a difference in F0 throughout the vowel, with the high vowel category having a higher F0 compared to the low vowel category. In Figure 1b, there is a difference in F0 between phonologically *voiced* and

---

[1] Other names include *intrinsic F0* or *intrinsic pitch* for vowel intrinsic F0 (Whalen and Levitt, 1995) and *obstruent intrinsic F0* (Hanson, 2009) or *pitch skip* (Francis et al., 2006) for consonant intrinsic F0. See also Di Cristo and Hirst (1986) for discussion of *intrinsic* and *co-intrinsic* effects.

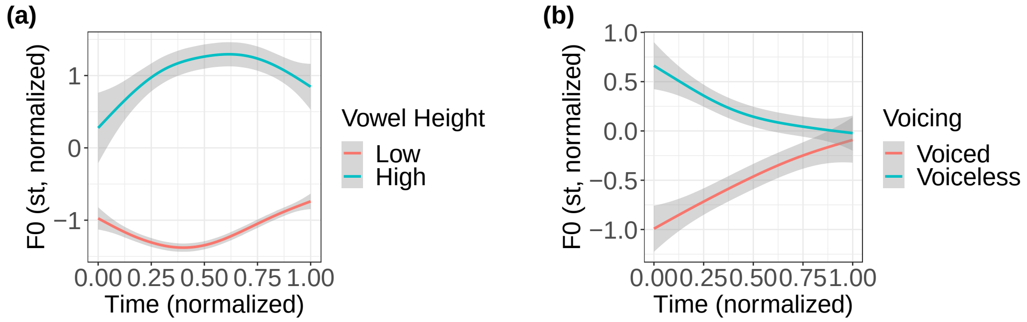


Figure 1: Illustration of pitch trajectories showing (a) the VF0 effect, where F0 for high vowels is overall higher than F0 for low vowels and (b) the CF0 effect, where F0 following voiceless categories is higher than F0 following voiced categories.

phonologically *voiceless* categories, particularly at the onset of the vowel, such that F0 is higher when the preceding obstruent is phonologically *voiceless*.

The vast literature investigating VF0 and CF0 effects (reviewed in Section 2) has shown that these effects are attested in almost all studied languages, raising the possibility that they stem from *automatic* consequences of articulation, i.e. truly intrinsic in the Wang and Fillmore sense (see Chen 2011, Hanson 2009, Whalen and Levitt 1995 for reviews). However, previous results show variability across studies in maximum effect size and how long the effect persists across the vowel, not completely explained by differences in phonetic realization. For CF0, the effect is generally robust regardless of the phonetic realization of 'voicing' (e.g. Dmitrieva et al., 2015; Hombert, Ohala and Ewan, 1979; Kingston and Diehl, 1994), and for VF0, the effect size (but not direction) varies across dialects and speakers of the same language despite similar phonetic realizations of the vowel height contrast (Hoole and Honda, 2011; Jacewicz and Fox, 2015). From such results, researchers have argued that IF0 effects are perhaps not entirely explained by an automatic physiological approach, and instead may be under speaker 'control' (Kingston and Diehl, 1994). In the *controlled* view, languages can have larger or smaller (or longer/shorter) effects in order to enhance or maintain contrasts, such as a greater CF0 effect to enhance a voicing contrast, a greater VF0 effect to enhance a vowel height contrast (Van Hoof and Verhoeven, 2011), or a smaller CF0 or VF0 effect to avoid impinging on a lexical tone contrast (Connell, 2002). Note that 'speaker control' does not refer to the active choice of individual speakers to produce larger or smaller IF0 effects, but represents speakers' or languages' selection of using particular IF0 settings out of a larger set of possibilities (Kingston and Diehl, 1994).

Other work has proposed a hybrid approach (e.g. Chen, 2011; Hanson, 2009; Hoole and Honda, 2011; Van Hoof and Verhoeven, 2011), combining both automatic and controlled proposals. Under a hybrid account, the underlying mechanisms of intrinsic F0 effects are automatic (i.e. originate from articulatory gestures required to make the target speech sound). However, speaker or language-specific control is explicitly not ruled out. Thus, a main prediction of a hybrid approach is that all languages will show intrinsic F0 effects in the expected direction, *and* that there can be variation across speakers and languages.

The extent of variation in the size and time-course of intrinsic F0 effects therefore has consequences for how we understand them, making cross-linguistic comparisons important. However, studies of intrinsic F0 effects have used a variety of methodologies, and are usually constrained to small groups of languages and speakers. Differences across studies make it difficult to conclude whether the variation across languages is a true reflection of the effects themselves or is the consequence of differing methodologies. Moreover, while variation in the entire trajectory of intrinsic F0 effects across languages is informative, most previous studies have focused on effect size, leaving the time-course of intrinsic F0 effects understudied. We therefore lack a clear picture of the *distribution* of intrinsic F0 effects—how much they differ and in what ways (as noted by Hanson, 2009; Kirby, 2018 for CF0, Van Hoof and Verhoeven, 2011 for VF0). In addition to the automatic/controlled question, understanding the distribution of IF0 effects is important in the context of sound change. It has been observed that CF0 effects are more likely to lead to sound change than VF0 effects. Examining the distribution of IF0 effects could help explain this asymmetry.

This paper explores the trajectories of both vowel and consonant intrinsic F0 effects across 20 typologically diverse languages using large corpora of read speech. This large-scale study allows us to go beyond the binary question of whether VF0 or CF0 effects in any given language are automatic or controlled. Instead, we can map out the distribution of VF0 and CF0 effects and, for the first time, compare their relative size and degree of variability. The hope is that the results will give us a better empirical basis to explore the sources of intrinsic F0 effects and their role in sound change.

A challenge in any large-scale typological study is how to decide which languages are comparable. For CF0, a major challenge is the diversity in type and number of laryngeal contrasts in the world's languages. We have chosen to include all of this diversity in our sample by including a language if we have the right type of data available, rather than based on the type of laryngeal contrasts it uses. We believe that more insight can be gained this way than is lost by 'lumping' together phonetically diverse languages. However not all researchers would agree with this choice (see e.g. Greenberg 2005 for discussion on 'lumping' vs. 'splitting'). Throughout the paper, we highlight which results depend on the 'lumping' strategy used in our analysis.

In the remainder of the paper, Section 2 provides background on intrinsic F0 effects and motivations for the current study. Our research questions and predictions are presented in Section 3, along with an overview of the corpus data and methods of analysis. Sections 4 and 5 present the results and discussion of our study. Section 6 concludes.

2. Background

As much of the literature has focused on either VF0 or CF0, we summarize previous work on vowel, and consonant, intrinsic F0 effects separately in Sections 2.1 and 2.2. Further motivations for the current study are provided in Section 2.3.

2.1. Vowel Intrinsic F0 Effects

The effect of vowel height on vowel F0 has been reported in studies dating back at least a century (e.g. Meyer, 1896; Shadle, 1985; Taylor, 1933). These findings have prompted researchers to posit that the effect is a direct consequence of physiological properties of vowel (tongue) height (Lehiste, 1976; Whalen and Levitt, 1995; Whalen et al., 1995). Variations of the so-called *tongue-pull hypothesis* generally assume that movements of the tongue during vowel articulation are tied to the amount of force exerted on the larynx, and consequently the amount of tension in the vocal folds (see e.g. Dyhr, 1982; Fischer-Jørgensen, 1990; Sapir, 1989). This would mean that the observed VF0 effects are primarily rooted in the articulation of vowels and are not strongly influenced by other factors (e.g. vowel inventory size, use of F0). However, different languages have different articulations of the same (underlying) vowel, and it is important to note that under an articulatory approach, we expect to find differences in VF0 effects given these differences in articulation.

The first VF0 meta-analysis by Whalen & Levitt (1995; henceforth W&L) examined the size of the effect in all of the languages reported prior to their study, covering 31 languages representing 11 of the world's 29 major language families. All languages showed an effect in the attested direction (higher F0 for high vs. low vowels). W&L took this as evidence that the VF0 effect is automatic. Several VF0 studies since then have reported similar results (e.g. Van Hoof and Verhoeven 2011 for Arabic and Dutch, Gonzales 2009 for Shona, Jacewicz and Fox 2015 across different English dialects). However, W&L's analysis and later studies also show a wide range of variation in the size of the VF0 effect across languages (with almost all studies showing effects of ∼0.2–3.7 semitones), which they argue could be largely 'noise' from methodological differences and small sample sizes.

Others suggest that languages have meaningfully different VF0 effects, whose size must be learned by speakers of the language. These differences could be arbitrary (and not solely determined by vowel articulation) (Jacewicz and Fox, 2015), or due to functional pressure from properties of the language (e.g. use of F0, vowel inventory size), or both. For example, Connell (2002) investigated the VF0 effect in four African tone languages: Ibibio, Kunama, Dschang, and Mambila. Only the first three showed a significant VF0 effect, and the effect size was smaller than typically reported ($<$ 1.1 semitones, compared to an average of 1.65 semitones reported by W&L). Furthermore, in the studies reviewed by W&L, 9 of the 10 tone languages showed that the VF0 effect disappeared for low tones (e.g. Hombert 1977 for Yoruba and Zee 1980 for Taiwanese). One explanation for these findings in tone languages is that the VF0 effect is not completely automatic, but that speakers are more restricted in lower ranges where further lowering of F0 is prevented by a hard pitch floor and raising of F0 creates ambiguity in pitch targets. This is compatible with findings in non-tonal languages like English, German, and Danish, in which VF0 differences are smaller or lost completely in non-prominent syllables (Ladd and Silverman, 1984; Petersen, 1978; Steele, 1986) or in lower pitch ranges (Ladd and Silverman, 1984; Whalen and Levitt, 1995). A larger VF0 effect could also enhance vowel contrasts in languages with larger, more crowded, vowel systems, as suggested by Van Hoof and Verhoeven (2011), who found that Dutch speakers

(12-vowel inventory) showed a VF0 effect three times larger than Arabic speakers (3-vowel inventory). Thus, multiple factors may affect intrinsic F0 effect size, in addition to potentially arbitrary differences, which cannot be easily teased apart without comparing across many languages.

Finally, most studies have examined overall size of the VF0 effect (i.e. differences in average F0 across the vowel) (cf. Lalhminghlui, Terhiija and Sarmah 2019). However, examining the full *trajectory* of IF0 effects could have consequences for how we understand them. If languages have larger or smaller effects in order to maintain contrasts, then we might expect those effects to also make up a longer or shorter proportion of the vowel, i.e. a positive correlation between effect size and temporal extent (see e.g. Francis et al. 2006 and Hombert 1977 for discussion of the extent of CF0 effects). On the other hand, we might expect these two variables to negatively correlate with one another, if greater reliance on one cue to a contrast allows less reliance on another cue. In this paper we examine both VF0 effect size (the maximum difference between high and low vowel trajectories) and the time-course of the effect, which we define as the proportion of the vowel over which there is a difference, henceforth *effect length*. This will allow comparison to CF0 effect size and length and let us see if there are ties between VF0 effect size and length cross-linguistically, which has not yet been systematically tested. Showing the overall F0 curve may also provide further insight if we interpret the shape of the curve as a representation of the articulatory and/or aerodynamic trajectory involved in the production of the segment, analogous to the logic for CF0 effects below.[2]

2.2. Consonant Intrinsic F0 Effects

It has long been known that a consonant's laryngeal class has an effect on the F0 of the following vowel. The traditional generalization refers to phonological voicing: phonologically voiceless obstruents are associated with higher F0 and phonologically voiced obstruents are associated with lower F0 on the following vowel (e.g. Hombert, 1978; Hombert, Ohala and Ewan, 1979; House and Fairbanks, 1953; Kingston and Diehl, 1994; Kohler, 1982; Lehiste and Peterson, 1961; Ohde, 1984). For languages with a 2-way laryngeal contrast, this pattern has been found to hold regardless of how the voicing contrast is phonetically realized (Dmitrieva et al., 2015). For example, the CF0 effect occurs in 'true voicing' languages such as French and Spanish (Caisse, 1981; Dmitrieva et al., 2015), where the voicing contrast is dependent on the presence/absence of (pre-)voicing, as well as in 'aspiration' languages such as English and Danish (Petersen, 1978), where the voicing contrast (word initially) is dependent on presence/absence of aspiration. Researchers have argued that this pattern is due to physiological and/or aerodynamic constraints inherent to consonant voicing production and is thus expected in all languages (Bell-Berti, 1975; Hyman, 1976; Ladd and Schmid, 2018; Löfqvist et al., 1989; Ohala, 1993). For example, the Vertical Tension Hypothesis (Hombert, Ohala and Ewan, 1979) attributes differences in F0 to larynx height, such that a raised larynx might result in high F0 (Hombert, Ohala and Ewan, 1979; Honda et al., 1999). Others have argued that the main source of the CF0 dichotomy is the biomechanical consequence

---

[2] We thank a reviewer for this suggestion.

of a gesture to inhibit or sustain voicing/phonation (Halle and Stevens, 1971; Hanson, 2009; Kirby and Ladd, 2016; Löfqvist et al., 1989). It remains an open question however, how to reconcile the consistent cross-linguistic pattern of CF0 that covers many types of laryngeal contrasts with a unifying articulatory account.

However, not all studies report findings in line with the generalization. For example, while some studies on Mandarin showed voiceless aspirated stops to have an overall higher F0 (Howie, 1976; Luo, 2018; see also Chen, 2011 for Shanghai Wu), other studies on the same language showed that voiceless aspirated stops *lower* F0 (Xu and Xu, 2003, see also Francis et al., 2006 for Cantonese). Yet others showed an interaction between the direction of the CF0 effect and lexical tone (Guo, 2020). The reported effect sizes in such cases are small, which suggest the possibility that the true effect size is small or null and that size and magnitude of the effect fluctuate between studies due to limited power and differing designs (Kirby and Sonderegger, 2018). Ladd and Schmid 2018 report that Zurich German fortis and lenis stops, which are both voiceless unaspirated but differ in length, pattern together in raising F0 (consistent with a voicing inhibition gesture) but might also have a small CF0 effect in the expected direction (if we consider the longer fortis stops to be "more voiceless"). This example highlights how simple phonetic categories based on voice onset time like 'voiceless unaspirated' may not properly capture phonological contrasts, and could group together articulations that are different. We return to this point below.

Many studies have looked for cross-linguistic differences in the size of the CF0 effect (e.g. Gandour, 1974; Jeel, 1975; Kagaya, 1974; Kagaya and Hirose, 1975; Zee, 1980), including recent work on a broader range of languages (Babinski, 2021; Chen, 2011; Coetzee et al., 2018; Dmitrieva et al., 2015; Gao and Arai, 2019; Hanson, 2009; Kirby, 2018; Kirby and Ladd, 2016; Kirby et al., 2020).[3] Importantly, unlike for VF0, there has not been a meta-analysis of the CF0 effect. We therefore do not have a clear picture of the distribution of the CF0 effect across languages.

One possible reason for this gap is that one must consider how to conduct a large scale analysis with typologically different languages while ensuring comparability across languages. In this paper, we will map out the distribution of CF0 effects by including languages with diverse laryngeal contrasts. We compare them by referring to 'voiced' vs. 'voiceless' categories in a broad sense (Keating, 1984) and we will assume that abstract phonological categories will broadly pattern similarly (in terms of effects on F0) regardless of their phonetic implementation. It's important to note that this procedure commits us to the "phonological" side of the longstanding debate: to what extent are CF0 effects due to phonological categories or phonetic realization. We will further assume that we can map such phonological categories onto VOT as a first approximation allowing us to examine the CF0 effect across languages, following a traditional view on the relationship between phonological and phonetic representations (Keating, 1984; Lisker and Abramson, 1964). Therefore, for all languages with a two-way contrast, we use 'voiced' and 'voiceless' to refer to the two categories in the language, using VOT

---

[3] All CF0 studies which have examined interspeaker differences have found they are present. While we will not directly consider interspeaker variability, it will be accounted for in our models.

as a rough measure of category voicelessness (i.e. the category with the higher VOT = voiceless). While this doesn't take into account other potentially relevant dimensions such as constriction duration, it does restrict us from making more intuitive decisions that might bias our results.

An alternative method would be to consider languages with different laryngeal systems separately. We consider both of these types of analyses methodologically valid, noting that choosing one or the other may have consequences for how the results are interpreted. In order to compare languages in separate typological analyses, much more data would be required than is available for the current study. To allow for different interpretations of the results, we provide colour-coding of our main findings in Section 4.3, splitting languages by different phonetic realizations of phonological categories.

For measures such as effect size and length, we believe that the results of our approach can be interpreted in the same way, regardless of whether a 'lumping' or 'splitting' system is implemented. That is, whether we consider languages in one large group vs. in many smaller groups, the direction of effect size remains the same. However, any results about the variability of CF0 effects may differ when considering 'lumping' vs. 'splitting' approaches, since any interpretation of variability is based on the particular set of languages under consideration. We hope that this study provides an example of one possible 'lumping' method of cross-linguistic investigation and inspires future large-scale explorations.

There is further complexity for any language that has more than 2 laryngeal categories (e.g. Korean and Madurese have 3-way contrasts, Bengali and Nepali have 4-way contrasts). The general pattern in 3-way contrasts seems to be that the stop with the longest VOT has higher F0 than the stop with the shortest VOT (e.g. Misnadin, Kirby and Remijsen 2015 for Madurese). In 4-way systems which contrast both voicing and aspiration orthogonally, F0 follows voicing and may sometimes be lowered slightly by aspiration (e.g. Clements and Khatiwada, 2007; Schertz and Khan, 2020). Therefore, it is important to acknowledge the complexity of voicing/laryngeal contrasts and remember that these distinctions involve interactions which do not form an explicit continuum from 'voiced' to 'voiceless aspirated' or a single dichotomy between 'true voicing' and 'aspirating' languages (see also Kirby and Ladd 2016 and Ladd and Schmid 2018 for related discussion). For a cross-linguistic analysis of CF0, if languages with more than 2 laryngeal categories are included, a consistent approach should be applied to each such language to determine which laryngeal categories will be considered.

In this study, the languages that have more than 2 laryngeal categories are: Hausa, Korean, Thai, and Vietnamese. We detail the choices for each language. Hausa has both pulmonic ('voiced' vs. 'voiceless') and glottalic (ejective, implosive) consonants (Schuh and Yalwa, 1993). We will consider only the pulmonic consonants, assuming a 'voicing' distinction between categories, which will be most comparable to the other languages in this study. Korean has a 3-way contrast between tense, unaspirated (or lax), and aspirated stops, with VOT generally increasing from tense to lax to aspirated in word-initial position (e.g. Han and Weitzman, 1965; Kagaya, 1974; Kang, 2014). We will consider tense and aspirated stops, as they have lowest and highest VOT. Similarly, Thai has a 3-way contrast between voiced, voiceless (or tenuis) and aspirated stops, with

VOT increasing from voiced to voiceless to aspirated. Thus, we will consider only the voiced and aspirated stops in Thai. Finally, Vietnamese has a 3-way contrast between pre-glottalized/voiced, unaspirated, and aspirated stops. Within this 3-way contrast, voiced stops have the shortest VOT and aspirated stops have the longest VOT. However, there is only one aspirated stop in the language. Thus, to have more comparable data, we will consider the unaspirated stops as the 'voiceless' category. While the voiced and unaspirated stop series in Vietnamese differ in VOT, the 'voiced' stops are pre-glottalized and sometimes imploded, making the comparison different from other languages. We include these four languages in our analysis and interpretation of the results to maximize the number of languages included of our study of CF0 effects. An alternative method, which can be applied in interpreting our results, would be to exclude these four languages.

Finally, like with VF0, most studies of CF0 effects have focused on effect size. Among the existing studies that have discussed effect length or duration, findings show variation. For example, (Francis et al., 2006) reported that the CF0 effect is restricted to the onset of the vowel (approximately 0-10ms) in Cantonese, while others have reported longer effects in other tonal languages (approximately 50-150ms; see Hombert, 1977 for Yoruba; Kenstowicz and Park, 2006 for Kyungsang Korean; Kirby 2018 for Thai and Vietnamese; and Lai et al., 2009 for Taiwanese Min). The effect has also been reported to last through the entire vowel, in both tonal (Jessen and Roux, 2002 for Xhosa) and non-tonal languages (Kirby, 2018 for Khmer; Xu and Xu, 2021 for English). One commonly proposed source of variation in CF0 effect length (as well as effect size), originating with Hombert (1977), is that CF0 effects will be smaller and shorter in tonal languages compared to non-tonal languages (see also Chen, 2011; Francis et al., 2006; Kirby, 2018 for discussion of the role of tonal and discourse context); this would predict a positive correlation between CF0 effect size and length across languages, as discussed for VF0 above. An alternative hypothesis would be that CF0 effect size and length are negatively correlated, such that greater reliance on one cue allows less reliance on the other. However, as has been noted repeatedly (e.g. Chen, 2011; Kirby, 2018; Pinget and Quené, 2023), comparing results from previous studies across different languages is difficult, given differences in methodology, sample size, and so on. Therefore, to better understand the range of consonant intrinsic F0 effects across (different kinds of) languages, the distribution of effect size and length should be examined using comparable data from a range of languages.

2.3. Motivations

Overall, previous studies investigating intrinsic F0 effects have shown a great deal of variation, which raises questions regarding any patterns of intrinsic F0 effects across languages. For one, are effects automatic, controlled (e.g. to preserve other contrasts), or a combination of the two (as proposed by Hoole and Honda 2011). Moreover, if they are not purely automatic, in what ways can they be controlled, and do we expect to find clear patterns between different groups of languages (e.g. tonal vs. non-tonal)? Our aim is to discuss the distribution of intrinsic F0 effects in the context of such possible explanations, focusing on those that have been discussed in the literature and leaving room for the possibility that more than one of these explanations are supported.

