Article

Language and Speech

Language *and* Speech
1–43

DOI: 10.1177/0023830920939132
journals.sagepub.com/home/las
SAGE

Disentangling the Effects of Position and Utterance-Level Declination on the Production of Complex Tones in Yoloxóchitl Mixtec

Christian DiCanio
Joshua Benn
University at Buffalo, USA

Rey Castillo García
Secretaría de educación pública (SEP), Chilpancingo, Guerrero, Mexico

Abstract
Phrase-final position is cross-linguistically the locus of both processes of phonetic reduction and processes of phonetic enhancement. In tone languages, phrasal position is a conditioning environment for processes of tone sandhi/allotony, though such patterns emerge from local processes of tonal enhancement or reduction. The current article examines the production of tone in Yoloxóchitl Mixtec, an endangered language of Mexico with nine lexical tones and fixed, stem-final stress, across phrasal and utterance positions via three experiments. In the first two experiments, the findings show that speakers lengthen syllables and expand the tonal F0 range in utterance-final position. The effect of this range expansion is high tone raising, low tone lowering, and falling contour lowering. Rising contour tones undergo substantial leveling when produced in a non-utterance-final context, similar to Taiwanese Mandarin. These findings suggest that postural changes in F0 range are *controlled*, intonational effects in tonal languages and not paralinguistic. In the third experiment, we examine utterance-level declination and raising within sentences consisting entirely of level tones. We show that utterance-level F0 changes are independent from local tonal hyperarticulation effects in phrase-final position. Together, the experiments largely support prosodically-conditioned phonetic undershoot as a control mechanism in tone production and demonstrate how tonal complexity may constrain universal tendencies in speech production.

Keywords
Tone, prosody, position, Oto-Manguean, fieldwork

Corresponding author:
Christian DiCanio, University at Buffalo, 609 Baldy Hall, Buffalo, NY 14260, USA.
Email: cdicanio@buffalo.edu

1 Introduction

Decades of research on prosody and intonation have demonstrated the complex ways that words can vary by their position in the utterance (Beckman & Edwards, 1990; Byrd et al., 2006; Byrd & Saltzman, 2003; Fletcher, 2010; Fougeron & Keating, 1997; Katsika et al., 2014; Keating et al., 2004; Ladd, 2008; Turk & Sawusch, 1997; Turk & Shattuck-Hufnagel, 2000). Words, and by extension phonological contrasts, may be hyperarticulated or strengthened at certain prosodic boundaries and hypo-articulated or weakened at others (Fougeron & Keating, 1997). In non-tonal languages, strengthening at prosodic boundaries typically coincides with intonational boundary tones. In tonal languages, which comprise at least 42% of the world's languages (Maddieson, 2013), boundary-related effects which influence F0 must necessarily compete with lexical tone. The result of this competition may be a change in the contour shape or height of the tone.

The current study investigates the interaction between utterance position and lexical tone in Yoloxóchitl Mixtec (YM, henceforth), an Oto-Manguean language spoken in Guerrero, Mexico (see Castillo García, 2007; DiCanio et al., 2014, 2018; Palancar et al., 2016). YM both possesses a large tonal inventory (9 tones) and utilizes tone to mark morphological contrasts. In the first study, 20 distinct tonal melodies on disyllabic words, produced by nine native speakers, were investigated in utterance-final and non-final contexts in a controlled elicitation experiment. In the second study, we evaluated the durational and tonal patterns observed in the elicitation data against a corpus of unscripted, spontaneous speech. In the third study, a set of utterances containing identical tones (i.e., all high (level) tone /4/), were examined to investigate patterns of utterance-level declination in declarative utterances. These results are examined in relation to those of the first two studies with the goal of disambiguating whether pre-pausal tonal effects are related directly to utterance-level patterns of declination or if the two are independent processes.

2 Background

Words in pre-pausal (or phrase-final) position often differ from those produced in non-final position. In terms of contrast, final position has been argued to be both a position of phonological neutralization, contrast maintenence, and contrast enhancement (see Barnes, 2006; Zhang, 2001). In terms of speech production, both processes of local weakening and local strengthening have been observed. For instance, phrase-final syllables are frequently longer than non-final syllables (e.g., in English (Byrd, 2000; Byrd et al., 2006; Cho, 2006; Edwards et al., 1991; Rakerd et al., 1987), Finnish (Nakai et al., 2009), Kipare (Herman, 1996), and Taiwanese (Peng, 1997)). This longer durational window ostensibly permits the hyperarticulation of phonetic targets and prevents articulatory undershoot (Lindblom, 1990; Parrell, 2014); and articulatory gestures have slower velocity (Krivokapić & Byrd, 2012). However, phrase-final position may also be the locus of glottalization (Huffman, 2005), devoicing (Wagner, 2002), and both a gradual decay in intensity and F0 (Gussenhoven, 2004; Ladd, 2008). These processes may obscure many phonological contrasts and, in particular, tonal contrasts.[1]

In a lexical tone language, tonal contrasts may be affected by either enhancing or neutralizing phonetic processes. There are three specific mechanisms that are traditionally responsible for changes in F0 contours in phrase-final position, outlined below:

1. **Phrase-final F0 shift**: local patterns of F0 lowering or raising will influence the tonal shape of the word(s) *only* in pre-pausal position.

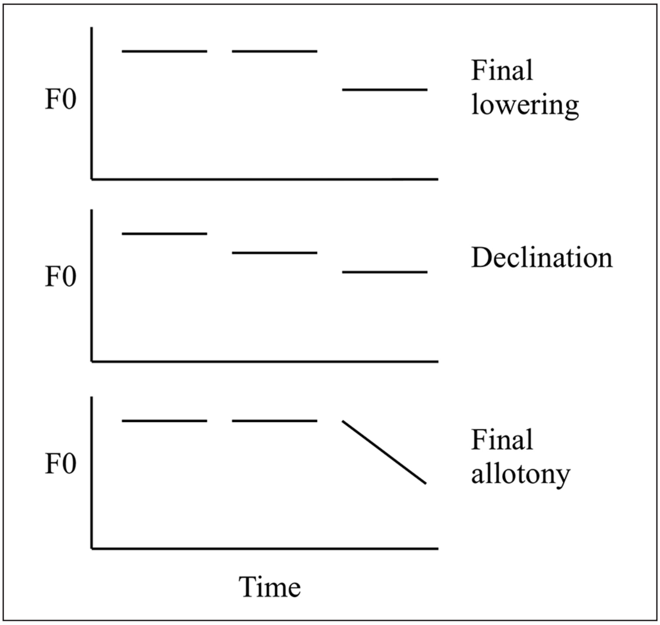

Figure 1. Hypothetical phonetic realizations of sequences of three high level tones in a phrase-final position.

2. **Positional allotony**: tones produced in pre-pausal position categorically differ in tonal shape from those produced elsewhere. This categorical difference is not immediately reducible to production factors.
3. **Utterance-level F0 shift**: global patterns of declination, downstep, or upstep across a phrase will influence tonal shape and/or height.

Each of these processes has the potential to enhance or neutralize lexical tonal contrasts. For instance, if final lowering were restricted to a lower tone, it would ostensibly enhance a contrast with a high tone in a given language. Moreover, though they are distinct, each process may also produce similar surface phonetic patterns on a phrase-final syllable. Consider one common acoustic measure of tone production—midpoint F0—in relation to the tonal sequences depicted in Figure 1.

One may observe similar, lowered values for the phrase-final H tone in Figure 1 but be unable to disambiguate whether the cause is a non-local intonational process, a local intonational process, or a phrase-level phonological alternation. Among those studies which have distinguished among these types of processes, considerable variation is found across languages, across tonal contrasts, and across individual speakers. The following subsections provide an overview of research on these three processes with respect to lexical tone languages. We limit the background discussion here mostly to research on the prosody of lexical tone languages. Though work on the prosody of non-tonal languages may be relevant, intonational pitch accents/tunes show limited competition with word-prosodic contrasts in such languages. Since we focus specifically on the issues of lexical tonal contrast enhancement and neutralization here, the literature on non-tonal languages is less relevant.

While we more generally seek to summarize the existing literature on the tone–intonation interface, note that a majority of the research on the interaction between lexical tone and intonation has focused on Niger-Congo languages (see Downing & Rialland, 2017). There is little work to date examining the interaction between tone and intonation in many language families with a large number of tone languages (Otomanguean, Tibeto-Burman, Tai-Kadai, Austroasiatic, Khoisan, etc.).

2.1 Local phrase-final F0 shift

Phrase-final and utterance-final position are frequently the locus of processes modifying F0 or F0 range. Research on lexical tone languages has not always distinguished phrase-final tonal effects from those spanning the utterance domain. We focus on the particular studies where researchers have done so. Within intonational phonology, general processes of F0 shift are represented via edge/boundary tones (L%, H%, HL%, etc.) (Gussenhoven, 2004; Ladd, 2008). If an intonational boundary tone occurs in a lexical tone language, it may either replace a lexical tone or cause the tone to shift in a particular direction, the latter being more common.

In Taiwanese, utterance-final tones in declarative utterances were lowered relative to non-utterance final tones, though the magnitude of this effect was smaller with the high level /55/ tone than with the other tones (/33, 24, 21, 51/) (Peng, 1997). In Tswana, final lowering is restricted to utterances where the preceding two tones are both L, suggesting that it is an enhancement feature for low tones (Zerbian, 2017). The observed asymmetry in these languages in the application of phrase-final lowering would seem to have the net effect of enhancing tonal distinctions in utterance-final position.

This tendency toward tonal contrast enhancement is not universal though. In several tone languages, all tones are lowered in phrase-final position in declarative utterances (e.g., in Kipare (Herman, 1996), Moro (Chung et al., 2016; Rose & Piccinini, 2017), Embosi (Rialland & Embanga Aborobongui, 2017), Chichewa (Downing, 2017; Myers, 1996), Tumbuka (Downing, 2017), Chimiini (Kisseberth, 2017), and Shingazidja (Patin, 2017)). In Mambila, a pattern of utterance-final lowering was consistently observed across speakers for high level tone /4/, variably for tone levels /3/ and /2/, and not at all for the lowest level tone /1/ (Connell, 2017). Though final lowering need not, a priori, result in obscuring tonal contrasts in phrase-final position, the phonetic patterns observed in the research on these languages almost universally depicts a shrinking of the tonal space here, at least in absolute terms. In Akan, final lowering completely neutralizes the distinction between H and L tones in declaratives (Kügler, 2017, pp. 101–103).

Note that final lowering is not limited to declarative utterances in lexical tone languages; it also occurs in polar questions in Moro (Rose & Piccinini, 2017) and Akan (Kügler, 2017). Polar questions may occur with a different boundary tone than that found in declarative utterances in lexical tone languages. In both Chichewa and Tumbuka, polar questions are marked with a phrase-final HL% contour (Downing, 2017). In polar questions in Shingazidja, an intonational LH* tone is attached to the phrase-final penult (often the locus of lengthening processes in many Bantu languages) and this is followed by a L% boundary tone. The net effect of final lowering in these languages (at least in declarative utterances) is to partially neutralize tonal distinctions in utterance-final position. In certain cases, such neutralization is complete. In Mesquital Otomi, the contrast between high and low tones is obscured by a phrase-final raising intonation (Sinclair & Pike, 1948, pp. 91, 96).

There are several languages where no final F0 shift is observed at all. For instance, final lowering is unattested in Yoruba, though a parallel process of *initial H tone raising* applies (Laniran & Clements, 2003). In Bàsàá, neither declarative nor interrogative utterances are produced with any boundary tone (Makasso et al., 2017). The overview here suggests that phrase-final lowering is not universal and may be restricted to certain tones in the language's inventory. Moreover, the number of tonal registers does not specifically predict whether phrase-final processes will be neutralizing or non-neutralizing. Certain languages with larger tonal inventories (like Mambila) may show a tendency towards neutralizing patterns of final lowering while others possessing smaller tonal inventories (like Tswana) may show a pattern towards maintaining tonal contrast in phrase-final position. Distinct mappings between prosody and tone are found for languages with even larger tonal inventories, like Cantonese (Sybesma & Li, 2007). The inventory size of the lexical tone system might be independent from the set of interactions tones may have with prosodic structure.

2.2 Utterance-level F0 shift

There are two different types of processes that span phrases/utterances and which can influence the production of a lexical tone: downstep and declination. We adopt the definition of *downstep* given in Connell (2017, 133):

> *Downstep* refers to the lowering of a High tone following another High tone with the effect that a new ceiling is established for subsequent High tones within a specifiable domain. Downstep is automatic when a surface Low tone triggers the lowering, and non-automatic when there is no surface trigger; in such cases it is often attributable to a floating L tone.

A more general term, *tonal terracing*, subsumes both processes of downstep and processes of upstep, whereby a high tone that follows another high tone is raised and a new ceiling is established for subsequent high tones. Upstep is rarer than downstep, but it is attested in Krachi (Niger-Congo) (Snider, 1990), Babanki (Niger-Congo) (Akumbu, 2015), and Acatlán Mixtec, a Northern Baja Mixtec language (Otomanguean: Mixtecan) (Pike & Wistrand, 1974).[2] Researchers typically represent tonal terracing processes autosegmentally with either a register tier or with floating tones (Hyman, 2007; Snider, 1990). In other words, it is represented with a discrete phonological specification. Declination, on the other hand, is normally considered to result from extrinsic scaling factors which are independent from the phonological specification of individual tones (Ladd, 2008) and it may reflect universal trends in speech production (see Gussenhoven, 2004, and his discussion of the *production code*). It is, par excellence, a phonetic process.

Declination across declarative utterances and rising (or high) F0 in polar questions have each been considered to be universal patterns or tendencies in human languages (Ohala, 1984, 1994; Gussenhoven, 2004). Ladd (2008, pp. 80–84) discusses several problems with this perspective, including the possibility that it may be based on "a fairly Eurocentric sample of languages" (p. 81). Another potential problem with declination as a universal process is that a vast majority of the world's tone languages (which themselves make up at least 42% of the world's languages) are under-described or undescribed. Marking lexical or morphological tone necesarily introduces constraints on the degree to which intonation may be used, but we know very little about how intonational patterns function in many of these languages.[3]

Similar to the discussion of local, phrase-final F0 shift above, one can examine patterns of declination or F0 rise in terms of how they may impact lexical tone contrasts. In Yoruba, declination occurs most consistently across a sequence of all L tones, but for some speakers, a sequence of H tones also undergoes declination (Laniran & Clements, 2003). Declination does not occur in sequences of high tones in Mandarin (Xu, 1999) or Taiwanese (Peng, 1997). The result of this restriction would be a tendency towards tonal contrast maintenance in phrase-final position. In Mambila, declination does not occur on sequences of low tone /1/, but a sequence of high tone /4/ does undergo declination across the utterance (Connell, 2017). Intermediate level tones (/2/, /3/) underwent variable declination. As above, the result of this restriction would be a tendency towards tonal contrast neutralization in phrase-final position. In certain tone languages, declination may be entirely absent (e.g., in Embosi (Rialland & Embanga Aborobongui, 2017) or Choguita Rarámuri (Garellek et al., 2015)). Variation across languages and speakers is typical in the implementation of declination in lexical tone languages.

