Onset pitch perturbations and the cross-linguistic implementation of voicing

Citation for published version:
Kirby, JP 2018, 'Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages', *Journal of Phonetics*, vol. 71, pp. 326-354.
https://doi.org/10.1016/j.wocn.2018.09.009

Digital Object Identifier (DOI):
10.1016/j.wocn.2018.09.009

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Phonetics

Onset pitch perturbations and the cross-linguistic implementation of voicing: evidence from tonal and non-tonal languages

James P. Kirby

School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, Scotland (U.K.)

Abstract

This paper investigates the relationship between Voice Onset Time (VOT) and onset f0 perturbations in three languages with a three-way laryngeal contrast between prevoiced, short-lag, and long-lag stops. To assess the relative contributions of aspiration and tonality to the realization of onset f0, a non-tonal language (Khmer) is compared to two tonal languages (Central Thai and Northern Vietnamese) using a common set of methods and materials. While the VOT distributions of the three languages are extremely similar, they differ in terms of their onset f0 behavior. Aspirated stops in general condition higher f0 on the following vowel, but this effect is mediated by tonal and sentential context: it is more prominent in citation forms than in connected speech, and for the tone languages, it is more visible with higher as opposed to lower tones. Examination of individual differences suggests that speakers may differ systematically in terms of their laryngeal adjustments for expressing voicelessness even while maintaining similar timing relations as indicated by VOT. Onset f0 differences may serve a useful complement to VOT, particularly when reasoning about the cross-linguistic implementation of voicing.

Key words: Voicing, onset f0, Khmer, Thai, Vietnamese

1. Introduction

The recent paper by Abramson & Whalen (2017), while noting that differences in Voice Onset Time (VOT) can be used to delimit voicing categories in many of the world's languages, also reminds us that this measure does not, and indeed was never intended to,

*Tel: +44 (0)131 650 3952; fax: +44 (0)131 651 3190.

Email address: `j.kirby@ed.ac.uk` (James P. Kirby)

quantify all aspects of voicing. The authors conclude with a call for further empirical study of the acoustic and articulatory aspects of the (de)voicing gestures underlying the distinction between stop consonants. This paper presents a case study of one such property, fundamental frequency at vowel onset, and suggests how it may prove useful in understanding the range of cross-linguistic and inter-speaker variability in the production of voicing.

1.1. Onset f0 perturbations (CF0)

It has long been known that differences in voicing are signaled by a range of acoustic cues in addition to VOT (Lisker, 1986). Fundamental frequency at vowel onset, or *co-intrinsic pitch* (aka 'pitch skip': Francis et al., 2006; Hanson, 2009), is one such cue that has received considerable study (House & Fairbanks, 1953 and much subsequent work). This paper uses the abbreviation CF0 to refer to this property, as distinct from *(vowel)-intrinsic pitch*, often abbreviated VF0 or IF0 (Di Cristo & Hirst, 1986; Kingston, 2007).

It is well established that, at least in languages with a two-way laryngeal contrast such as Spanish or English, CF0 is generally higher following phonologically voiceless stops, regardless of other aspects of their phonetic realization (Lea, 1973; Hombert et al., 1979; Kingston & Diehl, 1994; Hanson, 2009; Dmitrieva et al., 2015). It is also clear that CF0 can be perceptually relevant to listeners under certain conditions (Whalen et al., 1990, 1993; Francis et al., 2006). What remains debated is the ultimate source of the effect, and how it relates to the phonetic and phonological implementation of voicing. It has been suggested that CF0 is a purely mechanical by-product of the implementation of voicelessness (Kohler, 1982, 1984; Löfqvist et al., 1989), but also that it may be actively controlled by speakers to enhance a phonological contrast (Kingston & Diehl, 1994; Keyser & Stevens, 2006; Kingston, 2007), with some researchers suggesting both aspects may be in play (Hanson, 2009; Chen, 2011; Hoole & Honda, 2011).

A possible confound is that, in many of the languages in which CF0 effects have been studied most extensively (such as English or German), the nature of the relationship between the phonological specification (e.g., [voice] vs. [spread glottis]) and the phonetic implementation remains controversial. But as reviewed by Kingston & Diehl (1994, p. 434), even among languages in which the voicing contrast can uncontroversially be regarded as involving a difference in glottal aperture, findings are mixed: some studies show voiceless unaspirated stops as having higher CF0 than their aspirated counterparts, while other studies—even of the same language—may show the reverse (see Section 2.1).

One potential source of evidence bearing on this issue is languages which make a *three*-way contrast between prevoiced, voiceless unaspirated, and voiceless aspirated stops. This is because voiceless unaspirated stops in three-way languages are predicted to by phonologically 'inert' by many phonological theories of laryngeal contrast (e.g. Iverson & Salmons, 1995; Avery & Idsardi, 2001; Beckman et al., 2013). Despite their potential theoretical importance, however, empirical studies of CF0 in three-way languages are few and far between. This may be due in part to the overall typological rarity of such systems; out of the 451 languages in the UPSID database (Maddieson, 1984), just 38 have inventories containing a contrast among prevoiced, voiceless unaspirated, and voiceless aspirated coronal stops, and of these, 26 are tone or register languages, which investigators may disprefer in order to avoid an obvious experimental confound.

At the same time, the study of microprosody in tone languages provides another window into the extent to which CF0 is controlled. If CF0 is a primarily intrinsic effect, then f0 following obstruents in tone languages like Thai might be perturbed to a similar degree as in languages like English (Hombert, 1978). Alternatively, the magnitude of any such effect may be attenuated in tone languages, either to avoid interfering with the perception of lexical tone (Hombert, 1977; Francis et al., 2006), and/or because the realization of lexical tone requires demands finer control over the timing of f0 targets. Moreover, tonal context is likely to play an important role. The magnitude of CF0 effects has been shown to vary with intonational context in non-tonal languages, with more visible effects in high-pitch contexts (Hanson, 2009; Kirby & Ladd, 2016). While this is what might be predicted if congruent laryngeal settings produce synergistic effects on pitch (Kohler, 1982), as reviewed below, studies of CF0 in tonal languages have not always corroborated this hypothesis (Hombert, 1978; Xu & Xu, 2003; Chen, 2011).

Empirical clarity in both of these domains is important, given the implications that CF0 effects have for our understanding of the cross-linguistic implementation of voicing. As made abundantly clear by a large body of studies, the distribution of VOT alone does not tell the whole story. This should not be read as a slight against VOT as a measure; as noted by Abramson & Whalen (2017), this was not something that the measure was intended to do, nor did it ever claim to.[1] When reasoning about the ways in which different languages

[1] 'It must not be supposed...that we have ever claimed that even in utterance-initial position the dimension of laryngeal timing will explain every distinction of homorganic consonants that apparently involve laryngeal features of one sort or another' (Abramson, 1977, p. 299).

can realize a phonological voicing contrast, it is therefore important to remind ourselves that VOT may not reflect all possible aspects of laryngeal adjustment, even when similarly distributed (on this point see also Docherty, 1992, p. 129). CF0 may be an especially useful diagnostic to consider in this respect, because f0 is known to covary with gestures such as activation of the cricothyroid musculature (Löfqvist et al., 1989) and differences in larynx height (Honda et al., 1999)[2], which are also involved in implementation of voicing. Thus, regardless of whether CF0 is primarily the result of active or co-active gestures, it would be useful to know the extent to which it varies cross-linguistically (cf. Lisker & Abramson, 1964; Cho & Ladefoged, 1999 for VOT).

1.2. Goals and hypotheses

With this background, the primary goals of the present study were to improve the empirical picture of (1) how CF0 is realized in languages with three-way laryngeal contrasts, (2) how its realization varies between tonal and non-tonal languages with otherwise similar laryngeal contrasts, and (3) the degree to which CF0 is attenuated (or not) in different tonal contexts.

While a number of studies of CF0 in tonal languages have been conducted, the findings have been inconclusive and at times even contradictory, possibly due to the wide range methodologies employed over the years these studies were conducted (Chen, 2011). In an attempt to control for methodological confounds, the present study compares Standard Khmer, a non-tonal language, with two tone languages, Central Thai and Northern Vietnamese, using a closely matched set of experimental materials in a paradigm similar to that of Chen (2011) and Hanson (2009). We focus on three specific questions:

Q1: *How do CF0 effects pattern in languages with a three-way voicing contrast, particularly with respect to voiceless stops?* As noted above, CF0 behavior in two-way languages does not appear to be well predicted by the presence vs. absence of voicing lag. If CF0 behaves more consistently with respect to VOT in three-way languages, this would be an interesting finding in need of explanation.

[2]Hombert et al. (1979) suggest that larynx height could exert an effect on CF0 via changes in the *vertical* tension of the vocal folds. It must be noted that the actual mechanics of such an effect remain to be demonstrated. The study of Honda et al. (1999) shows instead how larynx height can change the *horizontal* tension of the folds, consistent with the hypothesis of Halle & Stevens (1971).

Q2: *How do CF0 effects differ in tonal vs. non-tonal languages?* If CF0 is actively attenuated to avoid interfering with lexical tone production or perception, or if there are difference in the timing of pitch targets, the magnitude of the CF0 effect should be smaller for Thai and Vietnamese (tonal languages) than for Khmer (a non-tonal language).

Q3: *How does the magnitude and extent of CF0 differ in different tonal contexts?* If the magnitude of the effect is simply a function of the local f0 environment, we might find that CF0 effects are strongest in high-tone contexts, as has been found for non-tonal languages such as English (Lea, 1973; Hanson, 2009) or French and Italian (Kirby & Ladd, 2016). This is consistent with the hypothesis that congruent laryngeal settings produce synergistic effects on pitch (Kohler, 1982). However, the oscillation of the vocal folds is also less likely to be perturbed by extrinsic factors when they are tense, as they would be for high pitch. Given that the change in f0 with subglottal pressure is smaller for high pitch than for low pitch (Titze, 1989), one might instead predict CF0 effects to be greatest in low tone contexts, especially if the effect is primarily an aerodynamic one (Xu & Xu, 2003).

The remainder of this paper is organized as follows. Sec. 2 briefly reviews previous studies of CF0 effects, with special attention to the preceding research questions. Sec. 3 provides some information on the languages considered in the present study, the experimental procedures, and the methods of analysis. Sec. 4 presents the empirical findings of the study, which are discussed in Sec. 5. Sec. 6 concludes.

2. Previous work on CF0

Studies of CF0 are relatively plentiful for languages like English, in which the duration of voicing lag is the principal acoustic correlate of phonological voicing. For a comprehensive review of this literature, the reader is referred to Hoole (2006), Hanson (2009), and Chen (2011). Data are less plentiful for so-called 'true voicing' languages, in which voicing lead is a crucial cue to the contrast (but see Dmitrieva et al. 2015; Kirby & Ladd 2016; Coetzee et al. 2018 for some recent studies). For present purposes, the critical take-aways from this literature are that (a) f0 is consistently *raised* following voiceless stops, regardless of other aspects of phonetic realization; but (b) f0 can also be *lowered* during (Kirby & Ladd, 2016) or after (Coetzee et al., 2018) the closure in languages with prevoiced stops.

2.1. *The (non-)relationship between voicing lag and CF0*

As noted above, there does not seem to be a consistent relationship between voice lag and CF0 across studies or languages. As shown in Table 1, some studies (e.g. Gandour, 1974; Xu & Xu, 2003; Francis et al., 2006) report that unaspirated stops are associated with greater CF0 perturbations than aspirated stops, while others (Erickson, 1975; Zee, 1980; Shimizu, 1994; Lai et al., 2009) report the opposite. As noted by Chen (2011), direct comparison is complicated, as the studies concerned were conducted over a period of several decades using a range of languages, recording equipment, methodologies and analysis procedures. In addition, many studies listed in Table 1 involve very small sample sizes, and/or report variation in the direction of the effect between participants.

2.2. *Differences in magnitude and temporal extent*

Studies also report differences in the magnitude and temporal extent of CF0 perturbations. In non-tonal languages, relatively large CF0 differences of 20 Hz or more between voiced and voiceless consonants are often reported to persist as much as 100 msec following vowel onset, much longer than would be predicted by purely aerodynamic factors (Hombert, 1978). In tone languages, early studies reported a tendency for CF0 differences, if present, to be on the order of just a few Hz and to converge much earlier in the vowel, compared to languages like French and English. For example, Hombert (1977) reports CF0 effects in Yoruba to last for just 20 to 30 msec; comparable durations are given by Gandour (1974) for Thai, Francis et al. (2006) for Cantonese, and Xu & Xu (2003) for Mandarin.

However, larger magnitudes have also been reported in tonal languages. In Shimizu's 1994 Thai data, f0 following voiceless aspirated stops was around 205 Hz on average, compared to 186 Hz for unaspirated stops and 183.5 Hz for voiced stops, while Vũ Thanh Phương (1981) reports the difference between /t/ and /t^{h}/ in Northern Vietnamese to be as great as 18 Hz for some speakers. While they do not give numerical estimates, figures in Xu & Xu (2003) suggest that CF0 differences between aspirated and unaspirated stops can be as great as 50 Hz in Mandarin Chinese (see Section 2.3 below). Similarly, the figures in Francis et al. (2006) suggest that the magnitude of the CF0 effect in Cantonese is comparable to that observed in European languages, although the f0 contour following aspirated and unaspirated voiceless stops converges much more quickly.

One possibility that does not appear to have been considered by most previous researchers is that *both* types of voiceless stops might exhibit a raising effect. That is, there

Language	C > C^h	C^h > C	C ~ C^h
Danish		Jeel (1975)	Fischer-Jørgensen (1968)
Hindi	Shimizu (1994)	Kagaya & Hirose (1975)	
Korean	Kagaya (1974)	Han & Weizman (1970); Kagaya (1974)	
Madurese	Cohn & Lockwood (1994)	Misnadin (2016)	
Burmese			Shimizu (1994)
Cantonese	Francis et al. (2006)	Zee (1980)	
Mandarin	Xu & Xu (2003)		Shimizu (1994)
Shanghainese		Chen (2011)	
Taiwanese Min		Lai et al. (2009) (F)	Lai et al. (2009) (M)
Thai	Gandour (1974); Erickson (1975)	Erickson (1975); Ewan (1976); Shimizu (1994)	
Vietnamese		Vũ Thanh Phương (1981)	Vũ Thanh Phương (1981)

Table 1: Summary of reported differences in CF0 effects across studies and languages. If a study is listed in more than one column, this means that the direction of the effect varied between participants. For the tone languages, some effects only obtain in particular tonal contexts; see Section 2.3 for discussion.

may be small numerical differences between voiceless aspirated and unaspirated stops, but these might both show a comparable magnitude of CF0 difference relative to voiced stops, or sonorants. In languages such as Mandarin and Cantonese, it is of course not possible to compare with truly voiced stops, but it is possible to compare with nasals or liquids, and as shown by Luo et al. (2016), both aspirated and unaspirated voiceless stops condition higher CF0 in both Cantonese and Mandarin relative to sonorants.

Many authors have reasoned that the differences in magnitude and extent may result from a general tendency in tone languages to minimize the CF0 effect, potentially in order to maximize perceptual distinctiveness. However, more robust CF0 effects have also been reported in tone languages: although the magnitude was rather small (~5-7 Hz), Lai et al. (2009) found CF0 effects to persist for up to 140 msec in Taiwanese Min, and Shimizu (1994) reports differences in Thai persisting over 100 msec following vowel onset. This may be because the duration of the perturbation effect is associated with a relative proportion of the tone, rather than an absolute duration; speaking rate may also be a factor, as may the tone of the syllable itself, as we now review.

2.3. Effect of tonal context

Both the magnitude and temporal extent of CF0 perturbations are known to interact with tonal context. In non-tonal languages, CF0 effects have been found to be most prominent in high-pitch environments, such as under focus (Lea, 1973; Hanson, 2009; Kirby & Ladd, 2016), although they are not always complete obliterated when pitch is low (Kohler, 1982).

In tone languages, the magnitude and extent of CF0 can depend on the tone of the vowel preceding, as well as following the obstruent. Figures in Xu & Xu (2003) suggest that CF0 differences range from about 8 Hz on syllables with falling tones to as much as 50 Hz for syllables with low tones. Similarly, Chen (2011) found that a preceding low rising tone allowed a much greater CF0 effect than either a high rising or high falling tone, both in terms of magnitude as well as temporal extent. In Taiwanese Min, CF0 effects appear to be greatest for low level, low rising, and high falling tones (Lai et al., 2009, Figure 5). CF0 effects were greatest for low and (low-)rising tones for the single Thai speaker studied by Gandour (1974). And in Cantonese, Francis et al. (2006) found CF0 perturbations to persist for just 20 msec on syllables with high tones, but up to 70 msec on falling tones.

Xu & Xu (2003) suggest that these differences may be due to the relationship between f0 and subglottal pressure: when intended pitch is low, the vocal folds are lax, and therefore more likely to be affected by aerodynamic factors. Their model assumes that there is an underlying pitch target for each tone, static for high and low tones and dynamic for the rising and falling tones. A slightly different account is put forth by Vũ Thanh Phương (1981, pp. 133–140), who proposes that the magnitude of the effect is the result of interpolation between the 'f0 loci' of the consonants–these a function of intrinsic articulatory or phonatory factors–and the 'pitch targets' of the tones themselves, which may not all be temporally distributed in an identical way. For instance, he suggests that a level tone might have a single pitch target at syllable midpoint, while a falling tone might have two, one at syllable onset and one at syllable offset. The degree of perturbation is then predicted to be a function of the amount of adjustment necessary to accommodate both targets, other factors being equal. Thus, voiceless onsets are predicted to perturb low tones more than high tones, while the opposite would be true for voiced onsets.