There is also considerable evidence that intrinsic F0 effects are a primary source of tonogenesis and subsequent tonal splits (e.g. Di Cristo, 1985; Hombert, Ohala and Ewan, 1979; Kingston, 2011; Matisoff, 1973), and that such sound changes arise much more commonly from CF0 than from VF0 effects. This process for CF0 effects is well-documented primarily in Southeast Asian languages (see e.g. Svantesson and House 2006 for Kammu) but has also been found in other languages (e.g. Coetzee et al. 2018 for Afrikaans, Kang 2014 for Korean, Kanwal and Ritchart 2015 for Punjabi). Similar processes forming due to F0 perturbations of intrinsic properties of vowels are more rare, but have been proposed (Bethin, 2006; Diffloth, 1982; Svantesson, 1989).[4]

It remains unclear why this asymmetry between intrinsic F0 effects exists in the first place, though several explanations have been offered. Ohala (1981) argues that for such a change to take place, F0 perturbations must be misattributed in perception. One possibility is that the size and length of these effects play an important role in such misattribution. That is, smaller/shorter effects may be more easily perceptually normalized while larger/longer effects are not, and are thus less likely to act as 'phonetic precursors' to sound change. Similarly, it's possible that more variable effects (in size and/or length) lead to greater misattribution, or the opposite: Kingston (2011) suggests that VF0 effects are *less* likely to phonologize because they seem to be more variable (within language) than CF0 effects as a function of prosodic context and tone, and hence offer a less consistent cue to listeners to misinterpret as tone. A final hypothesis, from the sound change literature, is that phonetic precursors which are more 'robust'—in the sense of larger effect size and/or less variability—will be more likely to lead to sound change (e.g. Ohala, 1994; see Moreton, 2008 and references therein).

All such explanations refer to the relative size and degree of variability of VF0 and CF0 effects across languages, which is unclear from previous work. If any of these explanations is on the right track, then mapping out the cross-linguistic distribution of production of VF0 and CF0 effects (which form the input to listeners' perception), including the relative sizes of effects within-language, could explain the asymmetrical behaviour of these effects as precursors to sound change.

Of the studies that contributed to the survey by W&L, more than half used five or fewer speakers. Other cross-linguistic comparisons (e.g. Babinski 2021; Connell 2002) have similar numbers of speakers. Moreover, the distribution of VF0 effect size in W&L was largest for those studies with few speakers and decreased as the number of speakers increased. Whalen and Levitt (1995) note that larger sample sizes are needed to bring out the true situation of VF0 effects in a given language. The same is likely true for CF0 effects. The closest antecedent to our own study is Babinski (2021), who extend the methodology of a preliminary version of the current study (Sonderegger, McAuliffe and Bang, 2017) to survey 16 Australian languages for both VF0 and CF0 effects. Babinski found that the CF0 effect showed a more consistent effect direction across languages compared to the VF0 effect, hence more 'robust'. It is important to note, however, that there were 10 or fewer speakers (average of 3) per language, and that in most languages

---

[4] Some have proposed a link between vowel height and tonal sound change (see e.g. Chen and Norman 1965, cited in Hombert, Ohala and Ewan 1979; Nitta 2001; Odden 2010, among others).

the effect of ‘voicing’ is allophonic and confounded with stress (Babinski, 2021, p.5). Thus these patterns remain to be tested across a larger number of speakers and across more typologically diverse languages.

3. The current study

3.1. Research Questions and Predictions

The current analysis explores intrinsic F0 effects across 20 languages using large corpora of read speech, with number of speakers ranging from 69 to 401 speakers across languages. This allows us to address the limitations of previous studies and explore additional questions that have not yet been empirically addressed in the literature. More specifically:

1. What is the **distribution** of intrinsic F0 effects across languages?
    - How (much) do languages differ?
    - In what ways do languages differ?
2. What is the **relationship** between intrinsic F0 effects across languages?
    - Is there a correlation between VF0 and CF0 effects?
    - Is there a correlation between effect size and length (within VF0, CF0)?

Some possible outcomes of these questions are presented in Figure 2. The consistency of intrinsic F0 effects observed in the current literature suggests that intrinsic F0 effects are universal in some sense. However, it’s not clear what ‘universal’ means. Under a strict view of ‘universality’, the effects might be considered to be unintended and automatic side effects (i.e. truly ‘intrinsic’). If this were the case, we might expect them to be (nearly) invariable across languages, as in Figure 2a. On the other hand, universality may reflect only the presence of an effect (in the same direction) in all languages, with some variation predicted as an automatic consequence of variability in articulation. For the purposes of this discussion, we will use ‘identical’ to describe the strict (invariable) view, and use the term ‘robust’ to describe a process that occurs consistently in the same direction but may show variation across languages, as in Figure 2b. The first main research question of this paper is to show the empirical distribution of both vowel and consonant intrinsic F0 effects and see whether the effects are robust (i.e. present and in the same direction), and how much variation exists across languages.

Another possibility is that the variation of these effects points to individual languages, or groups of languages, showing either a smaller or larger effect (or shorter or longer) relative to other languages. These differences could be socially-learned behaviour (Jacewicz and Fox, 2015), but most often they are assumed to exist for a functional reason: as a consequence of ‘enhancement’, where speakers increase CF0 or VF0 to maximize the consonant voicing or vowel height contrast (e.g. Diehl, 1991; Kingston and Diehl, 1994), or ‘suppression’, such as to maintain these contrasts in languages with tones or large vowel inventories (e.g. Connell, 2002; Van Hoof and Verhoeven, 2011). Note that the

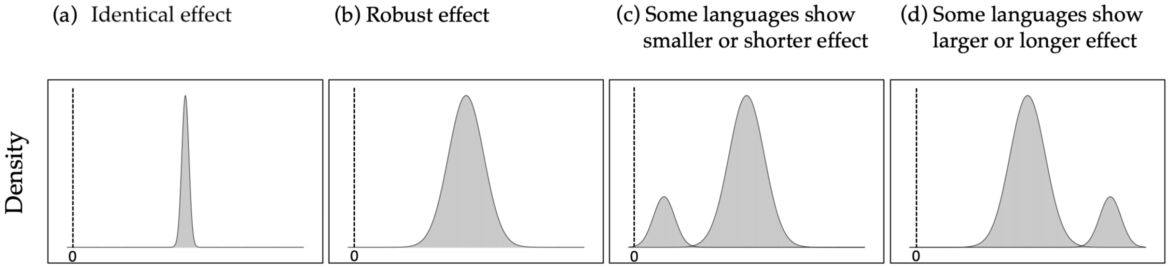


Figure 2: Possible distributions of intrinsic F0 effect size (or length) across languages. The horizontal axis could represent semitones in the case of effect size, or proportion of vowel duration in the case of effect length.

term 'enhancement' has various meanings in the literature, such as referring to enhancement of *features* (Stevens and Keyser, 1989) or *gestures* (Garrett and Johnson, 2013). Since the goal of this study is to present the overall distribution of effects of the languages examined here rather than advance theoretical proposals of enhancement and suppression, we will refrain from using these terms and will describe our results in terms of relative effect size and length across languages. Figure 2c reflects a scenario where some languages show a smaller or shorter effect, and Figure 2d reflects a pattern where some languages show a larger or longer effect which could eventually become phonologized (Coetzee et al., 2018; Kang, 2014; Kanwal and Ritchart, 2015; Kingston, 2011; Thurgood, 2002). We can imagine these possibilities for both effect size and length, separately.

Finally, we can explore relationships within and between vowel and consonant intrinsic F0 effects more generally. For example, if a language shows a greater effect *size* for an intrinsic F0 effect, does that effect persist for a greater or smaller *proportion* of the vowel? If a language shows a greater effect size for the VF0 effect, does it also then show a greater effect size for the CF0 effect? Although exploratory in nature, this comparison between different kinds of intrinsic F0 effects provides a starting point for discussing cross-linguistic patterns.

Some possible outcomes for these questions are presented in Figure 3. If we start by illustrating the relationship between VF0 and CF0 effect sizes, Figure 3a reflects a lack of correlation between the two, consistent with different proposed mechanisms for the two. Figure 3b reflects a positive correlation, consistent with proposals that languages will restrict any kind of F0 difference for functional reasons. Finally, figure 3c illustrates a positive correlation with some outlier languages (i.e. languages which show either a larger VF0 or CF0 effect), consistent with a proposal where languages are free to enhance one intrinsic F0 effect, or the other, or both (Hoole and Honda, 2011). We can also imagine these possibilities for the relationship between effect size and length. In that case, Figure 3c would indicate a trading relationship between effect size and length. This would be in line with the idea that increasing *both* effect size and length to support a contrast (e.g. vowel or consonant) would not be necessary if increasing *either* effect size or length is sufficiently salient to support the contrast.

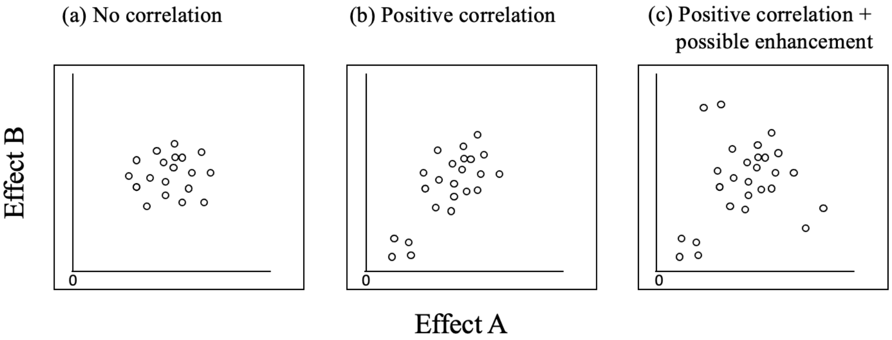


Figure 3: Possible outcomes for the relationship between intrinsic F0 effects across languages. Each dot represents effect size of an individual language.

3.2. Corpus data

Our data come from three read speech corpora, which were chosen based on availability and suitability for our analysis. The first is GlobalPhone (Schultz, Vu and Schlippe, 2013), a database of read speech from newspaper articles in 22 languages, with corresponding transcriptions and pronunciation dictionaries for most languages. Globalphone contains roughly 20 hours of speech for 100 speakers per language. We used data from 18 languages, excluding 4 that could not be force-aligned (Arabic, Japanese, Shanghai Wu, Tamil). The second corpus is LibriSpeech (Panayotov et al., 2015), which consists of ∼1000 hours of (mostly American) English read speech from audiobooks. We used a subset comparable to the amount of data (i.e. speakers, hours) per language included from the GlobalPhone corpus. The third corpus is IARPA Babel Cantonese (Andrus et al., 2016), from which we used the read speech portion. The current analysis includes data from 20 languages, listed in Table 1, using approximately 20 hours of read speech for each language.

3.3. Data processing and exclusions

The data for each language was force-aligned using the Montreal Forced Aligner (McAuliffe et al., 2017), using the pronunciation dictionary, orthographic transcriptions of each audio file, and utterance-level time-stamps provided with each corpus. Language-specific acoustic models were trained, then applied to time-align each transcript to the corresponding audio file at the phone and word levels. After alignment, the corpora were imported into PolyglotDB databases (McAuliffe et al., 2017), and pitch measurements were extracted every 10 msec using the PolyglotDB speaker-adapted algorithm, which involves two passes through Praat to calculate by-speaker pitch ranges for F0 and then re-extract pitch using these ranges.[5] A subset of the measurements was hand-checked to ensure minimal pitch extraction errors.

To compare across languages with different vowel inventories, the data was restricted to vowels which are roughly /a/, /i/, and /u/ in each language. For example, Turkish

---

[5] A full description can be found here: `https://polyglotdb.readthedocs.io/en/latest/acoustics_encoding.html#encoding-pitch`

Table 1: Descriptive statistics and family for the 20 languages included in the dataset. The number of trajectories and tokens listed are after exclusions, where ‘trajectories’ refers to individual vowels and ‘tokens’ refers to F0 measurement points.

Language	Family	*N*		
		Speakers	Trajectories	Tokens
Bulgarian	Indo-European	77	1942	14483
Cantonese	Sino-Tibetan	401	5295	54830
Croatian	Indo-European	92	1990	18244
Czech	Indo-European	102	1640	12714
English	Indo-European	131	2245	20576
French	Indo-European	100	1269	8786
German	Indo-European	77	3105	22214
Hausa	Afro-Asiatic	103	2952	28786
Korean	Korean	99	2953	21598
Mandarin	Sino-Tibetan	132	8408	75770
Polish	Indo-European	99	1814	14039
Portuguese (Brazilian)	Indo-European	82	750	6083
Russian	Indo-European	115	2174	17728
Spanish (Latin American)	Indo-European	100	1877	15126
Swahili	Niger-Congo	69	1706	15175
Swedish	Indo-European	98	1411	13355
Thai	Tai	98	3575	27130
Turkish	Turkic	100	3869	31279
Ukrainian	Indo-European	119	1838	15352
Vietnamese	Austro-Asiatic	129	3753	34296

does not have a low central vowel /a/, so we used its low back vowel /ɑ/ to represent the low vowel category. Although this ‘lumping’ method allows us to compare across languages, it does eliminate tenseness and length distinctions in certain languages. The dataset was further restricted to syllables beginning with a ‘voiced’ or ‘voiceless’ obstruent, as defined in Section 2.2 (see Appendix Table B.7). Although some previous studies have focused solely on stop consonants, others have found comparable CF0 effects with the inclusion of fricatives and affricates (Kirby and Ladd, 2016; Lehiste and Peterson, 1961; Xu and Xu, 2021), so the current study also considers these in addition to stop consonants.

The analysis included word-initial CV syllables from utterance-initial words, where an utterance occurs at the beginning of a file or is separated from previous speech by a force-aligned pause of at least 150 ms. We expect overall intonational context to moderate IF0 effects (e.g. Kirby, 2018; Shadle, 1985), and intonation would plausibly have a larger effect in utterance-initial position due to hyperarticulation (e.g. Cho and Keating, 2009). We restrict to utterance-initial context to control for intonation and focus on the context where we are most likely to see differences in intrinsic F0 effects. This control for prosody is only partial, as languages also differ in word-level prosody, which may lead to differences in intrinsic F0 effects when only initial position is considered (Kingston, 2011). We acknowledge this limitation, which we address with further discussion and post-hoc analysis in Section 4.5. In addition to vowel and consonant information, other factors expected to affect F0 trajectory were extracted: utterance length, tone (for tone languages), and speaker gender (if available).[6]

F0 measurements were transformed to semitones[7] and further data-processing and filtering steps were applied. First, to only consider with true F0 'trajectories', vowel measurements where pitch extraction could not be obtained for at least 50% of the vowel (∼13.5%) and vowel tokens shorter than 50ms (∼7.6%) were excluded. We then took steps to exclude suspicious F0 trajectories, when compared to the distribution of trajectories for the same language or speaker. Within languages, trajectories that included measurements with a sudden change in slope were excluded. Within speakers, trajectories that were two z-scores away from the speaker's mean were excluded. The outlier exclusion process is summarized in Appendix A and shown in full on the paper's OSF site.[8] The resulting datasets included 69–401 speakers and 0.7–8.4k vowels per language (see Table 1).

3.4. Analysis

For each language, the trajectory of F0 over time was modelled as a function of Voicing and Vowel using Generalized Additive Mixed Models, or GAMMs (Wood, 2017; see Sóskuthy, 2017 and Wieling, 2018 for introductions). GAMMs are suitable for analyzing data with non-linear relationships, allowing us to model F0 over the entire vowel without being restricted to a trajectory shape (e.g. Geissler et al., 2021; Li, Guan and Chen, 2020). This is particularly relevant for answering our research questions since we are interested in how languages differ in pitch trajectories, rather than just a particular point(s) in time. Models were built in R (R Core Team, 2022) using the `bam()` function (using the `fREML` method) from the *mgcv* package (Wood, 2017), with F0 (semitones) as the dependent variable, and Voicing (of the preceding consonant), Vowel, and Tone (where applicable) as (ordered) predictor variables.[9] Additionally, Gender and Utterance Length were included as control predictors; we expect female speakers to have overall higher pitch compared to male speakers, and longer utterances to have higher F0 at the beginning of the utterance (Cooper and Sorensen, 2012; t Hart, 1979).[10]

To model the time-course of F0 and differences between each level of Voicing and Vowel, smooth terms were included to (i) model the non-linear pattern of F0 over time

---

[6] Speaker gender was not available for Polish, so it was not included as a predictor in the model for Polish but was included for the other 19 languages.

[7] Using the formula $st = 12\log_2(Hz/100)$.

[8] https://osf.io/ehs6d.

[9] F0 measurements included in this analysis are normalized across all speakers and not normalized by speaker means. All variability across speakers, including in means, is accounted for with by-speaker random smooths in the statistical models.

[10] Gender was coded using the variable of this name provided in the corpora, with no further information regarding biological versus social gender.

(normalized between 0 and 1) and (ii) model F0 over time separately for the different levels of Voicing and Vowel. By-speaker differences were captured using factor smooths with a 'random reference/difference smooth structure' (Sóskuthy, 2021), analogous to random slopes and intercepts in linear mixed-effects models.[11] These non-linear random effects model a (potentially) non-linear difference over time with respect to the general time pattern for each speaker and for each level of Voicing and Vowel, allowing for the possibility that speakers differ in intrinsic F0 effects. A by-speaker random smooth of time by Consonant was also added (nested within Voicing), to account for differences between consonants beyond Voicing, and ensure that the modeled effect of Voicing is not biased by certain consonants being more frequent than others. Variation across words, beyond differences in Vowel/Consonant/Voicing, was accounted for by random intercepts, which allow for by-word differences in overall F0 trajectory height.[12] These term specifications allow us to model average CF0 and VF0 effects for each language while controlling for different kinds of variation.

Autocorrelation within F0 trajectories was controlled for by using an AR1 model design, by defining the starting point for each time series as the beginning of the vowel and using function `acf_resid()` from *itsadug* (van Rij et al., 2022) to estimate the autocorrelation parameter `rho`, which was used to fit a new GAMM accounting for autocorrelation of residuals. This step greatly reduced autocorrelation in the residuals. In the final models, deviance explained ranged from 81.8% to 98.8% across languages, suggesting good fits. Model diagnostics using `gam.check()` suggested appropriate levels of smoothing (`k` values). The model formula used for non-tonal languages is provided in (1). The same formula was applied to tonal languages with the addition of Tone as a parametric term, a by-Tone random smooth of time, and interactions of Tone with Voicing and Vowel.

```
(1) model = bam(F0 ∼ Voicing∗Vowel + gender + log_num_syllables.norm +
              s(time) + s(time, by= Voicing) + s(time, speaker, bs='fs', m=1) +
              s(time, speaker, by=Voicing bs='fs', m=1) +
              s(time, consonant, bs='fs', m=1) + s(time, by=Vowel) +
              s(time, speaker, by=Vowel, bs='fs', m=1) + s(word, bs='re'),
              data = data, method='fREML', rho = r2, AR.start = data$start)
```

After presenting the results of the GAMMs on cross-linguistic variation in VF0 and CF0, we will consider in post-hoc analyses whether properties of the languages (e.g. vowel inventory size, type of voicing contrast), shown in Table 2, can explain any of the variation. In this table, Swedish is labeled as having 'both' voicing contrast types since the voicing contrast is realized with aspirated and prevoiced stops (Helgason and Ringen, 2008). Similarly, Thai is labeled as 'both' since the aspirated and voiced categories are compared. See Appendix B for further details.

---

[11] Random smooths were parametrized with a non-linear penalty of order 1 (e.g. m = 1) rather than the default (2), to avoid under-smoothing by speaker differences, following Wieling (2018).

[12] We did not use by-word random smooths due to computational limitations (most words have only 1-2 tokens).

Table 2: Use of F0, type of voicing contrast, stress pattern, and number of vowels for each language.

Language	Use of F0	Voicing Contrast	Stress Pattern	Vowels			
				Oral (short)	Nasal	Long	Diphthongs
Bulgarian	Non-tonal	Voicing	Free	6			
Cantonese	Tonal	Aspirating	Tonal	7		1	11
Croatian	Word-accent	Voicing	Word-accent	5			
Czech	Non-tonal	Voicing	Initial	5		5	3
English	Non-tonal	Aspirating	Initial	14			5
French	Non-tonal	Voicing	Final	12	4		
German	Non-tonal	Aspirating	Penultimate/Final	10		7	3
Hausa	Tonal	Voicing	Tonal	5		5	2
Korean	Non-tonal	Other	Free	10			8
Mandarin	Tonal	Aspirating	Tonal	7			13
Polish	Non-tonal	Voicing	Penultimate	6	2		
Portuguese	Non-tonal	Voicing	Penultimate/Final	10	5		
Russian	Non-tonal	Voicing	Free	10			
Spanish	Non-tonal	Voicing	Final	5			5
Swahili	Word-accent	Aspirating	Penultimate	5			
Swedish	Non-tonal	Both	Word-accent	13		12	
Thai	Tonal	Both	Tonal	9		9	3
Turkish	Non-tonal	Voicing	Final	8			
Ukrainian	Tonal	Voicing	Penultimate/Final	6			
Vietnamese	Tonal	Voicing	Tonal	9			

4. Results

4.1. Estimated smooths

A model summary of GAMMs provides both *parametric* terms, indicating the overall height difference between two contours, and *smooth* terms, indicating the shape difference between two contours. The parametric terms allow us to compare overall F0 between levels of Vowel and Voicing, while the smooth terms provide information on differences in F0 trajectory shape between levels of Vowel and Voicing. However, interpreting GAMM results requires more than just referencing model outputs (Wieling, 2018). Moreover, the GAMM coefficients used in this analysis do not correspond straightforwardly to our research questions, which focus on comparing F0 trajectories over the entire vowel for each language and comparing similarities and differences in trajectory shapes across languages. We instead require a combination of visualizing (differences between) smooths and quantitative results derived from model predictions. In the remainder of this section, the effects of Vowel and Voicing on the F0 contour are assessed qualitatively using visualizations produced from the GAMM estimates.[13]

---

[13] Estimates were extracted, processed, and plotted using the *itsadug* and *tidymv* (Coretta, 2023) packages, among others. Difference smooths were plotted by obtaining estimated marginal means using *emmeans* (Lenth, 2024), calculating the estimated difference between *high* and *low* vowels at 0.1 intervals of normalized time, and marginalizing over irrelevant variables (Word, Consonant, Voicing, Speaker; Tone for tone languages); a similar procedure was used to calculate difference smooths for the estimated difference between *voiced* and *voiceless* consonant smooths. Code is available on the OSF site.