Another potential source for variation is the time-course of F0 lowering. Speakers of tonal languages may gradually lower F0 across the utterance, producing what is known as a "soft-landing approach" or they may lower F0 more abruptly, producing a "hard-landing approach" (Liberman & Pierrehumbert, 1984; Laniran & Clements, 2003). These distinct approaches, produced by different speakers, suggest that a single model for F0 declination may not capture individual strategies for F0 control with much accuracy. Rather, what is needed is a way to model individual patterns of final lowering distinctly from global trends in declination. In addition, the scaling factor for F0 has been shown to vary with constituent length more broadly—in longer phrases initial F0 begins higher than it does in shorter phrases, both in Mandarin (Shih, 2000) and Wenzhou Chinese (Scholz & Chen, 2014b). We explore this type of modelling as a function of sentence length in Section 5 where we examine patterns of declination in tone-controlled utterances.

2.3 Final allotony and lengthening

Tonal allophony, or allotony, is well-attested across different language families.[4] One of the most well-studied examples is Mandarin half-third sandhi whereby tone /213/ is realized as tone /21/ when it precedes another tone (Zhang & Lai, 2010).[5] In a typological survey of contour tone restrictions, Zhang (2001, 2004) found that 47 out of 187 languages demonstrated a preference for contour tones to surface only in word-final syllables. Though word-final and phrase-final positions are not identical, both are the locus of patterns of final lengthening cross-linguistically and the former typically arises via a generalization from the latter (Padgett & Myers, 2014). The general observation in domain-final positions is that tones which involve more complex dynamics in F0 movement will be faithfully produced but they will be simplified in non-domain-final positions. Such restrictions stem from differences in phonetic duration across an utterance; more complex contour tones are generally restricted to positions of greater duration (Zhang, 2004).

Many of the attested tonal patterns in utterance or phrase-final position reflect general articulatory constraints which co-occur with processes of durational shortening and lengthening. Phrase-final lengthening is a robust process across a variety of languages (in English (Beckman & Edwards, 1990; Byrd & Saltzman, 2003; Cambier-Langeveld & Turk, 1999; Rakerd et al., 1987), Norwegian (Lunden, 2006), Finnish (Nakai et al., 2009), and Vietnamese (Brunelle, 2017)). There are three possible phonetic mechanisms which may influence tones in non-utterance-final position: truncation, compression, and simplification (see Ladd, 2008, pp. 180–183) and Grønnum,1991 for a discussion of such processes in intonational systems). Given a shorter durational window, a lexical tone with a complex F0 trajectory may be truncated and lose a particular F0 target. For instance, in Thai, rising tones are realized without a rise in F0 in non-phrase final position but with a rise in phrase-final position (Morén & Zsiga, 2006). This pattern is non-neutralizing since the initial trajectory of a rising tone is distinct from that of either the level tones or the falling tone. However, like Mandarin third tone sandhi, it is an example of truncating tonal reduction.

Patterns of tonal compression also frequently involve an adjustment of F0 range whereby larger excursions are reduced to smaller ones. In Mixtepec Mixtec, Pike and Ibach (1978, p. 275) mention that the low tone is realized with a greater downglide in phrase-final position than in non-final position. A similar pattern is discussed in Huajuapan and Ayutla Mixtec (Pike & Cowan, 1967; Pankratz & Pike, 1967, p. 6). However, other types of changes are not so easily reducible to phonetic factors. For instance, high tones are realized as falling tones in phrase-final position in Diuxi Mixtec (Pike & Oram, 1976).[6]

Regarding patterns of tonal simplification, in fast speech in Taiwan Mandarin, the trajectory of rising tone (/35/) completely loses its positive slope and is produced as a level or falling contour (Cheng & Xu, 2015). In certain scenarios, processes of tonal compression may historically lead towards tonal simplification or contour leveling. In Shaoxing Wu, contour tones in the initial syllable of disyllabic compounds are simplified to level tones with intermediate tonal values, (i.e., /52 –> 33/, /31 –> 22/) (Zhang, 2006).[7] An extreme case is found in Zhangzhou Southern Min, where each of the eight tones have distinct allotones in phrase-final position when compared with non-phrase-final position (Huang, 2018).

The notion of just what qualifies as allotony is vague, however. Researchers have, by and large, assumed that either cases of contextual tonal neutralization or changes to F0 direction/shape qualify as allotony much in the same way as allophonic processes for vowels are described. Yet, subtle changes to F0 level or slope for tones due to paralinguistic or rate-induced factors have not typically been described as allotonic. For the purposes of this article, we define tonal allotony as abrupt changes in overall F0 shape (level vs. contour, concave vs. convex) or changes which seem to neutralize two tones. We do not attempt to resolve the status of the more subtle cases, though we refer readers to a recent treatment of the issue in Sun (2017).

2.4 Interim summary

The purpose of the current article is to determine the precise mechanisms which influence tone production in phrase-final and non-phrase final positions. Experiments 1 and 2 investigate local, phrase-final F0 shift and final allotony. The literature on phrase-final F0 shift suggests that F0 lowering occurs in declarative utterances, but may be specifically limited to the highest or lowest tones in a given inventory. Missing from the existing literature is a consideration of how different types of contour tones behave in phrase-final and non-phrase final position. This is explicitly addressed in the first two experiments.

The literature on utterance-level F0 shift suggests that patterns of declination may be limited to tones occurring at the ceiling or floor of the F0 range. The literature on final allotony and lengthening suggest that contour tones may be either truncated or compressed in non-phrase-final position (or in locations of shorter duration). We assess each of these points in relation to the tonal system of YM, below, and provide a set of language-specific, testable hypotheses. Experiment 3 attempts to disentangle local F0 shifts from more global patterns of declination.

2.5 Background: Yoloxóchitl Mixtec

YM is an Oto-Manguean language spoken in Guerrero, Mexico by approximately 4,000 people and first described in Castillo García (2007). "Mixtec" is an ethnolinguistic term which refers to a group of at least 12 mutually-unintelligible dialect clusters and which has approximately 2,000 years of internal diversification (Josserand, 1983). YM belongs to the Guerrero Mixtec cluster, which comprises four distinct dialects. All Oto-Manguean languages are tonal (see DiCanio & Bennett, Forthcoming) and YM possesses a large tonal inventory consisting of nine tonemes: /4, 3, 2, 1, 42, 32, 13, 14, 24/ (DiCanio et al., 2014, 2018).[8] All content roots in YM are minimally bimoraic, consisting of either two syllables with short vowels, CVCV, or one syllable with a long vowel, CVV. The TBU is the mora in YM and contour tones may freely occur on both moras in disyllabic or monosyllabic roots, for example, /k^{w}i^{14}i^{14}/ "is not peeling" versus /n^{d}a^{14}ta^{14}/ "to not split up" (DiCanio et al., 2014). Tone has a high functional load in YM—in addition to its lexical function, it is used to mark verbal aspect, negation, and person.

Glottalization is also contrastive on YM roots, though it is orthogonal to the tonal contrast. For instance, compare /n^{d}o^{1}ʔo^{4}/ “basket” versus /n^{d}o^{1}o^{4}/ “sugarcane”; or /na^{3}ʔma^{3}/ “thick” versus /na^{3}ma^{3}/ “to change.” Full glottal closure is observed in citation forms or in careful speech in such words, but the more typical realization of glottalization is as creaky phonation overlapping adjacent vowels or consonants, similar to that found in other Mixtecan languages (DiCanio, 2012; Gerfen & Baker, 2005). All syllables in YM are open and there is evidence that root-final syllables carry fixed stress (see DiCanio et al. (2018)). Consider first that tones in YM are asymmetrically distributed on roots. Whereas all tones may surface on the final mora of the root, only tones /4, 3, 1, 13, 14/ may surface on the initial mora.[9] Moreover, root-final syllables are also the locus of a greater number of vowel contrasts, such as the contrast between oral and nasal vowels and between mid and high back vowels. Finally, final syllables are also typically lengthened relative to non-final syllables (DiCanio et al., 2018, 2019). Together, these phenomena suggest that root-final syllables are stressed in YM.

YM is intonationally sparse as it does not possess pitch accents. Existing research on the tonal system (Castillo García, 2007; DiCanio et al., 2014, 2018) coupled with impressions from several documentation projects (see Amith & Castillo García, n.d.) have not shown evidence for boundary tones. Tones produced under contrastive focus in YM undergo F0 raising relative to narrow and sentential focus (DiCanio et al., 2018). Words produced under focus are also lengthened. DiCanio et al. (2018) argue that the F0 patterns associated with focus derive from focus-induced lengthening and not from the presence of a specific prosodic boundary. Yet, one of the ways in which lexical tone may interact with prosodic boundaries is via processes of allotony or phrase-final F0 adjustments (described above). Given the number of contrastive tones in YM, one predicts relatively little effect of phrasal position on tone production, as is argued for other 4-level tone systems like Mambila (Connell, 2017). However, this prediction should be carefully tested.

2.6 Hypotheses

Four specific hypotheses pertaining to the tone-prosody interface are evaluated in the current study. First, we anticipate some degree of utterance-final lengthening since it has been found for most of the languages where it has been investigated (Fletcher, 2010). This degree of lengthening might be further enhanced in utterance-final position since YM has fixed, final stress. Thus, our first hypothesis is that final syllables will have greater duration than non-final syllables, but this increased duration will be exaggerated in phrase-final position. Our second hypothesis is that tones whose offset approaches the upper limits of the tonal range (/4, 14, 24/) or the lower limits of the tonal range (/1/) will be more likely to undergo changes in phrase-final position than tones which do not (/3, 2, 32/), as per the findings in the literature. We anticipate significant effects of utterance position for tones whose offsets are at the limits of the tonal range and few or no effects for tones whose offsets are not at the limits of the tonal range. Our third hypothesis is that contour tones, especially rising tones, are more likely to undergo patterns of time-dependent tonal compression than level tones. Recall that F0 rises require more time than level or falling trajectories (Sundberg, 1979; Xu & Sun, 2002). Hypotheses (2) and (3) reflect how the mechanism of phrase-final F0 shift, discussed above, is constrained by the tonal inventory. Our fourth hypothesis relates to the mechanism of utterance-level F0 shift—we hypothesize that utterance-level patterns of F0 movement will affect tonal sequences at the limits of the tonal range (/1, 4/) moreso than tonal sequences in the middle of the tonal range, as per the findings in the literature.

Experiment 1 examines how tones are produced in phrase-medial and utterance-final position and evaluates hypotheses (1)–(3). Experiment 2 re-examines the results from Experiment 1 using corpus data. Experiment 3 examines global patterns of declination in sentences consisting of all level tones. It evaluates hypothesis (4) and also serves as a control whereby we distinguish between local and non-local patterns of F0 change.

3 Experiment 1: Positional effects on tone production

3.1 Methods

3.1.1 Design and stimuli. For the positional effect task, Castillo García, a native speaker and a linguist, produced the target sentence. After his production, each speaker repeated the target sentence four times. Speakers were instructed to pause briefly after each repetition. Given the complexity of the tonal system in YM and the absence of Mixtec literacy within the population, this method allowed us to elicit a large range of different tonal stimuli, control the morphological structure of target words, and carefully control the target context so that possible coarticulatory tonal patterns across words would be minimized. Yet, this task is not without potential problems as speakers may be more likely to mimic the speech they repeat too carefully. We discuss this issue at the end of Section 3.3 and in Section 6.4.

One additional characteristic of such repetitions is the possible use of "list intonation" (Ladd, 2008). Two characteristics of such intonation, if it occurs in YM, are that (a) it is fairly consistent across speakers and (b) usually only the first and last repetitions differ from the rest of the list (Himmelmann & Ladd, 2008). We believe that such characteristics justify the methods used in the current study, but to be certain, we explicitly examined whether native speakers used any list intonation across tone productions for the level tonal melody stimuli. These results are presented after the main experimental findings in Section 3.2.

Twenty tonal melodies occurring on disyllabic words were examined. Table 1 shows the tonal melodies and stimuli elicited within this study. For 16 of the tonal melodies, two stimuli words were selected. For four of the tonal melodies, only one stimulus word was included. These latter melodies were rarer than the others and it was not possible to find more than one noun within the sentence frame that was used. Note that a total of 27 tonal melodies surface on disyllabic words in YM (DiCanio et al., 2014, 2018). However, the additional 7 tonal melodies were difficult to include within the contexts devised for the current study. While it is preferable to include a greater number of words in a study like this rather than to rely on repetitions of the same word (see Winter, 2015), we were limited by the set of possible disyllabic nouns which could occur in the target syntactic context. Unlike languages like Mandarin Chinese, many tonal melodies in YM are restricted to specific parts of speech. Moreover, we were limited by the presence of contrastive glottalization in the language—many disyllabic nouns are glottalized and this will influence the measurement of F0.

Each of these target words was placed in a natural sentence where the local tonal contexts were controlled. For both the utterance-final and non-utterance-final conditions, the 3S.Masc clitic /=ɾa^{2}/ always preceded the target word. In the non-utterance-final condition, the target was always followed by a word with an initial tone /3/. In many sentences, this consisted of a final adverb expressing time or a noun expressing location of the action, for example, /βi^{3}ko^{4}/ "(at the) party,",/βeʔ^{3}e^{3}/ "(at the) house," /βi^{3}tĩ^{3}/ "now," / ɲũ^{3}ũ^{2}/ "(in the) town." This resulted in sentence pairs like /ʃa^{4}ʃi^{24}=ɾa^{2} ^{n}di^{3}ʃi^{4}/ "He is eating corn" and /ʃa^{4}ʃi^{24}=ɾa^{2} ^{n}di^{3}ʃi^{4} βi^{3}tĩ^{3}/ "He is eating corn now." For several sentences, the final word consisted of a post-nominal adjective which modified the preceding noun, for example, /ʃa^{4}ʃi^{24}=ɾa^{2} tʃi^{3}ta^{2}/ "He is eating a banana" and /ʃa^{4}ʃi^{24}=ɾa^{2} tʃi^{3}ta^{2} ʃeʔ^{3}e^{4}/ "He is eating an unripe banana." A total of 288 words (36 words × 2 conditions × 4 repetitions) were analyzed for each speaker, totalling 2,592 analyzed sentences for all nine speakers. Acoustic recordings with disfluencies were discarded prior to analysis, but these comprised a small set (approximately 1–2% of all sentences).