Discourse context influences CF0 in tonal languages as well. Chen (2011) demonstrated how CF0 effects become more salient under focus, with greater magnitude and longer temporal extent. She speculates this may be due to tonal gestures being exaggerated and therefore realized with an expanded pitch range. However, she notes that the effect of focus varied between and within speakers, and it is not completely clear how to reconcile the effect of expansion under focus with the empirical findings reported above that the greatest CF0 effects in tone languages often appear in incongruent contexts.

3. Materials and methods

3.1. Languages

The three languages examined in this study are Standard Khmer, Central (Bangkok) Thai, and Northern (Hanoi) Vietnamese. While Thai and Vietnamese are tone languages, Khmer is not. These three languages are in one sense a convenience sample, but they have certain properties that make them an especially appropriate comparison set. First, all three have a three-way voicing contrast between prevoiced, voiceless unaspirated, and voiceless aspirated stops; as noted in Section 1.1, this type of contrast appears to be comparatively rare, particularly among non-tonal languages, so Khmer is an important language to consider in this regard. Second, while Central Thai and Northern Vietnamese are both tonal languages, they have rather different tone systems from a phonetic standpoint, with phonation type playing a much more important role in Northern Vietnamese. Finally, it is relatively easy to find speakers of all of these languages who, if not strictly monolingual, are heavily dominant in their native idiom.

3.2. Speech materials

The basic speech materials were the same in all three languages: CV syllables consisting of one of the onsets /d t t^h n l/ with a low vowel /aː/ or /ɑː/. The coronal series was selected because /p/ is restricted to loanwords in Vietnamese, and because none of the languages have a prevoiced velar plosive /g/ (i.e., the full contrast is only present in the coronals and, for Khmer and Thai, bilabials). The exact place of articulation of these sounds varies from laminal-dentialveolar to apico-alveolar; see Harris (2006); Kirby (2011). Implosion of /d/, while possibly canonical in Khmer and Vietnamese, is non-contrastive in all languages. Other relevant differences between each language are noted below.

3.2.1. Standard (Central) Khmer

Most phonological inventories of Khmer, beginning at least with Henderson (1952), do not include an aspirated stop series, instead choosing to treat aspirated onsets as clusters of C + /h/. The primary evidence cited for this analysis is that application of the (productive) nominalizing infix *-am(n)-* to a form such as ខឹង /khəŋ/ 'angry' reliably yields forms like កំហឹង [kɑmhəŋ] 'anger'. However, the phonetic realization of aspiration in Khmer can be complex, especially in clusters (Huffman, 1972; Butler, 2014; Kirby, 2014a). For simplicity, the present work will treat the contrast as phonemic (/t/ vs. /t^h/), but with the caveat that the results here may not generalize neatly to all types of Khmer 'aspirated' stops.

Khmer orthography encodes an old historical voicing contrast, which in the modern language means that there are sometimes multiple graphs for a single phone (Huffman, 1970). Although there is some historical and comparative evidence suggesting that this may have at one point been a register contrast (Huffman, 1985; Wayland, 1997; Wayland & Jongman, 2002), there is no synchronic evidence for a phonetic register distinction in Central (Standard) Khmer (*pace* Henderson 1952, who later publicly recanted: see Wayland & Jongman, 2002:104). Nevertheless, as materials were being presented orthographically, items from both 'registers' were recorded for control purposes.

The Khmer materials thus consisted of 10 syllables: the 5 onsets in each of two vowels, yielding the syllables /dɑː tɑː t^hɑː lɑː nɑː dɔː tɔː t^hɔː lɔː nɔː/. Each syllable is the name of a letter in the Khmer alphabet, but most also have an additional lexical meaning.

Khmer participants were recorded in Phnom Penh, and also controlled the local vernacular of that city, which has certain unique phonetic features (Wayland & Guion, 2005; Filippi, 2006; Kirby, 2014b; Filippi & Heap, 2016); however, these features are not relevant for the current investigation.

3.2.2. Central (Bangkok) Thai

Central Thai has five open-syllable tones, usually described as *high*, *mid*, *low*, *rising*, and *falling*. However, these labels do not accurately reflect the phonetic realization of the modern-day language; as shown in Table 2, the falling tone has a higher pitch excursion than the so-called 'high' tone, the high tone rises from mid to high, and the rising tone is at least as low as the low tone for much of its excursion (see Potisuk et al., 1994; Morén & Zsiga, 2006; Teeranon, 2007 for further discussion). The values given in Table 2 are approximate and do not reflect the importance of the timing of the turning points, but will suffice for descriptive purposes.

Central Thai /t t^h/ are contiguous apico-dental lamino-alveolar ('denti-alveolar'), while /d n l/ are apico-alveolar (Harris, 2006). Harris also claims that Thai [t] is longer in duration than [d], with more extensive contact and a higher larynx position (2006:80).

Thai name	description	value
เสียงสามัญ	mid	33
เสียงเอก	low	21
เสียงตรี	high	35
เสียงโท	falling	453
เสียงจัตวา	rising	213

Table 2: Tones in Central (Bangkok) Thai. Values based on Morén & Zsiga (2006).

The Thai materials consisted of 15 items: the syllables /daː taː t^haː laː naː/ each bearing one of the low, mid, or falling tones. A few items had sonorant codas to avoid using nonwords. This subset of three tones was selected to provide a full range of f0 contexts while keeping the experiment as concise as possible (given that the high/mid tones and low/rising tones have similar starting points). Also importantly, these tones are not affected by any consonant-tone co-occurrence restrictions/generalizations (high and rising tones being largely restricted to syllables with sonorant or aspirated onsets in sonorant-final syllables).

3.2.3. Northern (Hanoi) Vietnamese

Northern (Hanoi) Vietnamese has six lexical tones in sonorant-final syllables, shown in Table 3. Note that, in the Northern dialect, at least two tones, broken-rising *ngã* /3ˀ5/ and checked *nặng* /31ˀ/ involve strong glottal constriction. The mid-falling tone *huyền* /32/ may be optionally breathy (Thompson, 1965; Phạm, 2003), while low-falling(-rising) tone *hỏi* /213/, usually realized as [21], may be optionally laryngealized (Nguyễn Văn Lợi & Edmondson, 1998; Michaud, 2004).

Vietnamese name	description	value
ngang	level	44
huyền	mid-falling	32
sắc	rising	24
hỏi	low-falling	21(3)
ngã	broken-rising	3ˀ5
nặng	checked	31ˀ

Table 3: Tones in Northern (Hanoi) Vietnamese.

Voiced plosives in Northern Vietnamese are canonically, but not consistently, imploded. Like Central Thai, initial /t tʰ/ are apico-dental, lamino-alveolar, or contiguous apico-dental lamino-alveolar, while /d n l/ are apico-alveolar (Kirby, 2011). Vowel length is generally considered contrastive only for /a/ and /ɤ/ (Thompson, 1965; Kirby, 2011; Emerich, 2012).

The Vietnamese wordlist was somewhat longer than that for Thai and Khmer, and consisted of 30 syllables: the 5 syllables /daː taː tʰaː laː naː/ crossed with all 6 tones. This was done in part to study the possible effects of voice quality on CF0; however, to facilitate comparison with Thai, the present paper will focus on the level *ngang* /44/, mid-falling *huyền* /32/, and rising *sắc* /24/ tones.

3.3. Participants

Participants were native speakers of Standard Khmer (14, 5 female; ages 28-38, mean=36), Central Thai (12, 6 female; ages 22-38, mean=27), or Northern Vietnamese (14, 6 female; ages 29-58, mean=30).

3.4. Recording procedure

Items were recorded three times each both (a) in isolation and (b) in an alternative-question carrier phrase appropriate for the language, with a fixed open syllable particle preceded the target. The carrier phrase sentences were all alternative questions of the form ‘Do {I, you, s/he} {say, read, write} X, or Y?’ The target was always in position X.[3] The alternative in Y was minimally different either in terms of vowel quality or onset type, so target /tʰɑː/ might be paired with /tɑː/ or /tʰɔː/, but not /tɔː/. To avoid monotony, the pronoun was randomly altered between a first, second, and third singular form, and different verbs with the meanings ‘say’, ‘read’, or ‘write’ were used.

Some example sentences include the following:

(1) Khmer: ខ្ញុំថាធឬថ

kɲom tʰaː **tʰɔː** rɨː tʰɑː
1SG say 19th-letter or 17th-letter

‘Did I say «tʰɔː» or «tʰɑː»?’ (target is /tʰɔː/)

(2) Thai: เขาอ่านคำว่าตาหรือตามนะ

kʰaw^{213} aːn^{21} kʰam^{33} waː^{453} **taː**^{33} rɯː^{213} taːm^{33} náʔ^{35}
3SG read word COMP eye or follow FP

‘Did s/he read «eye» or «follow»?’ (target is /taː^{33}/)

(3) Vietnamese: Cậu viết âm gì đây, âm “lá” hay âm “thá”?

kɤ̆w^{31ʔ} viət^{24} ɤ̆m^{44} zi^{32} ɗɤ̆j^{44} ɤ̆m^{44} **la**^{24} hăj^{44} ɤ̆m^{44} ta^{24}
2SG write sound which here sound leaf or sound dozen

‘What did you write here, «leaf» or «dozen»?’ (target is /la^{24}/)

For Khmer, the lexical item immediately preceding the target was always ថា /tʰaː/ ‘say, speak’, produced with normal (neutral) sentence intonation. For Thai, the preceding item was always the complementizer ว่า /waː^{453}/, cliticized as part of the disyllable คำว่า /kʰam^{33} waː^{453}/ ‘the word...’. Although this item has a high-falling tone in citation form, in

[3]Note that it is not obvious that the ‘Y’ position will have a lower pitch target in the languages studied here, unlike in the European languages in which this paradigm has previously been employed (Hanson, 2009; Kirby & Ladd, 2016).

non-final position in connected speech its tone is realized without a final fall (Potisuk et al., 1994; Tingsabadh & Deeprasert, 1997; Morén & Zsiga, 2006). This ensured a naturalistic utterance structure while being largely acoustically similar to the Vietnamese carrier phrases, in which the preceding item was always the syllable *âm* /ɤ̆m^{44}/ 'syllable, word', produced with a mid/high-level tone.

Although it was not possible to simultaneously control both the tonal context and the prosodic structure across the three languages, there are at least two reasons to doubt that the test words in the carrier sentence condition might potentially be produced with different prosodic boundary strengths (or phrasings). First, the sentence structures discourage the presence of a pause or break before the test items. Second, the syntactic structures of the carrier phrases are such that the test items are unlikely to coincide with an intonational phrase boundary: they are embedded within the VP in Khmer and Thai, and although the punctuation suggests the possibility of a full IP boundary in Vietnamese, the test syllable itself would still not be at this boundary.

Participants read 10 meaningful sentences not containing any of the target or filler items as a 'warm-up' and to ensure a good standard of recording. After the alternative questions, participants read each item three times in isolation. Order was randomized within block, but the isolation reading blocks always followed the alternative question blocks. There were 60 unique items for Khmer (5 onsets x 2 vowels x 3 repetitions x 2 conditions), 90 for Thai (5 onsets x 3 tones x 3 repetitions x 2 conditions), and 180 for Vietnamese (5 onsets x 6 tones x 3 repetitions x 2 conditions).

Khmer participants were recorded in a quiet room at the Buddhist Institute, Phnom Penh, Cambodia. Thai participants were recorded in a quiet office at the Department of Linguistics, Chulalongkorn University, Bangkok, Thailand. Vietnamese participants were recorded in a sound-treated booth at the MICA Centre, Hanoi University of Technology, Hanoi, Vietnam. In all cases, recordings were made using a Beyerdynamic Opus 55.18 Mk II omnidirectional headset condenser microphone, with a Focusrite USB 6 audio interface functioning as phantom power source, A/D convertor, and preamplifier. SpeechRecorder (Draxler & Jänsch, 2004) was used for data presentation and audio capture. All recordings were made in 2015.

3.5. Acoustic analyses

The segments of interest (consonant and following vowel) were manually labeled using Praat 5.4.08 (Boersma & Weenink, 2015) based on the periodicity in the acoustic waveform, supplemented by spectrographic evidence where necessary. Four primary acoustic landmarks were annotated: the onset and offset of oral constriction; the onset of periodic voicing; and the duration of following vowel. For /d n l/, onset of voicing was always coextensive with the onset of oral constriction. For /d/, there were a (very) small number of instances in carrier phrase context where voicing died out before the release of closure (instances of 'bleed' or 'trough' in the terminology of Davidson, 2016); for the purposes of measuring (negative) VOT, these were ignored, so that negative VOT as reported here is always equivalent to closure duration. However, in the event of voicing cessation (either during the closure or immediately following the burst) the post-release re-establishment of periodic voicing was also noted, and it was from this point that any acoustic measures of vowel quality were measured.

Following segmentation, acoustic measurements of the vowel were taken using Praat. Pitch analysis was performed using the autocorrelation method of Boersma (1993), with a Gaussian analysis window of 80 msec, a 5 msec frame duration, a pitch floor of 50 Hz, and a pitch ceiling of 500 Hz. Fundamental frequency was then measured at 10 msec intervals beginning 40 ms before release of closure to 100 msec following, the range in which CF0 effects are generally expected. Prior to further analysis, raw f0 measurements were transformed to semitones (relative to speaker-specific means) to facilitate comparison of total degree of pitch change across subjects and tokens.

VOT (Lisker & Abramson, 1964), defined here as the duration of the period from stop release to the onset of periodic voicing, was measured for the voiceless stops /t t^h/. Negative VOT (voicing lead) is reported for /d/ in isolation forms. For /n l/, as well as /d/ in carrier phrase contexts (where voicing frequently carries over unbroken from the preceding sonorant) closure duration is reported in lieu of VOT.

4. Results

Statistical analyses were performed in R (R Core Team, 2014) using the packages `lme4` (Bates et al., 2014) and `emmeans` (Lenth, 2018).

4.1. VOT

Out of 4405 total utterances, 2 instances of /d/ were produced without prevoicing (1 Thai, 1 Vietnamese), and 2 instances of /t/ or /t^h/ were produced with voicing lead. These 4 utterances were removed, leaving a total of 4401 utterances.

As seen in Figure 1 and Table 4, VOT clearly distinguishes the laryngeal categories within each language. The statistical similarity of the distributions were compared by first fitting a linear mixed-effects model using LANGUAGE, CONTEXT, and ONSET (and their interactions) as fixed effects, along with random intercepts for SPEAKER and speaker-specific slopes for ONSET. This model was then used to generate post-hoc pairwise comparisons between languages for each context and onset type using `emmeans` (Lenth, 2018; see Table 5). Voicing lead for Vietnamese /d/ was shorter than in both Thai and Khmer by about 30 msec in the carrier phrase context, while Vietnamese /t^h/ was longer than both Thai and Khmer /t^h/ by roughly the same amount. VOT for /t^h/ is also significantly less for Vietnamese in isolation than in Thai. Distributions for /t/ were statistically indistinguishable in both contexts.

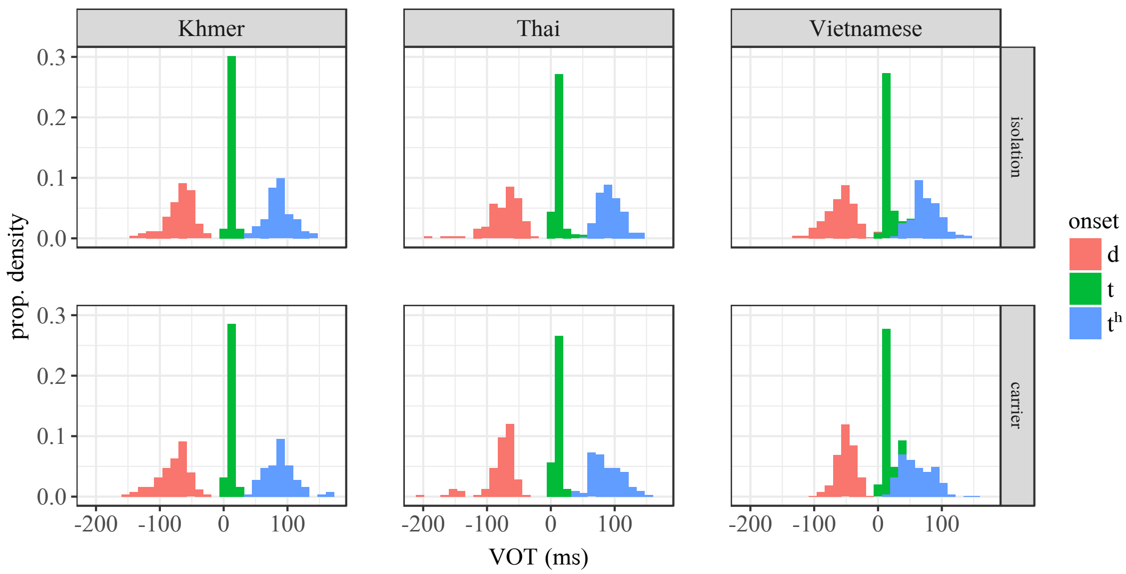

Figure 1: VOT distributions for voiced, voiceless unaspirated, and voiceless aspirated stops in each language averaged over all participants, in both isolation (top row) and alternative question carrier phrase (bottom row).