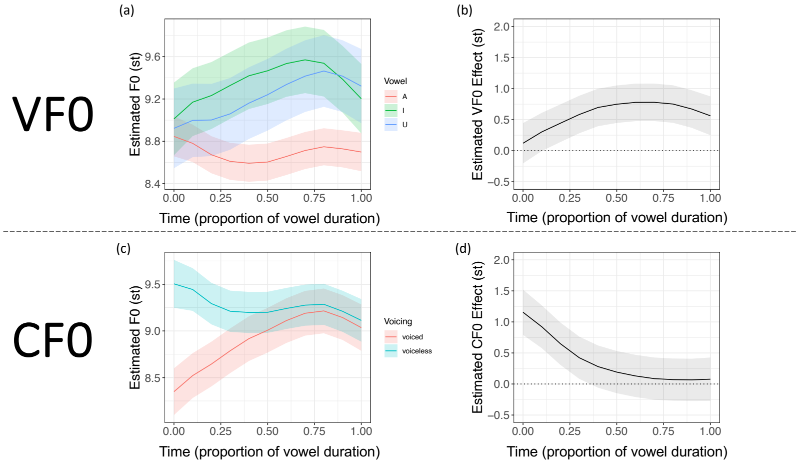


Figure 4: (a) Estimated F0 trajectories for levels of Vowel over normalized time in Croatian. Levels of Vowel are colour-coded: /i/ (green), /u/ (blue), /a/ (red). (b) Difference smooth for the effect of vowel height. (c) Estimated F0 trajectories for levels of Voicing (voiced: red, voiceless: blue) over normalized time (d) Difference smooth for the effect of consonant voicing. Shading shows 95% CIs.

Full model outputs with both parametric and smooth terms are provided in Appendix C.

Given the results of the GAMMs, for each language we can plot (i) *independent smooths* for each level of a given factor and (ii) a *difference smooth* between levels. For example, to interpret the Vowel F0 effect for a language, we can plot the individual Vowel smooths (i.e. /i, u, a/) and a difference smooth between *high* and *low* vowels, as shown in the top panels of Figure 4 for Croatian. Figure 4a plots the estimated smooths for Croatian vowels /i/, /u/, and /a/, separately, with pointwise 95% confidence intervals shown with shaded bands. To see the effect of vowel *height* on F0, we group high vowels /i/ and /u/ together and compare their averaged smooth to that of the low vowel /a/. The difference in F0 trajectory between high (/i/ and /u/) and low (/a/) vowels shows the vowel intrinsic F0 effect over time (Figure 4b). Empirical plots for each language, showing F0 trajectories for all vowels, are shown in Appendix D. Plotting the difference smooth allows us to see where and how the trajectories differ (Sóskuthy, 2017) and make cross-linguistic comparisons.

Figure 4a shows that /i/ and /u/ have an overall higher F0 than the low vowel /a/, with the largest difference in the middle of the vowel and convergence in F0 trajectories at the edges. This pattern is reflected in Figure 4b, where the estimated difference between *high* vs. *low* vowels is greatest (∼0.8 semitones) around the midpoint of the vowel.

Similarly, for the consonant F0 effect, Figure 4c and Figure 4d show the individual Voicing smooths and the difference smooth for Voicing. Figure 4c shows that the difference between the *voiceless* and *voiced* smooths is greatest at the vowel onset and decreases towards the end. This pattern is reflected in Figure 4d, which shows a maximum effect of Voicing (∼1 semitone) at the vowel onset, decreasing over time towards zero, indicating no difference between levels.

From these smooths, we can extract the maximum effect size (henceforth *effect size*) and effect length for each language. The effect size value is the maximum difference in F0 (semitones) over the trajectory for each language. For Croatian, for example, the VF0 effect size is ∼0.8 semitones and the CF0 effect size is ∼1.2 semitones. We define *effect length* as the fraction of the vowel for which the difference between levels is greater than 0. For Croatian, both the VF0 and CF0 effects persist through the entire vowel, meaning the estimated difference between *high* vs. *low* vowels and between *voiceless* vs. *voiced* obstruents is greater than 0 for the entire duration of the vowel. The effect length for both VF0 and CF0 is therefore 1 (i.e. 100% of the vowel). Similarly, 95% CIs for effect length are calculated as the proportion of the vowel for which the lower and upper CIs of the difference smooth are above 0. For Croatian, the lower CI for VF0 is above 0 for 82% of the vowel, while the upper CI is above 0 for the entire duration of the vowel.[14] These estimates are summarized in Table 3.

Table 3: Estimates and 95% CIs for effect size and length computed from the Croatian GAMM model for vowel (VF0) and consonant (CF0) intrinsic F0 effects.

	Effect Size	Length
VF0	0.78 [0.48, 1.08]	1.0 [0.82, 1.0]
CF0	1.16 [0.79, 1.52]	1.0 [0.36, 1.0]

4.2. VF0 effects across languages

Figure 5 shows the estimated difference in F0 (semitones) between high and low vowels (i.e. /i,u/ vs. /a/), over the course of the vowel for each of the 20 languages, illustrating where in the vowel the effect size is largest.

The results in Figure 5 suggest that the VF0 effect is robust: present in every language included in this study, across the entire vowel, though its effect size varies. The maximum F0 difference between high and low vowels varies across languages in magnitude (e.g. ∼1 for Polish, ∼0.7 for Russian) and location within the vowel. Some languages, such as Turkish and German, have a maximum F0 difference towards the center of the vowel, while others, such as English and Swedish, show a maximum difference at vowel onset. Other languages, such as Mandarin and Hausa, show an increase in F0 difference towards the end of the vowel. Figure 6 shows the distribution of VF0 effect size across the 20 languages (see Table 4). Languages have a VF0 effect size between 0.5–1.7 semitones, with one outlier (German) at 2.7 semitones. Colour-coding provided in Figure 6 categorizes languages based on vowel inventory size, allowing readers to examine the extent to which vowel inventory size affects VF0 size, although this categorization was not part of our original analysis plan.

---

[14] Our lower bound for effect length corresponds to the method used to calculate effect length in much previous work on CF0 effects (equivalent to where voiced/voiceless trajectories are no longer statistically significantly different). We report the proportion of the difference smooth above zero so that our estimate does not depend on sample size, which is desirable for any measure of effect size (Kelley and Preacher, 2012).

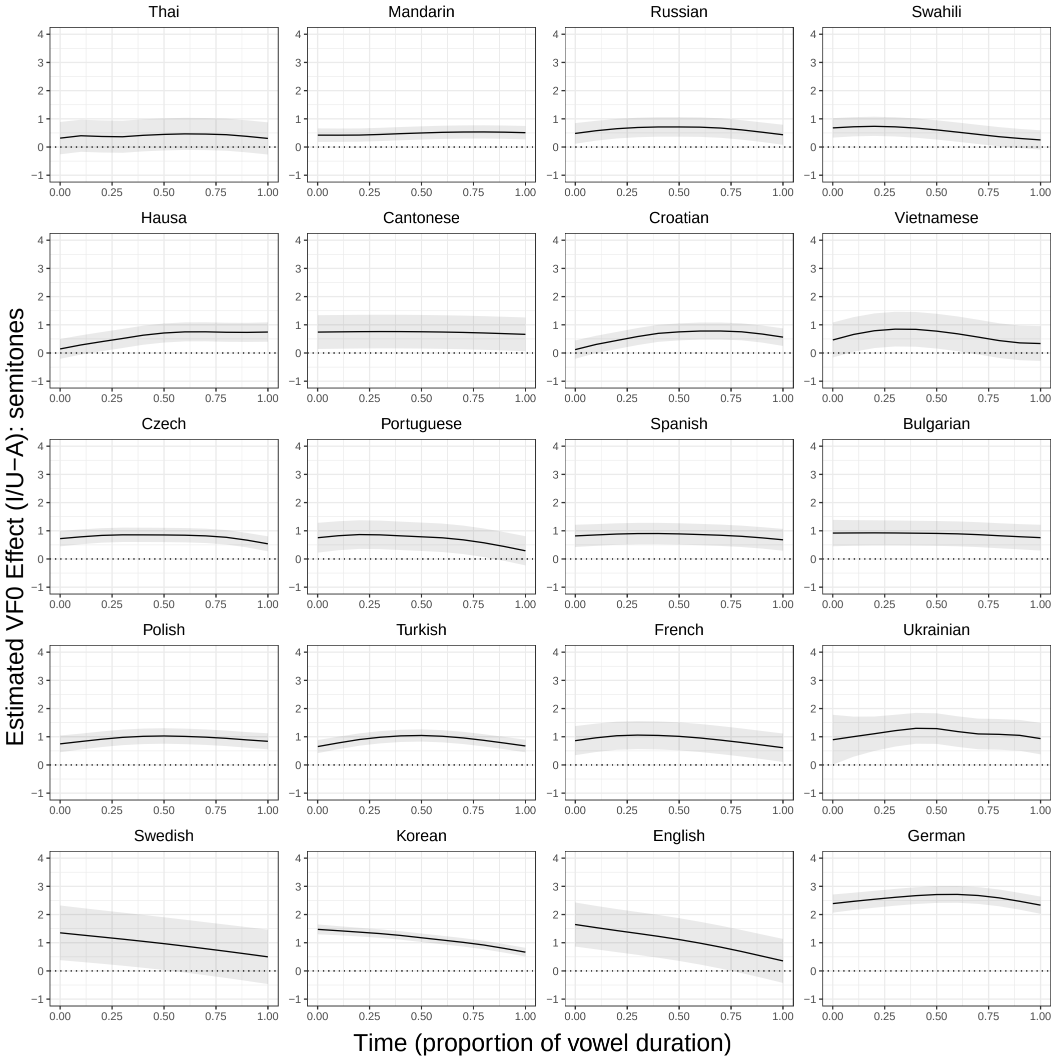


Figure 5: F0 difference smooths for vowel height (high vs. low) over time (normalized) for each of the 20 languages ordered by maximum effect size. Difference is calculated in semitones, pointwise 95% confidence intervals indicated by shaded bands.

Table 4: Estimates and 95% CIs for effect size and length from model results for the vowel intrinsic F0 effect.

Language	VF0 Effect Size		VF0 Effect Length	
	Estimate	95% CI	Estimate	95% CI
German	2.71	[2.42, 3.01]	1.00	[1.00, 1]
English	1.65	[0.86, 2.43]	1.00	[0.73, 1]
Korean	1.47	[1.30, 1.65]	1.00	[1.00, 1]
Swedish	1.35	[0.38, 2.32]	1.00	[0.54, 1]
Ukrainian	1.30	[0.75, 1.85]	1.00	[1.00, 1]
French	1.06	[0.56, 1.56]	1.00	[1.00, 1]
Turkish	1.05	[0.83, 1.26]	1.00	[1.00, 1]
Polish	1.03	[0.75, 1.30]	1.00	[1.00, 1]
Bulgarian	0.93	[0.48, 1.37]	1.00	[1.00, 1]
Spanish	0.90	[0.52, 1.28]	1.00	[1.00, 1]
Portuguese	0.87	[0.36, 1.37]	1.00	[0.82, 1]
Czech	0.86	[0.60, 1.11]	1.00	[1.00, 1]
Vietnamese	0.85	[0.23, 1.46]	1.00	[0.55, 1]
Croatian	0.78	[0.48, 1.08]	1.00	[0.82, 1]
Cantonese	0.76	[0.16, 1.36]	1.00	[1.00, 1]
Hausa	0.75	[0.41, 1.09]	1.00	[0.82, 1]
Swahili	0.73	[0.39, 1.08]	1.00	[0.82, 1]
Russian	0.71	[0.37, 1.06]	1.00	[1.00, 1]
Mandarin	0.54	[0.30, 0.77]	1.00	[1.00, 1]
Thai	0.47	[-0.10, 1.04]	1.00	[0.00, 1]

For VF0 effect length, Table 4 shows that all languages have a value of 1, which indicates that the effect persists through the entire vowel, though for some languages the 95% CI includes lower proportions (down to 0).

These results show that the VF0 effect is robust, with all languages showing a positive effect (F0 higher for high vs. low vowels). Compared to the predictions in Section 3.1, these results suggest that while most languages pattern similarly, certain individual languages (e.g. German) may have a larger VF0 effect size. In terms of effect length, the VF0 effect persisted through the entire vowel for all languages, thus no variability is reported and no correlation between VF0 effect size and length is possible.

4.3. CF0 effects across languages

We turn next to consonant intrinsic F0 effects. Figure 7 shows the estimated difference smooths between *voiced* and *voiceless* obstruents across the 20 languages, illustrating where in the vowel the effect size is greatest.

The results in Figure 7 suggest that the CF0 effect is fairly robust; it is present (and positive) at vowel onset in all languages examined here and usually tapers off over the course of the vowel. There is variation in both CF0 effect size and length across these languages. Some languages display a CF0 difference which persists through the entire vowel (e.g. Hausa, Polish), while others display a CF0 difference that is lost or diminished well before the end of the vowel (e.g. Swahili, Swedish). Figure 8 shows

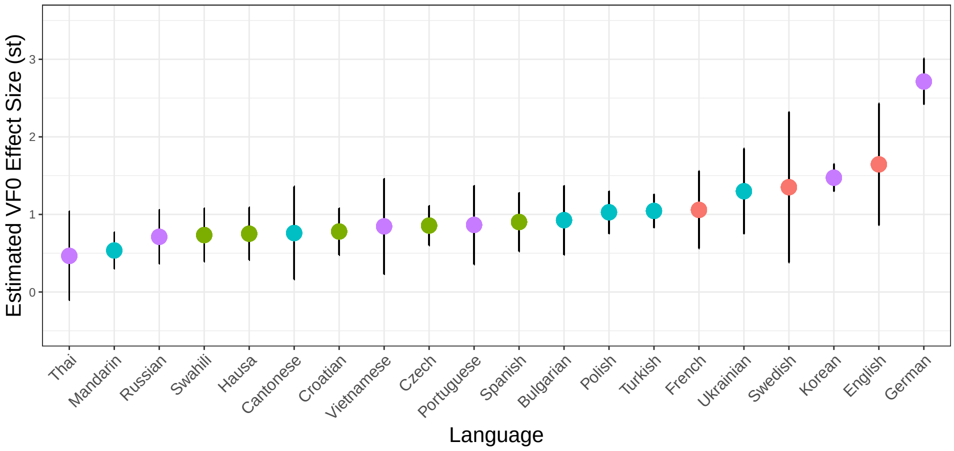


Figure 6: VF0 effect size for each of the 20 languages. 95% CIs indicated by vertical lines. Colours indicate vowel inventory size: Green = 5 or fewer, Blue = 6-8, Purple = 9-10, Red = 11 or more.

the distribution of CF0 effect size across the 20 languages (see Table 5), ranging from approximately 0.4–3.9 semitones. Different colours in Figure 8 indicate different phonetic realizations of the 'voicing' contrast. This allows readers to examine the extent to which the phonetic realization of the contrast affects CF0 size, given that our approach assumes a 'voicing' analysis lumping together phonetic realizations (see Section 2.2), although this categorization was not part of our original analysis plan.

For CF0 effect length, there is also considerable variation. 16 of the 20 languages have a value of 1, indicating an effect that persists through the entire vowel. However, for others the effect can taper off as early as half-way through the vowel, and for most languages, the 95% CI includes lengths below 1.

Compared to the predictions in Section 3.1, these results suggest that while most languages pattern similarly and show a CF0 effect within a certain range, some languages may have larger/longer or smaller/shorter CF0 effects in terms of size and/or length. This pattern is clearer for effect size, with most languages falling into the range of 0.9–3 semitones and individual languages showing relatively larger (Ukrainian) or smaller (Cantonese) effects. It is more difficult to determine whether there is any pattern in which languages have longer or shorter effects, given the greater uncertainty of effect length estimates, indicated by the wide confidence intervals in Table 5. Similar to the VF0 results, the CF0 results do not show a clear correlation between effect size and length, though there may be a weak positive relationship: all languages with estimated effect length lower than 1 are among the bottom 50% in effect size. We turn next to the relationship between the two kinds of intrinsic F0 effects.

4.4. The relationship between VF0 and CF0

Figure 9 compares effect size of VF0 and CF0 across the 20 languages, with identical values (VF0 = CF0) represented by the dashed blue line.

To interpret these results, we can return to the possible outcomes laid out in Section 3.1 (Figures 2 and 3). Figure 9 does not clearly suggest any overall correlation between

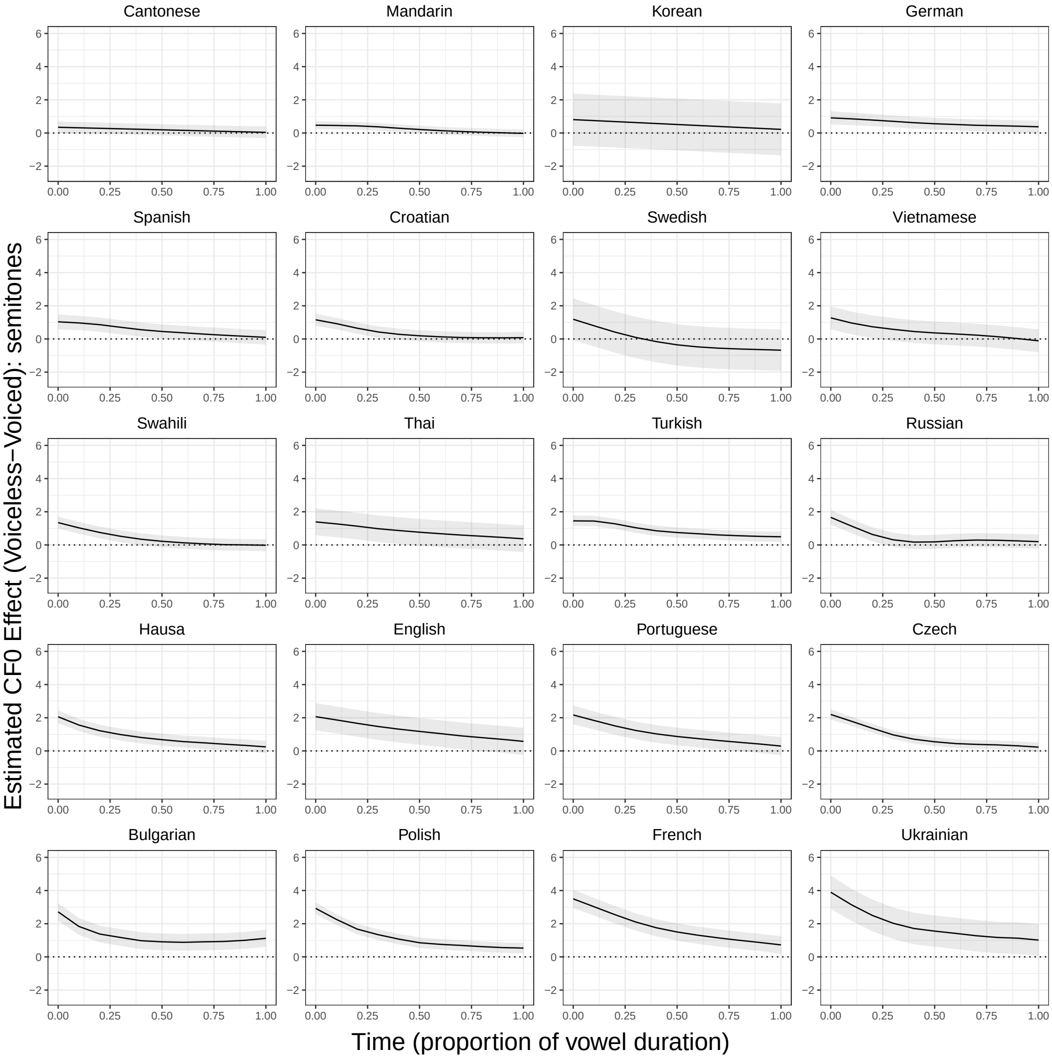


Figure 7: F0 difference smooths for VOICING (voiceless vs. voiced) over time (normalized) for each of the 20 languages ordered by maximum effect size. Difference is calculated in semitones, pointwise 95% confidence intervals indicated by shaded bands.