Table 1: Experimental stimuli - tonal melodies in disyllabic words. Note that tones are specified on individual syllables here, not words, for instance, /βa^{14n}gi^{4}/ "rainbow" has an initial rising tone followed by a high tone.

Melody	Word	Gloss	Melody	Word	Gloss
4.4	βa^{4}li^{4}	"little"	3.4	^{n}di^{3}ʃi^{4}	"corncob"
	ʃu^{4}ɲu^{4}	"pineapple"		βi^{3}ko^{4}	"party"
4.3	βi^{4}lu^{3}	"clay"	3.3	tʃi^{3}j ũ ^{3}	"work"
	^{n}di^{4}ko^{3}	"mother"		tʃi^{3}jo^{3}	"tile"
4.2	tʃi^{4}tu^{2}	"woodpecker"	3.2	tʃi^{3}ta^{2}	"banana"
	tʃa^{4}na^{2}	"grackle"		ta^{3}ta^{2}	"seed"
4.1	βi^{4}ʃi^{1}	"difficult"	3.42	βi^{3}ta^{42}	"soft"
	ta^{4}si^{1}	"wizard"		tʃa^{3}ko^{42}	"bearded"
13.3	ki^{13}ʃi^{3}	"come.PERF"	1.4	βi^{1}ko^{4}	"cloud"
	ʃi^{13}ʃi^{3}	"eat.PERF"		ta^{1}tã^{4}	"medicine"
13.2	βa^{13}βi^{2}	"knuckle"	1.3	βi^{1}ʃĩ^{3}	"cold"
	su^{13n}du^{2}	"doll"		ti^{1}ʃĩ^{3}	"belly"
14.3	tʃi^{14}k ũ ^{3}	"guamuchiles"	1.1	βi^{1}ka^{1}	"brush"
	tʃi^{14}ki^{3}	"prickly pear"		ka^{1}ta^{1}	"press (N)"
14.2	ja^{14}k ũ ^{2}	"be itchy"	1.32	mi^{1}nu^{32}	"epazote"
	^{n}da^{14}ku^{2}	"straight"		βi^{1}ʃi^{32}	"whiskers"
4.13	tʃi^{4}tʃi^{13}	"ripe"	4.24	ka^{4}ni^{24}	"long"
1.42	ta^{1}k^{w}i^{42}	"water"	14.4	βa^{14n}gi^{4}	"rainbow"

3.1.2 Speakers and recording. Nine speakers (7 females, 2 males) from the Yoloxóchitl community were recruited for this study. No speakers reported any speech or hearing difficulties. Each speaker was born in the town of Yoloxóchitl and spoke YM as their native language. Owing to the endangered status of the language, the average age of the participants was 48 years old, with an age range of 29–66 years old. All speakers were transported from Yoloxóchitl to the nearby town of San Luis Acatlán for recording purposes. Recording took place in a quiet room in San Luis Acatlán. The speaker and Castillo García were recorded on separate audio channels, each wearing a Shure SM10A head-mounted microphone. Acoustic recording was done on a Marantz PMD 661 solid state recorder at a 44.1 kHz sampling frequency.

3.1.3 Analytical techniques. All the recorded sentences were transcribed by the third author using ELAN (Wittenburg et al., 2006) and acoustically segmented by the second author using Praat (Boersma & Weenink, 2016). The target word in each sentence was segmented by hand on three segmentation tiers: lexical, syllabic, and segmental. Consonant and vowel durations were measured along with F0 on the vowel. Utterance-initial stops were excluded from the durational analysis. F0 measures were taken from each vowel at five equidistant time points with a script written for Praat (DiCanio, 2016). The F0 range used by the script was adjusted on an individual basis (i.e., higher for female speakers and lower for males). This script also extracted durational information for each segment within the word. Dynamic F0 measures were extracted from only those tokens that were 50 ms or longer in duration. For tokens shorter than 50 ms, no F0 information was extracted. The bases for excluding these tokens were concerns over the reliability of dynamic F0 measurements on short durations and concerns over extracting F0 on vowels which lack voicing (e.g., a word like /tʃi^{4}tu^{2}/ "woodpecker" may be produced as [tʃi̥^{4}tu^{2}] when spoken quickly). These shorter vowels comprised 4.2% of the total number of vowel tokens.

All statistics were calculated using R (R Development Core Team, 2017). For the duration data, words with each of the different tonal melodies were pooled together since they were all of the same prosodic shape, a disyllabic bimoraic word. Separate linear mixed effects models were examined for consonant duration and for vowel duration. Each model included two fixed effects (utterance position and word position) along with their interaction. As with the F0 data, Speaker and Word were treated as random intercepts and fully-crossed random slopes were included as well.

For the F0 data, all F0 values were converted to $\log_{10}$ values and statistically normalized (z-score normalization) to correct for individual speaker differences in F0 range and level prior to statistical analysis. Statistical analyses were done individually for each tonal melody; for example, separate analyses for melody 1.1, for melody 1.3, etc. Separating out each of the tonal melodies serves two purposes: (1) it allows one to control for possible differences in the direction of F0 movement and (2) it removes some of the inter-melodic variation from consideration in the statistics. With respect to (1), if lower tones undergo lowering and higher tones undergo raising, grouping such tones together might result in an inconsistent net effect when in fact the F0 range is expanding. With respect to (2), there is substantial phonetic variation between tonal melodies in YM. Tone /1/ is realized with a steeper, downward slope in a word like /βi^{4}ʃi^{1}/ "difficult" than in a word like /βi^{1}ka^{1}/ "brush." In the former, there is necessarily greater F0 movement across syllables than in the latter. Moreover, if the statistical models were based on individual tones, a different set of fixed effects would be necessary. Most of the contour tones surface only on the word-final syllable and, resultingly, word position would be excluded for these models. Treating tonal melodies as the units of analysis, rather than individual tones, allows us to focus on contextual effects to the exclusion of this melodic variability.

Linear mixed effects models were constructed for each tonal melody (the combinations listed in Table 1) with the following fixed effects: Utterance-position (Utterance-final, Non-final), Word-position (penultimate and unstressed, final and stressed), and time (5 points, centered at the vowel midpoint). Z-score normalized log F0 was the dependent variable. Speaker and Word were treated as random intercepts and we attempted to include random slopes for utterance position, word position, and an utterance by word position interaction as well (see Barr et al., 2013). However, many of these models either did not converge or resulted in singular fit errors. For tonal melodies /14.4, 4.13, 4.24/, only one target word was examined (see above). Thus, a random intercept for Word was not included here. For three tonal melodies, (/4.2, 1.4, 4.3/), the best model included only a random intercept for subject. For eleven tonal melodies, (/4.1, 14.2, 3.2, 3.3, 14.4, 1.3, 14.3, 4.24, 1.32, 3.42, 1.42/), the best model included a random slope for utterance position but not word position. For the remaining tonal melodies, (/1.1, 13.2, 4.4, 3.4, 4.13, 4.2, 1.4, 4.3/), random slopes for utterance position and for word position were included. All fixed effects were evaluated using *lmerTest* (Kuznetsova et al., 2017), a model which relies on the Satterthwaite method to approximate the degrees of freedom and reports both an F statistic and p-values via ANOVA, but only a lower bound on degrees a freedom. There is currently no way to approximate the upper bound on degrees of freedom for linear mixed effects models (Baayen, 2008).

The statistical results ought to be interpreted in the following way. First, a significant effect of utterance position on duration and a significant interaction between utterance position and word position on duration relate to the initial hypothesis regarding stress and prosodic lengthening. Namely, durational lengthening is a cue for word stress and this lengthening is more robust in utterance-final position. Second, in relation to the tonal data, significant interactions between word position and utterance position are considered positive evidence that these tones are realized differently in word-final, utterance-final position than in non-word-final and non-utterance-final positions. This relates to hypotheses (2) and (3) above. Yet, these results can illustrate what might be a process of local F0 shift in a pre-pausal position, final allotony, or a general declination pattern.

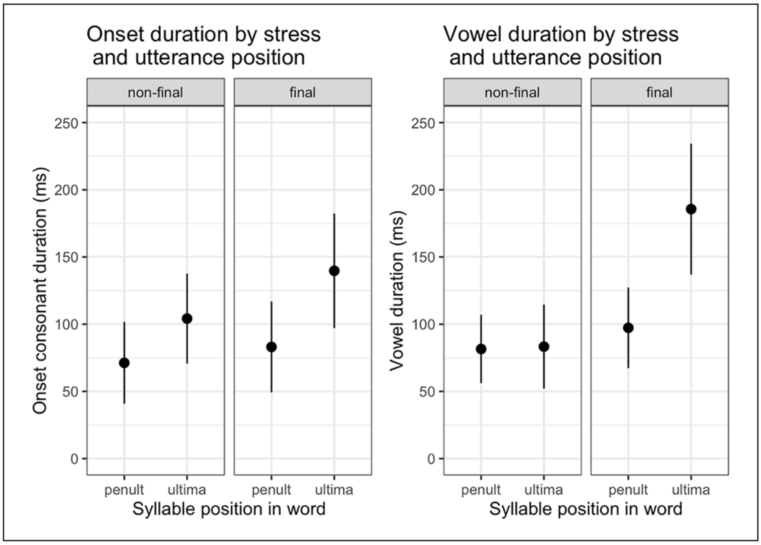

Figure 2. Effect of stress and utterance position on duration. Mean values are plotted with bars reflecting average standard deviation around mean.

Post-hoc analysis will allow us to distinguish between the first two mechanisms (by observing whether the observed patterns can be reduced to patterns of truncation or compression). Experiment 3 will allow us to distinguish between the first and third mechanisms—are observed patterns of tonal change across utterance positions reducible to global patterns of F0 shift?

3.2 Results

3.2.1 Duration results. Figure 2 shows the effect of word position and utterance position on the duration of onset consonants (left) and vowels (right) in YM. For the consonant duration data, there was a significant main effect of word position (stressed, unstressed) on onset duration, $t(8) = 12.8$, $p < .001$. Onset consonants appearing in the penultimate, unstressed syllable were shorter (77 ms) than those appearing in the ultimate, final syllable of the word (122 ms), a ratio of 1:1.58). A main effect of sentence position was also found, $t(8) = 8.9$, $p < .001$. Onset consonants appearing in an utterance-medial word were shorter (88 ms) than those appearing in an utterance-final word (111 ms), a 26% increase in duration). A significant interaction between word and utterance position was also found, $t(8.8) = 8.4$, $p < .001$. Onset consonants in penultimate syllables were lengthened only slightly in utterance-final words (71 vs. 83 ms, 17% lengthening) while those in final syllables were lengthened to a greater degree in utterance-final words (104 vs. 140 ms, 35% lengthening).

For the vowel duration data, there was a significant main effect of word position (stressed, unstressed) on onset duration, $t(8) = 11.5$, $p < .001$. Vowels appearing in the penultimate, unstressed syllable were shorter (90 ms) than those appearing in the ultimate, final syllable of the word (135 ms). A main effect of sentence position was also found, $t(8) = 15.1$, $p < .001$. Vowels appearing in an utterance-medial word were shorter (82 ms) than those appearing in an utterance-final word (142 ms). A significant interaction between word and utterance position was also found, $t(8) = 13.1$, $p < .001$. Vowels in penultimate syllables were lengthened only slightly in utterance-final words (82 vs. 97 ms, 18% lengthening) while those in final syllables were lengthened to a

Table 2. Summary of utterance-level effects. All values are given in normalized log standard deviations of F0 and reflect the differences across the utterance positions, averaged across each vowel. Negative numbers indicate final lowering. Positive numbers indicate final raising.

Melody	Penult	Ultima	*p*	Melody	Penult	Ultima	*p**
4.1	–0.16	–1.35	***	13.3	+0.12	+0.23	NS
1.1	–0.14	–0.37	***	4.4	+0.15	+0.68	***
4.2	–0.08	+0.07	NS	3.4	+0.11	+0.58	***
3.2	–0.06	–0.34	***	1.4	–0.09	+0.69	***
13.2	–0.07	–0.32	***	14.4	–0.04	+0.43	***
14.2	–0.17	–0.29	***	1.32	–0.16	–0.06	NS
4.3	+0.07	0	NS	1.42	–0.45	–0.56	NS
3.3	+0.06	+0.02	NS	3.42	–0.12	–0.45	***
1.3	–0.11	–0.02	NS	4.13	+0.21	–0.60	***
14.3	0	–0.32	***	4.24	–0.12	+0.41	***

Note: *P*-values are interpreted as follows: '***' 0.001 and '**' 0.01.

much greater degree in utterance-final words (83 vs. 186 ms, 224% lengthening). This interaction also reveals an important effect obscured by the main effects: there is little effect of word position on vowel duration in non-utterance-final words (82 in penult vs. 83 ms in ultima), but a very strong effect of word position in utterance-final words (97 in penult vs. 186 ms in ultima). The effect of word-position on duration is maintained for consonants across the two conditions, but neutralized for vowels in non-utterance-final position.

3.2.2 Results: Tonal melodies. Twenty tonal melodies were examined within the current study and we have divided the tonal results into sections corresponding with the final tone on the word. The reason for this organization is that one anticipates a stronger effect of utterance position for pre-pausal tones (at the end of the word) than for non-pre-pausal tones. The fixed factor *Word position* was examined as a main effect for only level tonal melodies (i.e., /1.1, 3.3, 4.4/). Significant main effects of word position were otherwise ignored since they reflect a comparison of different tones (i.e., examining whether tones /4/ and /1/ differ in a /4.1/ melody is irrelevant to the current study). For these melodies with varying tones, we focus more closely on the main effect of utterance position and any interactions with other factors. We present a main summary of the effects here followed by results specific to certain tonal melodies.

Summary of tonal effects by utterance and word position. Table 2 provides a summary of the patterns of tonal lowering and raising of each of the tonal melodies examined here. Two major patterns emerge from the data presented above. First, with few exceptions, tonal changes on the penultimate syllables were notably weaker than those on the ultima. Second, F0 changes by context differed by tonal category. In utterance-final and word-final position, tones /1/ and /2/ were consistently lowered, while tone /4/ was consistently raised. In this position, two different patterns were observed with tone /3/: (1) in melody /14.3/, it was lowered, (2) in all other melodies, it was unaffected. In terms of tonal dynamics, in utterance-final, word-final position tone /2/ was realized with a steeper falling contour and tone /4/ with a clear rising contour.