As can be inferred from Table 4, values are fairly consistent across speakers, especially

		isolation		*carrier*	
onset	language	mean	sd	mean	sd
	Khmer	-69	22	-79	25
/d/	Thai	-74	27	-80	27
	Vietnamese	-60	22	-51	15
	Khmer	12	4	11	4
/t/	Thai	12	7	11	5
	Vietnamese	14	5	13	8
	Khmer	87	21	88	25
/t^h/	Thai	94	19	88	24
	Vietnamese	75	21	61	25

Table 4: Summary statistics for VOT distributions (in msec), pooling over all tokens from all speakers.

for the voiceless unaspirated stop /t/. (See Supplementary Materials for individual VOT distributions.)

contrast	estimate	SE	df	t.ratio	p.value
onset = d, context = isolation					
Khmer - Thai	5.07	6.93	43.82	0.732	0.75
Khmer - Vietnamese	-8.78	6.60	42.46	-1.329	0.39
Thai - Vietnamese	-13.85	6.82	41.15	-2.031	0.12
onset = t, context = isolation					
Khmer - Thai	-0.31	2.14	280.62	-0.146	0.99
Khmer - Vietnamese	-1.93	1.89	203.48	-1.023	0.56
Thai - Vietnamese	-1.62	1.76	134.04	-0.918	0.63
onset = t^h, context = isolation					
Khmer - Thai	-6.48	6.61	44.81	-0.981	0.59
Khmer - Vietnamese	12.10	6.29	43.10	1.925	0.14
Thai - Vietnamese	18.58	6.50	41.81	2.861	0.02
onset = d, context = carrier					
Khmer - Thai	0.42	6.92	43.79	0.061	0.99
Khmer - Vietnamese	-28.56	6.60	42.43	-4.328	<0.001
Thai - Vietnamese	-28.98	6.82	41.10	-4.253	<0.001
onset = t, context = carrier					
Khmer - Thai	0.69	2.15	284.14	0.320	0.95
Khmer - Vietnamese	-1.51	1.88	200.56	-0.802	0.70
Thai - Vietnamese	-2.19	1.76	134.43	-1.245	0.43
onset = t^h, context = carrier					
Khmer - Thai	-0.37	6.61	44.81	-0.056	0.99
Khmer - Vietnamese	26.96	6.29	43.08	4.289	<0.001
Thai - Vietnamese	27.33	6.49	41.79	4.208	<0.001

Table 5: Between-language comparisons of VOT by onset and context. Fractional degrees of freedom computed using the Satterthwaite method, with *p*-values adjusted using the Tukey method for comparing a family of 3 estimates.

4.2. CF0

For each language aggregate results will be presented first, followed by a discussion of individual differences. Statistical comparisons in all languages are based on a linear mixed-effects model predicting f0 (in semitones) based on ONSET (/d n l t tʰ/), CONTEXT (isolation or carrier phrase) and their interaction, together with random slopes for subjects; models including subject-specific slopes did not converge. The Satterthwaite method was used to approximate degrees of freedom in all cases; *p*-value adjustments were made using the Tukey method. Statistical comparisons can be found in the Supplementary Materials, along with the data and code to reproduce the analyses and figures.

For Thai and Vietnamese, study of tonal context is restricted to just three tones in each language. This was done in part to try and compare similar tonal contexts in each language, but also to avoid potential interference from phonotactic consonant-tone co-occurrence restrictions (in Thai) and voice quality (in Vietnamese); see Sections 3.2.2 and 3.2.3.

4.2.1. Khmer

As seen in Figure 2, CF0 is raised for both types of voiceless stop in Khmer, with mean predicted f0 following voiceless stops 1-2 semitones higher than voiced stops and sonorants in both conditions. In addition, CF0 for voiceless aspirated /tʰ/ is the highest at all timepoints for both vowels and in both speech contexts. Interestingly, voiceless unaspirated /t/ also appears to raise f0, though not to the same extent as /tʰ/. The voiced stop /d/ and the sonorants /l n/ do not have different f0 profiles at any of the three timepoints considered (onset of vowel, 50 msec after vowel onset, 100 msec after vowel onset), although there does appear to be a lowering of f0 during the closure for /d/, as found in e.g. French and Italian (Kirby & Ladd, 2016).

In terms of the effects of orthographic register (= different vowel), the expected vowel-intrinsic f0 differences are observed (Whalen & Levitt, 1995). Marginal predicted mean differences of up to 0.77 st are found between vowels for a given onset type and context, though significant only for /ɑː/ and /ɔː/ in carrier phrase context (see Supplementary Materials). However, the difference (if any) is always in expected direction, i.e. /ɔː/ always has higher f0. Perhaps most important is the fact that the relative patterning holds across vowels and contexts: f0 is higher following aspirated /tʰ/ at all timepoints compared to voiced sounds, and often when compared to voiceless /t/ as well.

Figures 5 and 6 (see Appendix) show individual f0 trajectories for each speaker, averaged

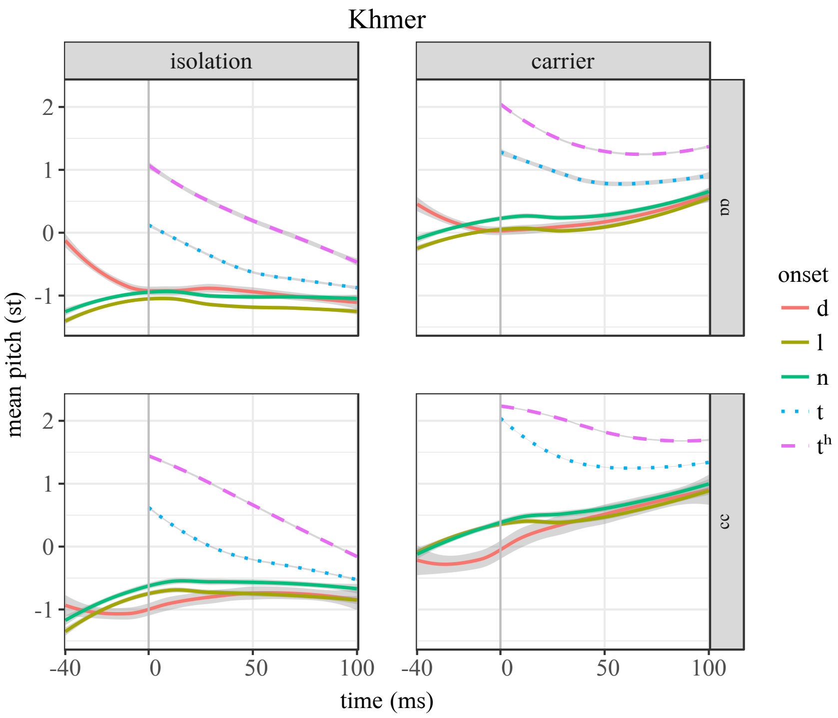

Figure 2: Mean smoothed f0 trajectories (in semitones) for Khmer data by vowel/orthographic register and context, averaged over speaker and repetition. Shading indicates 95% confidence interval.

over repetition and vowel. There are several things to note. First, the basic pattern seen in Figure 2 is largely apparent in both contexts for most speakers: /t^h/ conditions higher f0 than other segments, and f0 trajectories following voiced onsets are broadly similar. However, the behavior of /t/ is somewhat variable: for some speakers, it appears to pattern with the sonorants, while for others, it patterns closely with /t^h/. Second, the magnitude of the deflections is not, in general, attenuated in the carrier phrase context, consistent with the aggregate results presented in Figure 2. The third point of note concerns prevoiced /d/. Interestingly, /d/ does not always lower f0 during the closure; for some speakers, f0 actually *descends* during the closure, converging towards the baseline by the point of release. This might be expected if these are implosive realizations, which tend to pattern with voiceless stops in terms of their CF0 effects (Ohala, 1973; Nihalani, 1986; Wright & Shryock, 1993; Clements & Osu, 2002; Odden, 2007; Nagano-Madsen & Thornell, 2012). This possibility is discussed further in Section 5.2.

Finally, there is some variability with respect to how the f0 trajectory following /d/ patterns with respect to the sonorants /n l/, which are assumed to most faithfully reflect the globally specified intonation contour. For some speakers, f0 following /d/ can be either higher or lower than that of /n l/, and these differences can persist up to 100 msec into the vowel (and possibly further). For other speakers, however, there is no evidence that /d/ perturbs f0 away from the baseline at any point in its post-release trajectory.

4.2.2. Thai

Figure 3 shows the aggregate results for Thai. For the high-falling tone, f0 is significantly higher following /t^h/ compared to all other onsets for the first 100 msec of the vowel in the isolation condition, but cannot be statistically distinguished from most other stops in the carrier phrase context. At vowel onset in mid-tone syllables, f0 is higher following /t^h/ compared to /n l/ in isolation, but cannot be distinguished from /d/ or /t/. By 50 msec, /t^h/ can only be distinguished from /l/, and after 100 msec no differences are apparent. In carrier phrase contexts, onset type has no clear effect on CF0, an effect which is consistent across tonal contexts.

In isolated syllables bearing a low tone, f0 is higher following /t^h/ at vowel onset compared to /d n l/, but not compared to /t/. At 50 msec, /t^h/ is distinct only from /d l/, and by 100 msec only from /l/. In carrier phrase context, the relationships between syllables with low tone are more complex, but while /n/ has a numerically higher mean f0 than other onsets throughout the period examined, the f0 contours are overall highly similar, and no obvious perturbations are apparent.

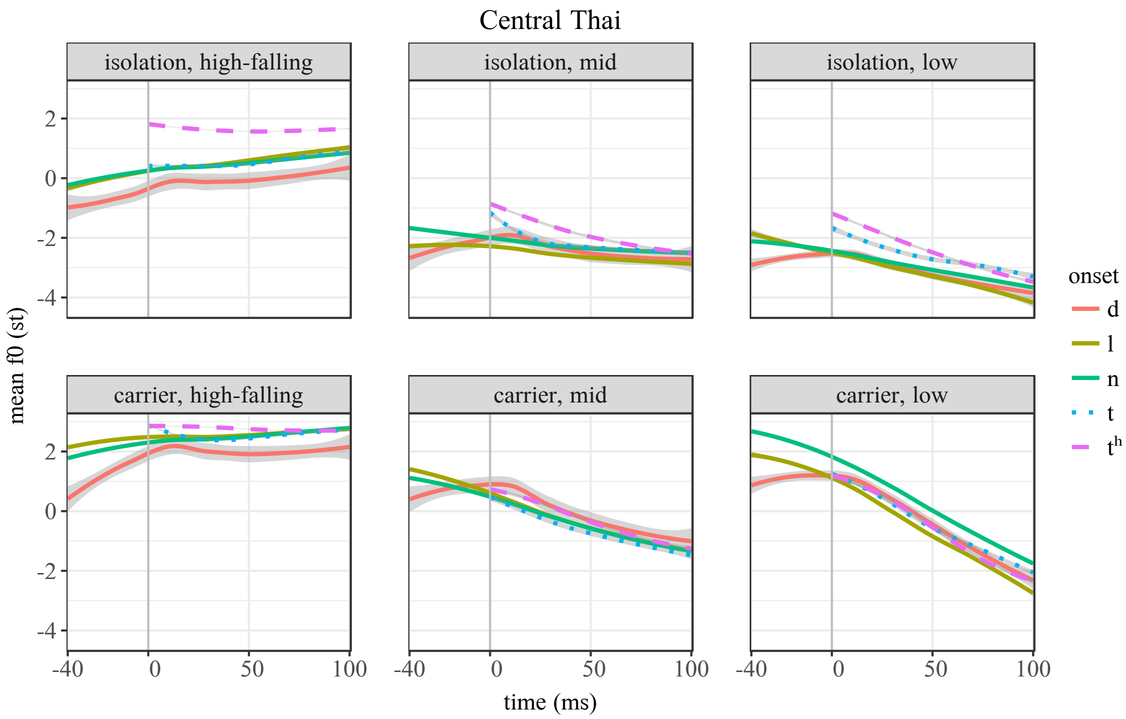

Figure 3: Mean smoothed f0 trajectories (in semitones) for Thai data by tone and context, averaged over speaker and repetition. Shading indicates 95% confidence interval.

The individual patterns in Thai (Figures 7-12 in Appendix) largely mirror the aggregate results. In particular, differences in f0 excursions are most apparent in isolation, with much more overlap in the carrier phrase contexts, especially for the mid tone. Unlike in Khmer, however, there are only a few speakers for whom /t/ patterns with /t^h/ with respect to CF0, and those speakers do not appear to be consistent across tones. For speaker TF5, for example, f0 following /t/ patterns with /d n l/ for the high-falling tone, but with /t^h/ for the other tones, while for speaker TM2, the effect for high and low tones is reversed. For some speaker-tone-context combinations, there are also instances where f0 is highest following /t/ or one of the sonorants /n l/ (see e.g. TM3, TF4, and TF6, mid tone, carrier phrase).

4.2.3. Vietnamese

Figure 4 shows aggregate results for Northern Vietnamese. Like Thai, the raising effect of voiceless stops is primarily visible in isolation; while clearly prominent on syllables bearing the high(-level) tone, the magnitude of the effect is fairly consistent across tones in this sentential context, on the order of 1-1.5 semitones. Unlike Thai, however, the temporal extent of the perturbations are generally much shorter (50 msec or less), consistent with the findings of Hombert (1977) for Yoruba and Francis et al. (2006) for Cantonese. By 100 msec after onset, only /d/ and /t^h/ can be statistically distinguished, and then only for the falling and rising tones. Also unlike Thai, voiceless unaspirated /t/ appears to initially pattern with aspirated /t^h/ for the level and falling tones in isolation. Finally, f0 following /d/ appears to more closely track the sonorant trajectory, without a clear dip during the closure phase (but see below). These observations are corroborated by the post-hoc comparisons (see Supplementary Materials).

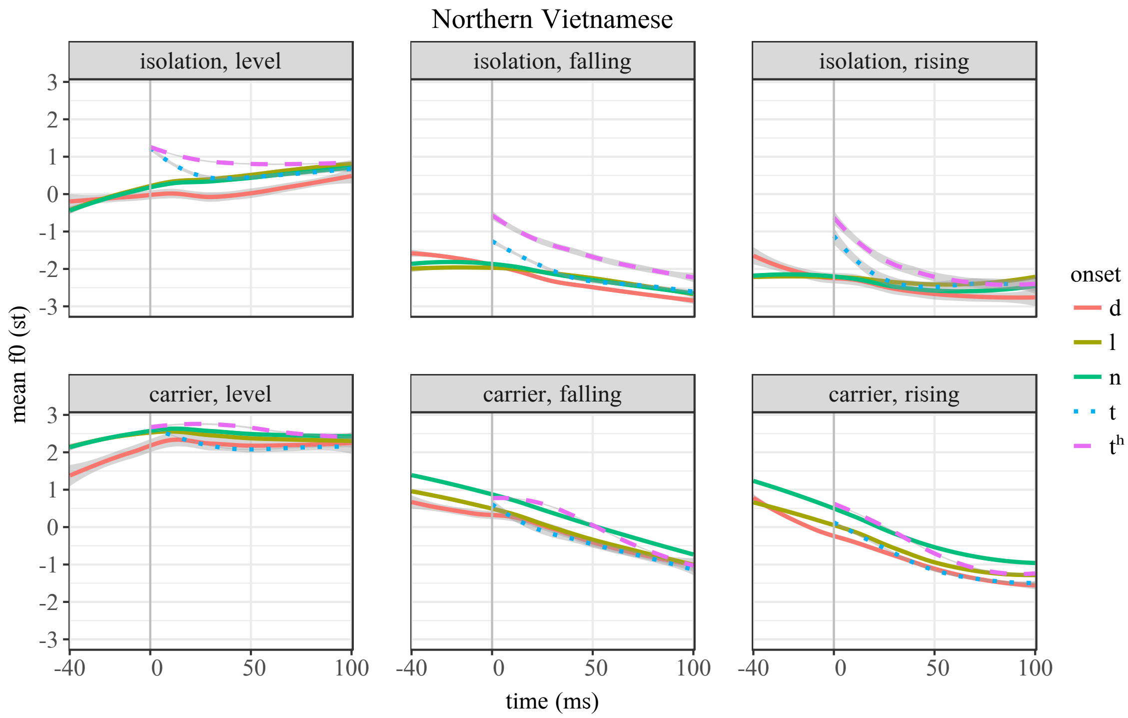

Figure 4: Mean smoothed f0 trajectories (in semitones) for Vietnamese data by tone and context, averaged over speaker and repetition. Shading indicates 95% confidence interval.

In terms of the behavior of the voiceless stops, the individual trajectories (Figures 13–18 in Appendix) largely mirror the aggregate results. As with Thai, there is some variation in the patterning of /t/ with respect to /t^h/, including a few instances where /t/ appears to condition higher f0 than /t^h/ (e.g. subject VM1, rising tone isolation condition, or subjects

VM3 and VF3 in level tone carrier phrase condition).

The variation in the trajectory of prevoiced /d/ during the closure is particularly interesting. While the aggregate plots in Figure 4 suggest that /d/ largely tracks the trajectory of /n l/, the individual plots cluster into two broad types. For speaker VM2, for example, f0 consistently starts high and falls throughout the closure, regardless of the lexical tone or sentential context. For speakers like VF4 or VF5, f0 rises throughout the closure, starting lower than the sonorant baseline and coming to meet it only at the onset of the vowel. As in Khmer, these differences are consistent with the descriptive observation that /d/ can be realized as an implosive (see Section 5.2), though this variability also highlights the existence of speaker-specific laryngeal adjustments which accompany prevoiced stops that are not be captured by the simple presence vs. absence of voicing lead.