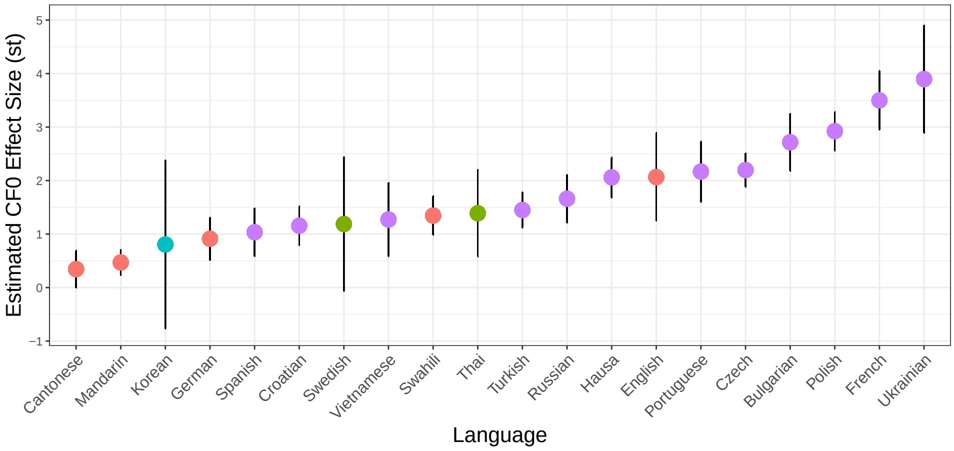


Figure 8: CF0 effect size for each of the 20 languages. 95% CIs indicated by vertical lines. Colours indicate phonetic realization of the ‘voicing’ contrast: Red = Aspirating, Purple = Voicing, Green = Both, Blue = Other.

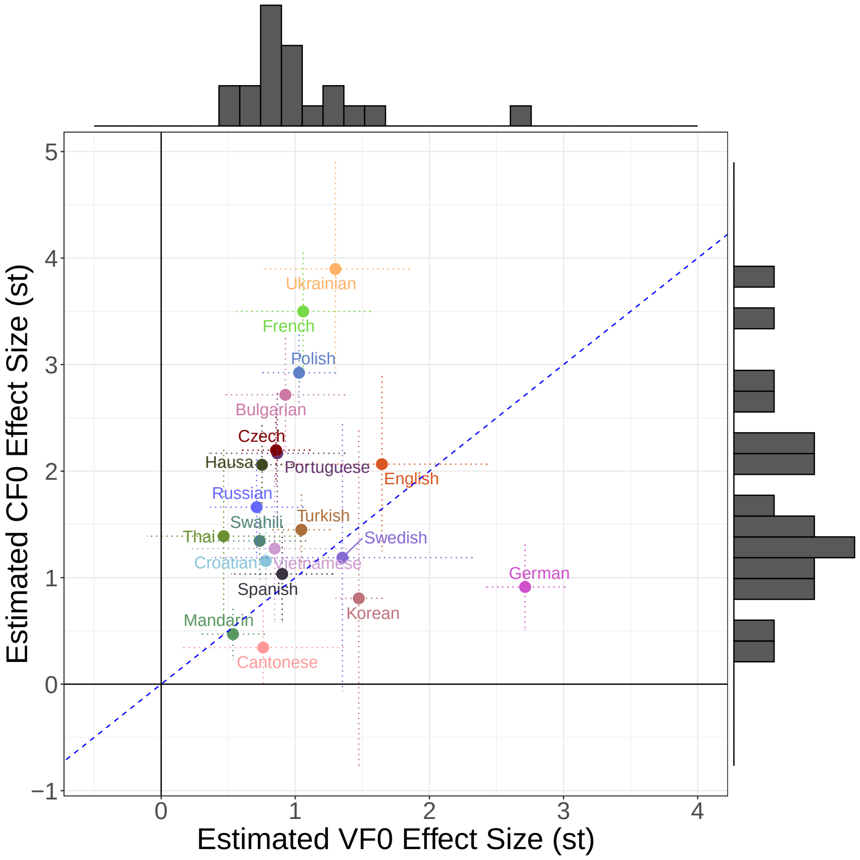


Figure 9: VF0 effect size versus CF0 effect size for each of the 20 languages. 95% confidence intervals indicated by dotted lines. Histograms show marginal distributions.

Table 5: Estimates and 95% CIs for effect size and length from model results for the consonant intrinsic F0 effect.

Language	CF0 Effect Size		CF0 Effect Length	
	Estimate	95% CI	Estimate	95% CI
Ukrainian	3.90	[2.89, 4.90]	1.00	[1.00, 1]
French	3.50	[2.95, 4.05]	1.00	[1.00, 1]
Polish	2.92	[2.56, 3.28]	1.00	[1.00, 1]
Bulgarian	2.72	[2.18, 3.25]	1.00	[1.00, 1]
Czech	2.20	[1.88, 2.51]	1.00	[0.91, 1]
Portuguese	2.17	[1.60, 2.73]	1.00	[0.73, 1]
English	2.07	[1.25, 2.89]	1.00	[0.73, 1]
Hausa	2.06	[1.68, 2.43]	1.00	[0.82, 1]
Russian	1.66	[1.21, 2.11]	1.00	[0.27, 1]
Turkish	1.45	[1.12, 1.78]	1.00	[1.00, 1]
Thai	1.39	[0.58, 2.20]	1.00	[0.46, 1]
Swahili	1.35	[0.98, 1.71]	0.91	[0.36, 1]
Vietnamese	1.27	[0.58, 1.96]	0.91	[0.27, 1]
Swedish	1.19	[-0.06, 2.44]	0.36	[0.00, 1]
Croatian	1.16	[0.79, 1.52]	1.00	[0.36, 1]
Spanish	1.03	[0.59, 1.48]	1.00	[0.54, 1]
German	0.91	[0.51, 1.31]	1.00	[1.00, 1]
Korean	0.81	[-0.77, 2.38]	1.00	[0.00, 1]
Mandarin	0.47	[0.23, 0.71]	0.91	[0.46, 1]
Cantonese	0.34	[-0.003, 0.69]	1.00	[0.00, 1]

the size of the two intrinsic F0 effects. However, some individual languages have substantially larger effect sizes for one or the other effect (Ukrainian and French for CF0, German for VF0), similar to Figures 2d and 3c. Mandarin and Cantonese stand out in having smaller effect sizes for both effects, similar to Figure 2c.

We can make two further observations. First, for most languages the CF0 effect size is larger than the VF0 effect size; only in German is the VF0 effect size clearly larger (considering 95% CIs). Second, CF0 effect sizes are more variable across languages than VF0 effect sizes. Since the effect persisted through the entire vowel for most languages, our conclusions regarding effect length are limited. We do see more variability in the CF0 effect, with some languages showing an effect that did not last for the entire vowel.

Thus, CF0 effects are larger (on average) and more variable across languages (in effect size, and perhaps length) than VF0 effects. These novel empirical findings could help explain why CF0 effects are more likely to lead to sound change compared to VF0 effects, as discussed further below.

4.5. Potential predictors of CF0 and VF0 size

We also considered whether properties of languages (use of F0, type of voicing contrast, stress pattern, and vowel inventory size) could motivate differences in intrinsic F0 effects. Only the first of these analyses (use of F0) was planned, while the other three were post-hoc, testing predictions we learned about after initial examination of

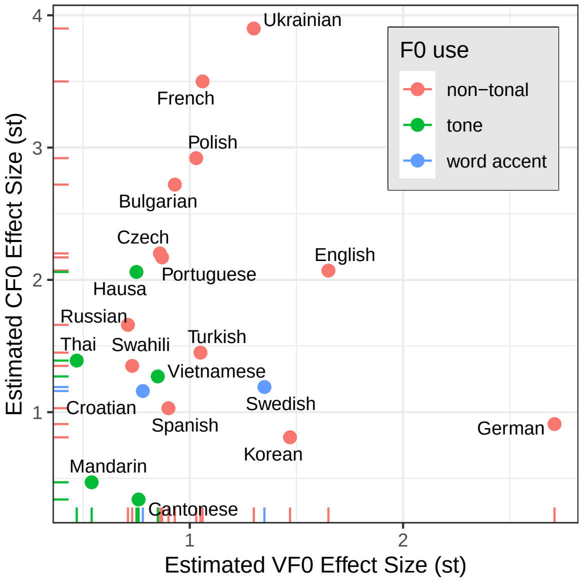


Figure 10: VF0 and CF0 as a function of F0 use in the language: word-accent language, tone language, or neither.

the data (vowel inventory size) or from reviewer suggestions (voicing contrast, stress pattern). Our conclusions from post-hoc analyses should be confirmed on a different cross-linguistic dataset.

4.5.1. Contrastive use of F0

Figure 10 shows VF0 and CF0 effect size, coloured by whether a language is tonal, has a word-accent system, or neither. These results moderately support the proposed trend that languages which use f0 contrastively have smaller VF0 and CF0 patterns. Mandarin and Cantonese (tonal languages) pattern together in having very small VF0 and CF0 effects, and tone or word-accent languages are closer to the lower left-hand corner (i.e. smaller IF0 effects) on average. However, there is no clear division by F0 use, with Spanish and Swahili (both non-tonal languages) having IF0 effects in the same range as the tone languages. A MANOVA found that the cross-linguistic variability predicted by language type is 21% and 11% (adjusted R^2), for VF0 and CF0, respectively. As an exploratory study, visualizing the results is more important than statistical tests, but we note that the MANOVA found that IF0 size (VF0, CF0 together) do not significantly differ by language type ($F(4, 34) = 2.5$, $p = 0.06$).

4.5.2. Laryngeal contrast realization

We then compared languages with different phonetic realizations of laryngeal contrasts, as discussed in Section 2.2, including "true voicing" languages contrasting voiced and voiceless stops, "aspirating" languages contrasting voiceless unaspirated (or devoiced) and aspirated stops, and languages with more complex laryngeal contrast systems. We present this analysis as one method of "splitting" the languages in our study.

Figure 8 presents CF0 effect size, coloured by type of laryngeal contrast, which shows no clear cut distinction between voicing and aspirating languages. For example, Spanish (voicing) and German (aspirating) have similar CF0 effects. On the other hand, an ANOVA found that 18% of the variability in CF0 effect size (adjusted R^2) is predicted by phonetic realization. This trend can also be seen in Figure 8 with the voiced languages generally having larger effect sizes. This effect was not statistically significant ($F(5,14) = 2.5$, $p = 0.10$), however this should not be interpreted too strongly as this was an exploratory analysis and our sample of non-voicing languages is relatively small.

We note that the classification of languages into "voicing" versus "aspirating" is not straightforward, and researchers may differ on how languages should be classified. For example, English and German have sometimes been analyzed as voicing languages, reflecting the realization of stops across dialects and in non-initial positions, as opposed to "true" aspirating languages like Mandarin (e.g. Hunnicutt and Morris, 2009; Jessen and Ringen, 2002; Lombardi, 1999). If English and German were classified as voicing rather than aspirating languages, the distribution of CF0 effect size in Figure 8 would show a somewhat clearer division between language types, though the number of aspirating languages would then be too small to draw firm conclusions. We must therefore leave room for different interpretations that can be drawn given different methodological and/or theoretical viewpoints, and using data from a larger number of languages.

4.5.3. Prominence

We also examined whether languages with different stress patterns pattern differently. Our dataset uses utterance-initial tokens, but investigating many languages introduces different prosodic systems, creating differences in the percentage of utterance-initial words having prominence. For example, words in Czech are almost always initial-stressed, while words in Polish are almost always penultimate-stressed. If differences in prosodic systems correlate with the magnitude of IF0 effects, languages with stress closer to initial position should show larger IF0 effects, since prominent syllables show larger IF0 effects than non-prominent syllables (Ladd and Silverman, 1984; Petersen, 1978; Steele, 1986), especially for VF0 (Kingston, 2011).

Figure 11 plots VF0 and CF0 effect size, coloured by whether a language has initial, final, penultimate, or free stress. While a larger sample might reveal a clearer pattern, the current distribution shows no robust patterns of intrinsic F0 effect size related to stress type. For example, there is no clear difference in effect sizes between "initial" and "penultimate/final" languages, which are expected to differ maximally in the percentage of utterance-initial prominent syllables.

For our dataset, we also calculated an approximate percentage of words with initial stress for 11 languages for which this was possible.[15] Figure 12 plots percentage of words

---

[15] These percentages are based on the subset of words in each dataset for which stress information could be determined, ranging from 30% to 100% of tokens. Stress information was extracted from corpus transcription when available (English, Portuguese); from pronunciation lexicons (German: Baayen, Piepenbrock and Gulikers, 1995; Spanish: Sebastián-Gallés, 2000; Swedish: Nordisk Sprakteknologi, 2020; Croatian, Russian, Ukrainian: Lee et al., 2020); or from assuming a default stress pattern (Czech, French, Polish).

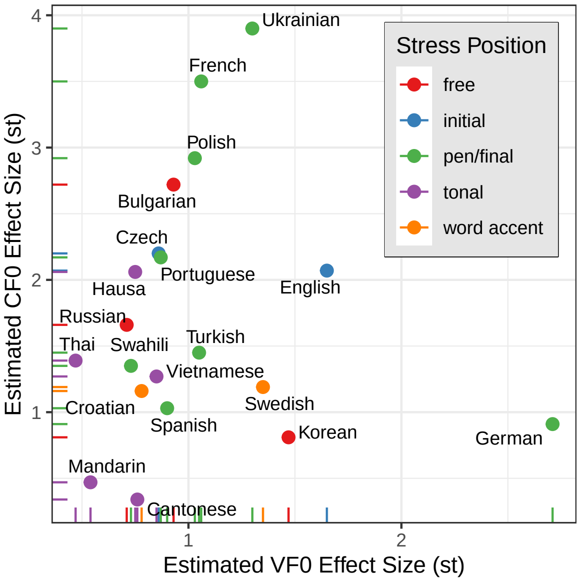


Figure 11: VF0 and CF0 as a function of stress pattern in the language: free, initial, penultimate/final, word-accent, tonal.

with initial prominence by VF0 and CF0 effect size, suggesting a positive correlation for VF0 (Spearman's $\rho = 0.342$, $p = 0.30$) and a negative correlation for CF0 ($\rho = -0.0865$, $p = 0.80$). While neither effect is statistically significant, they are 'moderate' (for VF0) and 'very small' by Cohen (1988)'s rules of thumb. The VF0 effect is in the predicted direction: languages with stress closer to initial position might show larger effect sizes. However, the CF0 effect trends in the opposite (unexpected) direction. There is thus tentative evidence for the expected relationship between stress pattern and VF0 size, with percentage of words with initial prominence explaining 12% ($=0.342^2$) of the cross-linguistic variation in VF0 size. Much less variation is explained for CF0 size ($0.7\% = -0.0865^2$). We acknowledge that our consideration of prosody is only a first approximation: we have not measured prosody acoustically or controlled for various prosodic factors that could affect F0 variability. Examining cross-linguistic differences in intrinsic F0 effects while carefully controlling for prosodic factors is an interesting avenue for future work.

4.5.4. Vowel inventory size

A final factor we consider is vowel inventory size. Previous work has noted that a larger vowel inventory may motivate greater VF0 effects in order to enhance contrasts between vowel categories (Van Hoof and Verhoeven, 2011). In contrast, there would be no reason to expect a relationship with the CF0 effect.

Following Whalen and Levitt (1995), we take vowel inventory size to be the number of monophthongal distinctive vowels (short oral vowels in Table 2), since F0 is hypothesized to enhance vowel height differences in F1/F2 space. (Nasal vowels, long vowels, and diphthongs would not be in need of enhancement.) Figure 13 shows that VF0, but not

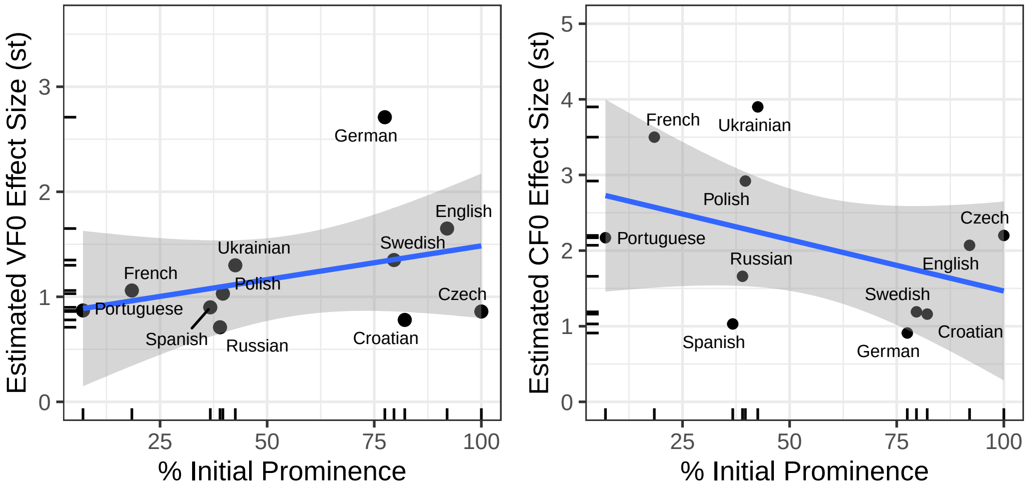


Figure 12: Effect size for VF0 (left) and CF0 (right) as a function of percentage of initial-stressed words in 11 languages.

CF0, effect size seems positively correlated with vowel inventory size, with Spearman's ρ of 0.406 ($p = 0.08$) and 0.0715 ($p = 0.76$), respectively, though statistical tests should be taken with caution in this exploratory study. These are 'moderate' and 'very small' effects, by Cohen's (1988) rules of thumb. Thus, there is tentative evidence for the expected relationship between vowel inventory size and VF0 size, with vowel inventory size explaining 16% (= 0.406^2) of the cross-linguistic variation in VF0 size. Much less variation is explained for CF0 size (0.5% = 0.0715^2).

4.5.5. Summary

Interpreting the results for individual languages within the broader distribution of intrinsic F0 effects across languages, it is clear there is substantial variation among languages that cannot be explained by any single factor. Some of this variation is probably an automatic consequence of different articulations of the 'same' underlying phoneme, but we leave open the question of how much. More generally, to what extent the observed variation can be attributed to language-specific differences, and the ways these effects are due to intrinsic versus extrinsic influences, are topics we leave for future work. Overall, as both VF0 and CF0 effects were found to be robust and variable across languages, our results support a hybrid approach (e.g. Hoole and Honda, 2011; Keating, 1990; Maddieson, 1997): While intrinsic F0 effects may originate in the articulatory properties of vowel and voicing production, the effects can be in part controlled by speakers, whether to maximize a contrast or as arbitrary language-specific differences. Furthermore, while there are probably systematic reasons why languages vary, related to their phonological systems (e.g. use of F0, vowel inventory size), these seem to be weak effects individually, leaving much of the variation across languages unexplained.

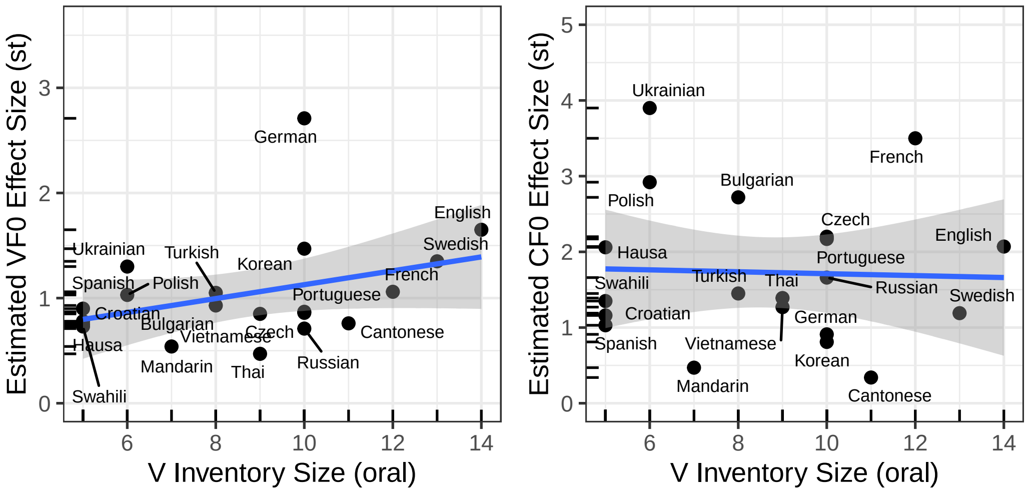


Figure 13: Size of VF0 (left) and CF0 (right) effects as a function of number of oral vowels in the language (from Table 2).

5. Discussion

The current analysis explored variation in intrinsic F0 effects across 20 languages using large corpora of read speech. The primary goal was a better understanding of the distribution of both vowel and consonant intrinsic F0 effects across languages, while addressing limitations of previous studies and exploring questions that have not been empirically addressed in the literature. We examined the distribution of intrinsic F0 effects across languages in terms of both effect size and length, using GAMMs to visualize the full trajectory of F0 across vowels, and probed the relationship between vowel and consonant intrinsic F0 effects within and across languages.

We found that intrinsic F0 effects are robust across the languages examined in this study, with positive effects present in all languages, but with variation observed for both vowel and consonant intrinsic F0 effects. VF0 effect size was concentrated around 1 semitone (mean=0.95, median=0.87), with German having a substantially larger effect and Mandarin and Thai showing relatively smaller effects. CF0 effect size for most languages fell within 1–3 semitones (mean=1.65, median=1.42), with French and Ukrainian showing larger effects and Cantonese and Mandarin showing smaller effects. We note that our report of CF0 effect size mean and range here assumes that the languages are comparable enough to evaluate variability. We discuss implications of this approach below. Regarding the time-course of intrinsic F0 effects, VF0 effects peaked near the middle of the vowel and persisted through the entire vowel for all 20 languages, while CF0 effects peaked at vowel onset and persisted through the entire vowel for 16 of the 20 languages.