Final contour tones also underwent changes by utterance-position. Both tones /42/ and /13/ underwent F0 lowering in utterance-final, word-final position (though tone /32/ was unaffected). This lowering is a byproduct of changes to tonal shape in this utterance position. Tone /42/ involved less of a fall in utterance non-final position. As a result, the average F0 of the contour was higher when not utterance-final. Tone /13/ was realized with an altogether different contour (as a fall) in utterance non-final position than in utterance-final position, where it was realized as a rise. Finally, tone /24/ underwent raising in utterance-final, word-final position. This was again a consequence of contour simplification; the rising tone was realized as a level tone in non-utterance-final, word-final position.

Melodies with final tones /1/ and /2/. Six of the tonal melodies contain a final tone which is in the lower range: /4.1, 1.1, 4.2, 3.2, 13.2, 14.2/. Figure 3 shows the effect of word and utterance position on the production of these tonal melodies in YM. For the level tonal melody, **/1.1/**, there was a significant main effect of word position on F0, $t(8.1) = -8.1, p < .001$. The tone on the final syllable of the word was lowered by 0.71 *sd* relative to the penult.

There was a significant main effect of utterance position on F0 for tonal melodies /4.1/, $t(8.1) = -10.3, p < .001$, /1.1/, $t(8.2) = -6.9, p < .001$, /14.2/, $t(8.0) = -5.4, p < .01$, /13.2/, $t(8.0) = -2.7, p < .05$, and /3.2/, $t(8.0) = -3.0, p < .05$. No main effect was observed for tonal melody /4.2/. Tones on these words were lowered (between 0.2–0.25 *sd*) in utterance-final position relative to non-utterance-final position.

A significant interaction between utterance and word position was found for tonal melodies /4.1/, $t(1170) = -20.6, p < .001$, /1.1/, $t(1173) = -7.3, p < .001$, /13.2/, $t(1410) = -7.5, p < .001$, /3.2/, $t[1237] = -9.8, p < .001$, /14.2/, $t(1311) = -3.6, p < .001$. No significant interaction between utterance and word position was found for tonal melody /4.2/. For each of these tonal melodies showing a significant main effect, the final tone in the word was lowered relative to the penultimate tone. In some cases, the asymmetry in F0 lowering across the word was large. For tonal melody /4.1/, relatively little F0 lowering was observed on the penultimate syllable in utterance-final position (0.16 *sd*), but much more substantial lowering was observed on the final syllable with tone /1/ (1.35 *sd*). For the other tonal melodies, the asymmetry was smaller; the F0 on the final syllable was approximately 0.2–0.3 *sd* lower than that observed on the penult in utterance-final position. While tonal lowering occurs across the word when words are in utterance-final position, it is significantly stronger on the final, pre-pausal syllable.

A significant three-way interaction between Time, Word position, and Utterance position was found for tonal melody /3.2/, $t(1235) = -3.2, p < .01$, /13.2/, $t(1409) = -9.0, p < .001$, and /1.1/, $t(1166) = -3.5, p < .001$. While tones /2/ and /1/ were produced with a falling trajectory in both utterance contexts, their slopes in the final syllable were lowered (resulting in a sharper fall) in utterance-final position relative to non-final position.

Melodies with final tones /3/ and /4/ Nine of the tonal melodies contain a final tone which is in the higher range: /4.3, 3.3, 1.3, 14.3, 13.3, 4.4, 3.4, 1.4, 14.4/. Figure 4 shows the effect of word and utterance position on the production of these tonal melodies in YM. A significant effect of Word position on F0 was found for level tonal melodies /3.3/, $t(1124) = -9.5, p < .001$, and /4.4/, $t(8.0) = 4.9, p < .01$. In the former melody, tone /3/ lowers by 0.18 *sd* between the penult and the final tone, while in the latter, tone /4/ raises by 0.46 *sd*.

There was a significant main effect of utterance position on F0 for all tonal melodies ending in the highest tone /4/: /4.4/, $t(8.0) = 5.3, p < .001$, /3.4/, $t(8.0) = 4.8, p < .01$, /1.4/, $t(1401) = 12.4, p < .001$, and /14.4/, $t(8.0) = 2.7, p < .05$. Across the board, tones on words with these melodies were raised in utterance-final position relative to utterance non-final position. In addition, for each of these tonal melodies, a significant interaction between utterance position and word position was also observed: /4.4/, $t(1405) = 17.8, p < .001$, /3.4/, $t(1362) = 12.8, p < .001$, /1.4/, $t(1401) = 16.5, p < .001$, and /14.4/, $t(661) = 6.4, p < .001$. For these tonal melodies, the final syllable's tone was raised to a greater degree in utterance-final position than the tone on the penultimate syllable. Significant three-way interactions between utterance position, word position, and time were observed several of these tonal melodies as well: /4.4/, $t(1401) = 5.6, p < .001$, /3.4/, $t(1361) =$

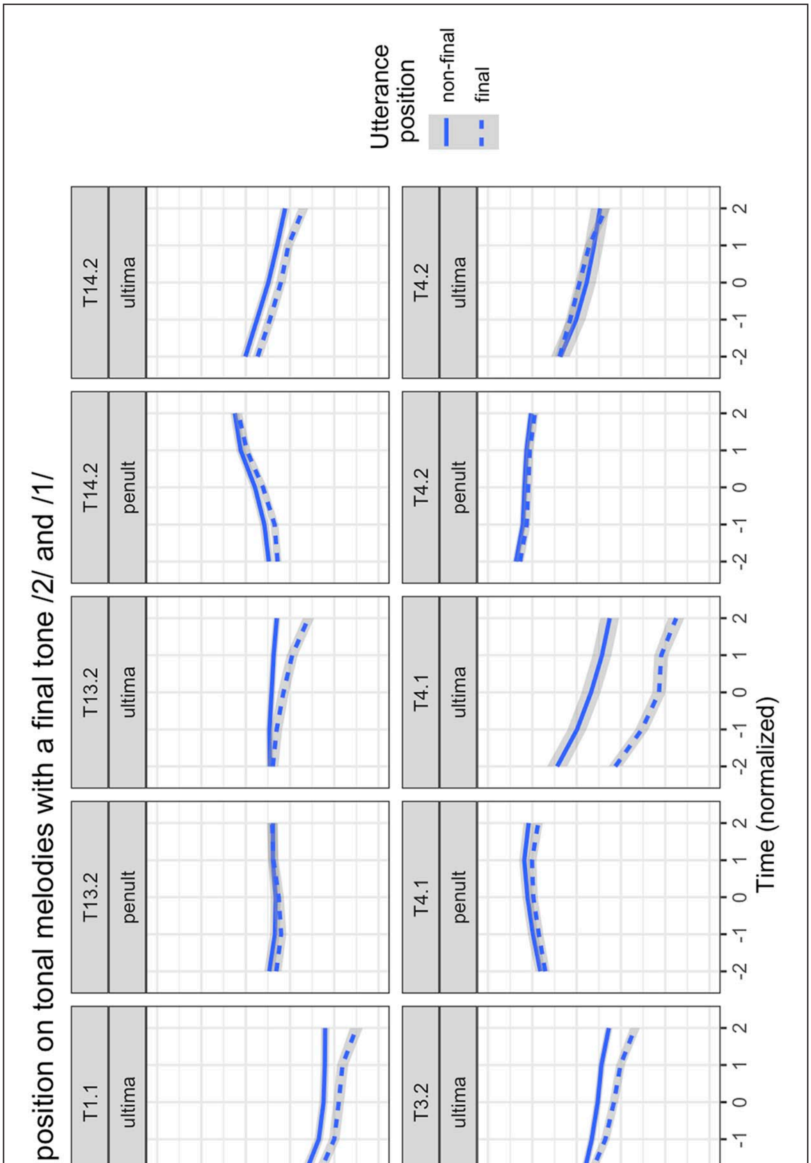

Figure 3. Effect of utterance position on F0 in disyllabic words. Gray bars around mean values reflect standard error.

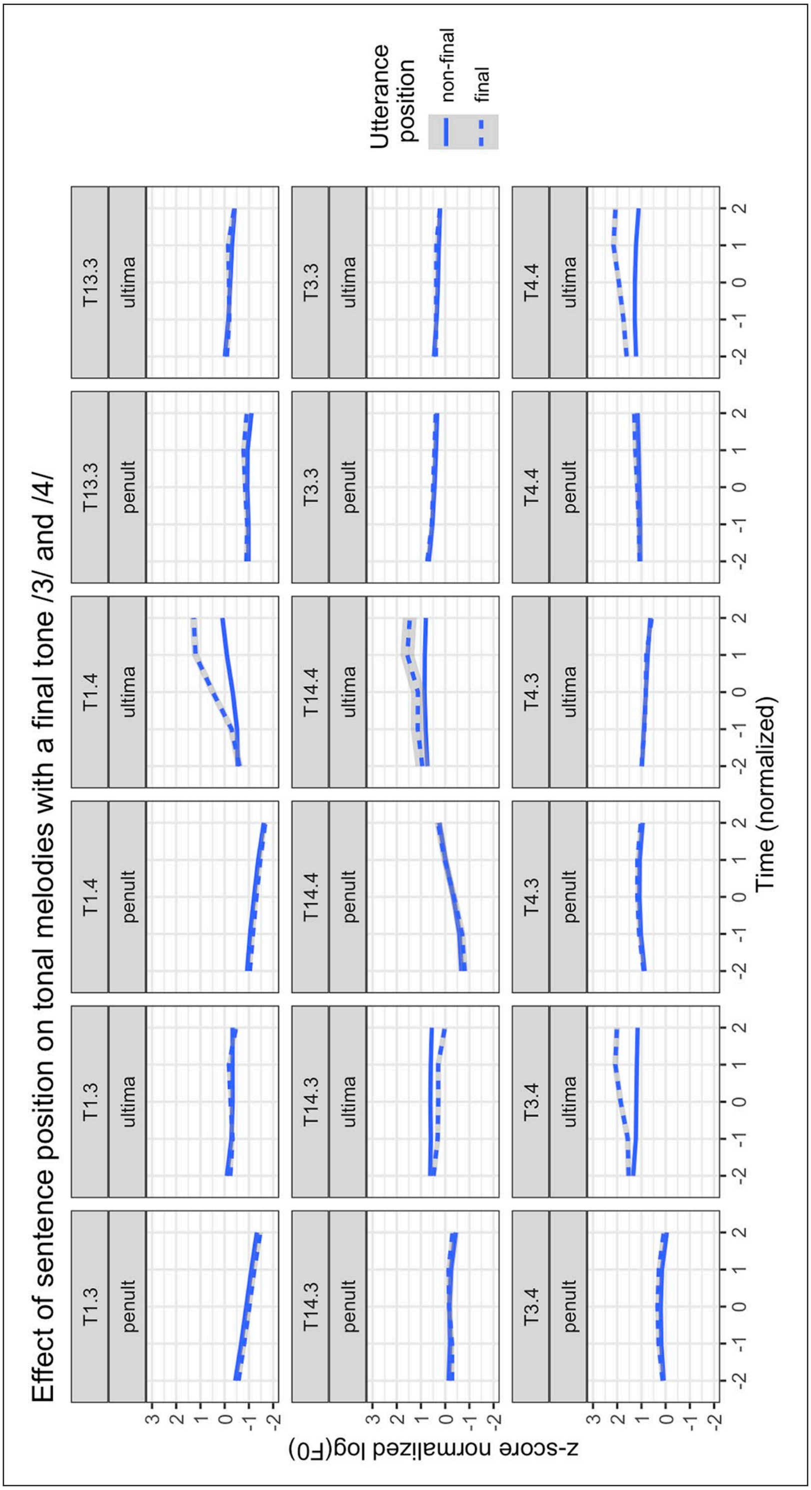

Figure 4. Effect of utterance position on F0 in disyllabic words. Gray bars around mean values reflect standard error.

6.4, $p < .001$, and /1.4/, $t(1401) = 10.0$, $p < .001$. The final tone /4/ was realized with a significant F0 rise in utterance-final, word-final position that was absent in non-utterance final position.

No significant main effects of utterance position on F0 were found for any tonal melody possessing a final tone /3/ (/4.3, 3.3, 1.3, 13.3, 14.3/). No significant interactions between utterance and word position were observed for any tonal melody with a final tone /3/, but for tonal melody /14.3/ significant three-way interaction between Time, Word position, and Utterance position was found, $t(1218) = -5.2$, $p < .001$. The final tone /3/ in this melody was realized with more of a falling F0 slope in utterance-final position. No three-way interactions were observed for any other of these tonal melodies.

Melodies with final contour tones. Five of the tonal melodies contain a final tone which is a contour: /4.13, 4.24, 1.32, 1.42, 3.42/. Figure 5 shows the effect of word and utterance position on the production of these tonal melodies in YM. A significant main effect of utterance position on F0 was observed for only tonal melodies /1.42/, $t(7.8) = -7.3$, $p < .001$, and /3.42/, $t(8.0) = -5.4$, $p < .001$. In both cases, the tones realized on these words were lowered in utterance-final position relative to non-final position.

A significant interaction between utterance and word position was observed for tonal melodies /3.42/, $t(1449) = -7.9$, $p < .001$, /4.13/, $t(562) = -11.1$, $p < .001$, and /4.24/, $t(671) = 10.0$, $p < .001$. In the two former melodies, the final tone was lowered to a greater degree than the penultimate tone in utterance-final position. For tonal melody /4.24/, the penultimate tone /4/ was slightly lowered when the word appeared in utterance-final position.

A significant three-way interaction between utterance position, word position, and time was observed for tonal melodies /4.13/, $t(7) = 3.5$, $p < .05$, /4.24/, $t(8.4) = 5.1$, $p < .001$, /1.32/, $t(1342) = -4.3$, $p < .001$, and /1.42/, $t(596) = -2.8$, $p < .01$. In each case, the phrase-final, utterance-final tone was realized with a more dynamic contour (falling or rising) when compared with the same tone in non-utterance-final position.

3.2.3 List intonation. To assess whether list intonation occurred across utterances, we examined F0 production of three tonal melodies /1.1, 3.3, 4.4/ produced across the four repetitions for all speakers. The results are shown in Figure 6.