5. Discussion

The goals of this study were to understand how CF0 patterns in languages with three-way voicing contrasts, especially with respect to voiceless stops (Q1); how CF0 effects differ, or don't, between tonal and non-tonal languages (Q2); and how the magnitude and extent of CF0 differs with tonal context (Q3). The basic findings can be summarized as follows:

- Q1: *How do CF0 effects pattern in languages with a three-way voicing contrast, particularly with respect to voiceless stops?* The VOT distributions of the three languages are extremely similar in both isolation and carrier phrase contexts, but the effect of onset type on f0 varied by language, tonal context, and speaker. Following the voiceless aspirated stop /t^h/, f0 was generally raised relative to /d n l/ (at least in isolated speech). The behavior of /t/ was more variable, sometimes patterning with /t^h/ and sometimes with /d n l/. This variability was observed between speakers as well as contexts (see Section 5.1). There are also differences evident in the realization of /d/, both within and across languages, consistent with descriptions of Vietnamese and Khmer which indicate that this segment may be canonically imploded (see Section 5.2).

- Q2: *How do CF0 effects differ in tonal vs. non-tonal languages?* The magnitude and temporal extent of CF0 was greatest and most consistent in non-tonal Khmer. However, effects were also observed in both tone languages in a range of tonal contexts, at least in isolation forms. In carrier phrase context, microprosodic effects were not detectable in either Thai or Vietnamese (see Section 5.3).

Q3: *How does the magnitude and extent of CF0 differ in different tonal contexts?* In Thai, CF0 effects were greatest in the high-falling tonal context. In Vietnamese, CF0 was more reliably visible with (high-)level and mid-falling tones, compared to the low-rising tone. In both languages, speakers tended to be consistent across tonal contexts; that is, if a speaker is a '/t^h/ raiser' in one tonal context, s/he is likely to show a similar effect in all contexts (e.g. VM1, VF1, TF3, TF5). These issues are discussed further in Section 5.3 below.

5.1. Variation in /t/: differences in vocal fold tension?

The range of observed individual differences suggests that there may be significant, and systematic, variability in how relevant laryngeal gestures are programmed by different speakers (Flege, 1982; Docherty, 1992): some speakers may implement the same type of gesture to support devoicing for both /t/ and /t^h/, while others may distinguish /t/ from /t^h/ both in terms of timing differences as well as devoicing gesture.

In general, /t^h/ tended to condition higher f0 than other segments, while f0 trajectories following voiced onsets were broadly similar for all three languages. Conversely, CF0 following /t/ was more variable: for some speakers, it appears to pattern with the sonorants, while for others, it patterns closely with /t^h/. If this is due to a by-speaker difference in vocal fold tension, we might expect a correlation between the difference in pitch between /t/ and the sonorants /n l/ and the difference in an acoustic measure correlating with vocal fold tension (such as H1-H2 or H1-A1), since for a tenser voice, the expectation is that higher-order harmonics should dominate the spectrum. If we assume that the e.g. H1-H2 value for sonorants represents a relatively modal voice quality, then if /t/ is tense, the (positive) difference in H1-H2 between sonorants and /t/ should be proportional to the (negative) difference in pitch between these same segments. In other words, if $\Delta(f0_t - f0_{sonorant})$ is positive, then $\Delta(\text{H1-H2}_t - \text{H1-H2}_{sonorant})$ should be negative; and as the former quantity approaches 0, so too should the latter. To take a specific example, we might expect the voice quality measures for /t/ and /n l/ at vowel onset in the isolation context to be very similar for speaker KF2, since the f0 differences are minimal, while /t/ would be expected to have a smaller or negative spectral slope compared to /n l/ for speaker KM4, where the differences are larger (see Figure 5).

This was tested by comparing the speaker mean difference in pitch (in semitones) with H1*-H2* and H1*-A1* (in dB) for /t/ and /d n l/ at 10 msec following closure release, separately in isolation and carrier phrase contexts. Amplitudes were corrected for the effect of formant bandwidths using the method of Iseli & Alwan (2004), as implemented in PraatSauce (Kirby, 2018). The results for Khmer (isolation context) and Thai (high-falling tone isolation context) are given in Tables 6 and 7. In general, the predicted pattern does not emerge. For example, the difference in f0 for KF2 is indeed quite small, but /t/ has a smaller spectral slope than /n l/; while for KM4 and KM5, the f0 difference is nearly 2 semitones, but there is virtually no difference in H1*-H2*. Similarly, for the Thai examples, spectral slope is generally less for /t/ than for /n l/ regardless of how f0 patterns. See Supplementary Materials additional plots and data.

Δ f0	Δ H1*-H2*	Δ H1*-A1*	subject
-0.05	-3.76	-3.50	KF2
0.25	0.13	-8.99	KM8
0.48	-0.24	1.67	KM6
0.56	-4.82	-3.23	KF3
0.57	-3.91	-1.89	KF1
0.64	0.02	-4.19	KM1
0.75	1.80	-3.21	KM3
0.76	-1.81	-0.45	KF5
0.79	-0.41	-3.73	KM9
1.04	-3.53	-3.46	KF4
1.21	3.72	-1.22	KM2
1.91	-0.01	-2.99	KM4
1.97	-1.60	-2.83	KM7
2.28	-0.04	-4.95	KM5

Table 6: Mean differences in f0 (in semitones) and voice quality measures (in dB) between /t/ and /n l/ at 10 msec following vowel onset, Khmer speakers, isolation context. Rows are sorted in ascending order of f0 difference magnitude.

Δ f0	Δ H1*-H2*	Δ H1*-A1*	subject
-0.96	-5.60	-4.98	TM5
-0.83	-0.91	-0.81	TM3
-0.72	-1.48	-1.74	TM1
-0.56	-5.04	-6.22	TF2
0.04	-4.90	-2.68	TF1
0.18	-3.78	-8.06	TF6
0.32	-0.53	-4.30	TF5
0.43	1.41	-3.58	TM6
0.80	0.20	-2.74	TM4
1.00	-1.57	-0.37	TM2
2.10	-1.25	-2.09	TF3

Table 7: Mean differences in f0 (in semitones) and voice quality measures (in dB) between /t/ and /n l/ at 10 msec following vowel onset, Thai speakers, high-falling tone isolation context. Rows are sorted in ascending order of f0 difference magnitude.

5.2. Implosive realizations

There are also differences evident in the realization of the prevoiced stop /d/, both within and across languages, consistent with descriptions of Vietnamese and Khmer which indicate that this segment may be canonically imploded. Speakers appear to vary (semi-)systematically in whether or not they realized prevoiced /d/ as implosive [ɗ]. Based on visual inspection of the acoustic waveforms, those instances of /d/ with falling f0 during the closure phase are indeed implosive, as evidenced by an increase in voicing amplitude during the closure phase (Lindau, 1984). The f0 differences might be accounted for by the aerodynamic model of Ohala (1976), whereby a rapidly descending larynx would result in a higher rate of transglottal airflow and a correspondingly greater pressure difference across the glottis[4]. If this were the case, we might expect to see the effects manifest most strongly late in the closure, when they would contribute most to maintaining the transglottal pressure drop. In the present data, however, the f0 peaks occur at the midpoint of the closure or

[4] *Pace* Wright & Shryock (1993), Ohala's model does not predict higher f0 following the release of an implosive (Hombert, 1978, p. 91), although it then remains to be explained why implosives often pattern historically with other f0-raising segments.

even earlier (cf. Nagano-Madsen & Thornell, 2012, where pitch appears to peak in the last 1/3 or 1/4 of the closure).

Another possibility is that the drop is caused by a raised, and subsequently lowered larynx. To the extent that f0 is correlated with larynx height (Honda et al., 1999), a raised larynx may result in high(er) f0 (although Honda et al., 1999 observed little vertical larynx movement in the high f0 range). If the larynx were raised above its normal resting position before it was lowered, this would have the effect of increasing the possible distance it can move downwards, thereby maximizing the rarefaction of the air in the oral cavity. This is admittedly speculation, but would be consistent with the f0 patterns observed here. Note also that, for those speakers for whom f0 descends during the closure of /d/, they are not always consistent across contexts, suggesting that they are not producing /d/ in the same way in connected speech as in isolation forms. However, the data do not include any instances where speakers produce implosives in the carrier phrase but not in isolation.

5.3. CF0 in tone languages and implications for CF0 control

The present results suggest that if there is in fact attenuation of CF0 in tone languages (Hombert, 1977; Francis et al., 2006), such attenuation is not across-the-board, but depends on language-particular (and possibly speaker-specific) implementation of laryngeal maneuvers to support (de)voicing and achieve lexical pitch targets. It seems reasonable to think that, in languages with fewer lexical tones, timing of laryngeal gestures relegating tonal pitch targets may be less precise and/or more flexible (Hombert, 1977; Lai et al., 2009). This leads to the prediction that CF0 effects in languages with fewer tones—for example, many African tone languages, or East/Southeast Asian languages which have developed tone while losing final fricatives and glottal stops, but which still retain a laryngeal contrast in the onsets—may be more extensive than those in languages which have undergone subsequent tone splits, but less so than in non-tonal languages such as English. One might also expect differences in listeners' perceptual sensitivity to these effects. These hypotheses await experimental investigation.

Diachronic considerations also give reason to doubt that CF0 is actively suppressed. According to the tonogenetic model of Haudricourt (Haudricourt, 1954; Matisoff, 1973), tone *qua* tone tends to come into being due to loss of final consonants, whereas initial voicing contrasts tend to condition tonal splits (Haudricourt, 1961). If CF0 effects are actively suppressed in tone languages, this is unexpected, as speakers would need to be actively suppressing the very perturbations which are ostensibly the source of the split. Rather than suppressing CF0 effects, it is also possible that speakers instead compensate for them, by making an articulation to raise or lower f0 in the opposite direction of the effect. If different speakers compensate to different degrees, due to different expectations about the necessary extent of perceptual compensation (Sonderegger & Yu, 2010), perturbations might persist in the productions of some speakers, but not others. This suggests interesting possibilities for future investigations of the production/perception relationship of CF0.

The attenuation of CF0 in tone languages, and in certain intonational contexts in non-tonal languages, could also have a more general explanation in terms of f0 control. In non-tonal languages, CF0 effects are most prominent in high-pitch, focused conditions (Lea, 1973; Hanson, 2009; Kirby & Ladd, 2016), although they are not always complete obliterated in low-pitch conditions (Kohler, 1982). In the Thai and Vietnamese data shown in the present study, CF0 effects were most prominent in isolated words, but were visible in all tonal contexts for at least some speakers. In carrier phrase context, however, CF0 effects were frequently attenuated or completely absent.

The higher visibility of CF0 effects in the isolated word context is potentially related to the more general tendency for pitch range expansion under focus in both tonal (e.g. Xu, 1999; Chen, 2009) and non-tonal languages (e.g. Xu & Xu, 2005; Ladd, 2008). It is also reasonable to suspect that prosodic structure may play a role, as it is known to influence the realization of suprasegmental features (Shattuck-Hufnagel & Turk, 1996; Cho, 2011); however, the study of Mandarin Chinese by Yan et al. (2011) found articulatory evidence of spatio-temporal domain-initial strengthening, but no acoustic evidence of pitch range expansion associated with hierarchical prosodic structure. Furthermore, as noted in Section 3.4, the test syllables never occurred on an IP boundary in any of the three languages. If phrasing itself impacts CF0, one might have expected Khmer and Thai to pattern together, but instead Thai and Vietnamese were broadly more similar. That said, the possible role of phrasal position on CF0 remains to be explored more fully.

The disappearance of CF0 effects in carrier phrase contexts might also be explained with reference to constraints on the timing of programming and executing f0 gestures (Xu & Sun, 2002; Kuo et al., 2007; Gao, 2008, 2009). If gestures to achieve lexical tone targets are privileged, the laryngeal devoicing adjustment (which, under other circumstances, would raise f0) is not executed; but since this gesture is independent of the gesture associated with the consonant, there are no surface differences in VOT. To the extent that isolation forms most faithfully reflect the intended gestural targets for a lexical item, however, the intrinsic effects of the gesture responsible for CF0 appears to be realized in parallel with the target(s) for the lexical tone (cf. Xu & Xu, 2005). Helpful in understanding the nature of this relationship will be explicit theories of speech melody which seek to model, rather than simply smooth over, CF0 perturbations.

Finally, the differences between the carrier phrase and isolation contexts may have to do at least in part with the nature of the items used in the focus condition. Although the target item was always in focus, the contrasting item was always minimally different. This could be construed as a 'clarificational' environment, in which speakers might well hyperarticulate the target (Lindblom, 1990; Moon & Lindblom, 1994). If speakers are actively suppressing CF0 effects, any such suppression might be greater when the contrasting items are a minimal pair in terms of tone, and weaker when they contrast in terms of voicing (where CF0 would be a useful secondary cue). Since the contrastive carrier phrases in the present study almost never contained items contrasting in tone or onset voicing, it is difficult to test this explicitly with the current data, but this possibility is clearly worth pursuing.

Previous studies of CF0 in Chinese languages have have found the magnitude and temporal extent of perturbations to be greater in low tone environments. Xu & Xu (2003) found greater perturbation with Mandarin Low and Rising tones vs. High and Falling tone; Francis et al. (2006) found the temporal extent of the perturbation to last longest for the Cantonese low falling tone Lai et al. (2009); and Chen (2011) found greater perturbation following low-rising as opposed to high-falling or high-level tones in Shanghai Chinese. In contrast, in the present study the greatest CF0 effects for Thai were found in the high-falling tone environment, consistent with what has been found for non-tonal European languages. In Vietnamese, the results are less clear-cut, but there appear to be at least some speakers for whom aspiration-related CF0 persists for up to 100 msec for all three tones. While methodological explanations cannot be ruled out, it seems possible that these diverging results reflect real differences in the way that languages, or speakers, program and execute laryngeal gestures.

5.4. VOT, CF0, and feature systems for three-way languages

Given the three-way laryngeal contrasts of the languages studied here, and given the robust prevoicing and strong aspiration which characterizes /d/ and /t^h/ in all three languages, it seems clear that any feature system for these languages requires at least two distinct gestures, such as [slack] and [spread] (e.g. Halle & Stevens, 1971). However, it is not the case that voiceless unaspirated stop are simply 'unmarked' in terms of their voicing behavior, at least not for all speakers or in all contexts: in all three languages examined here, voiceless unaspirated /t/ had f0-raising properties for some, but not all speakers, and was not always raised to the same degree in both sentential contexts. This type of variation is clearly not something that is visible only by considering VOT, as VOT for voiceless unaspirated stops is extremely consistent across speakers and contexts in the present sample (see Table 4). While this in no way diminishes the utility of VOT as a measure of differences between voicing categories *within* a given language, it calls for caution in the *between*-language interpretation of categories such as 'voiceless unaspirated stop', as similarity in VOT does not necessarily imply similarity in other laryngeal adjustments (Docherty, 1992; Kingston & Diehl, 1994; Francis et al., 2006; Hanson, 2009; Ladd, 2009; Hoole & Honda, 2011; Ridouane et al., 2011).

5.5. Aerodynamic and perceptual considerations

While the direction of the differences observed in the present study ($C^h > C$) are consistent with the 'aerodynamic hypothesis' of Hombert et al. (1979, p. 42), they do not conform to the predictions of Ohala's mathematical model of speech aerodynamics (Ohala, 1974, 1976, 1978). According to this model, upon release on an aspirated stop, the rate of transglottal airflow is high, and Ps decreases rapidly during the aspiration phase. Conversely, upon release of an unaspirated stop there is little transglottal airflow, and Ps changes more slowly. As a result, Ps is expected to be lower following aspirated as opposed to unaspirated stops; and all else being equal, lower Ps should mean lower transglottal pressure, leading to lower f0 following aspirated stops.

While the present results do not corroborate this prediction (cf. the studies of Chinese languages reviewed above), this should not be taken as evidence that aerodynamic factors *never* contribute to CF0. As noted by Kohler (1982, 1985), aerodynamic effects may be masked by the magnitude of vocal fold tension adjustments in any particular utterance. Unless the vocal-fold tension component is carefully controlled, the aerodynamic effect, if any, may not be visible. In the present study, this may mean that the aerodynamic effect of CF0 raising following /t/ (relative to /d/) is not visible due to the control of vocal-fold tension for intonational and/or lexical purposes, but it would appear to leave unexplained the relatively consistent and temporally persistent raising of f0 following /t^h/ in the languages investigated here. The magnitude and consistency of this effect (at least in isolation forms) suggests instead that the laryngeal devoicing maneuver has a greater effect on CF0 than do aerodynamic factors, which in any case do not seem able to account for the temporal persistence of the perturbation observed in many languages (Hombert et al., 1979, but see e.g. Rose, 1996 for a counter-example). Although engagement of the cricothyroid (CT) musculature has been suggested as a more plausible culprit (Löfqvist et al., 1989), it must be noted that studies have not always found a clear relationship between f0 and CT activity (e.g. Erickson & Abramson, 2013); moreover, the temporal delay between peak CT activity and f0 raising is currently not well explained (Kingston & Diehl, 1994; Hoole, 2006).

The relative fragility and variability of CF0 effects may lead one to question the potential perceptual relevance of such a feature, particularly in tone languages, where f0 is already recruited for lexical functions (but see Section 5.3 above). Perceptual studies of CF0 in tonal languages are unfortunately few and far between. A notable early study is that of Abramson & Erickson (1978) on Thai. The interpretation of the Abramson & Erickson (1978) results is somewhat complicated, but their findings appear to indicate that CF0 may impact the boundary between /b/ and /p/ in Thai, but not /p/ and /p^h/. In light of the current findings, this might be expected to vary on a per-speaker or per-listener basis, depending on whether or not a particular speaker's laryngeal adjustments for /p/ cause it to pattern primarily with /b/ or /p^h/ with respect to CF0. Francis et al. (2006) found that, while Cantonese listeners can use CF0 perturbations to distinguish between voiceless aspirated and unaspirated stops, they can only do so when the temporal extent of the perturbations is far greater than what occurs in natural Cantonese speech. This finding, together with the present finding that CF0 effects seem to be largely neutralized in carrier phrase contexts in tone languages, raise some doubts as to the perceptual relevance of these perturbations in running speech (see also Abramson & Whalen, 2017). Nevertheless, as noted above, CF0 perturbations would appear to be necessarily perceptible in order to condition tone splits. Clearly much more work in this area is necessary.