Comparing our VF0 results to the meta-analysis of Whalen and Levitt (1995), we find differences in both the overall effect size and the degree of variation across languages. Our average VF0 effect of around 1 semitone is smaller than W&L's report of 1.65

semitones. We also found a smaller range of VF0 effect sizes across languages: 0.47–1.65 semitones (excluding German), compared to 0.5–3.5 semitones for W&L (for studies with at least 10 speakers). This may come down to methodological differences. We examined read speech, which is associated with less careful speech and potentially less active "control" of intrinsic F0 effects compared to words spoken in isolation, as examined in most studies reported in W&L. This would be consistent with findings that VF0 differences are smaller or lost in non-prominent syllables (Ladd and Silverman, 1984; Petersen, 1978; Steele, 1986). Furthermore, our study uses more comparable data across languages, which should decrease variation compared to the different methodologies across studies in W&L. Our data is broadly consistent with W&L's hypothesis that the VF0 effect is largely automatic and variance comes from methodological differences, though German is an important outlier.

Individual differences may also contribute to the cross-linguistic variation in intrinsic F0 effects observed here. As Hoole and Honda (2011) and Kirby (2018) have found, individual speakers can vary greatly in effect size, due in part to different articulatory strategies, and the distribution of strategies used could conceivably differ by language. Individual characteristics such as gender and age could also contribute to observed variation. Addressing the individual and joint influence of potential factors and probing individual differences in intrinsic F0 effects are interesting avenues for future work.

After mapping the distribution of effect size and length separately for VF0 and CF0 effects, we investigated whether there might be a relationship between effect size and length. For VF0, there was no variation in effect length, so we could not detect any relationship. We found weak evidence for a relationship between size and length of CF0 effects. The four languages with an effect lasting shorter than the full vowel all tended to have smaller effect sizes. This may point to a tendency to minimize CF0 effects in these languages, or it may simply reflect the fact that smaller effects are harder to reliably detect throughout the full vowel.

A novel empirical contribution of our study is examination of the relationship between vowel and consonant intrinsic F0 effects across languages. We tested whether certain languages have larger and smaller intrinsic F0 effects (across CF0 and VF0). We found that languages can show a larger IF0 effect for vowels or consonants, but not for both simultaneously. Languages that were outliers in one intrinsic effect were not outliers in the other. This is consistent with the view that larger intrinsic F0 effects reflect some independent control by the speaker, again aligning with a hybrid approach where articulatory mechanisms are fundamentally responsible for intrinsic F0 effects, but articulatory differences can be reinforced by active adjustments (Hoole and Honda, 2011), whether for enhancing phonological contrasts or as arbitrary learned differences (Jacewicz and Fox, 2015; Kingston and Diehl, 1994).

At the other extreme, Cantonese and Mandarin had the smallest CF0 effect sizes, while Thai and Mandarin had the smallest VF0 effect sizes among the languages examined. Mandarin was also one of the few languages where the CF0 effect did not last throughout the entire vowel. Although previous work, which has focused more on Mandarin than other tone languages, has suggested that Mandarin's exceptionally small IF0 effects are due to it being a tonal language, our study shows that tone alone cannot explain Mandarin's small CF0 and VF0 effects, as tone languages only have slightly smaller IF0 effects on average compared to non-tone languages.

One possible explanation for this distribution of a larger effect as needed versus minimizing across the board is that enhancement of a particular contrast does not necessarily require enhancement of other contrasts. In contrast, having a small intrinsic F0 effect applies to the language more generally, thus requiring all intrinsic F0 effects to be smaller. While this is just one possible hypothesis based on our results, we emphasize that only by mapping out the *distribution* of IF0 effects across languages can we generate such hypotheses. We leave more comprehensive investigations of such hypotheses for future work.

Finally, a further question we could address by examining the distribution of both consonant and vowel driven effects was their potential roles as precursors to sound change. We found that CF0 effects were larger and more variable than VF0 effects, on the whole, in effect size and to some extent length. We also noted a tendency for the languages with a shorter CF0 effect to also have a smaller CF0 effect size. Thus we have evidence of languages with larger CF0 effects as well as other languages having shorter and smaller CF0 effects. VF0 effects on the other hand had variation only in size of the effect, and less so.

These results suggest a possible explanation for the observation that CF0 effects lead to sound change far more often than VF0 effects. CF0 effects may be a more likely precursor to sound change both because they are (i) larger and (ii) more variable across languages. Both factors make it more likely that a given language at a given time will have an unusually large CF0 effect. In a randomly sampled set of languages, high mean and variance of CF0 effects will sometimes result in a large CF0 effect, while low mean and variance of VF0 effects will rarely produce a large VF0 effect. If only sufficiently large effects can be phonologized through listener misperception, we expect CF0 to be phonologized more often. While (i) lines up with the notion from the sound change literature that 'more robust' phonetic precursors are more likely to be phonologized (e.g. Ohala, 1994; Moreton, 2008), we are not aware of work claiming (ii).

Our logic in linking (i) and (ii) is similar to Hombert (1977) and Hombert, Ohala and Ewan (1979) in pointing to CF0 as a 'more perceptible' phonetic precursor than VF0, but differs in the source of this perceptibility difference: Hombert, Ohala and Ewan argue that the CF0 effect is more perceptually salient because it is separable from its consonantal trigger, while the VF0 effect is inherently tied to the vowel; we claim that large VF0 effects simply do not occur. Both explanations could play a role: the input perceivers receive and speech perception itself are sources of the 'phonetic bias' which are at the root of much sound change (Garrett and Johnson, 2013).

This study demonstrates how mapping out the cross-linguistic distribution of 'intrinsic' effects (i.e. possible phonetic precursors) is possible with modern tools and speech corpora, and is relevant for understanding sound change. While our 'lumping' approach has certain limitations, we believe this study forms a good starting point for cross-linguistic analysis of intrinsic F0 effects. Our approach assumes that languages are comparable enough for variability in intrinsic F0 effect size to be meaningfully interpreted. A 'splitting' approach would examine smaller groups of languages with phonetically more-similar contrasts, potentially leading to different conclusions. Thus, our variation-based explanation for the asymmetry between CF0 and VF0 effects leading to sound change depends on our 'lumping' approach. Future work could consider alternative methods of analyzing intrinsic F0 effects across languages to confirm the proposed hypotheses, and examine other pairs of intrinsic effects which do and do not commonly lead to sound change (e.g. different kinds of assimilation), to test the hypothesis that size and variability across languages can help explain why only certain phonetic precursors commonly lead to change. We also note that our exploratory analyses, while limited by sample size, do point to some potential factors that could account for variation in effect size across languages. Each analysis accounted for between 0.5% and 21% of the variation between languages, with contrast type and F0 use being most predictive of CF0 and F0 use, vowel inventory size and prominence type being the most predictive for VF0. These are potentially important predictors, and future work should explore them further. Thus while this paper doesn't achieve a real understanding of the cross-linguistic distribution of intrinsic F0 effects, we hope that it prepares it.

6. Conclusion

This study investigated vowel and consonant intrinsic F0 effects across 20 languages using large corpora of read speech. We found that both vowel and consonant effects are robust: present and positive in all languages examined, with some variation in effect size. A few languages had substantially larger vowel or consonant intrinsic F0 effects, while some showed relatively small effects, particularly for CF0. Most cross-linguistic variation was unexplained by any single factor. These results are in line with a 'hybrid' approach (Chen, 2011; Hanson, 2009; Hoole and Honda, 2011; Van Hoof and Verhoeven, 2011), suggesting that intrinsic F0 effects have both 'automatic' and 'controlled' components. The CF0 effect was more variable than the VF0 effect across languages, at least under the 'lumping' approach used in this study, and in almost every language the CF0 effect was larger than the VF0 effect. This pattern correlates with the cross-linguistic tendency for CF0 effects to lead to sound change more than VF0 effects. These results contribute to our understanding of intrinsic F0 effects, their cross-linguistic distribution, and their role as precursors to sound change.

Appendix A. Outlier data exclusion procedure

For each trajectory, we calculated the maximum length of consecutively skipped points, both as a number of points and as a percentage of the trajectory, as well as the slope over the skipped points. The goal was to exclude trajectories with numerous skipped points and/or large slopes over any number of skipped points. However, excluding trajectories based solely on the number of skipped points, regardless of slope, would have excluded trajectories that were otherwise suitable for analysis (e.g. an overall stable trajectory). To avoid this, we excluded only trajectories that had *any* pitch skip *and* a large slope. The slope value for exclusion was chosen for each language by testing a range of values and visually inspecting the trajectories that would be excluded, to choose a value that would not eliminate trajectories suitable for analysis.

For each trajectory, we also calculated the minimum distance between the trajectory (at any time point) and the speaker's mean pitch as well as the overall range of the trajectory. Trajectories with a high minimum difference (i.e. very different from the speaker's mean pitch) and a low range (i.e. relatively stable) were excluded. This was done using z-scored pitch within-speaker to make finding outlier trajectories more comparable across speakers.

Finally, we used first and second derivative values to approximate pitch changes within a trajectory, aiming to exclude trajectories with large pitch differences even when no pitch skips were involved. This was done independently for each language to determine appropriate values that would retain trajectories suitable for the analysis. The percentage of excluded data for each language based on these procedures is summarized in Table A.6, and the full procedure for each language is shown in the OSF project.

Table A.6: Percentage of data excluded from each language as outliers. 'Tokens' refers to F0 measurement points.

Language	% Excluded	
	Tokens	Trajectories
Bulgarian	3.1	2.9
Cantonese	2.0	1.9
Croatian	2.9	2.1
Czech	2.6	1.6
English	3.3	2.1
French	2.4	1.9
German	2.0	1.5
Hausa	2.6	1.2
Korean	3.0	2.4
Mandarin	1.8	1.2
Polish	2.2	1.9
Portuguese	1.7	1.8
Russian	2.3	1.6
Spanish	2.9	2.1
Swahili	2.5	2.1
Swedish	4.1	2.4
Thai	1.9	1.5
Turkish	1.9	1.4
Ukrainian	2.2	1.8
Vietnamese	2.8	1.9

Appendix B. Vowel and obstruent phones for each language

Table B.7 lists the vowels and obstruents that were included in the analysis for each language. Obstruents which had no utterance-initial tokens followed by i/a/u, such as English /z/, are not listed.

In order to restrict to two stop series per language, 'voiced' and 'voiceless', the following obstruents were excluded:

- Hausa: ejectives and implosives (ts', c', k', ɓ, ɗ)
- Korean: lax stops (p, t, k)
- Swahili: pre-nasalized stops (mb, mv, nd, ŋg, ndʒ, nz)
- Thai: tenuis obstruents (p, t, c, k, ʔ)
- Vietnamese: aspirated stop (t^h)

Table B.7: List of vowels and obstruents included for each language. Obstruent labels for languages from the GlobalPhone corpus are taken from the GlobalPhone documentation, which assumes a [voice] analysis, except for languages where the 'voicing' contrast involves differences in aspiration (German, Mandarin, and Swedish). Labels for Swahili and Ukrainian, for which there was no GlobalPhone documentation available, are taken from Polomé (1967), Mohamed (2001), and Pompino-Marschall, Steriopolo and Żygis (2017). Labels for English are based on conventional descriptions. Labels for Cantonese are based on IARPA documentation.

Language	Vowels	Obstruents
Bulgarian	i, a, u	p, b, t, d, k, ɡ, f, v, s, z, ʃ, ʒ, ts, tʃ
Croatian	i, a, u	p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, x, ts, tʃ, dʒ, dʑ, tɕ
Cantonese	ɐ, iː, ɑː, uː	f, p, ts, s, t, tsʰ, tʰ, pʰ, k, kʰ
Czech	ɪ, a, u, iː, aː, uː	p, b, t, d, k, ɡ, f, v, s, ʃ, ʒ, ts, tʃ
French	i, a, u	p, b, t, d, k, g, f, v, s, ʃ, ʒ
English	i, u, ɑ	pʰ, p, tʰ, t, kʰ, k, f, ð, s, ʃ, tʃ, dʒ
German	i, a, u, iː, aː, uː	pʰ, p, tʰ, t, kʰ, k, f, v, s, z, ʃ, ç, x, ts
Hausa	i, a, u, iː, aː, uː	s, z, t, p, g, b, ʃ, dʒ, d, k, d, ɸ
Korean	i, a, u	s, tɕ, pʰ, tʰ, tɕʰ, kʰ, p͈, t͈, k͈, tɕ͈, s͈
Mandarin	i, a, u	p, pʰ, t, tʰ, k, kʰ, f, s, ts, tsʰ, tʃ, ɕ, tɕ, ʈʂ, ʈʂʰ ʂ
Polish	i, a, u	p, b, t, d, c, k, g, f, v, s, z, ʃ, ʒ, tʃ, tɕ, ʑ, tʑ, dʑ
Portuguese	i, a, u	p, b, t, d, k, g, f, v, s, z, ʃ
Russian	i, a, u	p, b, t, d, g, k, f, v, s, z, ʃ, ʒ, ts, tɕ, tʃ, ʃtʃ, x
Spanish	i, a, u	p, b, t, d, k, g, f, s, x, ɣ, tʃ
Swahili	i, a, u	p, b, t, d, k, g, f, v, s, z, dʒ
Swedish	i, a, u, iː, aː, uː	pʰ, b, tʰ, d, kʰ, g, f, v, s, ʃ, ç
Thai	i, a, u, iː, aː, uː	pʰ, b, tʰ, d, cʰ, kʰ, kʰw
Turkish	i, u, ɑ	p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, tʃ, dʒ
Ukrainian	i, a, u	p, b, t, d, k, f, s, z, ʃ, ʒ, ts, tʃ, x
Vietnamese	i, a, u, aː	ɓ, k, ʈ, x, v, s, ɗ, tɕ, ɣ, s, t, f, ɟ

- All applicable languages (Cantonese, English, German, Hausa, Korean, Mandarin, Polish, Swedish, Turkish, Vietnamese): voiceless glottal fricative /h/

Appendix C. Full Model Summaries

Summaries are shown for the parametric terms and smooth terms for the `bam()` model fitted to each language.

BULGARIAN - PARAMETRIC TERMS

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	7.892	0.345	22.881	0.000	7.216	8.568
Voicing.ordvoiceless	0.785	0.334	2.351	0.019	0.131	1.439
base_vowel.ordI	0.921	0.326	2.830	0.005	0.283	1.559
base_vowel.ordU	0.294	0.629	0.468	0.640	-0.938	1.527
gender_male	-9.232	0.502	-18.385	0.000	-10.216	-8.248
log_utterance_num_syllables.norm	0.033	0.049	0.680	0.497	-0.063	0.129
Voicing.ordvoiceless:base_vowel.ordI	0.045	0.391	0.116	0.908	-0.721	0.811
Voicing.ordvoiceless:base_vowel.ordU	1.033	0.666	1.550	0.121	-0.273	2.338

BULGARIAN - SMOOTH TERMS

Term	edf	Ref.df	*F*	*p*
s(norm_time)	3.621	4.070	4.515	0.001
s(norm_time):Voicing.ordvoiceless	7.742	8.199	15.391	0.000
s(norm_time,speaker)	196.280	691.000	32.389	0.000
s(norm_time,speaker):Voicing.ordvoiceless	222.241	692.000	3.010	0.000
s(norm_time,consonant)	53.831	115.000	11.002	0.117
s(norm_time):base_vowel.ordI	1.000	1.001	1.923	0.166
s(norm_time):base_vowel.ordU	1.861	2.269	0.791	0.467
s(norm_time,speaker):base_vowel.ordI	94.186	557.000	2.463	0.000
s(norm_time,speaker):base_vowel.ordU	64.710	290.000	0.862	0.003
s(word)	386.205	503.000	5.662	0.000

CANTONESE - PARAMETRIC TERMS

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	11.527	0.200	57.629	0.000	11.135	11.919
Voicing.ordvoiceless	0.161	0.217	0.741	0.458	-0.264	0.586
base_vowel.ordI	0.198	0.258	0.770	0.441	-0.306	0.703
base_vowel.ordU	0.376	0.269	1.399	0.162	-0.151	0.902
tone.ord2	-3.386	0.350	-9.671	0.000	-4.072	-2.700
tone.ord3	-2.691	0.257	-10.470	0.000	-3.194	-2.187
tone.ord4	-4.172	0.311	-13.406	0.000	-4.782	-3.562
tone.ord5	-5.532	1.326	-4.171	0.000	-8.131	-2.932
tone.ord6	-3.358	0.300	-11.187	0.000	-3.946	-2.769
gender_male	-8.292	0.251	-33.095	0.000	-8.783	-7.801
Voicing.ordvoiceless:base_vowel.ordI	0.570	0.272	2.095	0.036	0.037	1.103
Voicing.ordvoiceless:base_vowel.ordU	0.208	0.326	0.640	0.522	-0.430	0.847
Voicing.ordvoiceless:tone.ord2	-1.006	0.437	-2.302	0.021	-1.862	-0.149
Voicing.ordvoiceless:tone.ord3	-0.222	0.322	-0.690	0.490	-0.853	0.409
Voicing.ordvoiceless:tone.ord4	0.000	0.000	NaN	NaN	0.000	0.000
Voicing.ordvoiceless:tone.ord5	0.000	0.000	NaN	NaN	0.000	0.000
Voicing.ordvoiceless:tone.ord6	-0.156	0.294	-0.529	0.597	-0.733	0.421
base_vowel.ordI:tone.ord2	0.406	0.469	0.865	0.387	-0.513	1.325
base_vowel.ordU:tone.ord2	-0.179	0.476	-0.375	0.707	-1.113	0.755
base_vowel.ordI:tone.ord3	0.599	0.339	1.767	0.077	-0.065	1.263
base_vowel.ordU:tone.ord3	-0.745	0.496	-1.502	0.133	-1.717	0.227
base_vowel.ordI:tone.ord4	-0.390	0.431	-0.905	0.365	-1.234	0.454
base_vowel.ordU:tone.ord4	-0.177	0.606	-0.292	0.770	-1.366	1.011
base_vowel.ordI:tone.ord5	2.002	1.481	1.351	0.177	-0.901	4.905
base_vowel.ordU:tone.ord5	1.377	2.229	0.618	0.537	-2.992	5.747
base_vowel.ordI:tone.ord6	0.558	0.377	1.480	0.139	-0.181	1.297
base_vowel.ordU:tone.ord6	-0.813	0.451	-1.802	0.072	-1.697	0.071

CANTONESE - SMOOTH TERMS

Term	edf	Ref.df	*F*	*p*
s(norm_time)	5.819	6.877	21.182	0.000
s(norm_time):Voicing.ordvoiceless	1.001	1.002	26.926	0.000
s(norm_time):tone.ord2	4.540	5.729	27.019	0.000
s(norm_time):tone.ord3	3.259	4.211	7.873	0.000
s(norm_time):tone.ord4	2.465	3.174	77.889	0.000
s(norm_time):tone.ord5	3.447	4.446	8.594	0.000
s(norm_time):tone.ord6	3.523	4.512	35.208	0.000
s(norm_time,speaker)	830.390	3604.000	16.086	0.000
s(norm_time,speaker):Voicing.ordvoiceless	427.052	3585.000	1.614	0.000
s(norm_time,consonant)	5.857	88.000	0.995	0.119
s(norm_time):base_vowel.ordI	1.002	1.004	9.051	0.003
s(norm_time):base_vowel.ordU	2.675	3.345	2.357	0.063
s(norm_time,speaker):base_vowel.ordI	1464.501	3265.000	3.349	0.000
s(norm_time,speaker):base_vowel.ordU	533.715	2265.000	1.836	0.000
s(word)	983.041	1248.000	5.947	0.000

CROATIAN - PARAMETRIC TERMS

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	8.440	0.283	29.812	0.000	7.885	8.995
Voicing.ordvoiceless	0.465	0.208	2.233	0.026	0.057	0.873
base_vowel.ordI	0.807	0.284	2.838	0.005	0.250	1.365
base_vowel.ordU	0.632	0.314	2.010	0.044	0.016	1.248
gender_male	-9.415	0.510	-18.461	0.000	-10.415	-8.415
log_utterance_num_syllables.norm	0.582	0.063	9.290	0.000	0.459	0.704
Voicing.ordvoiceless:base_vowel.ordI	-0.221	0.341	-0.650	0.516	-0.890	0.447
Voicing.ordvoiceless:base_vowel.ordU	-0.174	0.380	-0.457	0.648	-0.918	0.571

CROATIAN - SMOOTH TERMS

Term	edf	Ref.df	*F*	*p*
s(norm_time)	4.235	4.841	3.328	0.006
s(norm_time):Voicing.ordvoiceless	4.732	5.487	11.048	0.000
s(norm_time,speaker)	302.031	826.000	44.334	0.000
s(norm_time,speaker):Voicing.ordvoiceless	165.570	810.000	1.665	0.000
s(norm_time,consonant)	36.857	157.000	0.716	0.000
s(norm_time):base_vowel.ordI	4.401	5.263	7.039	0.000
s(norm_time):base_vowel.ordU	3.448	4.301	4.615	0.001
s(norm_time,speaker):base_vowel.ordI	235.597	742.000	3.505	0.000
s(norm_time,speaker):base_vowel.ordU	138.334	714.000	1.757	0.000
s(word)	677.325	842.000	6.417	0.000

Czech - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	7.224	0.259	27.879	0.000	6.716	7.732
Voicing.ordvoiceless	0.609	0.233	2.620	0.009	0.154	1.065
base_vowel.ordI	0.537	0.187	2.862	0.004	0.169	0.904
base_vowel.ordU	0.819	0.255	3.207	0.001	0.319	1.320
gender_male	-10.234	0.413	-24.807	0.000	-11.042	-9.425
log_utterance_num_syllables.norm	0.305	0.054	5.686	0.000	0.200	0.410
Voicing.ordvoiceless:base_vowel.ordI	0.445	0.251	1.770	0.077	-0.048	0.937
Voicing.ordvoiceless:base_vowel.ordU	-0.003	0.317	-0.008	0.994	-0.624	0.619