Two linear mixed effects model were constructed and compared to examine if a list intonation effect was found. In the first model, the dependent variable was the normalized log F0 of the vowel midpoint and the fixed effects were Word-Position (penult, ultima), Utterance-Position (non-final, final), Tonal Melody (3 levels), and Utterance repetition (1–4). The fixed effects were fully crossed so that possible interactions among tone, position, and utterance repetition could be included. The random effects included the by-subject intercept and the fully crossed set of random slopes matching the fixed effects. The second model was identical, but it did not contain the fixed effect of utterance repetition nor the random slope for utterance repetition. These models were compared using log-likelihood comparison (for a discussion of these methods applied to LME models, see Matuschek et al., 2017). The model containing the fixed effect of utterance position (and its interactions) was significantly better than the one excluding it, AIC[26] = 747 versus AIC[14] = 755, $\chi^2 = 32.5$, $p < .01$.

Investigating the improved model, it was found that the main effect of utterance repetition was not itself significant, $t = -0.2$, but the interaction between utterance repetition and melody /4.4/ was, $t = -2.6$. Thus, there was a significant effect of list intonation on F0 in YM. However, the magnitude of this effect varied by tone. For tonal melody /1.1/, virtually no difference in mean F0 was observed across repetitions ($<.05$ *sd* change in F0 across utterances). For tonal melody /3.3/, the mean difference in F0 between the first and final utterance was 0.24 *sd* (and not statistically

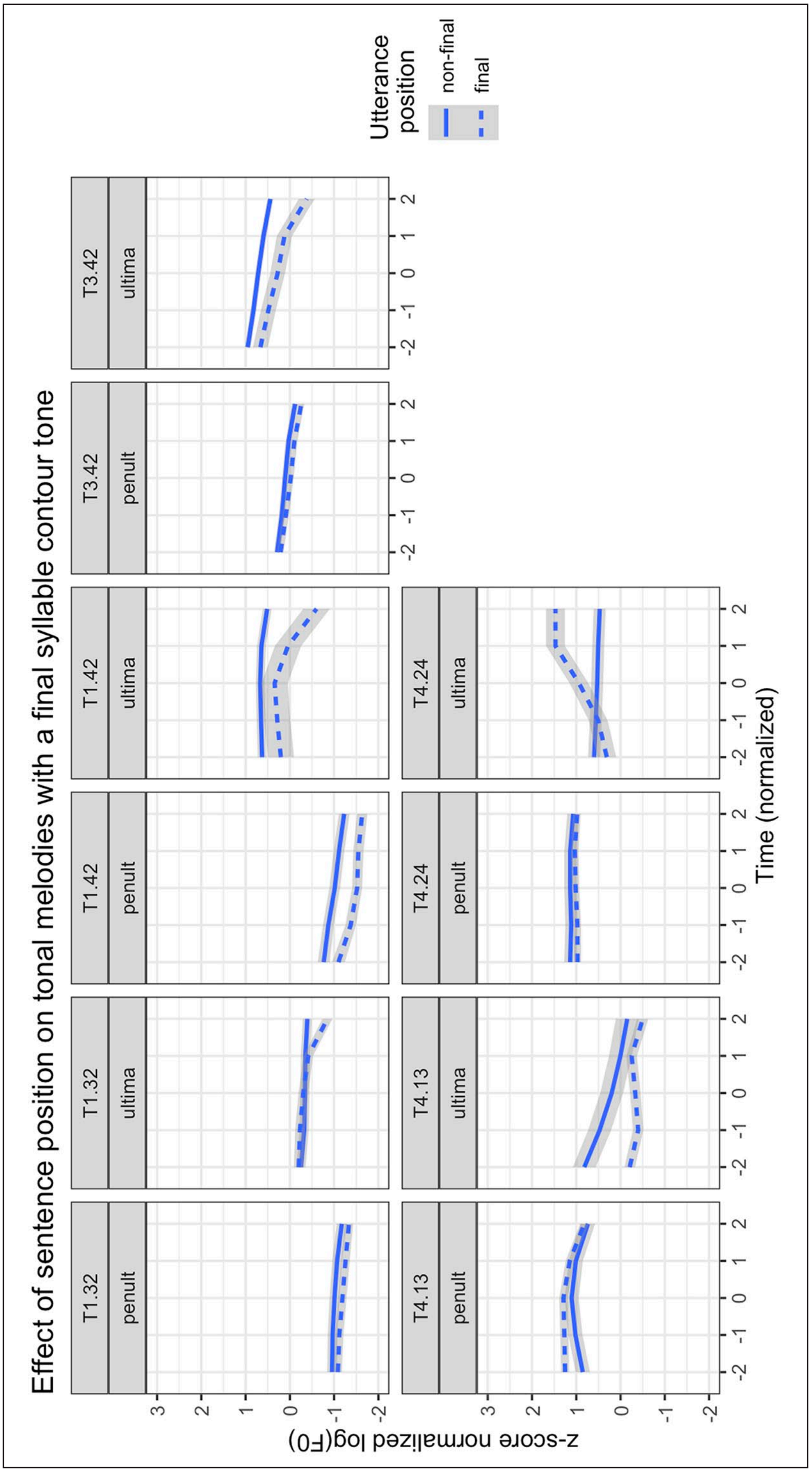

Figure 5. Effect of utterance position on F0 in disyllabic words. Gray bars around mean values reflect standard error.

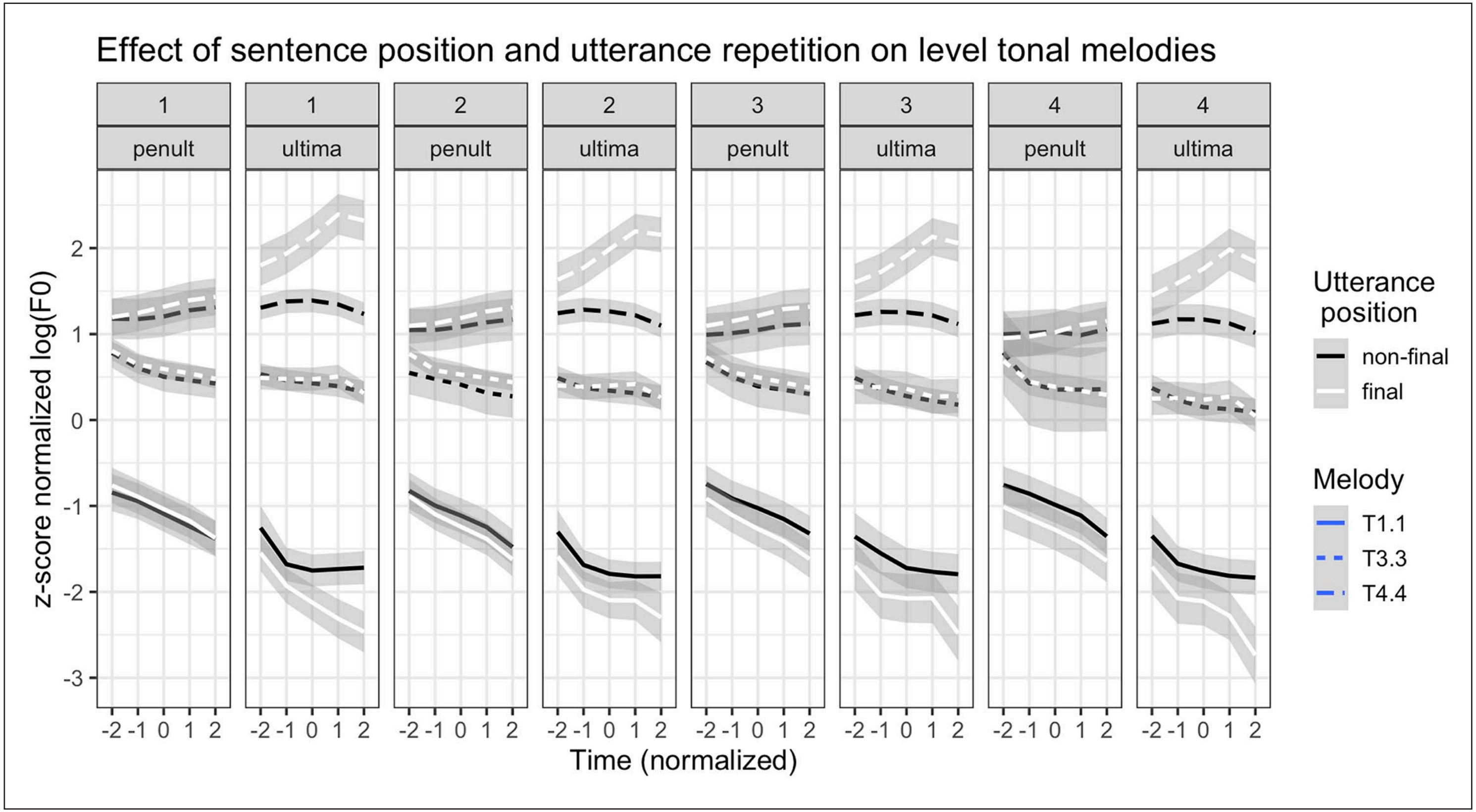

Figure 6. Average effect of utterance repetition on the production of three target level tone melodies across all speakers. Numbers reflect the utterance repetition. The tonal contours on the penult are presented before the contours on the ultima for each repetition. The grey bars indicate a 95% confidence interval around the mean values, assuming a normal distribution.

significant). For tonal melody /4.4/, it was 0.27 *sd*. This change (for melody /4.4/) corresponded to a difference of just 2 to 4 Hz across all repetitions for male speakers and 8 to19 Hz across repetitions for female speakers. The effects of utterance number of F0 were most pronounced between the first and second repetition and between the third and final repetitions. There was virtually no change in F0 between the middle two utterances.

3.3 Discussion: positional effects on tone production

The data examined in Experiment 1 demonstrate a significant effect of utterance position on duration and the production of lexical tone in YM. At the word-level, onset consonants in word-final syllables were lengthened relative to onset consonants in word-initial syllables. This finding mirrors that of previous research (DiCanio et al., 2018, 2019) showing that word-medial position (the onset of stressed syllables) is the locus of stress-related lengthening in YM. While we generally predict that stress will be realized via vowel lengthening, languages for which there is contrastive vowel length may illustrate distinct patterns. In Swedish, for instance, focal lengthening is associated with consonant lengthening before short vowels (Heldner & Strangert, 2001).In Pirahã, stress induces durational lengthening in onset consonants more than in vowels and there is a vowel length contrast (Everett, 1998). Yoloxóchitl Mixtec does not possess a typical vowel length contrast as roots are simply minimally bimoraic /CVV/ or /CVCV/ (with two short vowels)—there is no contrast between CV and CVV roots. Yet, there may be a necessity for listeners to distinguish between prefixed monosyllabic roots (/CV-CVV/) and non-prefixed disyllabic (/CVCV/) roots. This necessity may constrain the domain of stress-induced lengthening on the vowel. Final syllable lengthening in Yoloxóchitl Mixtec is characteristically distinct from a fairly common pattern of penultimate stress in Mixtec languages. In Southeastern Nochixtlán Mixtec and Ixpantepec Nieves Mixtec, stress is penultimate and the penultimate syllable or vowel undergoes lengthening under focus in these languages (McKendry, 2013; Carroll, 2015).[10]

Stress-related lengthening also interacts with utterance position. Final syllables in utterance-final position were lengthened more than penultimate syllables. This finding confirms our first hypothesis regarding stress-related lengthening; it is exaggerated in phrase-final position. However, it is worth noting that the type of domain-final lengthening observed here is qualitatively different than the focus-induced lengthening observed in DiCanio et al. (2018). Focus-induced lengthening similarly targets the stressed, final syllable in polysyllabic words in YM, but it occurs primarily on the onset consonant, not on the vowel. Domain-final lengthening in the present study targeted the final vowel to a much greater degree than focus-induced lengthening. The qualitative differences in lengthening found between the current study and DiCanio et al. (2018) suggest a slightly revised analysis of the latter. Focused constituents are left-dislocated in YM and one possible analysis of such constituents is that they occur at a prosodic boundary (Féry, 2013). However, the pattern of focus-related lengthening in YM is qualitatively dissimilar from final-lengthening. This finding accords well with studies of prosodic lengthening in other tonal languages, such as in Mandarin and Wenzhou Chinese, where boundary-related prosodic effects are characteristically distinct from those related to focus (Chen, 2006; Scholz & Chen, 2014a). In sum, the differences between these studies on Yoloxóchitl Mixtec lend support to the view that focus-related lengthening is not simply the result of a change in constituent alignment.

The utterance-final position in YM was not characterized by a single effect on F0, but by two parallel processes, each of which occurred on the word-final syllable. First, the results were compatible with a general process of F0 range expansion in utterance-final position. Recall that the highest level tone /4/ underwent utterance-final raising while the lower level tones /2/ and /1/ each underwent greater lowering. Final tone /3/ did not vary in its production across utterance contexts.

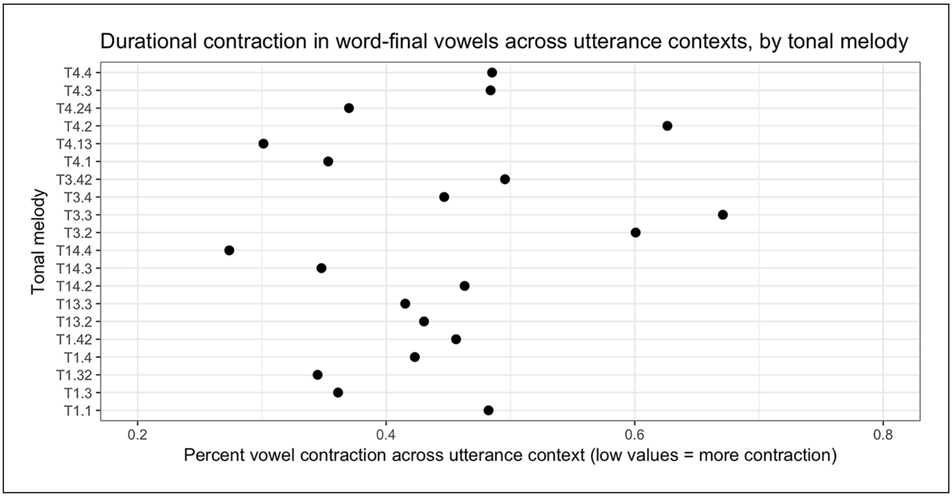

Figure 7. Average vowel contraction by tonal melody. Values here indicate the percent by which the final vowel was contracted in non-utterance-final contexts in comparison with utterance-final contexts.

The high falling tone /42/ was also systematically lowered in utterance-final position. These findings confirm our second hypothesis—tones whose offsets approach the upper and lower F0 range were more likely to undergo changes in phrase-final position.

However, the observed effects were also compatible with a process of *syntagmatic tonal contrast enhancement*. Tone /3/ lowered in the context when it followed tone /14/, but raised in the context when it followed tone /13/. Tone /1/ lowered to a much greater degree (1.35 *sd* vs. 0.37 *sd*) when it followed tone /4/ than when it followed tone /1/.[11] These latter effects are not explainable via a general process of range expansion. Instead, similar to findings in DiCanio et al. (2018), tonal contrasts across disyllabic words were enhanced in utterance-final position.