6. Conclusion

This paper has examined the effects of consonant type on the realization of onset-induced pitch perturbations (CF0) in Khmer, a non-tonal language, and Northern Vietnamese and Central Thai, two tonal languages. In general, voiceless aspirated /t^h/ was correlated with higher post-release f0 in all three languages, while f0 following /d/ largely patterned with sonorants. However, voiceless unaspirated /t/ also raised CF0 for at least some speakers in all three languages, illustrating once again that CF0 not simply a function of VOT, but is instead a language- and speaker-specific function of how, or whether, a laryngeal adjustment to support active devoicing is present. It was also not always the case that CF0 was attenuated in the tonal languages; rather, the extent and magnitude of attenuation was found to vary with language, sentential, and to some extent tonal context.

In terms of distinguishing between plosives within a given language, the VOT measure has certainly stood the test of time, and as stated by Abramson & Whalen (2017), it continues to serve its intended purpose. But care must be exercised when using VOT solely, or even primarily, to reason about the cross-linguistic implementation of laryngeal contrasts. For this, it can be useful to supplement VOT with other relatively simple acoustic measures, such as CF0, in order to shed light on how laryngeal adjustments differ across languages and speakers.

Acknowledgements

This research was funded in part by a grant from the University of Edinburgh PPLS Pilot Fund. Thanks to Bob Ladd for many inspiring discussions as well as to John Kingston; editor Taehong Cho; two anonymous reviewers; and audiences at BAAP 2016, Lancaster and LabPhon 2016, Cornell for many insightful comments and suggestions. I am immensely grateful to my gracious hosts Sor Sokny of the Buddhist Institute, Phnom Penh; Pittayawat Pittayaporn of Chulalongkorn University, Bangkok; and Alexis Michaud and Trần Đỗ Đạt of the MICA Speech Communication department at the Hanoi University of Technology. Finally, thanks to Jasmine 良瑀 Sung for research assistance. The author is solely responsible for any errors of fact or interpretation.

Appendix: f0 trajectories by individual

Khmer

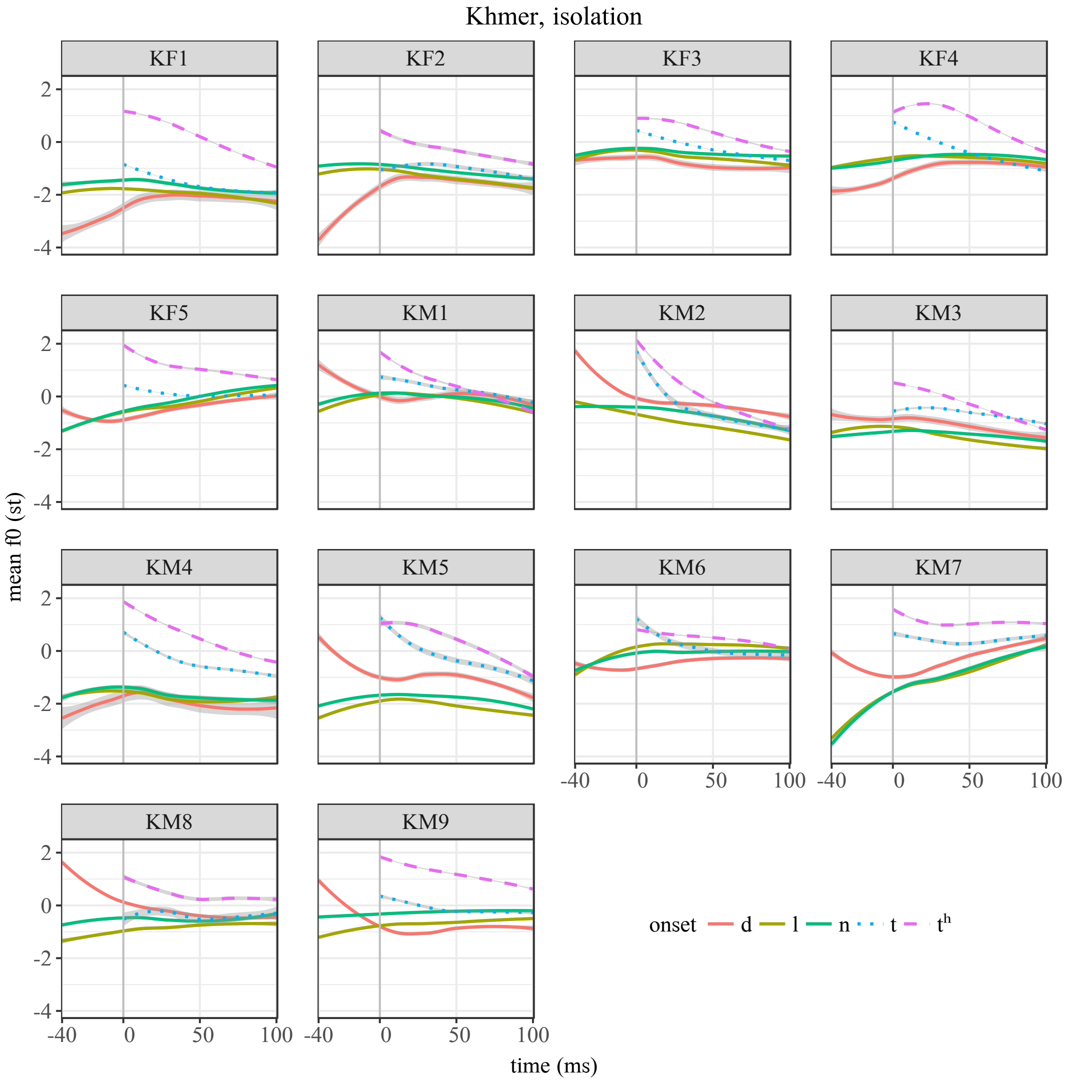

Figure 5: Mean smoothed f0 trajectories (in semitones) for Khmer data by speaker, averaged over repetition and vowel/orthographic register, isolation context. Shading indicates 95% confidence interval.

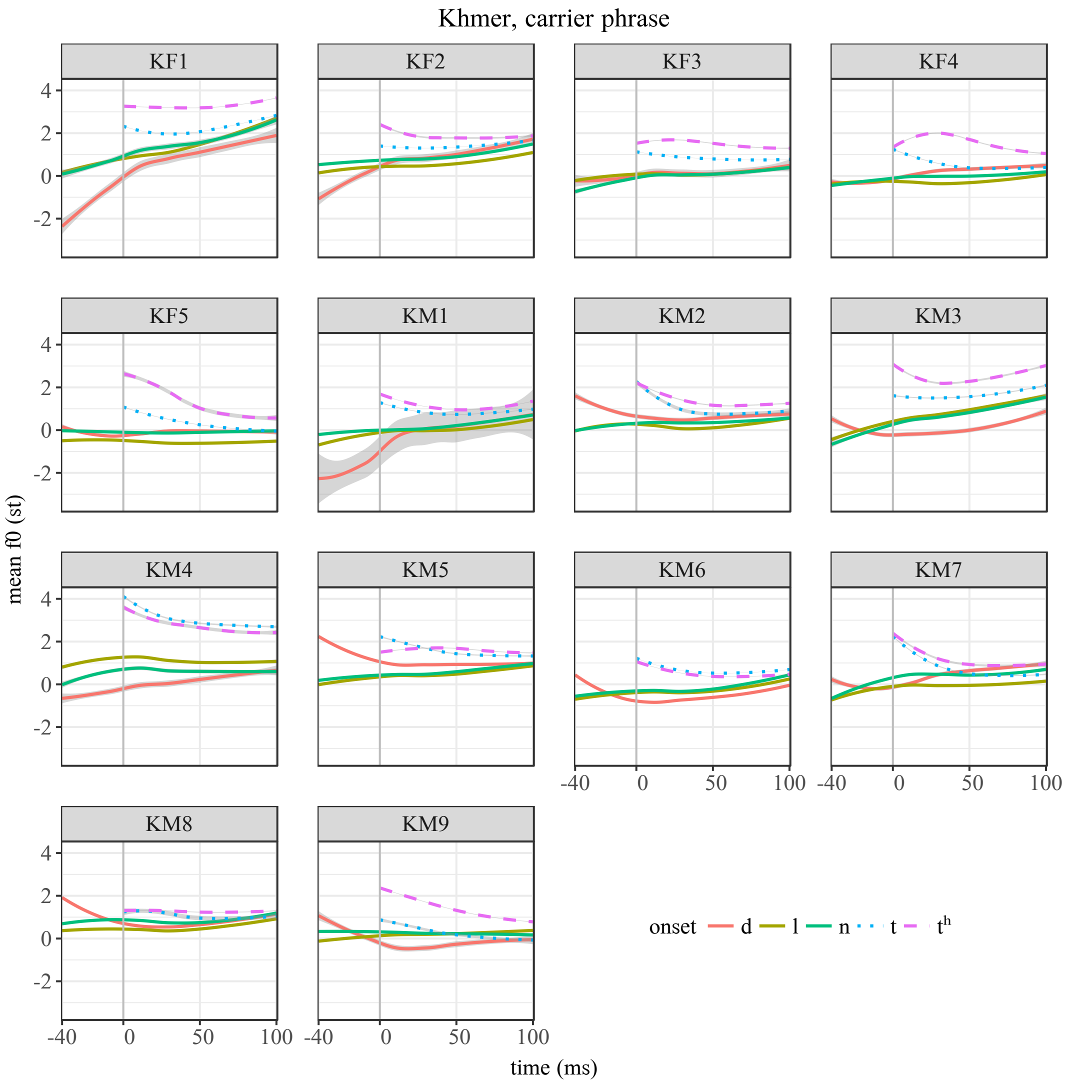

Figure 6: Mean smoothed f0 trajectories (in semitones) for Khmer data by speaker, averaged over repetition and vowel/orthographic register, carrier phrase context. Shading indicates 95% confidence interval.

Thai

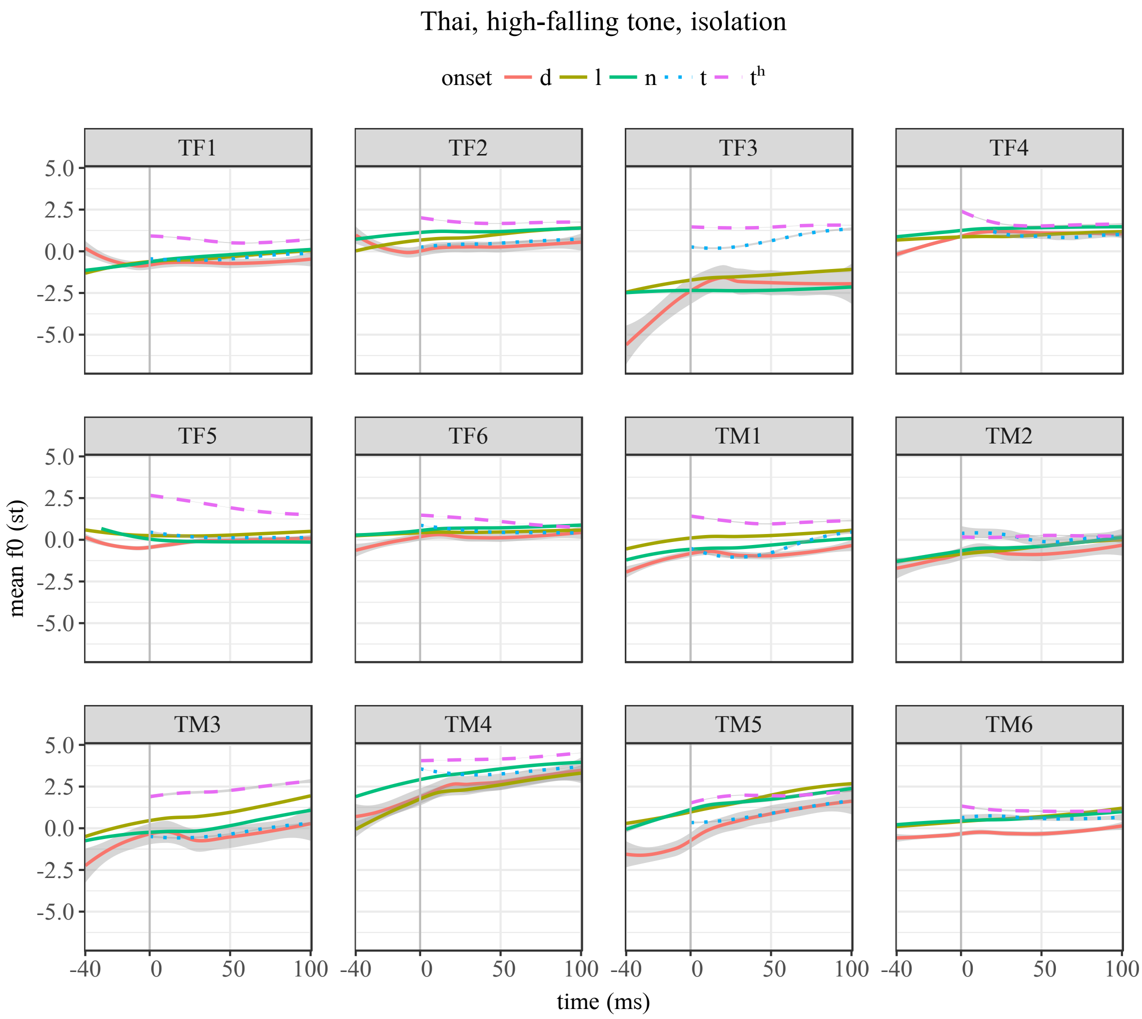

Figure 7: Mean smoothed f0 trajectories (in semitones) for Thai falling tone data by speaker, averaged over repetition, isolation context. Shading indicates 95% confidence interval.

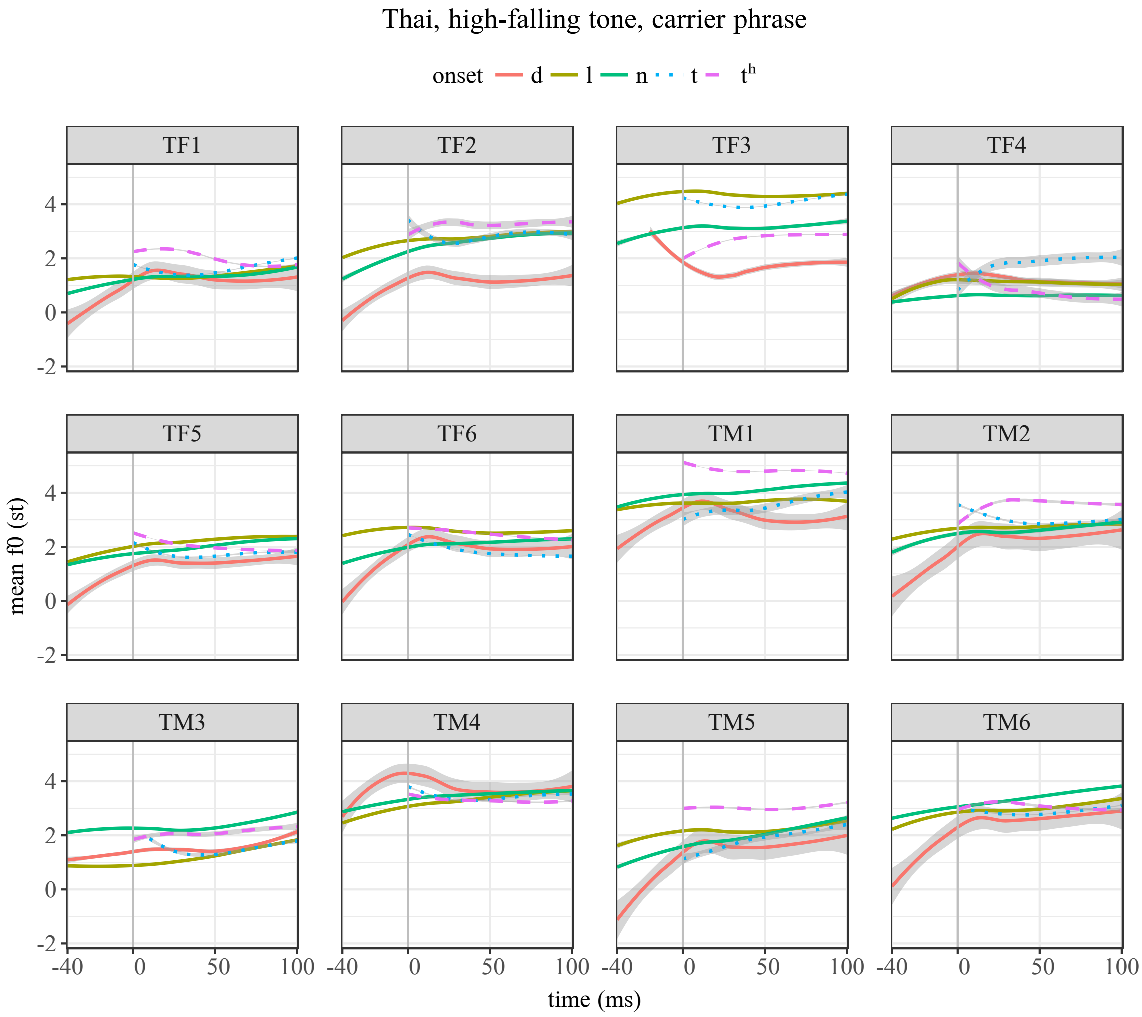

Figure 8: Mean smoothed f0 trajectories (in semitones) for Thai falling tone data by speaker, averaged over repetition, carrier phrase context. Shading indicates 95% confidence interval.