Czech - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	1.000	1.000	6.247	0.012
s(norm_time):Voicing.ordvoiceless	5.766	6.599	32.931	0.000
s(norm_time,speaker)	319.464	914.000	15.041	0.000
s(norm_time,speaker):Voicing.ordvoiceless	263.826	881.000	1.939	0.000
s(norm_time,consonant)	30.086	115.000	0.998	0.247
s(norm_time):base_vowel.ordI	4.846	5.772	9.413	0.000
s(norm_time):base_vowel.ordU	3.799	4.733	3.157	0.013
s(norm_time,speaker):base_vowel.ordI	84.182	906.000	0.819	0.000
s(norm_time,speaker):base_vowel.ordU	82.802	840.000	0.721	0.000
s(word)	484.159	661.000	3.624	0.000

English - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	8.498	0.494	17.200	0.000	7.530	9.467
Voicing.ordvoiceless	1.308	0.566	2.310	0.021	0.198	2.418
base_vowel.ordI	0.674	0.665	1.013	0.311	-0.630	1.977
base_vowel.ordU	1.557	0.793	1.964	0.050	0.003	3.112
gender_male	-8.486	0.498	-17.027	0.000	-9.462	-7.509
log_utterance_num_syllables.norm	1.066	0.082	12.924	0.000	0.904	1.227
Voicing.ordvoiceless:base_vowel.ordI	0.636	0.789	0.806	0.421	-0.911	2.183
Voicing.ordvoiceless:base_vowel.ordU	-0.838	1.040	-0.806	0.420	-2.878	1.201

English - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	3.553	4.168	7.128	0
s(norm_time):Voicing.ordvoiceless	2.317	2.708	18.818	0
s(norm_time,speaker)	298.756	1171.000	2.405	0
s(norm_time,speaker):Voicing.ordvoiceless	110.889	1161.000	0.640	0
s(norm_time,consonant)	29.245	106.000	0.876	0
s(norm_time):base_vowel.ordI	2.368	2.813	6.959	0
s(norm_time):base_vowel.ordU	2.037	2.492	18.076	0
s(norm_time,speaker):base_vowel.ordI	206.633	1158.000	0.483	0
s(norm_time,speaker):base_vowel.ordU	142.632	1024.000	0.527	0
s(word)	180.920	244.000	5.042	0

French - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.323	0.371	25.114	0.000	8.595	10.050
Voicing.ordvoiceless	0.991	0.369	2.690	0.007	0.269	1.714
base_vowel.ordI	0.359	0.393	0.912	0.362	-0.412	1.130
base_vowel.ordU	0.557	0.594	0.938	0.348	-0.607	1.720
gender_male	-9.631	0.452	-21.300	0.000	-10.517	-8.744
log_utterance_num_syllables.norm	0.057	0.062	0.924	0.355	-0.064	0.178
Voicing.ordvoiceless:base_vowel.ordI	0.698	0.509	1.373	0.170	-0.298	1.695
Voicing.ordvoiceless:base_vowel.ordU	1.116	0.676	1.651	0.099	-0.209	2.441

French - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	4.495	5.223	14.344	0.000
s(norm_time):Voicing.ordvoiceless	4.784	5.439	34.019	0.000
s(norm_time,speaker)	92.175	894.000	3.763	0.000
s(norm_time,speaker):Voicing.ordvoiceless	170.983	888.000	0.564	0.000
s(norm_time,consonant)	31.411	97.000	1.325	0.000
s(norm_time):base_vowel.ordI	4.750	5.825	7.444	0.000
s(norm_time):base_vowel.ordU	2.301	2.858	3.692	0.015
s(norm_time,speaker):base_vowel.ordI	65.610	787.000	0.831	0.000
s(norm_time,speaker):base_vowel.ordU	146.676	836.000	0.517	0.000
s(word)	203.088	260.000	11.238	0.000

GERMAN - PARAMETRIC TERMS

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.060	0.489	18.521	0.000	8.101	10.018
Voicing.ordvoiceless	0.998	0.286	3.491	0.000	0.438	1.558
base_vowel.ordI	2.560	0.201	12.767	0.000	2.167	2.953
base_vowel.ordU	3.216	0.310	10.380	0.000	2.609	3.823
gender_male	-10.379	0.899	-11.549	0.000	-12.140	-8.618
log_utterance_num_syllables.norm	0.320	0.037	8.738	0.000	0.248	0.392
Voicing.ordvoiceless:base_vowel.ordI	-0.494	0.313	-1.580	0.114	-1.107	0.119
Voicing.ordvoiceless:base_vowel.ordU	-0.701	0.403	-1.738	0.082	-1.491	0.089

GERMAN - SMOOTH TERMS

Term	edf	Ref.df	*F*	*p*
s(norm_time)	4.481	5.116	6.212	0.000
s(norm_time):Voicing.ordvoiceless	1.426	1.591	1.920	0.095
s(norm_time,speaker)	232.271	691.000	40.221	0.000
s(norm_time,speaker):Voicing.ordvoiceless	34.827	636.000	0.320	0.000
s(norm_time,consonant)	49.904	121.000	25.359	0.008
s(norm_time):base_vowel.ordI	4.965	5.960	15.501	0.000
s(norm_time):base_vowel.ordU	3.317	4.207	3.373	0.008
s(norm_time,speaker):base_vowel.ordI	227.423	692.000	1.386	0.000
s(norm_time,speaker):base_vowel.ordU	41.814	527.000	0.610	0.000
s(word)	256.292	381.000	7.517	0.000

Hausa - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.307	0.280	33.288	0.000	8.759	9.855
Voicing.ordvoiceless	0.331	0.306	1.082	0.279	-0.268	0.931
base_vowel.ordI	1.169	0.330	3.540	0.000	0.522	1.816
base_vowel.ordU	1.578	0.332	4.760	0.000	0.928	2.228
tone.ordL	0.219	0.075	2.936	0.003	0.073	0.366
tone.ordT1	0.478	0.194	2.464	0.014	0.098	0.859
tone.ordT2	-0.031	0.098	-0.316	0.752	-0.223	0.161
gender_male	-6.963	0.398	-17.511	0.000	-7.742	-6.183
log_utterance_num_syllables.norm	0.485	0.038	12.725	0.000	0.410	0.559
Voicing.ordvoiceless:base_vowel.ordI	0.411	0.398	1.033	0.301	-0.369	1.192
Voicing.ordvoiceless:base_vowel.ordU	-0.030	0.363	-0.083	0.934	-0.742	0.682
Voicing.ordvoiceless:tone.ordL	0.526	0.222	2.368	0.018	0.091	0.961
Voicing.ordvoiceless:tone.ordT1	0.664	0.260	2.556	0.011	0.155	1.173
Voicing.ordvoiceless:tone.ordT2	0.083	0.265	0.314	0.753	-0.436	0.602
base_vowel.ordI:tone.ordL	-0.411	0.309	-1.332	0.183	-1.016	0.194
base_vowel.ordU:tone.ordL	-0.676	0.279	-2.425	0.015	-1.222	-0.130
base_vowel.ordI:tone.ordT1	-1.255	0.324	-3.869	0.000	-1.891	-0.619
base_vowel.ordU:tone.ordT1	-0.809	0.270	-3.003	0.003	-1.338	-0.281
base_vowel.ordI:tone.ordT2	-1.604	0.322	-4.977	0.000	-2.236	-0.972
base_vowel.ordU:tone.ordT2	-2.113	0.696	-3.036	0.002	-3.478	-0.749

Hausa - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	1.000	1.000	68.632	0.000
s(norm_time):Voicing.ordvoiceless	6.480	7.088	22.196	0.000
s(norm_time):tone.ordL	2.981	3.835	2.003	0.073
s(norm_time):tone.ordT1	1.001	1.002	2.297	0.129
s(norm_time):tone.ordT2	1.001	1.002	0.030	0.864
s(norm_time,speaker)	354.653	925.000	20.779	0.000
s(norm_time,speaker):Voicing.ordvoiceless	190.665	914.000	2.105	0.000
s(norm_time,consonant)	52.481	97.000	10.728	0.017
s(norm_time):base_vowel.ordI	4.316	5.174	9.444	0.000
s(norm_time):base_vowel.ordU	2.884	3.429	3.647	0.010
s(norm_time,speaker):base_vowel.ordI	258.519	817.000	0.935	0.000
s(norm_time,speaker):base_vowel.ordU	360.389	867.000	1.843	0.000
s(word)	449.759	536.000	10.183	0.000

Korean - Parametric Terms

term	estimate	std.error	statistic	p.value	95% CI	
(Intercept)	10.835	0.600	18.064	0.000	9.660	12.011
Voicing.ordaspirated	0.706	0.742	0.952	0.341	-0.748	2.159
base_vowel.ordI	0.661	0.784	0.844	0.399	-0.874	2.197
base_vowel.ordU	1.623	0.757	2.145	0.032	0.140	3.107
gender_male	-10.643	0.428	-24.866	0.000	-11.482	-9.804
log_utterance_num_syllables.norm	0.385	0.045	8.572	0.000	0.297	0.473
Voicing.ordaspirated:base_vowel.ordI	0.243	0.785	0.310	0.757	-1.295	1.781
Voicing.ordaspirated:base_vowel.ordU	-0.937	0.760	-1.233	0.218	-2.426	0.553

Korean - Smooth Terms

term	edf	ref.df	statistic	p.value
s(norm_time)	2.355	2.713	1.487	0.337
s(norm_time):Voicing.ordaspirated	1.000	1.000	11.796	0.001
s(norm_time,speaker)	222.831	878.000	31.290	0.000
s(norm_time,speaker):Voicing.ordaspirated	180.081	872.000	3.180	0.000
s(norm_time,consonant)	31.896	97.000	3320.025	0.000
s(norm_time):base_vowel.ordI	2.573	3.051	7.220	0.000
s(norm_time):base_vowel.ordU	2.220	2.740	15.273	0.000
s(norm_time,speaker):base_vowel.ordI	235.214	666.000	21.642	0.000
s(norm_time,speaker):base_vowel.ordU	108.935	493.000	31.180	0.000
s(word)	1001.295	1166.000	11.287	0.000

Mandarin - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	14.128	0.313	45.180	0.000	13.515	14.741
Voicing.ordvoiceless	0.399	0.267	1.495	0.135	-0.124	0.923
base_vowel.ordI	0.797	0.264	3.023	0.003	0.280	1.313
base_vowel.ordU	0.455	0.313	1.454	0.146	-0.158	1.068
tone.ord2	-3.403	0.343	-9.931	0.000	-4.075	-2.731
tone.ord3	-5.219	0.260	-20.076	0.000	-5.728	-4.709
tone.ord4	0.148	0.223	0.664	0.507	-0.289	0.585
gender_male	-8.454	0.458	-18.454	0.000	-9.352	-7.556
log_utterance_num_syllables.norm	0.987	0.025	38.759	0.000	0.937	1.036
Voicing.ordvoiceless:base_vowel.ordI	-0.226	0.314	-0.718	0.473	-0.842	0.390
Voicing.ordvoiceless:base_vowel.ordU	0.045	0.254	0.176	0.861	-0.452	0.542
Voicing.ordvoiceless:tone.ord2	-0.470	0.305	-1.542	0.123	-1.067	0.127
Voicing.ordvoiceless:tone.ord3	0.196	0.286	0.685	0.494	-0.365	0.756
Voicing.ordvoiceless:tone.ord4	-0.180	0.245	-0.735	0.462	-0.659	0.300
base_vowel.ordI:tone.ord2	0.244	0.328	0.742	0.458	-0.400	0.887
base_vowel.ordU:tone.ord2	0.613	0.404	1.517	0.129	-0.179	1.405
base_vowel.ordI:tone.ord3	0.387	0.330	1.173	0.241	-0.260	1.035
base_vowel.ordU:tone.ord3	-0.551	0.364	-1.514	0.130	-1.264	0.162
base_vowel.ordI:tone.ord4	-0.705	0.273	-2.584	0.010	-1.240	-0.170
base_vowel.ordU:tone.ord4	-0.738	0.331	-2.230	0.026	-1.386	-0.089

Mandarin - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	6.553	7.113	9.821	0.000
s(norm_time):Voicing.ordvoiceless	3.138	3.558	6.158	0.000
s(norm_time):tone.ord2	6.185	7.399	64.020	0.000
s(norm_time):tone.ord3	6.033	7.265	60.873	0.000
s(norm_time):tone.ord4	6.358	7.500	307.913	0.000
s(norm_time,speaker)	455.582	1186.000	47.808	0.000
s(norm_time,speaker):Voicing.ordvoiceless	253.117	1187.000	1.435	0.000
s(norm_time,consonant)	79.488	142.000	49.825	0.000
s(norm_time):base_vowel.ordI	2.121	2.604	0.725	0.422
s(norm_time):base_vowel.ordU	3.816	4.627	4.197	0.002
s(norm_time,speaker):base_vowel.ordI	352.731	1187.000	1.604	0.000
s(norm_time,speaker):base_vowel.ordU	453.182	1169.000	1.930	0.000
s(word)	1103.873	1374.000	7.186	0.000

Polish - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.759	0.533	18.321	0.000	8.715	10.803
Voicing.ordvoiceless	0.685	0.189	3.621	0.000	0.314	1.055
base_vowel.ordI	0.649	0.199	3.258	0.001	0.259	1.040
base_vowel.ordU	0.622	0.327	1.904	0.057	-0.018	1.263
log_utterance_num_syllables.norm	0.531	0.055	9.746	0.000	0.425	0.638
Voicing.ordvoiceless:base_vowel.ordI	0.491	0.291	1.686	0.092	-0.080	1.061
Voicing.ordvoiceless:base_vowel.ordU	0.724	0.363	1.996	0.046	0.013	1.435

Polish - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	1.000	1.000	23.446	0.000
s(norm_time):Voicing.ordvoiceless	6.457	7.160	28.597	0.000
s(norm_time,speaker)	266.843	890.000	452.912	0.000
s(norm_time,speaker):Voicing.ordvoiceless	245.051	843.000	5.225	0.000
s(norm_time,consonant)	64.960	160.000	3.075	0.000
s(norm_time):base_vowel.ordI	3.221	3.791	3.220	0.017
s(norm_time):base_vowel.ordU	2.850	3.494	2.604	0.043
s(norm_time,speaker):base_vowel.ordI	352.477	771.000	4.127	0.000
s(norm_time,speaker):base_vowel.ordU	191.774	668.000	11.164	0.000
s(word)	793.283	962.000	6.848	0.000

Portuguese - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.534	0.431	22.097	0.000	8.688	10.380
Voicing.ordvoiceless	0.031	0.407	0.077	0.939	-0.767	0.829
base_vowel.ordI	0.185	0.511	0.362	0.718	-0.817	1.187
base_vowel.ordU	-0.161	0.551	-0.293	0.769	-1.241	0.918
gender_male	-8.829	0.589	-14.985	0.000	-9.984	-7.674
log_utterance_num_syllables.norm	0.408	0.115	3.551	0.000	0.183	0.633
Voicing.ordvoiceless:base_vowel.ordI	1.239	0.598	2.072	0.038	0.067	2.411
Voicing.ordvoiceless:base_vowel.ordU	1.491	0.611	2.438	0.015	0.292	2.689

Portuguese - Smooth Terms

Term	edf	Ref.df	F	p
s(norm_time)	1.000	1.001	42.171	0.000
s(norm_time):Voicing.ordvoiceless	4.203	5.026	21.030	0.000
s(norm_time,speaker)	131.236	721.000	6.949	0.000
s(norm_time,speaker):Voicing.ordvoiceless	84.953	699.000	0.630	0.000
s(norm_time,consonant)	14.098	97.000	0.417	0.133
s(norm_time):base_vowel.ordI	4.262	5.371	5.042	0.000
s(norm_time):base_vowel.ordU	2.018	2.441	5.912	0.001
s(norm_time,speaker):base_vowel.ordI	27.617	475.000	0.789	0.000
s(norm_time,speaker):base_vowel.ordU	170.788	637.000	1.520	0.000
s(word)	238.525	322.000	5.044	0.000

Russian - Parametric Terms

Term	Estimate	SE	t	p	95% CI	
(Intercept)	8.794	0.354	24.853	0.000	8.101	9.488
Voicing.ordvoiceless	0.539	0.370	1.459	0.145	-0.185	1.264
base_vowel.ordI	0.457	0.342	1.337	0.181	-0.213	1.128
base_vowel.ordU	0.992	0.353	2.807	0.005	0.299	1.684
gender_male	-9.822	0.460	-21.332	0.000	-10.725	-8.920
log_utterance_num_syllables.norm	0.443	0.056	7.901	0.000	0.333	0.553
Voicing.ordvoiceless:base_vowel.ordI	-0.134	0.409	-0.328	0.743	-0.936	0.668
Voicing.ordvoiceless:base_vowel.ordU	-0.236	0.426	-0.555	0.579	-1.071	0.598

Russian - Smooth Terms

Term	edf	Ref.df	F	p
s(norm_time)	4.107	4.682	15.541	0.000
s(norm_time):Voicing.ordvoiceless	6.615	7.293	19.239	0.000
s(norm_time,speaker)	343.989	1033.000	26.620	0.000
s(norm_time,speaker):Voicing.ordvoiceless	354.732	1031.000	2.173	0.000
s(norm_time,consonant)	49.679	151.000	12.454	0.010
s(norm_time):base_vowel.ordI	3.692	4.669	6.090	0.000
s(norm_time):base_vowel.ordU	2.723	3.418	2.310	0.068
s(norm_time,speaker):base_vowel.ordI	55.454	721.000	1.124	0.000
s(norm_time,speaker):base_vowel.ordU	162.583	883.000	1.193	0.000
s(word)	583.394	762.000	5.493	0.000

Spanish - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	8.103	0.317	25.547	0.000	7.481	8.724
Voicing.ordvoiceless	0.211	0.292	0.721	0.471	-0.362	0.784
base_vowel.ordI	0.265	0.306	0.867	0.386	-0.335	0.865
base_vowel.ordU	1.028	0.500	2.056	0.040	0.048	2.007
gender_male	-9.384	0.441	-21.299	0.000	-10.248	-8.521
log_utterance_num_syllables.norm	0.334	0.063	5.330	0.000	0.211	0.457
Voicing.ordvoiceless:base_vowel.ordI	0.948	0.376	2.519	0.012	0.210	1.685
Voicing.ordvoiceless:base_vowel.ordU	-0.183	0.528	-0.347	0.729	-1.218	0.852

Spanish - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	4.910	5.676	1.851	0.133
s(norm_time):Voicing.ordvoiceless	3.801	4.450	8.092	0.000
s(norm_time,speaker)	190.275	898.000	14.302	0.000
s(norm_time,speaker):Voicing.ordvoiceless	140.828	899.000	1.957	0.000
s(norm_time,consonant)	23.812	97.000	1.214	0.172
s(norm_time):base_vowel.ordI	3.816	4.713	2.600	0.021
s(norm_time):base_vowel.ordU	2.047	2.530	2.384	0.061
s(norm_time,speaker):base_vowel.ordI	135.625	876.000	1.246	0.000
s(norm_time,speaker):base_vowel.ordU	201.668	802.000	2.708	0.000
s(word)	593.012	719.000	6.091	0.000

Swahili - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	8.892	0.403	22.072	0.000	8.103	9.682
Voicing.ordvoiceless	0.285	0.289	0.987	0.323	-0.281	0.852
base_vowel.ordI	0.123	0.302	0.408	0.684	-0.468	0.714
base_vowel.ordU	0.844	0.432	1.951	0.051	-0.004	1.691
gender_male	-9.642	0.672	-14.338	0.000	-10.960	-8.324
log_utterance_num_syllables.norm	0.177	0.030	5.812	0.000	0.117	0.236
Voicing.ordvoiceless:base_vowel.ordI	0.819	0.338	2.425	0.015	0.157	1.481
Voicing.ordvoiceless:base_vowel.ordU	-0.570	0.452	-1.260	0.208	-1.457	0.317

Swahili - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	2.616	3.027	26.415	0.000
s(norm_time):Voicing.ordvoiceless	4.866	5.479	21.257	0.000
s(norm_time,speaker)	262.119	618.000	59.022	0.000
s(norm_time,speaker):Voicing.ordvoiceless	164.759	616.000	5.535	0.000
s(norm_time,consonant)	34.690	97.000	7.143	0.025
s(norm_time):base_vowel.ordI	4.378	5.354	8.839	0.000
s(norm_time):base_vowel.ordU	2.978	3.548	6.730	0.000
s(norm_time,speaker):base_vowel.ordI	120.795	575.000	6.329	0.000
s(norm_time,speaker):base_vowel.ordU	253.893	593.000	7.238	0.000
s(word)	523.964	593.000	18.911	0.000

Swedish - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	8.901	0.316	28.179	0.000	8.282	9.520
Voicing.ordvoiceless	0.025	0.311	0.080	0.936	-0.585	0.634
base_vowel.ordI	0.182	0.322	0.565	0.572	-0.449	0.812
base_vowel.ordU	1.964	1.411	1.393	0.164	-0.800	4.729
gender_male	-9.924	0.471	-21.085	0.000	-10.846	-9.001
log_utterance_num_syllables.norm	0.662	0.074	8.925	0.000	0.517	0.807
Voicing.ordvoiceless:base_vowel.ordI	0.386	0.385	1.003	0.316	-0.368	1.141
Voicing.ordvoiceless:base_vowel.ordU	-0.918	1.843	-0.498	0.618	-4.529	2.694