As shown by the number of three-way interactions with time, tonal melodies with final contour tones are more prone to temporally-induced changes in F0 slope across the utterance. Grosso modo, final contours were flattened in utterance-non-final position relative to utterance-final position. However, the two rising tones (/13/, /24/) underwent a qualitatively different set of changes in utterance-final position than level and falling tones did. Tone /13/ was realized with a shallow F0 rise in utterance-final position but with a flat contour in non-final position. Tone /24/ was realized with a rise in utterance-final position but with a falling contour in non-final position. A closer look at the durational differences by tone reveal a relevant interaction. As previously discussed, word-final vowels in non-utterance-final position were 0.44 times as long as vowels in utterance-final position (224% lengthening in utterance-final position).

Though rather than conceptualizing this as a process related to final *lengthening*, it is more useful to think of it as a process related to durational *compression*. Certain tones underwent greater compression than others. Figure 7 shows the average contraction of final vowels in non-utterance-final contexts for each tonal melody. A number of different tones were shortened to less than 0.4 times the duration of the vowels in utterance-final position, including tones /13/ and /24/. The effect of this compression was an extreme durational shortening. Final syllable vowels with tone /13/ in utterance-final position were 184 ms long (on average), but the same vowels in non-utterance-final position were 55 ms long. Final syllable vowels with tone /24/ in utterance-final position were 216 ms long (on average), but the same vowels in non-utterance-final position were 80 ms long.[12]

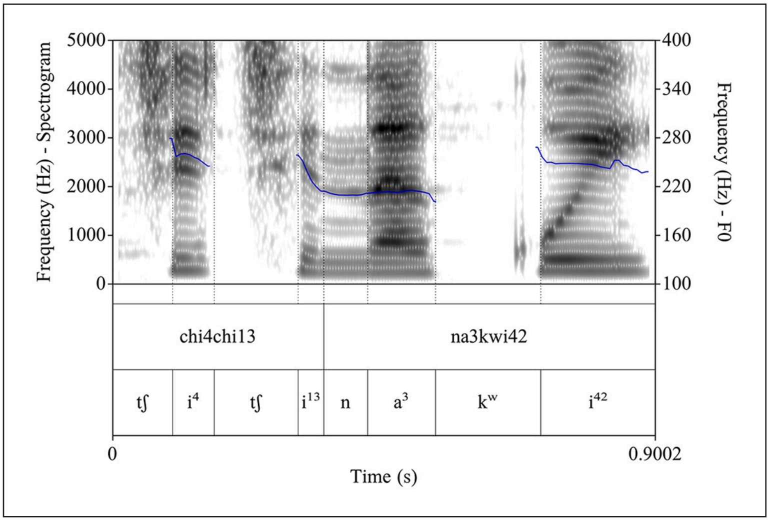

Figure 8. Production of non-utterance-final tone /13/. Observe the extreme shortening of the final syllable and the falling F0 trajectory, shown in blue.

The durational data suggest that durational compression resulted in a phonetic leveling of rising contours in YM. These results closely match data observed in Cheng and Xu (2015) for Taiwanese Mandarin. Rising tones are more likely to undergo contour leveling in contexts where duration is compressed than other contour tones are, confirming our third hypothesis. Moreover, rising tones on final syllables in YM only surface after a preceding tone /4/ (i.e., /4.13, 4.24, 4.14, 14.13, 14.24, 14.14/) are all possible but */3.13, 1.13, 1.14/ do not occur (DiCanio et al., 2014, 2018). This phonological restriction in the language creates what is termed a *conflicting* context for contour tone production (Xu, 1994) since all final rising tones only surface after a preceding high tone and never after a lower tone. It is difficult to produce a rising contour when the preceding target must always be high. While there is no phonological neutralization in YM here, the phonetic patterns suggest a diachronic pathway towards a process of contour simplification, such as those discussed in Hyman (2007).

Note that this tonal contraction does not result in a process whereby the rising tone spreads to the following word. That is, durational contraction in YM does not result in tonal spreading. Figure 8 demonstrates that tone /13/ is reduced to a slight falling tone when the vowel duration is shorter (in non-utterance-final position), but it does not influence the following tone.

The utterance repetition effects examined in 3.2.3 suggest that there is minimal list intonation in YM and the effects that occur are limited to the highest tone. Given the number of tonal contrasts in the language, one possible explanation is that speakers may inhibit a natural tendency towards F0 declination across utterances (see Gussenhoven, 2004). However, it is also possible that what is known as list intonation is not as universal as previously assumed and instead a learned behavior more common in languages where such changes in F0 might not lead to a neutralization of lexical contrast. For the current study, these data show that a repetition task does not introduce strong effects of list intonation for YM.

As mentioned above, one limitation of Experiment 1 is its reliance on sentence repetition for eliciting tones in controlled contexts. There is a possibility that the speakers in a study like this may imitate the observed effects in the process of repeating the sentences. Though there is no research to date on imitation in tonal languages, research where speakers are explicitly told to imitate prosody in repetition tasks has shown that they are mostly successful in doing so (Cole & Shattuck-Hufnagel, 2011; D'Imperio et al., 2014).

Examining the true breadth of the YM tonal inventory in detail is virtually impossible within a spontaneous speech corpus because the tonal melodies have quite different distributions. Level tones are quite common but more complex contour melodies; for example, /3.42/ or /13.3/ occur more rarely. Moreover, a large percentage of utterances in YM spontaneous speech end with function words; for example, /^{n}di^{4}/ "that" or /tã^{3}/ "and, but" (Amith & Castillo García, 2019). This makes finding words matched for tonal, morphological, and prosodic structure at different types of utterance boundaries extremely difficult. To address this concern and the methodological confound, we examined target words in utterance-final and utterance-medial position in a corpus of spontaneous speech in the following study.

4 Experiment 2: Evaluating durational and tonal effects in a spontaneous speech corpus

4.1 Methods

4.1.1 Design and stimuli. As a comparison for the results in Experiment 1, we analyzed the *Yoloxóchitl Mixtec Frog Story* speech corpus. This corpus consists of six different speakers producing an unscripted narrative while viewing pages of "frog, where are you?" (Mayer, 1969), a book consisting of 32 pages of a visual narrative with no text. To find words which occurred within the corpus both in utterance-medial and utterance-final position, we initially labelled all utterance-final content words for an individual speaker. We excluded function words because the target words in Experiment 1 were all content words. We then searched for these words in utterance-medial position within that speaker's recording. If three or more repetitions of a word surfaced in both utterance-final and utterance-medial positions, this word was segmented and analyzed. Finding words occurring in both positions resulted in limiting the word list to those words possessing only the most common tonal melodies: /1.1, 3.2, 3.4/. Within a transcribed speech corpus containing 462,770 bimoraic roots, roots consisting of two level tones (e.g., 1.1, 3.4, 1.3, 3.3. . .) comprise 88.7% of all these words (Amith & Castillo García, 2019).

For two of these tonal melodies, only one representative word occurred in each of the different speakers' recordings in both utterance-medial and utterance-final position: /ha^{1}na^{1}=ɾa^{1}/ "pet=3S. Masc" for tone melody /1.1/ and /i^{3}tũ^{4}/ "tree" for tone melody /3.4/. Many words possessing tone melody /3.2/ occurred in both prosodic positions though. The examined words alongside the number of observations for all speakers is given in Table 3. A total of 515 words were analyzed across the six speakers: 388 in utterance-medial position and 127 matched words in utterance-final position.

4.1.2 Speakers and recording. The third author recruited and recorded seven speakers producing a "frog, where are you?" narrative from the Yoloxóchitl community. This included the third author himself. No speakers reported any speech or hearing difficulties. Each speaker was born in the town of Yoloxóchitl and spoke YM as their native language. Speaker age was recorded for four of the participants and their average age at the time of recording was 51 years. All speakers were transported from Yoloxóchitl to the nearby town of San Luis Acatlán for recording purposes. Recording took place in a quiet room in San Luis Acatlán. The speaker and Castillo García were recorded on separate audio channels, each wearing a Shure SM10A head-mounted microphone Acoustic recording was done on a Marantz PMD 661 solid state recorder at a 44.1 kHz sampling frequency. The third author transcribed and translated each of the recordings made by the speaker

Table 3. Stimuli chosen from corpus.

Melody	Word	Gloss	N (utterance-medial)	N (utterance-final)
3.4	i³tũ⁴	"little"	79	20
3.2	i³na²	"dog"	12	4
	i³su²	"deer"	25	13
	sa³ˀβa²	"frog"	98	37
	ja³βi²	"hole"	21	5
	jo³ko²	"wasp"	20	6
1.1	ha¹na¹=ɾa¹	"pet=3SING.MASC"	133	42

using ELAN (Wittenburg et al., 2006). The average duration of the speaker's telling of the frog story was 9 minutes 13 seconds. Though seven speakers were recruited, upon later analysis, we excluded one of the speaker's recordings since his excessively hoarse voice quality throughout the recording caused substantial problems for F0 analysis. A total of 55.3 minutes of speech was analyzed for the six speakers.

4.1.3 Analytical techniques. The methods for analyzing both F0 contour and duration were identical to those discussed in Experiment 1 in Section 3.1.3, with one exception. In Experiment 1, dynamic F0 measures were extracted from only those tokens that were 50 ms or longer in duration. In Experiment 2, the limit was set to 40 ms instead. This choice was motivated by the faster speech rate observed in running speech. This resulted in excluding F0 trajectory data from 3/515 final syllables (0.6%) and 19/511 penultimate syllables (3.7%). In addition, a percentage of tokens were excluded from analysis due to variable devoicing of vowels (especially in utterance-final position). Vowels were realized with excessive creaky phonation or were devoiced in 46/515 (8.9%) word-final syllables and 49/511 penultimate syllables (9.6%). In sum, F0 trajectories were examined from a total of 443 penultimate syllables and 466 final syllables in disyllabic YM words.

The statistical tests were similar to those discussed in Experiment 1. For the duration data, separate LME models for each dependent variable (vowel duration, consonant duration) were run. For both models item/word was excluded as a random effect since neither model converged with its inclusion. For the vowel duration model, two crossed fixed effects were specified: position in word (penult, final) and position in utterance (non-final/final). Random slopes for this fully crossed model were specified in the random effects structure. For the consonant duration model, the same fixed effects were specified but only a random slope for utterance position was specified in the random effects structure. A more fully-specified random effects structure resulted in the model not converging.

For all statistical models of tonal melodies, the dependent variable of normalized log F0 was examined in relation to three fixed effects: utterance position, word position, and centered, normalized time. For all models, the maximal model which converged included a random intercept for speaker and a random utterance position by speaker slope. Models containing random slope effects for word position or for normalized time either did not converge or resulted in singular model fits. Since the number of observations per speaker varied in the corpus, certain speakers produced too few repetitions to support a complex random effects structure.

4.2 Results

4.2.1 Duration results. Figure 9 shows the effect of word position and utterance position on the duration of onset consonants (left) and vowels (right). For the consonant duration data, there was a

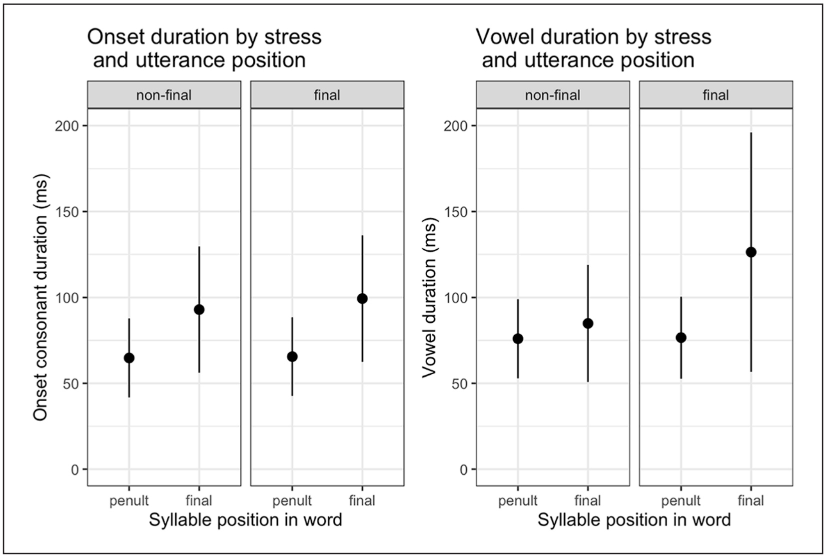

Figure 9. Effect of stress and utterance position on duration. Mean values are plotted with bars reflecting average standard deviation around mean.

significant main effect of word position (penult, final) on onset duration, $t(4.8) = 3.4$, $p < .05$. Onset consonants in the penultimate, unstressed syllable were shorter (65 ms) than those appearing in the ultimate, final syllable of the word (94 ms). The main effect of utterance position was not significant and no significant interactions were found. For the vowel duration data, there was a significant main effect of word position, $t(5.1) = 5.9$, $p < .01$, and utterance position, $t(4) = 3.9$, $p < .05$. Vowels appearing in the penultimate syllable were shorter (76 ms) than those appearing in the final syllable (95 ms). Vowels appearing in a word in utterance-final position were longer (101 ms) than those appearing in a word in non-utterance-final position (80 ms). A significant interaction between syllable and word position was also found, $t(1004) = 8.3$, $p < .001$. Vowels in the final syllable of non-utterance-final words were just 9 ms longer than those in penults (76 vs. 85 ms), while vowels in the final syllable of utterance-final words were 50 ms longer than those in penults (77 vs. 126 ms).

4.2.2 Results: Tonal melodies. Figure 10 shows the influence of word and utterance position the production of three tone melodies (/1.1, 3.2, 3.4/) in the YM corpus data. For tonal melody /1.1/, there was a significant main effect of utterance position on normalized F0, $t(4.9) = 4.1$, $p < .01$. Tone /1/ produced in utterance-final words was lowered approximately 0.45 *sd* relative to tone /1/ produced in non-utterance-final words. A significant main effect of word position was also observed, $t(2086) = -11.3$, $p < .001$. Tone /1/ in stem-final syllables was 0.72 *sd* lower than the same tone on the penultimate syllable. A significant interaction between these main effects was also observed, $t(2109) = 3.0$, $p < .01$. The effect of utterance-final lowering was stronger in stem-final syllables (and incidentally, with the final, tone /1/ clitic pronoun too) than in penultimate syllables (lowering of 0.53 *sd* vs. 0.3 *sd*).