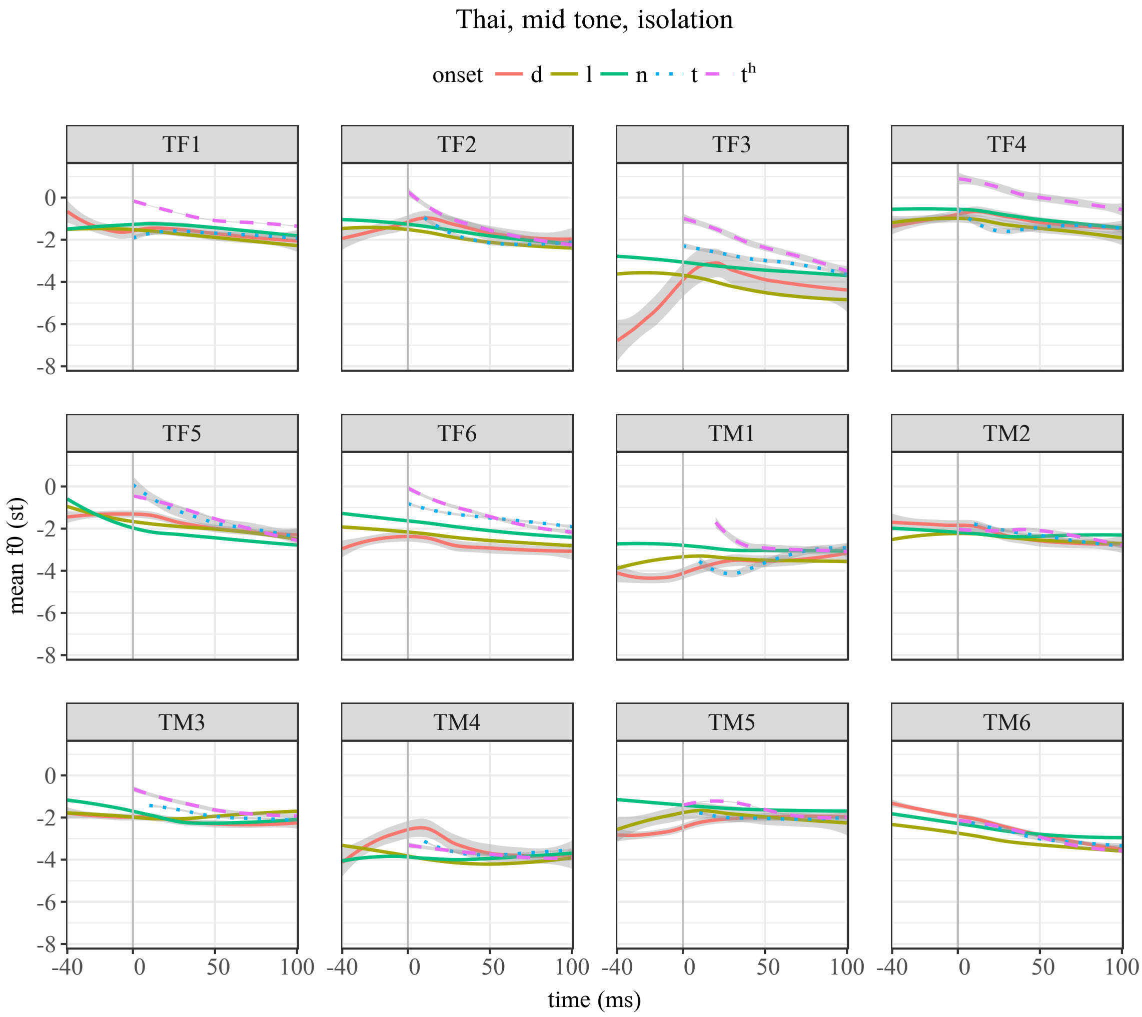

Figure 9: Mean smoothed f0 trajectories (in semitones) for Thai mid tone data by speaker, averaged over repetition, isolation context. Shading indicates 95% confidence interval.

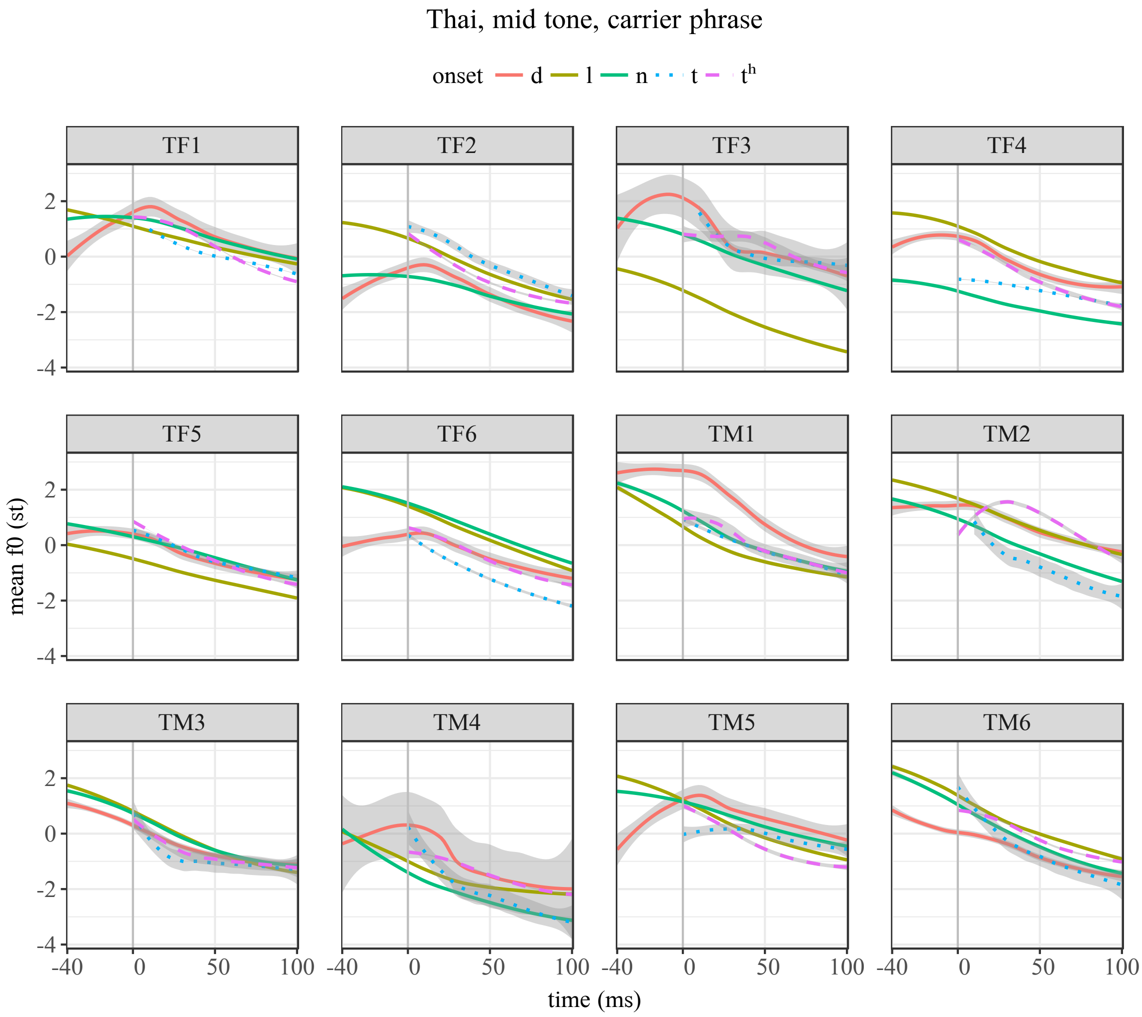

Figure 10: Mean smoothed f0 trajectories (in semitones) for Thai mid tone data by speaker, averaged over repetition, carrier phrase context. Shading indicates 95% confidence interval.

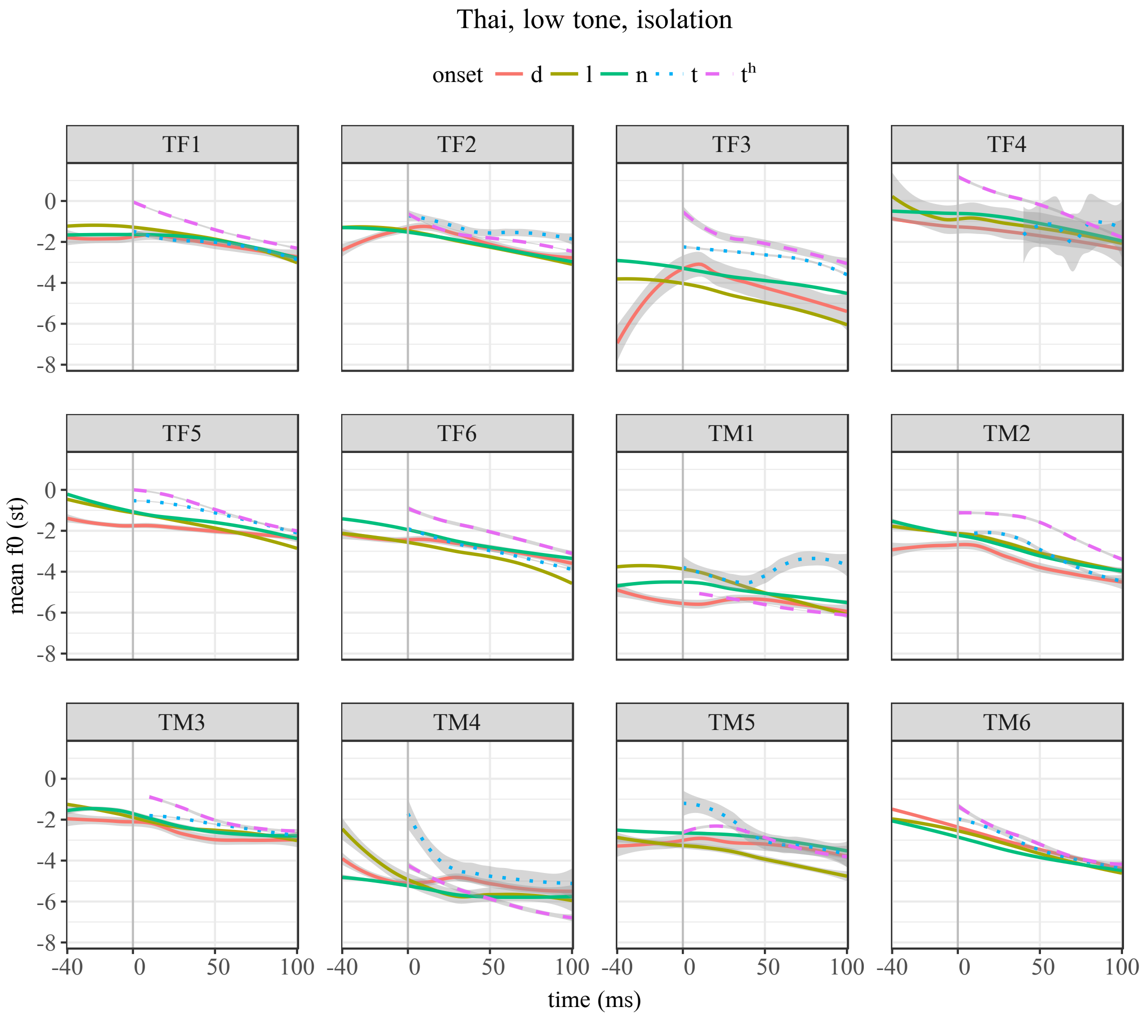

Figure 11: Mean smoothed f0 trajectories (in semitones) for Thai low tone data by speaker, averaged over repetition, isolation context. Shading indicates 95% confidence interval.

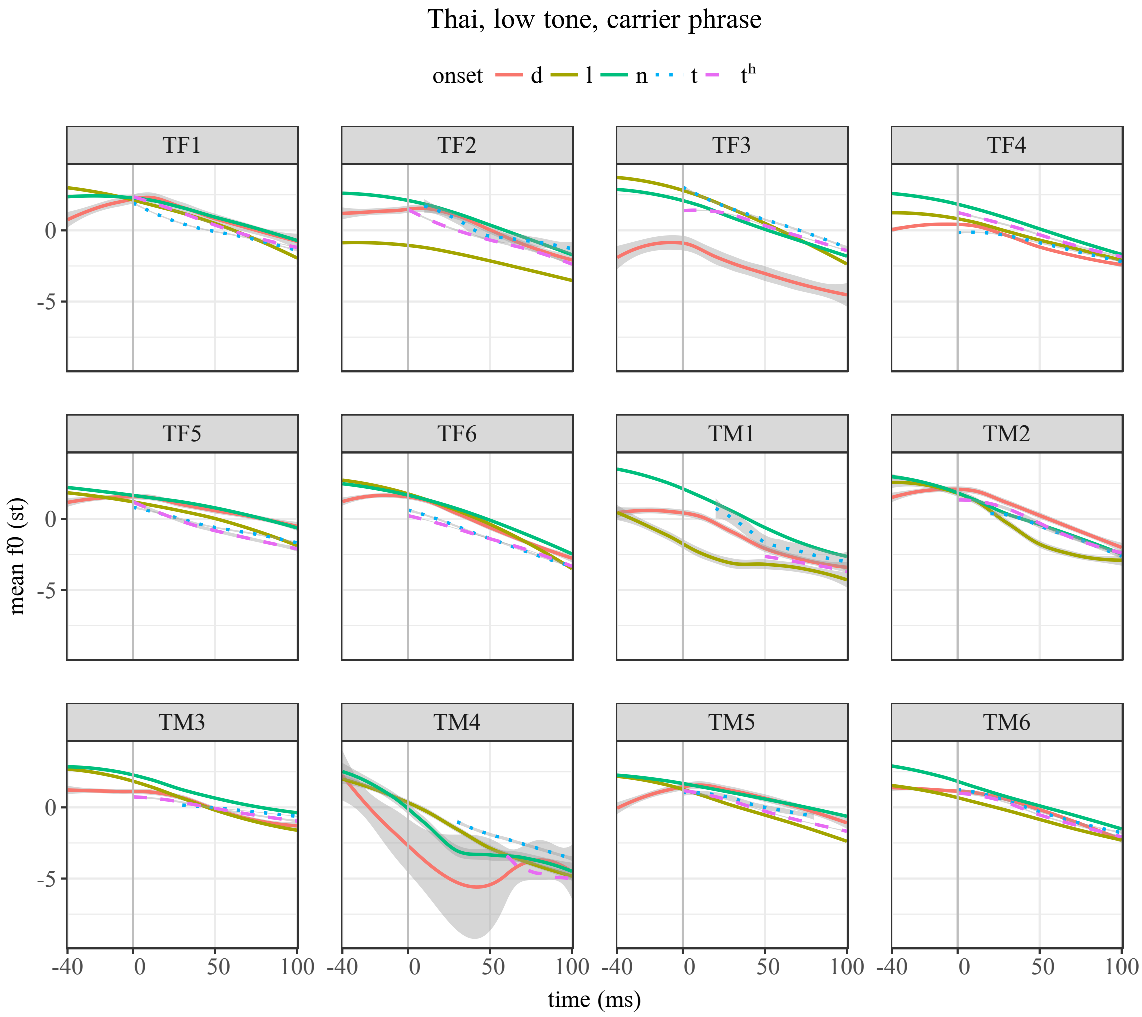

Figure 12: Mean smoothed f0 trajectories (in semitones) for Thai low tone data by speaker, averaged over repetition, carrier phrase context. Shading indicates 95% confidence interval.