Swedish - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	5.139	6.147	11.331	0.000
s(norm_time):Voicing.ordvoiceless	5.311	6.488	57.152	0.000
s(norm_time,speaker)	229.682	880.000	10.930	0.000
s(norm_time,speaker):Voicing.ordvoiceless	80.704	877.000	2.425	0.000
s(norm_time,consonant)	0.647	97.000	0.007	0.271
s(norm_time):base_vowel.ordI	2.054	2.580	5.151	0.003
s(norm_time):base_vowel.ordU	1.000	1.000	7.960	0.005
s(norm_time,speaker):base_vowel.ordI	201.365	878.000	2.216	0.000
s(norm_time,speaker):base_vowel.ordU	5.170	92.000	1.964	0.021
s(word)	409.477	508.000	5.234	0.000

Thai - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.412	0.354	26.550	0.000	8.717	10.106
Voicing.ordvoiceless	0.380	0.347	1.094	0.274	-0.301	1.061
base_vowel.ordI	0.518	0.667	0.776	0.438	-0.790	1.826
base_vowel.ordU	0.342	0.537	0.638	0.524	-0.710	1.394
tone_syll1.ord1	-0.012	0.613	-0.019	0.985	-1.213	1.190
tone_syll1.ord2	2.352	0.482	4.885	0.000	1.408	3.296
tone_syll1.ord3	-0.904	1.535	-0.589	0.556	-3.913	2.104
gender_male	-10.080	0.384	-26.267	0.000	-10.832	-9.328
log_utterance_num_syllables.norm	0.538	0.029	18.800	0.000	0.482	0.594
Voicing.ordvoiceless:base_vowel.ordI	-0.466	0.833	-0.559	0.576	-2.098	1.167
Voicing.ordvoiceless:base_vowel.ordU	0.992	0.606	1.636	0.102	-0.196	2.181
Voicing.ordvoiceless:tone_syll1.ord1	-0.449	0.644	-0.697	0.486	-1.711	0.813
Voicing.ordvoiceless:tone_syll1.ord2	-0.068	0.546	-0.125	0.900	-1.138	1.001
Voicing.ordvoiceless:tone_syll1.ord3	1.514	1.517	0.998	0.318	-1.459	4.486
base_vowel.ordI:tone_syll1.ord1	0.259	1.017	0.255	0.799	-1.734	2.253
base_vowel.ordU:tone_syll1.ord1	-1.651	0.644	-2.564	0.010	-2.913	-0.389
base_vowel.ordI:tone_syll1.ord2	0.602	0.622	0.969	0.333	-0.616	1.821
base_vowel.ordU:tone_syll1.ord2	-0.478	0.526	-0.908	0.364	-1.510	0.554
base_vowel.ordI:tone_syll1.ord3	0.448	0.629	0.713	0.476	-0.784	1.681
base_vowel.ordU:tone_syll1.ord3	-0.438	0.600	-0.730	0.465	-1.613	0.737

Thai - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	4.420	5.006	9.862	0.000
s(norm_time):Voicing.ordvoiceless	2.744	3.099	15.905	0.000
s(norm_time):tone_syll1.ord1	1.403	1.696	23.347	0.000
s(norm_time):tone_syll1.ord2	5.053	6.157	89.953	0.000
s(norm_time):tone_syll1.ord3	3.785	4.822	5.141	0.000
s(norm_time,speaker)	151.478	876.000	6.191	0.000
s(norm_time,speaker):Voicing.ordvoiceless	278.688	868.000	1.183	0.000
s(norm_time,consonant)	23.750	61.000	2.932	0.000
s(norm_time):base_vowel.ordI	5.064	6.237	8.974	0.000
s(norm_time):base_vowel.ordU	7.266	8.130	2.538	0.024
s(norm_time,speaker):base_vowel.ordI	81.175	807.000	0.644	0.000
s(norm_time,speaker):base_vowel.ordU	245.347	809.000	1.414	0.000
s(word)	198.867	243.000	9.866	0.000

Turkish - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.416	0.306	30.780	0.000	8.816	10.016
Voicing.ordvoiceless	0.386	0.216	1.785	0.074	-0.038	0.809
base_vowel.ordI	0.710	0.158	4.505	0.000	0.401	1.019
base_vowel.ordU	0.383	0.207	1.855	0.064	-0.022	0.789
gender_male	-9.884	0.540	-18.316	0.000	-10.942	-8.827
log_utterance_num_syllables.norm	0.373	0.042	8.953	0.000	0.291	0.454
Voicing.ordvoiceless:base_vowel.ordI	0.627	0.254	2.471	0.013	0.130	1.124
Voicing.ordvoiceless:base_vowel.ordU	0.782	0.279	2.802	0.005	0.235	1.329

Turkish - Smooth Terms

Term	edf	Ref.df	*F*	*p*
s(norm_time)	3.265	3.845	11.224	0.000
s(norm_time):Voicing.ordvoiceless	6.379	7.200	9.947	0.000
s(norm_time,speaker)	280.583	898.000	25.918	0.000
s(norm_time,speaker):Voicing.ordvoiceless	208.898	899.000	1.348	0.000
s(norm_time,consonant)	36.165	124.000	4.282	0.005
s(norm_time):base_vowel.ordI	4.547	5.560	8.892	0.000
s(norm_time):base_vowel.ordU	4.318	5.194	7.854	0.000
s(norm_time,speaker):base_vowel.ordI	140.452	883.000	0.625	0.000
s(norm_time,speaker):base_vowel.ordU	269.085	895.000	1.363	0.000
s(word)	810.363	1153.000	4.775	0.000

Ukrainian - Parametric Terms

Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	9.875	1.002	9.851	0.000	7.910	11.839
Voicing.ordvoiceless	0.785	1.171	0.671	0.503	-1.510	3.080
base_vowel.ordI	0.000	0.000	NaN	NaN	0.000	0.000
base_vowel.ordU	1.019	0.564	1.808	0.071	-0.086	2.124
gender_male	-8.695	0.454	-19.155	0.000	-9.585	-7.805
log_utterance_num_syllables.norm	-0.003	0.065	-0.040	0.968	-0.130	0.125
Voicing.ordvoiceless:base_vowel.ordI	1.759	0.658	2.673	0.008	0.469	3.049
Voicing.ordvoiceless:base_vowel.ordU	0.708	0.673	1.053	0.292	-0.610	2.026

UKRAINIAN - SMOOTH TERMS

Term	edf	Ref.df	*F*	*p*
s(norm_time)	1.001	1.001	34.388	0.000
s(norm_time):Voicing.ordvoiceless	5.582	6.266	11.507	0.000
s(norm_time,speaker)	224.715	1069.000	8.240	0.000
s(norm_time,speaker):Voicing.ordvoiceless	372.486	1057.000	0.936	0.000
s(norm_time,consonant)	58.613	122.000	14.398	0.009
s(norm_time):base_vowel.ordI	4.375	5.150	2.451	0.027
s(norm_time):base_vowel.ordU	2.051	2.335	0.609	0.511
s(norm_time,speaker):base_vowel.ordI	302.918	730.000	1.399	0.000
s(norm_time,speaker):base_vowel.ordU	527.290	913.000	2.457	0.000
s(word)	145.241	207.000	5.972	0.000

VIETNAMESE - PARAMETRIC TERMS

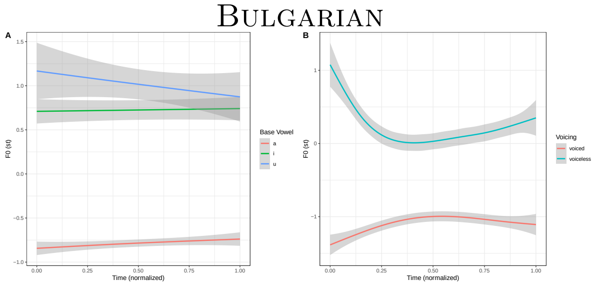
Term	Estimate	*SE*	*t*	*p*	95% CI	
(Intercept)	11.034	0.529	20.860	0.000	9.997	12.070
Voicing.ordvoiceless	0.967	0.542	1.784	0.074	-0.095	2.030
base_vowel.ordI	1.379	0.731	1.886	0.059	-0.054	2.812
base_vowel.ordU	1.066	0.742	1.436	0.151	-0.389	2.521
tone.ordT2	-2.381	0.657	-3.626	0.000	-3.668	-1.094
tone.ordT3	-0.007	0.553	-0.013	0.990	-1.091	1.077
tone.ordT4	-2.619	0.713	-3.670	0.000	-4.017	-1.220
tone.ordT5	-4.004	1.092	-3.667	0.000	-6.144	-1.864
tone.ordT6	-3.306	0.707	-4.675	0.000	-4.691	-1.920
gender_male	-8.809	0.554	-15.911	0.000	-9.894	-7.724
log_utterance_num_syllables.norm	0.665	0.038	17.266	0.000	0.589	0.740
Voicing.ordvoiceless:base_vowel.ordI	-0.569	0.614	-0.926	0.354	-1.772	0.635
Voicing.ordvoiceless:base_vowel.ordU	-1.561	0.627	-2.491	0.013	-2.790	-0.333
Voicing.ordvoiceless:tone.ordT2	-0.329	0.764	-0.430	0.667	-1.827	1.169
Voicing.ordvoiceless:tone.ordT3	0.458	0.635	0.721	0.471	-0.787	1.703
Voicing.ordvoiceless:tone.ordT4	-1.722	0.819	-2.104	0.035	-3.326	-0.118
Voicing.ordvoiceless:tone.ordT5	1.541	1.481	1.041	0.298	-1.361	4.443
Voicing.ordvoiceless:tone.ordT6	0.909	0.816	1.113	0.266	-0.691	2.509
base_vowel.ordI:tone.ordT2	-0.952	0.870	-1.095	0.274	-2.658	0.753
base_vowel.ordU:tone.ordT2	-0.187	0.884	-0.211	0.833	-1.919	1.545
base_vowel.ordI:tone.ordT3	-1.212	0.714	-1.698	0.089	-2.610	0.187
base_vowel.ordU:tone.ordT3	0.212	0.683	0.310	0.756	-1.127	1.551
base_vowel.ordI:tone.ordT4	-0.556	0.873	-0.636	0.524	-2.267	1.155
base_vowel.ordU:tone.ordT4	1.242	0.923	1.345	0.179	-0.568	3.052
base_vowel.ordI:tone.ordT5	0.156	1.539	0.101	0.919	-2.860	3.172
base_vowel.ordU:tone.ordT5	0.272	1.854	0.147	0.883	-3.362	3.906
base_vowel.ordI:tone.ordT6	-0.025	1.005	-0.025	0.980	-1.995	1.945
base_vowel.ordU:tone.ordT6	0.382	0.878	0.435	0.664	-1.339	2.102

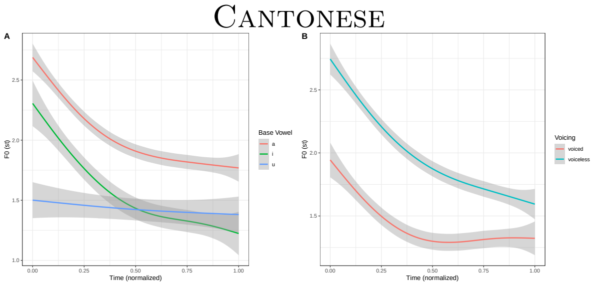
VIETNAMESE - SMOOTH TERMS

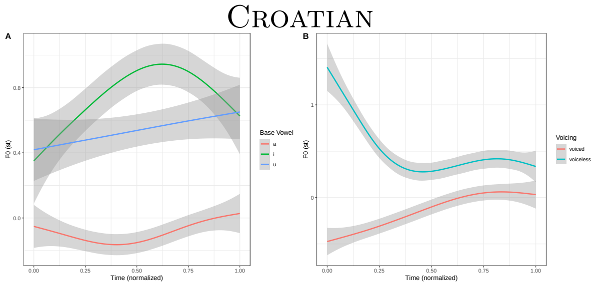
Term	edf	Ref.df	F	p
s(norm_time)	1.006	1.009	10.851	0.001
s(norm_time):Voicing.ordvoiceless	4.528	5.120	8.621	0.000
s(norm_time):tone.ordT2	1.000	1.001	111.430	0.000
s(norm_time):tone.ordT3	5.133	6.226	47.156	0.000
s(norm_time):tone.ordT4	4.923	6.123	25.192	0.000
s(norm_time):tone.ordT5	6.716	7.841	49.293	0.000
s(norm_time):tone.ordT6	1.000	1.001	148.554	0.000
s(norm_time,speaker)	518.585	1159.000	8.187	0.000
s(norm_time,speaker):Voicing.ordvoiceless	480.059	1160.000	3.764	0.000
s(norm_time,consonant)	51.283	115.000	3.034	0.000
s(norm_time):base_vowel.ordI	3.104	3.754	2.575	0.035
s(norm_time):base_vowel.ordU	5.509	6.438	7.192	0.000
s(norm_time,speaker):base_vowel.ordI	446.498	1147.000	2.665	0.000
s(norm_time,speaker):base_vowel.ordU	431.785	1074.000	3.873	0.000
s(word)	275.178	314.000	15.730	0.000

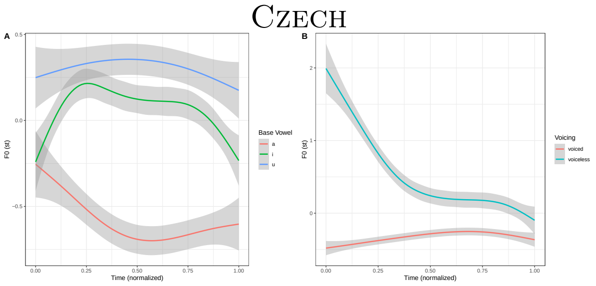
Appendix D. Empirical IF0 effects

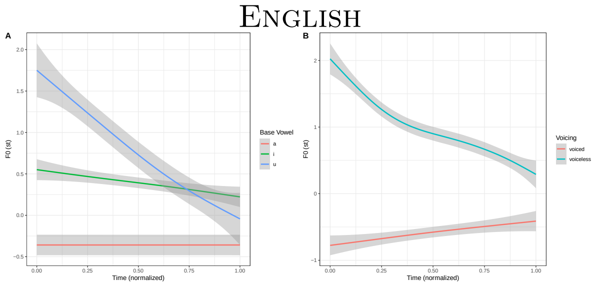
Empirical plots of F0 (semitones) over normalized time for each level of (a) vowel and (b) voicing, for each language. Bands represent 95% confidence intervals (from a GAM smoother).

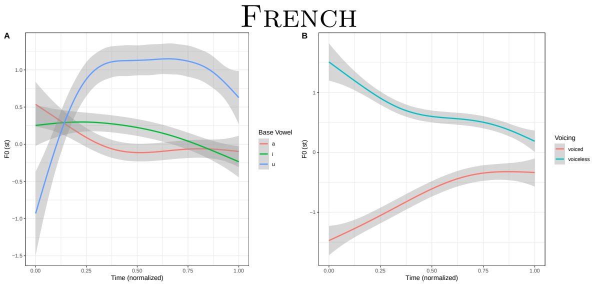

BULGARIAN

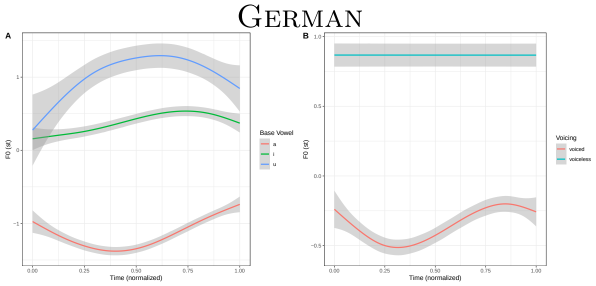

CANTONESE

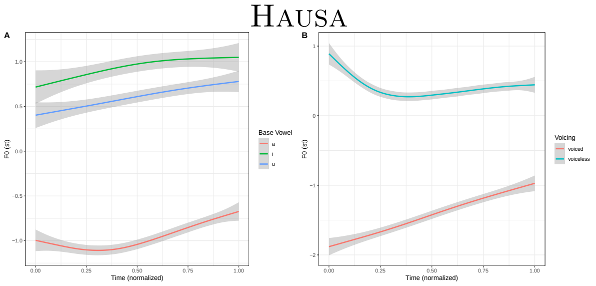

CROATIAN


CZECH

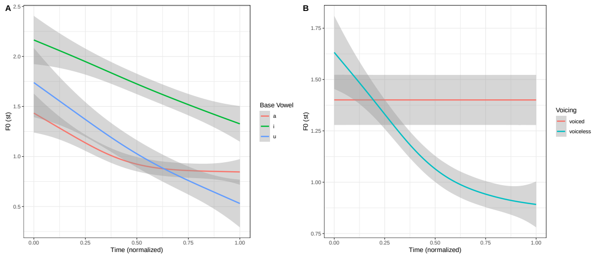

ENGLISH


FRENCH

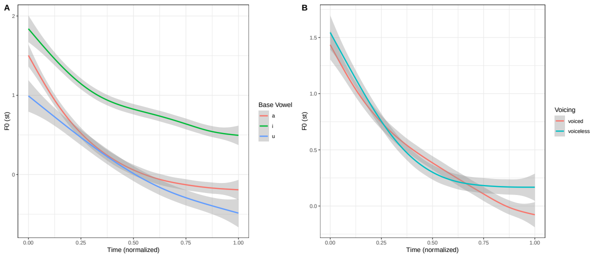

GERMAN


HAUSA

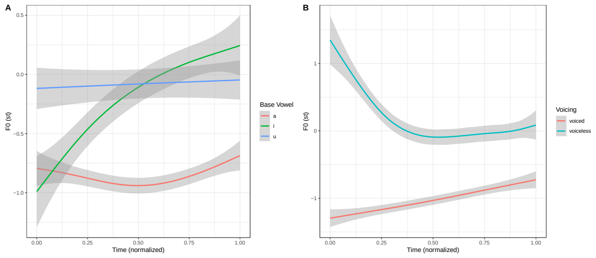
Korean



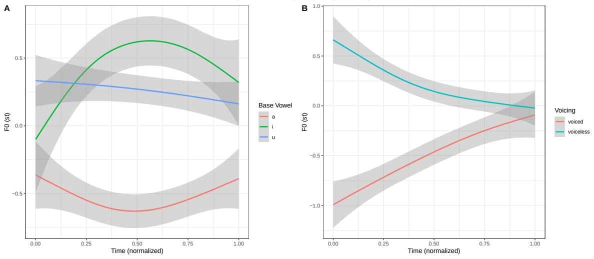
Mandarin



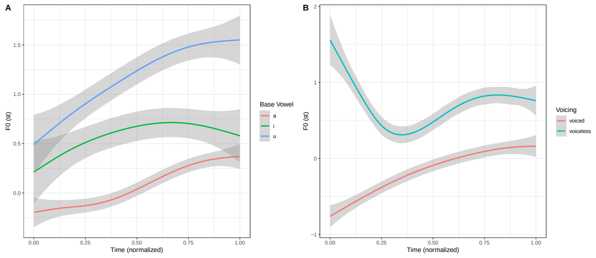
Polish



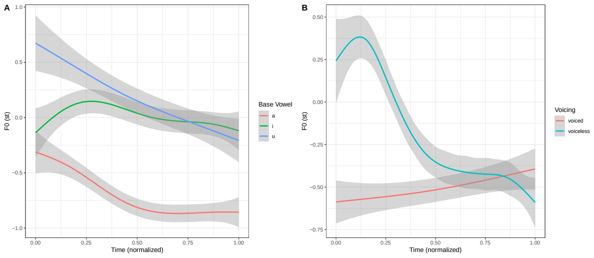
Portuguese



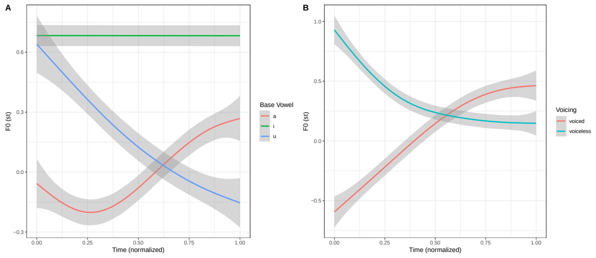
Russian



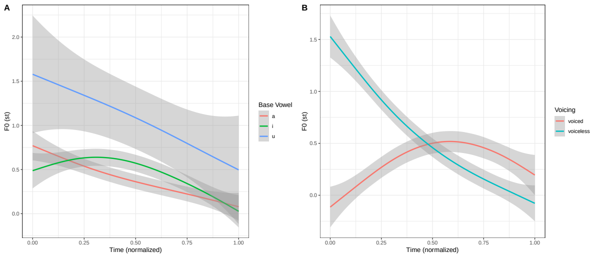
Spanish



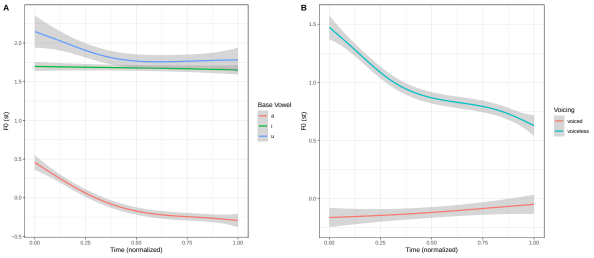
Swahili



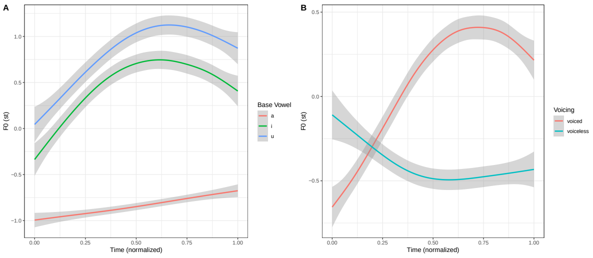
Swedish



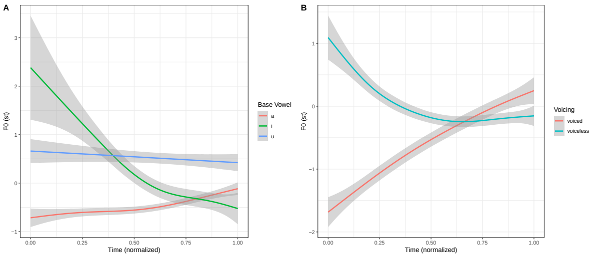
Thai



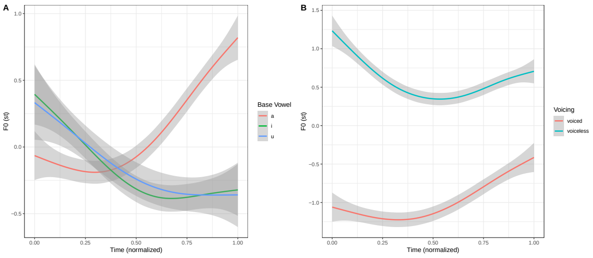
Turkish



Ukrainian



Vietnamese



References

Andrus, Tony, Eyal Dubinski, Jonathan G. Fiscus, Breanna Gillies, Mary Harper, T.J. Hazen and Jamie Wong. 2016. *IARPA Babel Cantonese Language Pack.* Philadelphia: Linguistic Data Consortium.