For tonal melodies /3.2/ and /3.4/, we are not interested in the main effect of syllable position on tone because the tones differ across the syllables within the word, but we are interested in the interaction between word and utterance position. For tonal melody /3.2/, the main effect of utterance

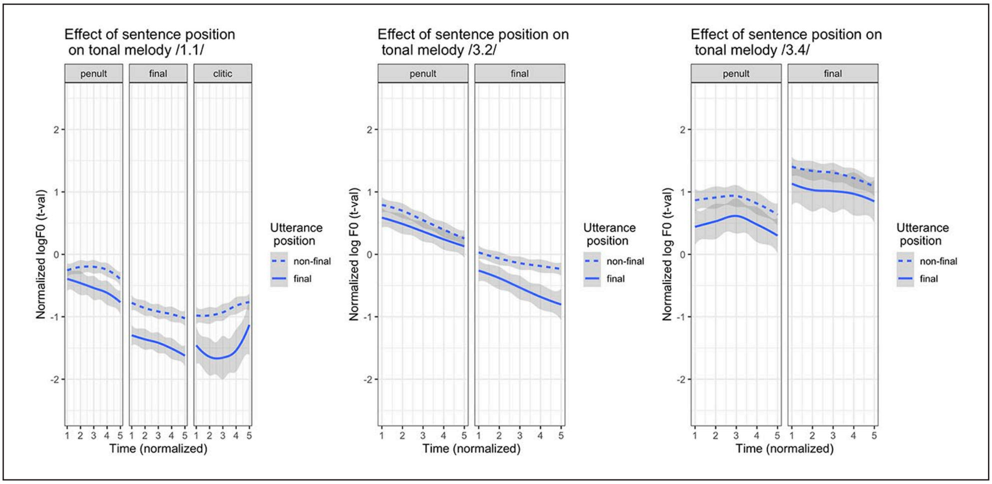

Figure 10. Effect of utterance position on F0 in disyllabic words. Gray bars around mean values reflect standard error.

position was not significant but there was a significant interaction between word and utterance position, $t(1951) = 3.4$, $p < .001$. While the initial tone /3/ did not differ significantly between utterance positions, the final tone /2/ was significantly lower in utterance-final position than in utterance non-final position. For tonal melody /3.4/ the main effect of utterance position was not significant and there were no significant interactions between utterance and word position on normalized F0.

4.3 Discussion: spontaneous corpus data

The duration data examined in the corpus in Experiment 2 resemble the findings from the repetition task in Experiment 1. First, onset consonants in word-final syllables were longer relative to those in penultimate syllables. Like the results in Experiment 1, this mirrors findings in DiCanio et al. (2018, 2019) showing that word-medial position (the onset of stressed syllables) is the locus of stress-related lengthening in YM. Unlike Experiment 1, there does not appear to be an effect of utterance position on onset duration. Second, the vowel duration patterns are virtually identical to those observed in Experiment 1. Vowels are lengthened in stem-final, stressed syllables but this lengthening is much stronger in utterance-final position than in utterance-medial position. Taken together, the results here add additional confirmation of our hypothesis regarding stress-related lengthening—it is greater in phrase-final position.

Methodologically, it also appears that stress-related lengthening is robust across different recording procedures. It is found in words under narrow, contrastive, and broad focus in response to questions following a short narrative (DiCanio et al., 2018), in responses to repetition tasks elicited by native and non-native speakers (DiCanio et al., 2019, Experiment 1), and here, in spontaneous speech. While repetition tasks might result in an exaggeration of the degree of prosodic lengthening, they do not seem to influence the underlying phonetic pattern for YM speakers.

The F0 data in Experiment 2 also resemble the findings from Experiment 1. First, tones /1/ and /2/ undergo a pattern of utterance-final lowering which was stronger in the final syllable of the word than in the penult. Second, like the results in Experiment 1, tone /1/ produced in the penultimate syllable was 0.7 *sd*. higher than the same tone in the final syllable. These findings suggest that utterance-final lowering for the lower tones (/2, 1/) is robust across different recording contexts (repetition, unscripted narrative).

However, unlike the results in Experiment 1, the corpus data did not reveal any significant pattern of tonal raising in utterance position for tone /4/. One possible reason for this is that the target word for this tone, /i³tũ⁴/ "little," was often exclusively found in reference to the main character of "frog, where are you?," a little boy. Here it functioned as a kind of pronominal referent throughout the text—"the little (one)." Though we excluded overt function words from our analysis because they tend to be reduced in running speech (Pitt et al., 2005), the syntactic function of this word may have inhibited patterns of phrase-final tonal raising. While more work on this question is needed, function words generally are weak targets for patterns of prosodic strengthening (pitch accent placement, boundary-adjacent lengthening) (German et al., 2006).

5 Experiment 3: Declination in tone-controlled utterances

One unanswered question thus far is the extent to which the observed positional effects are strictly local or reflect broader utterance-level F0 shifts. In order to distinguish these two types of effects in a lexical tone language, one accepted method is to examine F0 movement across utterances consisting of identical level tones (Connell, 2011, 2017; Lindau, 1986).

5.1 Methods

5.1.1 Design and stimuli. The elicitation methods were identical to those used in the previous experiment. For the declination effect task, Castillo García, a native speaker and a linguist, produced the target sentence. After his production, each speaker repeated the target sentence four times. Ten tone-controlled sentences were constructed with one of three tonal patterns: all tone /4/, all tone /3/, and all tone /1/. With the exception of the sentences consisting entirely of tone /4/, four sentences were elicited for each tonal pattern. Only two sentences with tone /4/ were elicited. It was difficult to construct sentences in YM consisting only of tone /4/ since most verbs include a lower tone, many pronomial clitics include lower tones, and relatively few nouns contain tone /4/. The sentences used for the declination task are given in the Appendix. The elicited sentences varied slightly in overall length. With one exception, all were between 5 to 7 syllables long and minimally contained three morphemes. Each utterance was repeated four times by each speaker (10 sentences $\times$ 4 repetitions), totalling 360 analyzed sentences for all nine speakers.

5.1.2 Speakers and recording. The same nine speakers that completed the first experiment took part in the second experiment. All speakers were transported from Yoloxóchitl to the nearby town of San Luis Acatlán for recording purposes. Recording took place in a quiet room in San Luis Acatlán. The speaker and Castillo García were recorded on separate audio channels, each wearing a Shure SM10A head-mounted microphone. Acoustic recording was done on a Marantz PMD 661 Solid state recorder with a 16 bit sampling rate at a 44.1 kHz sampling frequency.

5.1.3 Analytical techniques. The recording and transcription procedures for experiment 3 were identical to those from experiment 1. All statistics were calculated using R (R Development Core Team, 2017). Two types of analyses were undertaken: (1) analyses to determine global patterns of declination in the target sentences and (2) analyses which specifically examine the factors contributing to F0 changes. The first type answers the research question "Is there global declination?" while the second answers the research question "Is this declination explained by local processes of utterance-final F0 shift observed in Experiment 1 or by additional factors?" For both analysis types, all F0 data were converted to $\log_{10}$ values and statistically normalized (z-score normalization) to balance

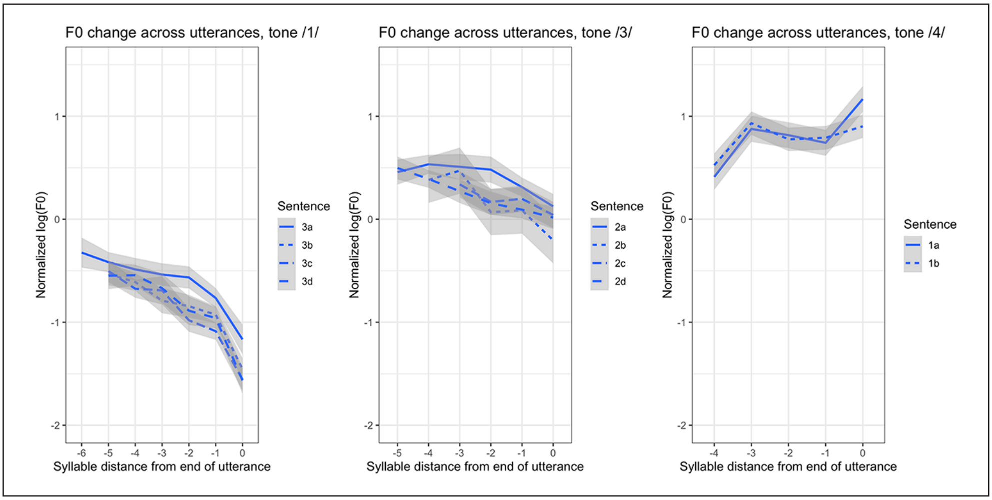

Figure 11. Utterance-level F0 measurements for sentences consisting entirely of level tones (/1, 3, 4/). F0 measurements here reflect the average F0 for the medial 60% of each vowel in the utterance. Sentence reflects different target sentences (items) of varying length. The x-axis shows the number of syllables in the utterance, measured from right to left (i.e., a value of "−2" means it is the antepenultimate syllable in the utterance).

individual speaker differences in F0 range and level prior to statistical analysis. We extracted F0 at five time points and those values across the medial 60% of the vowel duration (time points 2–4) were averaged for each vowel and treated as the dependent variable. This method ensured that little or no effect of glottalization or onset consonant would influence the F0 measurements. Moreover, transitional movements between different level tones are minimized this way.

For the first type of analysis, the data were analyzed using a linear mixed effects model. For the level tone sequences, the fixed effects were tone (/1, 3, 4/) and the distance of the given syllable from the end of the utterance (0–6 syllables away). The maximal model which converged here possessed a random intercept for speaker and a random intercept for item/sentence (1a, 1b, 1c, etc.). Models with random slopes did not converge. The fixed effects were evaluated using *lmerTest* (Kuznetsova et al., 2017), a model which relies on the Satterthwaite method to approximate the degrees of freedom and reports both an F statistic and p-values via the ANOVA wrapper function, but only a lower bound on degrees a freedom. There is currently no way to approximate the upper bound on degrees of freedom for linear mixed effects models (Baayen, 2008). For the second type of analysis, model comparison, was done to determine whether syllable finality (utterance-final vs. utterance-non-final) improved the initial model. Model comparison is described in greater detail within its sub-section below.

5.2 Results: Patterning across the utterance

A plot of the level tone melodies is shown in Figure 11. The results of the linear mixed effects model showed a significant main effect distance from utterance endpoint on the F0 of the level tonal patterns, $t(3931) = -22.9$, $p < .001$. In comparison with the sequence of tone /1/ (the model intercept), there was a significant main effect of tone as well, for tone /3/, $t(9.3) = 20.9$, $p < .001$, and tone /4/, $t(9.2) = 29.8$, $p < .001$. A significant interaction between tonal sequence and distance

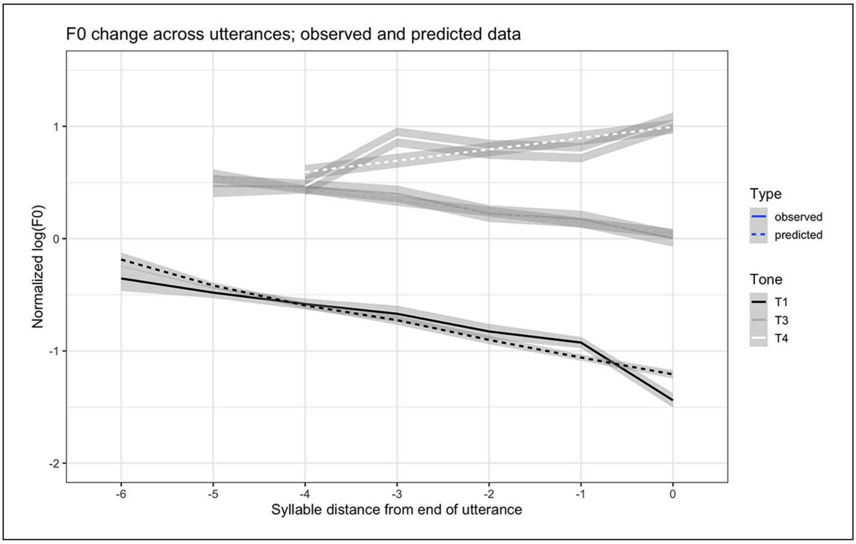

Figure 12. Averaged utterance-level F0 measurements for sentences consisting entirely of level tones (/1, 3, 4/). F0 measurements here reflect the average F0 for the medial 60% of each vowel in the utterance. The line labelled *observed* reflects the average F0 trajectory for the three level tone conditions. The line labelled *predicted* shows the predicted F0 trajectories from the full mixed effects model containing F0 values up to the end of the utterance for each of the tonal sequences. The x-axis shows the number of syllables in the utterance, measured from right to left (i.e., a value of "−2" means it is the antepenultimate syllable in the utterance).

was also found for tone /3/, $t(3935) = 3.7$, $p < .001$, and tone /4/, $t(3929) = 16.8$, $p < .001$. Whereas the sequence consisting of tone level /1/ involved substantial F0 declination across the utterance (−1.15 *sd*), the sequence consisting of tone level /3/ involved less declination (−0.47 *sd*) and the sequence consisting of tone level /4/ involved an F0 rise (0.57 *sd*). With respect to the tone /4/ sentences, a note of caution is necessary - while the utterance-initial tone /4/ is lower than the following syllables and the utterance-final tone /4/ is higher than the preceding syllables, the intervening syllables are fairly level in F0 (falling slightly, −0.14 *sd*).

5.3 Results: Modeling declination

In Figure 12, we see the average F0 trajectories for the three level tones alongside the predicted values from the linear mixed effects model described above. Note that the observed and predicted lines correspond closely for the sequence consisting of tone level /3/. This suggests that the processes influencing F0 production locally in phrase-final position are an extension of more global process of declination across the utterance; one can not distinguish these two processes. For sequences of tone level /4/, a different pattern is observed. The predicted fit closely matches utterance-initial and utterance-final F0 values, but does not match the observed plateau between these two points. This suggests that the processes affecting the starting and endpoint F0 values for these sentences are independent, local processes and not related to an utterance-level pattern of F0 raising. Rather, utterance-initial tone /4/ is lower than subsequent identical tones and, as we observed in Section 3.2.2, tone /4/ undergoes a local process of phrase-final F0 raising.

Table 4. Comparison of linear mixed effect models. The original model lacks a two-way categorical factor of Utterance-Finality (final/non-final).