Vietnamese

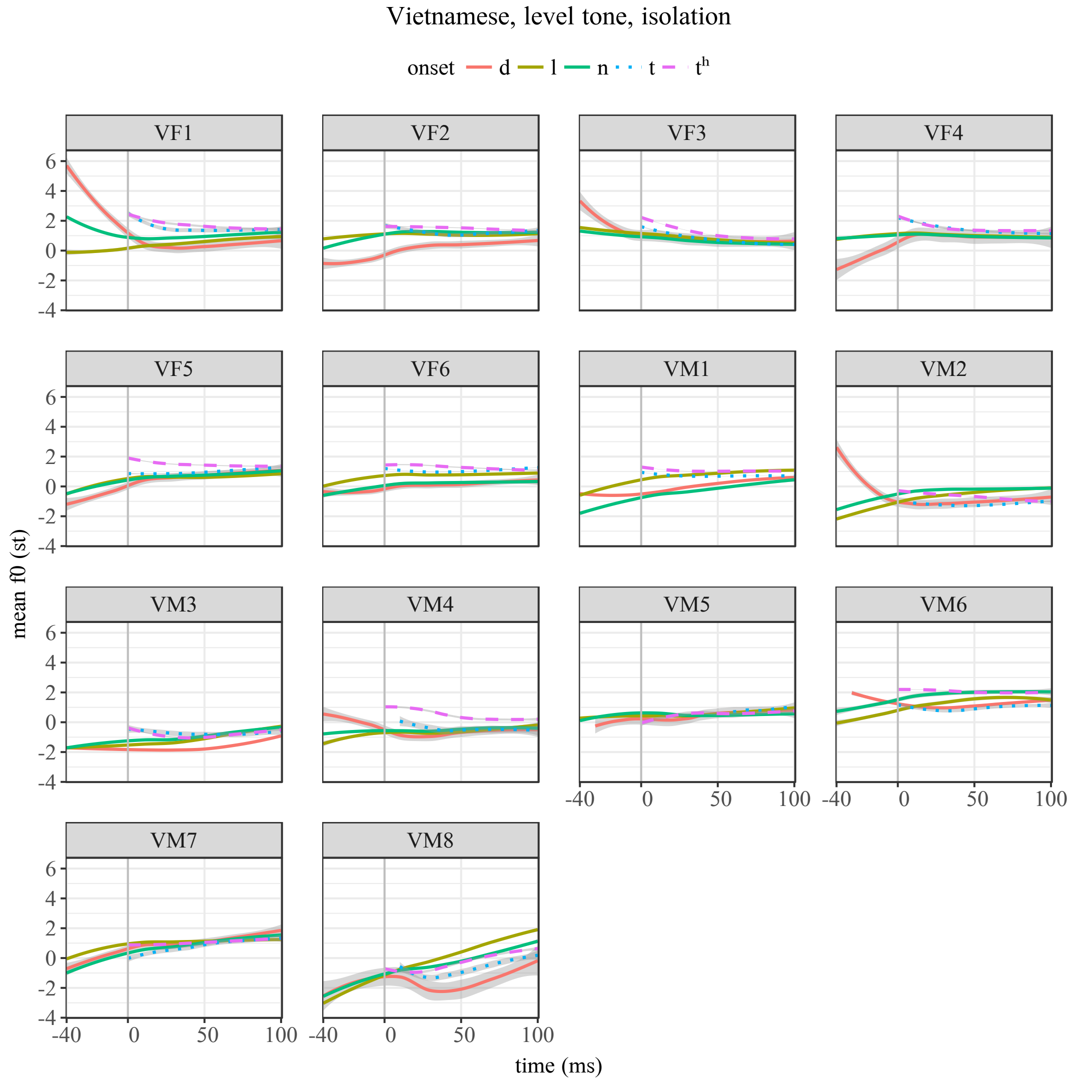

Figure 13: Mean smoothed f0 trajectories (in semitones) for Vietnamese level tone data by speaker, averaged over repetition, isolation context. Shading indicates 95% confidence interval.

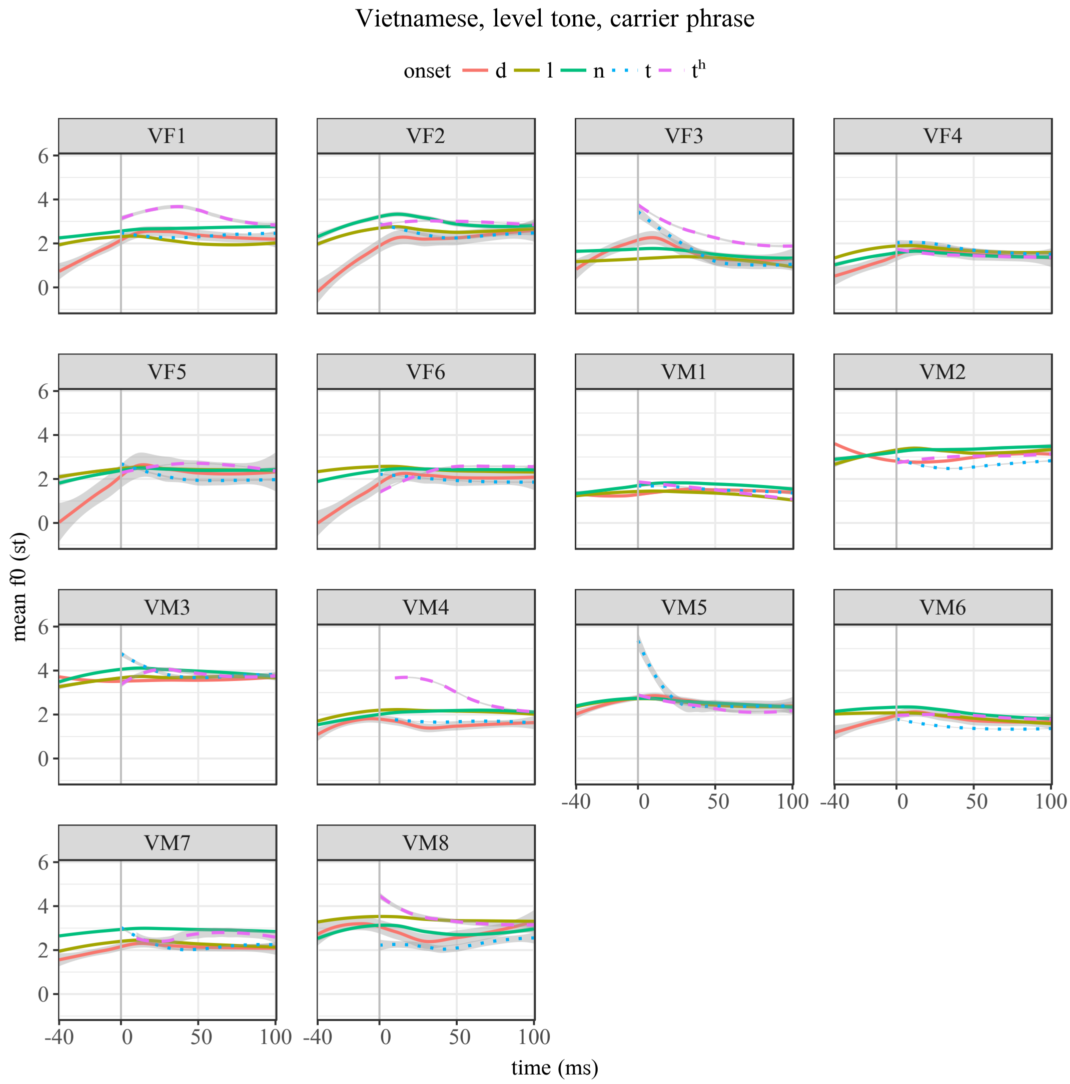

Figure 14: Mean smoothed f0 trajectories (in semitones) for Vietnamese level tone data by speaker, averaged over repetition, carrier phrase context. Shading indicates 95% confidence interval.

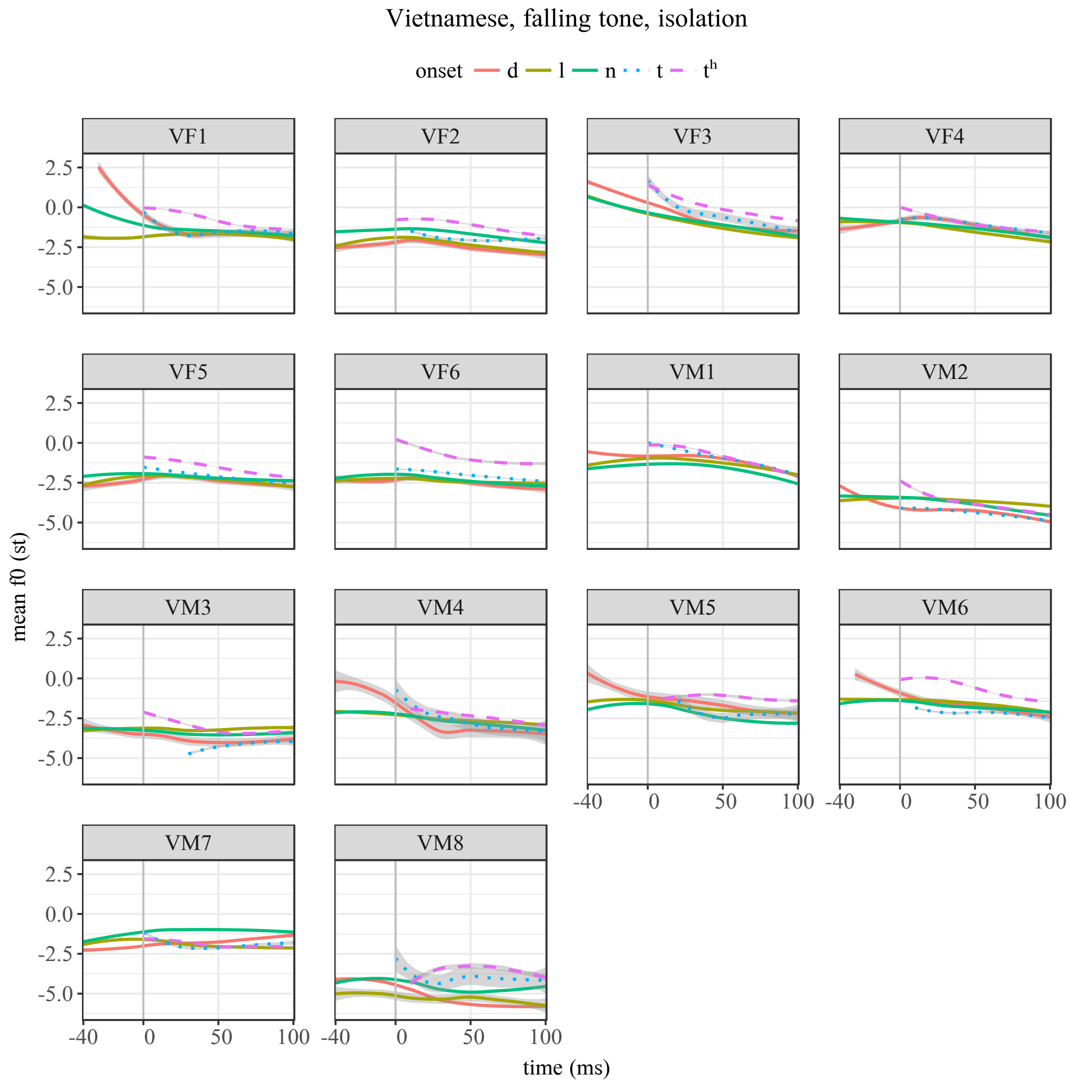

Figure 15: Mean smoothed f0 trajectories (in semitones) for Vietnamese falling tone data by speaker, averaged over repetition, isolation context. Shading indicates 95% confidence interval.

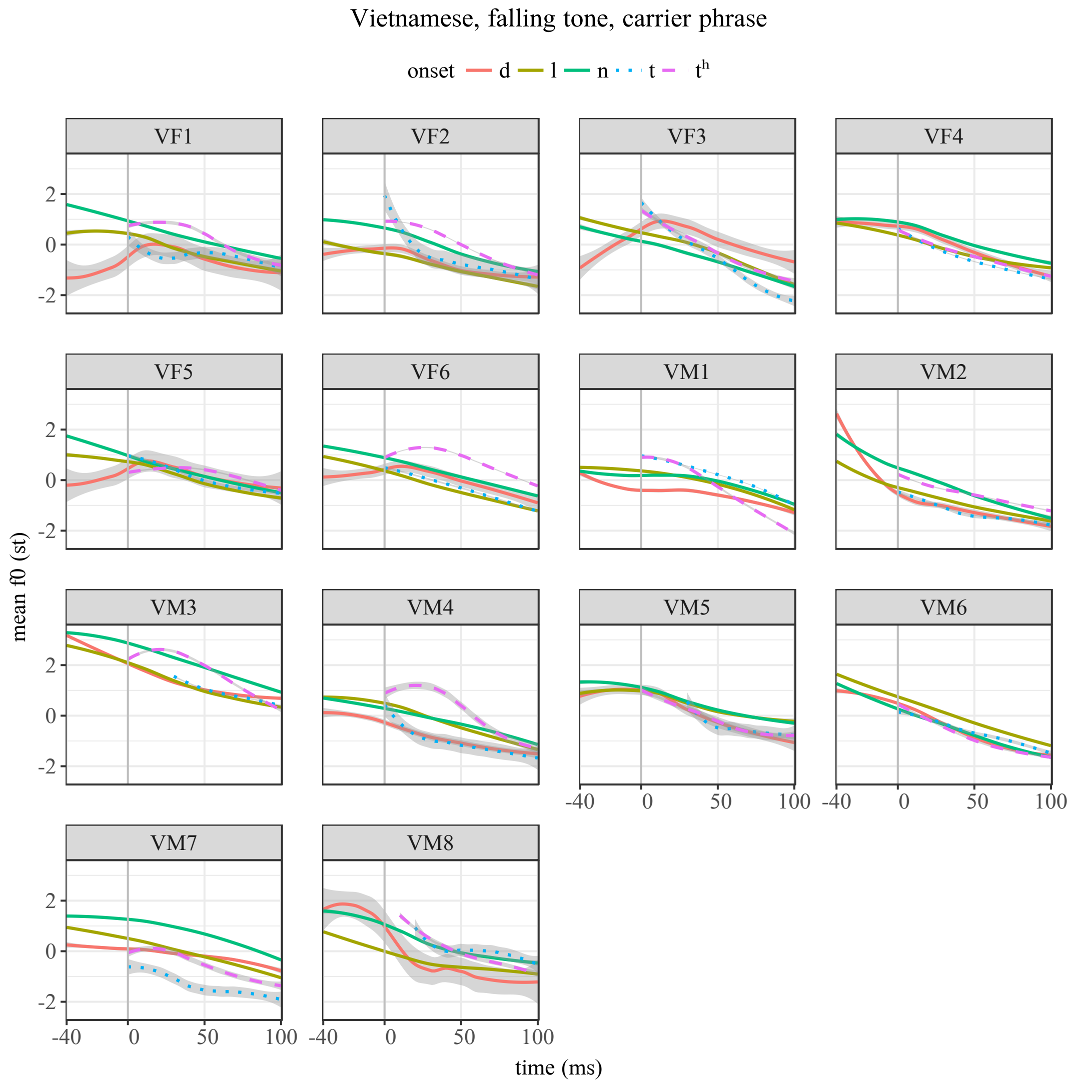

Figure 16: Mean smoothed f0 trajectories (in semitones) for Vietnamese falling tone data by speaker, averaged over repetition, carrier phrase context. Shading indicates 95% confidence interval.

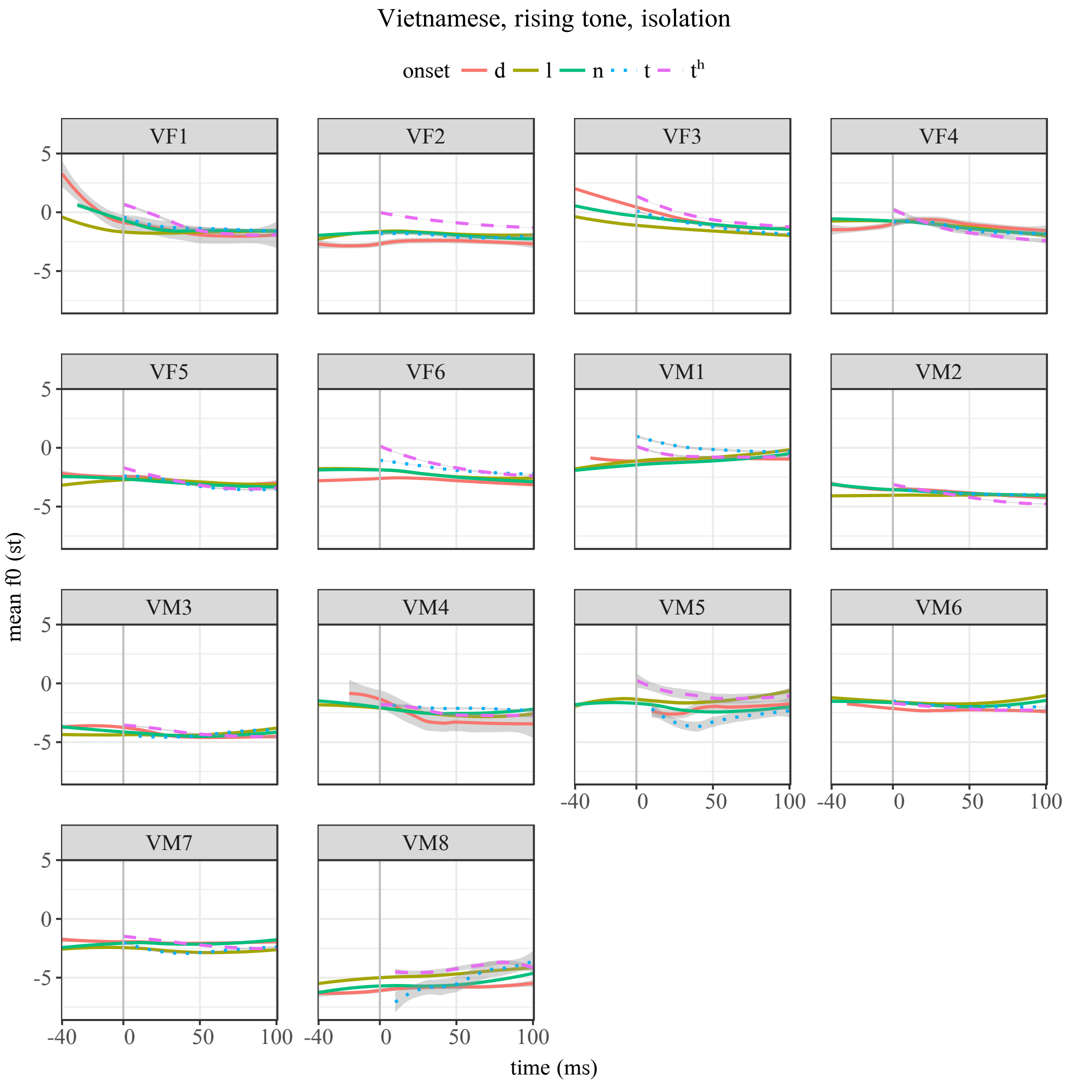

Figure 17: Mean smoothed f0 trajectories (in semitones) for Vietnamese rising tone data by speaker, averaged over repetition, isolation context. Shading indicates 95% confidence interval.