Baayen, R H., R Piepenbrock and L Gulikers. 1995. *CELEX2.* Philadelphia: Linguistic Data Consortium.

Babinski, Sarah. 2021. Intrinsic F0 and sound change: Evidence from Australian languages. In *Proceedings of the 2020 Annual Meeting on Phonology*, ed. Ryan Bennett, Richard Bibbs, Mykel Loren Brinkerhoff, Max J. Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel and Maya Wax Cavallaro. pp. 1–9.

Bell-Berti, Fredericka. 1975. "Control of pharyngeal cavity size for English voiced and voiceless stops." *Journal of the Acoustical Society of America* 57:456–461.

Bethin, Christina Y. 2006. "Stress and tone in East Slavic dialects." *Phonology* 23(2):125–156.

Caisse, Michelle. 1981. "Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops." *Journal of the Acoustical Society of America* 70:S76–S77.

Chen, Leo and Jerry Norman. 1965. *An introduction to the Foochow dialect.* San Francisco: San Francisco State College.

Chen, Yiya. 2011. "How does phonology guide phonetics in segment–f0 interaction?" *Journal of Phonetics* 39:612–625.

Cho, Taehong and Patricia Keating. 2009. "Effects of initial position versus prominence in English." *Journal of Phonetics* 37(4):466–485.

Clements, George N. and Rajesh Khatiwada. 2007. Phonetic realization of contrastively aspirated affricates in Nepali. In *Proceedings of the 16th International Congress of Phonetic Sciences*, ed. J. Trouvain and W. J. Barry. pp. 629–632.

Coetzee, Andries W., Patrice S. Beddor, Kerby Shedden, Will Styler and Daan Wissing. 2018. "Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis." *Journal of Phonetics* 66:185–216.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences.* Second ed. New York: Routledge.

Connell, Bruce. 2002. "Tone languages and the universality of intrinsic F0: Evidence from Africa." *Journal of Phonetics* 30:101–129.

Cooper, William E. and John M. Sorensen. 2012. *Fundamental frequency in sentence production.* Springer-Verlag: New York.

Coretta, Stefano. 2023. *tidymv: Tidy model visualisation for generalised additive models.* R package version 3.4.2.
URL: *https://CRAN.R-project.org/package=tidymv*

Di Cristo, Albert. 1985. *De la microprosodie à l'intonosyntaxe.* France: Publications de l'Université d'Aix.

Di Cristo, Albert and Daniel J Hirst. 1986. "Modelling French micromelody: Analysis and synthesis." *Phonetica* 43(1-3):11–30.

Diehl, Randy L. 1991. "The role of phonetics within the study of language." *Phonetica* 48(2-4):120–134.

Diffloth, Gérard. 1982. "Registres, dévoisement, timbres vocaliques: Leur histoire en Katouique." *Mon-Khmer Studies* 11:47–82.

Dmitrieva, Olga, Fernando Llanos, Amanda A. Shultz and Alexander L. Francis. 2015. "Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English." *Journal of Phonetics* 49:77–95.

Dyhr, Niels. 1982. "The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels." *Phonetica* 47:141–154.

Fischer-Jørgensen, Eli. 1990. "Intrinsic F0 in tense and lax vowels with special reference to German." *Phonetica* 47:99–140.

Francis, Alexander L., Valter Ciocca, Virginia K. M. Wong and Jess K. K. Chan. 2006. "Is fundamental frequency a cue to aspiration in initial stops?" *Journal of the Acoustical Society of America* 120:2884–2895.

Gandour, J. 1974. "Consonant types and tone in Siamese." *Journal of Phonetics* 2:337–350.

Gao, Jiayin and Takayuki Arai. 2019. "Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery." *Journal of Phonetics* 77:10932.

Garrett, Andrew and Keith Johnson. 2013. Phonetic bias in sound change. In *Origins of Sound Change: Approaches to Phonologization*, ed. Alan C. L. Yu. Vol. 1 Oxford: Oxford University Press pp. 51–97.

Geissler, Christopher, Jason Shaw, Fang Hu and Mark Tiede. 2021. Eccentric C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts. In *Proceedings of the 12th International Seminar on Speech Production*, ed. Mark Tiede, Douglas. H. Whalen and Vincent Gracco. pp. 1–4.

Gonzales, Ashleigh. 2009. Intrinsic F0 in Shona Vowels: A Descriptive Study. In *Selected Proceedings of the 39th Annual Conference on African Linguistics*. pp. 145–155.

Greenberg, Joseph H. 2005. On lumping and splitting in linguistics and biology. In *Genetic Linguistics: Essays on Theory and Method*, ed. William Croft. Oxford: Oxford University Press pp. 115–117.

Guo, Yuting. 2020. Production and perception of laryngeal contrasts in Mandarin and English by Mandarin speakers. PhD thesis George Mason University, Fairfax, VA.

Halle, Morris and Kenneth N. Stevens. 1971. "A note on laryngeal features." *Quarterly Progress Report, MIT Research lab. of Electronics* 101:198–213.

Han, Mieko S. and Raymond S. Weitzman. 1965. *Studies in the phonology of Asian languages III: Acoustic characteristics of Korean stop consonants.* Acoustic Phonetics Research Laboratory, University of Southern California.

Hanson, Helen M. 2009. "Effects of obstruent consonants on fundamental frequency at vowel onset in English." *Journal of the Acoustical Society of America* 125:425–441.

Helgason, Pétur and Catherine Ringen. 2008. "Voicing and aspiration in Swedish stops." *Journal of Phonetics* 36(4):607–628.

Hombert, Jean-Marie. 1977. "Consonant types, vowel height and tone in Yoruba." *Studies in African Linguistics* 8:173–190.

Hombert, Jean-Marie. 1978. Consonant types, vowel quality, and tone. In *Tone: A Linguistic Survey*, ed. Victoria. A. Fromkin. New York: Academic Press pp. 77–111.

Hombert, Jean-Marie, John J. Ohala and William G. Ewan. 1979. "Phonetic explanations for the development of tones." *Language* 55:37–58.

Honda, Kiyoshi, Hiroyuki Hirai, Shinobu Masaki and Yasuhiro Shimada. 1999. "Role of vertical larynx movement and cervical lordosis in f0 control." *Language and Speech* 42:401–411.

Hoole, Phil and Kiyoshi Honda. 2011. Automaticity vs. feature-enhancement in the control of segmental F0. In *Where do Phonological Features Come From?: Cognitive, Physical and Developmental Bases of Distinctive Speech Categories*, ed. NickClements and Rachid Ridouane. Amsterdam and Philadelphia: John Benjamins pp. 133–171.

House, Arthur S. and Grant Fairbanks. 1953. "The influence of consonant environment upon the secondary acoustical characteristics of vowels." *Journal of the Acoustical Society of America* 25:105–113.

Howie, John M. 1976. *Acoustical studies of Mandarin vowels and tones.* Cambridge: Cambridge University Press.

Hunnicutt, Leigh and Paul A. Morris. 2009. Prevoicing and Aspiration in Southern American English. In *Proceedings of the 39th Annual Penn Linguistics Conference*. pp. 215–224.

Hyman, Larry. 1976. Phonologization. In *Linguistic Studies Presented to Joseph H. Greenberg*, ed. Alphonse Juilland. Saratoga: Anma Libri pp. 407–418.

Hyman, Larry. 2013. Enlarging the scope of phonologization. In *Origins of Sound Change: Approaches to Phonologization*, ed. Alan C. L. Yu. Oxford: Oxford University Press pp. 3–28.

Jacewicz, Ewa and Allen F. Fox. 2015. "Intrinsic fundamental frequency of vowels is moderated by regional dialect." *Journal of the Acoustical Society of America* 138:405–410.

Jeel, Vivi. 1975. "An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants." *Annual Report of the Institute of Phonetics* 9:191–211.

Jessen, Michael and Catherine Ringen. 2002. "Laryngeal features in German." *Phonology* 19:189–218.

Jessen, Michael and Justus C. Roux. 2002. "Voice quality differences associated with stops and clicks in Xhosa." *Journal of Phonetics* 30:1–52.

Kagaya, Ryohei. 1974. "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives." *Journal of Phonetics* 2:161–180.

Kagaya, Ryohei and Hajime Hirose. 1975. "Fiberoptic, electromyographic, and acoustic analyses of Hindi stop consonants." *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 9:27–46.

Kang, Yoonjung. 2014. "Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study." *Journal of Phonetics* 45:76–90.

Kanwal, Jasmeen and Amanda Ritchart. 2015. An experimental investigation of tonogenesis in Punjabi. In *Proceedings of the 18th International Congress of Phonetic Sciences*, ed. The Scottish Consortium for ICPhS 2015.

Keating, Patricia A. 1984. "Phonetic and phonological representation of stop consonant voicing." *Language* 60:286–319.

Keating, Patricia A. 1990. "Phonetic representations in a generative grammar." *Journal of Phonetics* 18:321–334.

Kelley, Ken and Kristopher J. Preacher. 2012. "On effect size." *Psychological Methods* 17(2):137–152.

Kenstowicz, Michael and Chiyoun Park. 2006. "Laryngeal features and tone in Kyungsang Korean: A phonetic study." *Studies in Phonetics, Phonology and Morphology* 12:247–264.

Kingston, John. 2007. Segmental influences on f0: Automatic or controlled? In *Tones and Tunes Vol 2: Experimental Studies in Word and Sentence Prosody*, ed. Carlos Gussenhoven and Tomas Riad. Berlin: Mouton de Gruyte pp. 171–210.

Kingston, John. 2011. Tonogenesis. In *The Blackwell Companion to Phonology*, ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice. Malden: Wiley Online Library pp. 1–30.

Kingston, John and Randy L. Diehl. 1994. "Phonetic knowledge." *Language* 70:419–454.

Kirby, James and D. Robert Ladd. 2016. "Effects of obstruent voicing on vowel f0: Evidence from "true voicing" languages." *Journal of the Acoustical Society of America* 140:2400–2411.

Kirby, James, Felicitas Kleber, Jessica Siddins and Jonathan Harrington. 2020. Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German. In *Proceedings of the 10th Annual International Conference on Speech Prosody*, ed. Nobuaki Minematsu. pp. 210–214.

Kirby, James and Morgan Sonderegger. 2018. "Mixed-effects design analysis for experimental phonetics." *Journal of Phonetics* 70:70–85.

Kirby, James P. 2018. "Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages." *Journal of Phonetics* 71:326–354.

Kohler, Klaus J. 1982. "F0 in the production of lenis and fortis plosives." *Phonetica* 39:199–218.

Ladd, Robert D. and Kim E.A. Silverman. 1984. "Vowel intrinsic pitch in connected speech." *Phonetica* 41:31–40.

Ladd, Robert D. and Stephan Schmid. 2018. "Obstruent voicing effects on f0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops." *Journal of Phonetics* 71:229–248.

Lai, Yuwen, Christopher Huff, Joan Sereno and Allard Jongman. 2009. The raising effect of aspirated prevocalic consonants on F0 in Taiwanese. In *Proceedings of the 2nd International Conference on East Asian Linguistics*, ed. J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva and S. Morton et al.

Lalhminghlui, Wendy, Viyazonuo Terhiija and Priyankoo Sarmah. 2019. Vowel-tone interaction in two Tibeto-Burman languages. In *Proceedings of the 20th Annual Interspeech*, ed. Gernot Kubin and Zdravko Kacic. pp. 3970–3974.

Lee, Jackson L., Lucas F.E. Ashby, M. Elizabeth Garza, Yeonju Lee-Sikka, Sean Miller, Alan Wong, Arya D. McCarthy and Kyle Gorman. 2020. Massively Multilingual Pronunciation Modeling with WikiPron. In *Proceedings of the 12th Language Resources and Evaluation Conference*, ed. Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk and Stelios Piperidis. European Language Resources Association pp. 4223–4228.

Lehiste, Ilse. 1976. "Influence of fundamental frequency pattern on the perception of duration." *Journal* *of Phonetics* 4:113–117.

Lehiste, Isle and Gordon E. Peterson. 1961. “Some basic considerations in the analysis of intonation.” *Journal of the Acoustical Society of America* 33:419–423.

Lenth, Russell V. 2024. *emmeans: Estimated marginal means, aka least-squares means.* R package version 1.10.2.
URL: *https://CRAN.R-project.org/package=emmeans*

Li, Bin, Yihan Guan and Si Chen. 2020. Carryover Effects on Tones in Hong Kong Cantonese. In *Proceedings of the 10th Annual International Conference on Speech Prosody*, ed. Nobuaki Minematsu. pp. 489–493.

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

Lombardi, Linda. 1999. “Positional faithfulness and voicing assimilation in Optimality Theory.” *Natural Language and Linguistic Theory* 17:267–302.

Luo, Qian. 2018. Consonantal effects on F0 in tonal languages. PhD thesis Michigan State University, East Lansing, MI.

Löfqvist, Anders, Thomas Baer, Nancy S. McGarr and Robin S. Story. 1989. “The cricothyroid muscle in voicing control.” *Journal of the Acoustical Society of America* 85:1314–1321.

Maddieson, Ian. 1997. Phonetic Universals. In *Handbook of Phonetic Sciences*, ed. John Laver and William J. Hardcastle. Oxford: Blackwells.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In *Consonant Types and Tones*, ed. Larry M. Hyman. Vol. 1 Los Angeles: Linguistics Program, University of Southern California pp. 71–95.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner and Morgan Sonderegger. 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Annual Interspeech*, ed. Gernot Kubin and Zdravko Kacic. pp. 498–502.

Meyer, Ernst A. 1896. “Zur tonbewegung des vokals im gesprochenen und gesungenen einzelwort.” *Phonetische Studien (Beiblatt zu der Zeitschrift Die Neuren Sprachen)* 10:1–21.

Misnadin, James Kirby and Bert Remijsen. 2015. Temporal and spectral properties of Madurese stops. In *Proceedings of the 18th International Congress of Phonetic Sciences*, ed. The Scottish Consortium for ICPhS 2015.

Mohamed, Mohamed Abdulla. 2001. *Modern Swahili Grammar.* Nairobi: East African Publishers.

Moreton, Elliott. 2008. “Analytic bias and phonological typology.” *Phonology* 25(1):83–127.

Nitta, Tetsuo. 2001. The accent system in the Kanazawa dialect: The relationship between pitch and sound segments. In *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Japanese Accentology, and Other Topics.* Tokyo University of Foreign Studies pp. 153–185.

Nordisk Sprakteknologi. 2020. “NST Norwegian ASR Database (16 kHz) – Reorganized.” The Language Bank: oai:nb.no:sbr-54.

Odden, David. 2010. “Features impinging on tone.” *Tones and Features: Phonetic and Phonological Perspectives* 107:81–107.

Ohala, John J. 1981. The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior*, ed. Carrie S. Masek, Roberta A. Hendrick and Mary. F. Miller. Chicago: Chicago Linguistic Society pp. 178–203.

Ohala, John J. 1993. The phonetics of sound change. In *Historical Linguistics: Problems and Perspectives*, ed. Charles Jones. London: Longman pp. 235–278.

Ohala, John J. 1994. “Hierarchies of environments for sound variation; plus implications for ‘neutral’ vowels in vowel harmony.” *Acta Linguistica Hafniensia* 27(1):371–382.

Ohde, Ralph N. 1984. “Fundamental frequency as an acoustic correlate of stop consonant voicing.” *Journal of the Acoustical Society of America* 75:224–230.

Panayotov, Vassil, Guoguo Chen, Daniel Povey and Sanjeev Khudanpur. 2015. Librispeech: An asr corpus based on public domain audio books. In *Proceedings of the 40th International Conference on Acoustics, Speech and Signal Processing*, ed. Doug Gray and Douglas Cochran. IEEE pp. 5206–5210.

Petersen, Niels Reinholt. 1978. “Intrinsic fundamental frequency of Danish vowels.” *Journal of Phonetics*

6:177–189.

Pinget, Anne-France and Hugo Quené. 2023. "Effects of obstruent voicing on vowel fundamental frequency in Dutch." *The Journal of the Acoustical Society of America* 154(4):2124–2136.

Polomé, Edgar C. 1967. *Swahili language handbook.* ERIC.

Pompino-Marschall, Bernd, Elena Steriopolo and Marzena Żygis. 2017. "Ukrainian." *Journal of the International Phonetic Association* 47(3):349–357.

R Core Team. 2022. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.
URL: *https://www.R-project.org/*

Sapir, Shimon. 1989. "Theoretical, physiological, and clinical considerations." *Journal of Voice* 18:44–51.

Schertz, Jessamyn and Sarah Khan. 2020. "Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu." *Journal of Phonetics* 81:1–25.

Schuh, Russell G. and Lawan D. Yalwa. 1993. "Hausa." *Journal of the International Phonetic Association* 23:77–82.

Schultz, Tanja, Ngoc Thang Vu and Tim Schlippe. 2013. GlobalPhone: A multilingual text & speech database in 20 languages. In *Proceedings of the 38th International Conference on Acoustics, Speech and Signal Processing.* IEEE pp. 8126–8130.

Sebastián-Gallés, Núria. 2000. *LEXESP: Léxico informatizado del español.* Barcelona: Edicions Universitat Barcelona.

Shadle, Christine H. 1985. "Intrinsic fundamental frequency of vowels in sentence context." *Journal of the Acoustical Society of America* 78:1562–1567.

Sonderegger, Morgan, Michael McAuliffe and Hye-Young Bang. 2017. "Segmental influences on F0: Cross-linguistic and interspeaker variability of phonetic precursors.". Presented at the 4th Workshop on Sound Change.

Sóskuthy, Márton. 2017. "Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction." *arXiv preprint arXiv:1703.05339* .

Sóskuthy, Márton. 2021. "Evaluating generalised additive mixed modelling strategies for dynamic speech analysis." *Journal of Phonetics* 84:101017.

Steele, Shirley A. 1986. "Interaction of vowel f0 and prosody." *Phonetica* 43:92–105.

Stevens, Kenneth N. and Samuel Jay Keyser. 1989. "Primary features and their enhancement in consonants." *Language* pp. 81–106.

Svantesson, Jan-Olaf. 1989. "Tonogenetic mechanisms in northern Mon-Khmer." *Phonetica* 46:60–79.

Svantesson, Jan-Olaf and David House. 2006. "Tone production, tone perception and Kammu tonogenesis." *Phonology* 23:309–333.

t Hart, Johan. 1979. "Explorations in automatic stylization of f0 curves." *IPO Annual Progress Report* 14:61–65.

Taylor, Harold C. 1933. "The fundamental pitch of English vowels." *Journal of Experimental Psychology* 16:565–582.

Thurgood, Graham. 2002. "Vietnamese and tonogenesis: Revising the model and the analysis." *Diachronica* 19:333–363.

Van Hoof, Sarah and Jo Verhoeven. 2011. "Intrinsic vowel f0, the size of vowel inventories and second language acquisition." *Journal of Phonetics* 39:168–177.

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen and Hedderik van Rijn. 2022. "itsadug: Interpreting time teries and autocorrelated data using GAMMs.". R package version 2.4.1.

Wang, William S-Y. and Charles J. Fillmore. 1961. "Intrinsic cues and consonant perception." *Journal of Speech and Hearing Research* 4:130–136.

Whalen, Douglas H. and Andrea G. Levitt. 1995. "The universality of intrinsic f0 of vowels." *Journal of Phonetics* 23:349–366.

Whalen, Douglas H., Andrea G. Levitt, Pai-Ling Hsiao and Iris Smorodinsky. 1995. "Intrinsic f0 of vowels in the babbling of 6-, 9-, and 12-month-old French and English-learning infants." *Journal of the Acoustical Society of America* 97:2533–2539.

Wieling, Martijn. 2018. “Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English.” *Journal of Phonetics* 70:86–116.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R.* Second ed. Chapman and hall/CRC.

Xu, Ching X. and Yi Xu. 2003. “Effects of consonant aspiration on Mandarin tones.” *Journal of the International Phonetic Alphabet* 33:165–181.

Xu, Yi and Anqi Xu. 2021. “Consonantal f0 perturbation in American English involves multiple mechanisms.” *Journal of the Acoustical Society of America* 149:2877–2895.

Zee, Eric. 1980. “Tone and vowel quality.” *Journal of Phonetics* 8:247–258.