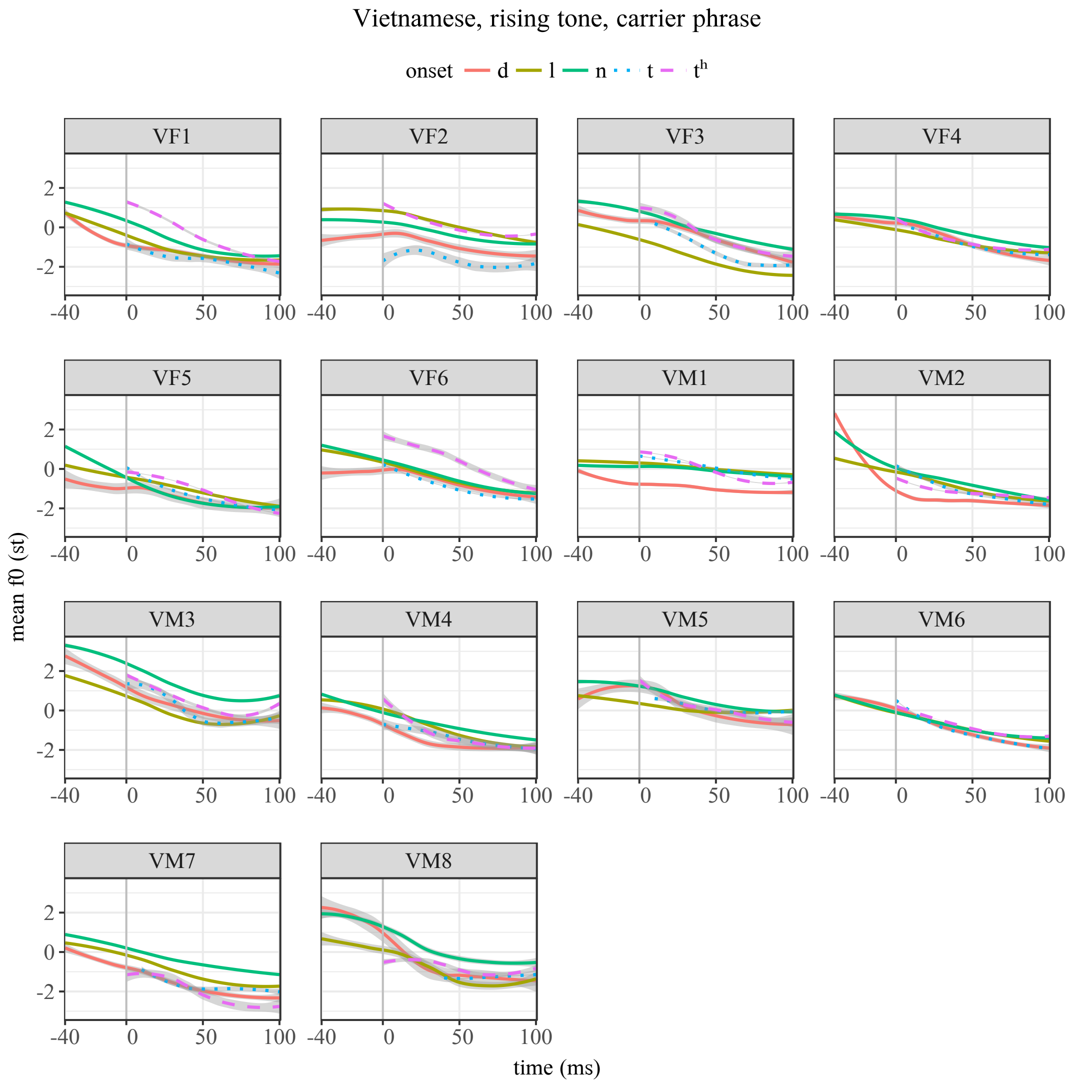

Figure 18: Mean smoothed f0 trajectories (in semitones) for Vietnamese rising tone data by speaker, averaged over repetition, carrier phrase context. Shading indicates 95% confidence interval.

Supplementary materials

Additional plots, post-hoc comparisons, code, and data files for carrying out all analyses in this paper can be found at `http://dx.doi.org/10.7488/ds/2418`.

References

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica*, *34*, 295–303.

Abramson, A. S., & Erickson, D. M. (1978). Diachronic tone splits and voicing shifts in Thai: some perceptual data. *Haskins Laboratories Status Report on Speech Research SR-53*, *2*, 85–96.

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of phonetics*, *26*, 3–25.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics*, *32*, 177–197.

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In T. A. Hall (Ed.), *Distinctive feature theory* (p. 41–70). Berlin: De Gruyter.

Bates, D., Maechler, M., & Bolker, B. (2014). *lme4: Linear mixed-effects models using Eigen and S4.*.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*, 259–284.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*, 97–100.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.2.08, retrieved 19 December 2014 from http://www.praat.org.

Butler, B. A. (2014). *Deconstructring the Southeast Asian sesquisyllable: A gestural account.* Ph.D. thesis Cornell University.

Chen, Y. (2009). Prosody and information structure mapping: Evidence from Shanghai Chinese. *Chinese Journal of Phonetics*, *2*, 123–133.

Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, *39*, 612–625.

Cho, T. (2011). Laboratory phonology. In N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *The Continuum Companion to Phonology* (pp. 343–368). London: Continuum.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.

Clements, G. N., & Osu, S. (2002). Explosives, implosives, and nonexplosives: the linguistic function of air pressure differences in stops. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology 7* (p. 299–350). Mouton de Gruyter.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216.

Cohn, A. C., & Lockwood, K. (1994). A phonetic description of Madurese and its phonological implications. *Working Papers of the Cornell Phonetics Laboratory*, *9*, 67–92.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, *54*, 35–50.

Di Cristo, A., & Hirst, D. J. (1986). Modelling french micromelody: analysis and synthesis. *Phonetica*, *43*, 11–30.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (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.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents.* Berlin: Foris.

Draxler, C., & Jänsch, K. (2004). SpeechRecorder - a universal platform independent multi-channel audio recording software. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)* (p. 559–562). Lisbon.

Emerich, G. H. (2012). *The Vietnamese vowel system*. Ph.D. thesis University of Pennsylvania.

Erickson, D. (1975). Phonetic implications for an historical account of tonogenesis in Thai. In J. G. Harris, & J. R. Chamberlain (Eds.), *Studies in Tai Linguistics in Honor of William J. Gedney* (p. 100–111). Bangkok: Central Institute of English Language.

Erickson, D. M., & Abramson, A. S. (2013). F0, EMG and tonogenesis in Thai. *Journal of Nagoya Gakuin University*, *24*, 1–13.

Ewan, W. G. (1976). *Laryngeal behavior in speech*. Ph.D. thesis University of California, Berkeley.

Filippi, J.-M. (2006). Norme et phonétisme dans le Cambodge moderne. *Udaya (Journal of Khmer Studies)*, *7*, 1–31.

Filippi, J.-M., & Heap, C. V. (2016). *Khmer pronouncing dictionary: Standard Khmer and Phnom Penh dialect*. UNESCO, Éditions KAM.

Fischer-Jørgensen, E. (1968). Voicing, tenseness and aspiration in stop consonants, with special reference to french and danish. *Annual Report of the Institute of Phonetics, University of Copenhagen*, *3*, 63–114.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial english stops. *Journal of Phonetics*, *10*, 177–192.

Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *Journal of the Acoustical Society of America*, *120*, 2884–2895.

Gandour, J. T. (1974). Consonant types and tone in Siamese. *Journal of Phonetics*, *2*, 337–350.

Gao, M. (2008). *Mandarin tones: An articulatory phonology account*. Ph.D. thesis Yale University.

Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics*, *2*, 43–50.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT Quarterly Progress Report*, *101*, 198–212.

Han, M. S., & Weizman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /ph, th, kh/. *Phonetica*, *22*, 112–128.

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

Harris, J. G. (2006). A palatographic study of places of articulation in Thai and some other Southeast Asian languages: Dentals, alveolars, and palatals. In J. E. Harris (Ed.), *Readings in Articulatory Phonetics, Volume 1: Consonants and Phonation Types* (pp. 63–91). Bangkok: Ek Phim Thai Co.

Haudricourt, A.-G. (1954). De l'origine des tones en viêtnamien. *Journal Asiatique*, *242*, 69–82.

Haudricourt, A.-G. (1961). Bipartition et tripartition des systèms de tons dans quelques langues d'Extrême-Orient. *Bulletin de la Société de Linguistique de Paris*, *56*, 163–80.

Henderson, E. J. A. (1952). The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies, University of London*, *14*, 149–174.

Hombert, J.-M. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, *8*, 173.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 77–111). New York: Academic Press.

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

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in f0 control. *Language and Speech*, *42*, 401–411.

Hoole, P. (2006). *Experimental studies of laryngeal articulation*. Habilitation thesis LMU Munich.

Hoole, P., & Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental f0. In N. Clements, & R. Ridouane (Eds.), *Where do phonological features come from?: Cognitive, physical and developmental bases of distinctive speech categories* Language Faculty and Beyond (LFAB): Internal and External Variation in Linguistics (p. 133–171). Amsterdam: John Benjamins.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment on secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *25*, 105–113.

Huffman, F. E. (1970). *Cambodian system of writing and beginning reader*. New Haven: Yale University Press.

Huffman, F. E. (1972). The boundary between the monosyllable and the disyllable in Cambodian. *Lingua*, *29*, 54–66.

Huffman, F. E. (1985). Vowel permutations in Austroasiatic languages. In *Linguistics of the Sino-Tibetan area: The state of the art: Papers presented to Paul K. Benedict for his 71st birthday* (pp. 141–145). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Proc. ICASSP* (pp. 669–672). Montreal.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in germanic. *Phonology*, *12*, 369–396.

Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Annual Report of the Institute of Phonetics, University of Copenhagen*, *9*, 191–221.

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

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

Kanwal, J., & Ritchart, A. (2015). An experimental investigation of tonogenesis in punjabi. In *Proceedings of the 18th International Congress of Phonetic Sciences*. The University of Glasgow.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language*, *82*, 33–63.

Kingston, J. (2007). Segmental influences on f0: Automatic or controlled? In C. Gussenhoven, & T. Riad (Eds.), *Tones and tunes, volume 2: Experimental studies in word and sentence prosody* (p. 171–201). Berlin: Mouton de Gruyter.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Kirby, J. (2011). Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association*, *41*, 381–392.

Kirby, J. (2014a). Acoustic transitions in Khmer word-initial clusters. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production (ISSP)* (p. 234–237).

Kirby, J. (2014b). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*, *43*, 69–85.

Kirby, J. (2018). Praatsauce: Praat-based tools for spectral analysis [computer program]. Retrieved February 18, 2018 from https://github.com/kirbyj/praatsauce.

Kirby, J., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel f0: evidence from "true voicing" languages. *The Journal of the Acoustical Society of America*, *140*, 2400–2411.

Kohler, K. J. (1982). F0 in the production of fortis and lenis plosives. *Phonetica*, *39*, 199–218.

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, *41*, 150–174.

Kohler, K. J. (1985). F0 in the perception of lenis and fortis plosives. *The Journal of the Acoustical Society of America*, *78*, 21–32.

Kuo, Y. C., Xu, Y., & Yip, M. (2007). The phonetics and phonology of apparent cases of iterative tonal change in standard chinese. In C. Gussenhoven, & T. Riad (Eds.), *Tones and tunes, Volume II: Phonetic and Behavioural Studies in Word and Sentence Prosody* (p. 211–237). Mouton de Gruyter.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press.

Ladd, R. (2009). Phonetics in phonology. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *Handbook of Phonological Theory (2nd edition)*. Cambridge: Wiley-Blackwell.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on F0 in Taiwanese. In J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva, S. Morton, & A. Rimrott (Eds.), *Proceedings of the 2nd International Conference on East Asian Linguistics*.

Lea, W. A. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In L. Hyman (Ed.), *Consonant types and tone* (pp. 17–70). Los Angeles: University of Southern California volume 1 of *Southern California Occasional Papers in Linguistics*.

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.1.2.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics*, *12*, 147–55.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*, 3–11.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

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

Luo, Q., Durvasula, K., & Lin, Y.-H. (2016). Inconsistent consonantal effects on f0 in cantonese and mandarin. In *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL-2016)* (p. 52–55).

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), *Consonant types and tone* (pp.

71–95). Los Angeles: University of Southern California.

Michaud, A. (2004). Final consonants and glottalization: new perspectives from Hanoi Vietnamese. *Phonetica*, *61*, 119–146.

Misnadin (2016). *The phonetics and phonology of the three-way laryngeal contrast in Madurese*. Ph.D. thesis University of Edinburgh.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, *109*, 1181–1196.

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory*, *24*, 113–178.

Nagano-Madsen, Y., & Thornell, C. (2012). Acoustic properties of implosives in bantu mpiemo. *FONETIK 2012*, (p. 73).

Nguyễn Văn Lợi, & Edmondson, J. A. (1998). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, *28*, 1–18.

Nihalani, P. (1986). Phonetic implementation of implosives. *Language and Speech*, *29*, 253–262.

Odden, D. (2007). The unnatural tonology of zina kotoko. In T. Riad, & C. Gussenhoven (Eds.), *Tones and tunes volume 1: typological studies in word and sentence prosody* (p. 63–89). Mouton de Gruyter volume 1.

Ohala, J. J. (1973). The physiology of tone. In L. Hyman (Ed.), *Consonant types and tone* (pp. 1–14). USC: SCOPIL.

Ohala, J. J. (1974). A mathematical model of speech aerodynamics. In G. Fant (Ed.), *Proceedings of the Speech Communication Seminar, Stockholm* (pp. 65–72). volume 2.

Ohala, J. J. (1976). A model of speech aerodynamics. *Report of the phonology laboratory (Berkeley)*, *1*, 93–107.

Ohala, J. J. (1978). The production of tone. In V. A. Fromkin (Ed.), *Tone: a linguistic survey* (pp. 5–40). New York: Academic Press.

Phạm, A. H. (2003). *Vietnamese tone: a new analysis*. New York: Routledge.

Potisuk, S., Gandour, J. T., & Harper, M. P. (1994). F0 correlates of stress in Thai. *Linguistics of the Tibeto-Burman Area*, *17*, 1–27.

R Core Team (2014). *R: A language and environment for statistical computing.*. R Foundation for Statistical Computing.

Ridouane, R., Clements, N. G., & Khatiwada, R. (2011). Language-independent bases of distinctive features. In J. Goldsmith, E. Hume, & L. Wetzels (Eds.), *Tones and Features: Phonetic and Phonological Perspectives* (pp. 264–291).

Rose, P. (1996). Aerodynamic involvement in intrinsic f0 perturbations - evidence from Thai-Phake. In P. McCormak, & A. Russell (Eds.), *Proceedings of the 6th Australian International Conference on Speech Science and Technology* (p. 593–598). Australian Speech Science and Technology Association.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.

Shimizu, K. (1994). F0 in phonation types of initial-stops. In *Proceedings of the 5th Australasian International Conference on Speech Science and Technology* (p. 650–655). Perth volume 2.

Sonderegger, M., & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 375—380). Austin, TX: Cognitive Science Society.

Teeranon, P. (2007). The change of standard thai high tone: an acoustic study and a perceptual experiment. *SKASE Journal of Theoretical Linguistics*, *4*, 1–17.

Thompson, L. C. (1965). *A Vietnamese grammar*. Seattle: University of Washington.

Tingsabadh, K. M., & Deeprasert, D. (1997). Tones in standard thai connected speech. In A. S. Abramson (Ed.), *Southeast Asian Linguistics Studies In Honor of Vichin Panupong* (pp. 297—308). Chulalongkorn University Press.

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America*, *85*, 901–906.

Vũ Thanh Phương (1981). *The acoustic and perceptual nature of tone in Vietnamese*. Ph.D. thesis Australian National University Sydney.

Wayland, R. (1997). *Acoustic and perceptual investigation of breathy and clear phonation in Chanthaburi Khmer: Implications for the history of Khmer phonology*. Ph.D. thesis Cornell University.

Wayland, R., & Jongman, A. (2002). Registrogenesis in Khmer: A phonetic account. *Mon-Khmer Studies*, *32*, 101–115.

Wayland, R. P., & Guion, S. G. (2005). Sound changes following the loss of /r/ in Khmer: A new tonogenetic mechanism? *Mon-Khmer Studies*, *35*, 55–82.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica*, *47*, 36–49.

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

Whalen, D. H., Lisker, L., Abramson, A. S., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, *93*, 2152–2159.

Wright, R., & Shryock, A. (1993). The effects of implosives on pitch in SiSwati. *Journal of the International Phonetic Association*, *23*, 16–23.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*, 165–181.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, *111*, 1399–1413.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*, 159–197.

Yan, Q., Lehnert-LeHouillier, H., & McDonough, J. (2011). Articulatory-acoustic interactions in the prosodic strengthening of domain-initial vowels in Mandarin Chinese. In *Proc. ISSP 9* (p. 385–392).

Zee, E. (1980). The effect of aspiration of the f0 of the following vowel in Cantonese. *UCLA Working Papers in Phonetics*, *49*, 90–97.