	Df	AIC	BIC	logLik	deviance	χ^2	Df	*p*
Original model	9	2056.9	2113.5	1019.47	2038.9			
$Model_2$	10	2008.6	2071.5	−994.32	1988.6	50.287	1	<.001
$Model_3$	12	1957.3	2032.7	−966.64	1933.3	55.354		<.001

Df, degrees of freedom; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; logLik, Log-likelihood ratio; χ^2, Chi-squared statistic; p, probability.

For sequences of tone level /1/, the predicted fit closely matches the observed values until the end of the utterance. In utterance-final position, the predicted values remain higher than the observed values. This suggests that final F0 lowering for these sequences is not simply an extension of an utterance-level pattern of F0 declination, but an independent process affecting phrase-final tones, as discussed in Section 3.2.2.

While Figure 12 suggests that local processes of lowering and raising contribute to F0 production for sequences consisting of high (/4/) and low (/1/) tones, we may test this observation more explicitly. Two additional statistical models were compared to the model described above. $Model_2$ contained an additional two-level fixed effect of Utterance Finality (non-final / final). $Model_3$ was identical to $Model_2$ but it also contained an Utterance Finality by Tone interaction term. These models were compared to the model shown above using log-likelihood comparison, using maximum-likelihood estimation. The model with the lowest AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) was chosen as the best fit to the observed data (for a discussion of similar model comparison methods, see Matuschek et al., 2017). The model results are shown in Table 4

The models containing the categorical factor of finality (final/non-final) were significantly better than the original model which excluded this factor. The results of the statistical comparison demonstrate that the third model best fits the observed data, with AIC and BIC values much lower than those found in the original model. Alongside the continuous predictor of syllable distance (from boundary), both the main effect of finality and the interaction of finality with tone were significant predictors.

5.4 Discussion: Experiment 3

The results from Experiment 3 show a significant pattern of declination across utterances in Yoloxóchitl Mixtec, though this effect was limited to tones /3/ and /1/. The pattern observed with the highest tone was distinct. Similar to Mandarin Chinese (Xu, 1999) and Taiwanese (Peng, 1997), utterance-level declination does not occur with high tones in Yoloxóchitl Mixtec. For certain tones, the utterance-level declination patterns are distinct from local processes producing F0 range expansion in utterance-final position. For tone /4/, utterance-final raising is not an extension of an utterance-level process (where there is an F0 plateau), but it is strictly local. For tone /1/, utterance-final lowering is also not simply an extension of a general declination process. Though declination occurs in sentences consisting of tone /1/, the slope of the F0 trajectory abruptly decreases between the penultimate and ultimate syllables of the utterance. For tone /3/, however, the pattern of utterance-final lowering appears to be an extension of a subtle utterance-level process. Broadly, these results demonstrate how local and global intonational effects independently influence the F0 of tone in Mixtec words.

6 General discussion

6.1 The prosody-tone interface in Yoloxóchitl Mixtec

The results from the experiments here confirm the hypotheses discussed in Section 2.6. First, word-final syllables were lengthened regardless of utterance position but these syllables underwent even greater lengthening in utterance-final position than in utterance non-final position. This was observed in both experiments 1 and 2. As discussed above, this lengthening was qualitatively distinct from the lengthening observed under focus (DiCanio et al., 2018). In the current study, lengthening occurred predominantly on the final vowel instead of on the final syllable's onset. Domain-final lengthening is well-attested across languages (Barnes, 2006; Edwards et al., 1991; Nakai et al., 2009; Rakerd et al., 1987). Similar to the findings in DiCanio et al. (2018); Heldner and Strangert (2001) for Yoloxóchitl Mixtec and Swedish, respectively, the onset consonant duration was consistently lengthened under stress.

Second, the results of Experiments 1 and 2 confirm that tones with final offsets at the limits of the tonal range (/4, 14, 24, 1, 14, 13/) are more likely to undergo F0 changes in phrase-final position than other tones were (/3, 2, 32/). These results are in agreement with previous studies showing that phrase-final F0 changes are restricted to either the highest or lowest tones in a speaker's range (Peng, 1997; Zerbian, 2017). Unlike many of the tonal languages surveyed where all tones lowered in phrase-final position (Kipare (Herman, 1996), Moro (Chung et al., 2016; Rose and Piccinini, 2017), Embosi (Rialland & Embanga Aborobongui, 2017), Chichewa (Downing, 2017; Myers, 1996), Tumbuka (Downing, 2017), Chimiini (Kisseberth, 2017), and Shingazidja (Patin, 2017)), Yoloxóchitl Mixtec speakers avoid neutralizing lexical tones in phrase-final position. Grosso modo, tonal contrasts are enhanced, in agreement with Zhang (2001).

Third, the results confirmed that contour tones (specifically rising tones) were more likely to undergo patterns of time-dependent tonal compression than level tones were, in accordance with predictions from Sundberg (1979); Xu and Sun (2002); Zhang (2001). Both tones /13/ and /24/ underwent prosodically-conditioned leveling in non-phrase-final position, a finding consistent with research on Taiwanese Mandarin (Cheng & Xu, 2015). This process of *tonal reduction* is also methodologically important. Since Yoloxóchitl Mixtec possesses no pitch accents, examining the degree of tonal reduction among contour tones (or other complex tones) may serve as a metric for degree of prosodic prominence in the language.

The results from the third experiment partially confirmed the hypothesis that utterance-level patterns of F0 movement would affect tonal sequences at the limits of the tonal range (/1, 4/) moreso than tonal sequences in the middle of the tonal range. Utterance-level declination occurred across sentences consisting of tone /1/ and tone /3/ but not for sentences of tone /4/. This finding closely coincides with previous findings for Mandarin (Xu, 1999) and Taiwanese (Peng, 1997), where sequences of high tones do not undergo declination. For the tones /4/ and /1/, the local phrase-final changes to F0 are independent from utterance-level declination effects. For tone /3/, there is no distinct process affecting tones in phrase-final position (i.e., any F0 changes between the penult and final syllables of the utterance are simply an extension of declination over a larger constituent—here, a sentence/utterance).

One thing absent from Experiment 3 is a critical examination of the time-course of global declination processes. In languages like Wenzhou Chinese, they have been shown to be sensitive to syntactic constituents (Scholz & Chen, 2014b). Wenzhou speakers will reset F0 declination when a new major syntactic constituent occurs in the utterance. Each of the utterances in Experiment 3 are fairly short and each is a simple clause. This is partly for practical reasons - given the complex tonal inventory and the tonal morphology of the language, it is difficult to construct long utterances in YM consisting of the same tone. Moreover, it may be more difficult for listeners reproducing such sentences to recall or reproduce the entire utterance if it were longer. We have currently examined declination as a global, utterance-level process, but it may be sensitive to different types of syntactic boundaries as well.

6.1.1 Articulatory control mechanisms. The major findings within the current study can be understood to result from changes to dynamic parameters within a task dynamics model of speech production (Saltzman and Munhall, 1989; Saltzman et al., 2008). In particular, changes in tonal F0 corresponding with utterance position co-occur with adjustments in duration between these positions. One dynamic parameter, *shrinking*, reflects "a change in both target (amplitude) and stiffness which are scaled proportionally." (Cho, 2006, p. 6). This particular parameter is arguably the control mechanism for prosodically-conditioned kinematic variation (Cho, 2006; Byrd, 2000) and it seems to closely correspond with the observed patterns in the Yoloxóchitl Mixtec data. Under this control parameter, the magnitude of the F0 excursion is correlated with the size of the durational window (for a discussion of this linkage in relation to segmental lenition, see Parrell, 2014).

The mechanism controlling the patterns of tonal reduction is arguably distinct. The rising tones, as shown in Figure 5, have a rising F0 trajectory in utterance-final position but appear to undergo leveling when non-utterance-final. Extreme tonal contraction is not uncommon in speech production, especially in contexts where multiple tonal targets must be reached in a short time window (Cheng & Xu, 2015; Xu, 1994). Moreover, F0 rises require more time than level or falling trajectories. Thus we might expect rising tones to be restricted to contexts with longer phonetic duration, such as in phrase-final position (Sundberg, 1979; Zhang, 2004).

Yet, curiously, F0 leveling occurs for some speakers even in utterance-final contexts. The rising tone data are examined in more depth in Figure 13 for individual speakers. For tonal melody /4.24/, tone /24/ undergoes leveling for all nine speakers in utterance non-final position. Though, for two out of nine speakers, the slope of this tone is level even in utterance final position (speakers CTB and CTC). For tonal melody /4.13/, tone /13/ is realized with a rise for only three of the six speakers in utterance final position.[13]

The process affecting the final rising tones may be a case of emerging tonal change for some speakers and allotonic, durationally-induced F0 leveling for others. Why might speakers produce a final rising tone as a level tone even in phrase-final contexts? Articulatory undershoot induces phonetic leveling in the segmental domain (Parrell, 2014; Mücke & Grice, 2014), and a similar process appears to take place here, where neither the low nor the high F0 targets are reached. This variation across speakers appears similar to cases of segmental lenition which occur in longer durational contexts, as discussed in Parrell (2014), where the author states ". . . Prosodically-conditioned undershoot can lead to sound change if learners misattribute conditioned variability to phonological control instead of prosodic influence." (p. 97). Yoloxóchitl Mixtec speakers may choose to level rising tones in phrase-final contexts even though a durational window is present for producing a clear F0 rise. The current findings suggest that prosodically-conditioned durational compression may be a trigger, but speakers are actively producing levelled tonal variants in contexts with no trigger. While more research on this particular type of variation is required, the reduction and variation here parallels processes leading to sound change in the segmental domain.

6.2 F0 range effects

Across tonal languages, the shape and size of individual tonal inventories varies substantially. This cross-linguistic variation leads speakers of different tonal languages to exert different degrees of control on how a tone is realized. For instance, in a language with many level tones, there may be

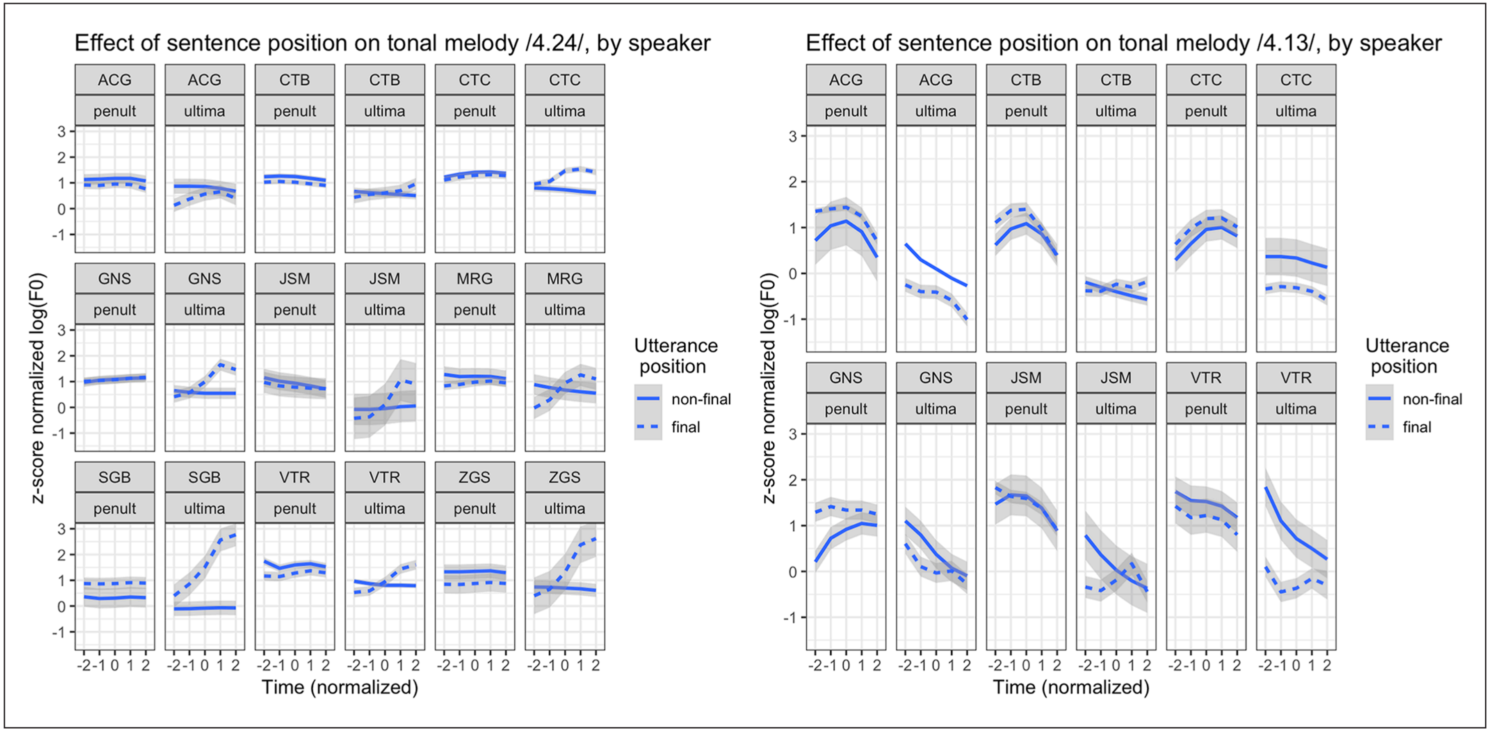

Figure 13. Effect of utterance position on F0 of melodies with final rising tones, organized by speaker. Melodies with a final /24/ tone are shown on the left and melodies with a final /13/ tone are shown on the right.

less flexibility for speakers to shift tones upward or downward. In a language with more contour tones, tonal shape/contour may be maintained at the expense of F0 height (for a computational model prioritizing contour shape, see Xu & Prom-on, 2014). As a consequence of this variation, one can not predict which tones in a given language will vary with prosodic contexts; each language must be independently examined.

Within the context of the current study, we observed a range of distinct patterns of tonal variation related to utterance-finality. In relation to F0 height, we observed that tones with offsets at the edges of a speaker's F0 range (tones /4/ and /1/ here) were more strongly influenced by utterance position than those with offsets at intermediate values (tones /3/ and /2/ here). The highest tones were raised in utterance-final position while the lowest tones were lowered. The global effect of this is an expansion of the tonal space for individual speakers.

This global F0 expansion effect may be considered either a target of the speech production mechanism for YM speakers or an unintentional byproduct of local processes of tonal hyperarticulation. With respect to the former perspective, note that in existing theories of tone production there is no single mechanism responsible for maximizing the relative distance of articulatory-acoustic targets in the phonetic space, even though communicatively-driven or stylistically-driven hyperarticulation seems to do precisely this, especially in vowel systems (see Cho et al., 2011; DiCanio et al., 2015; DiCanio & Whalen, 2015; Wagner et al., 2015). Instead, F0 range expansion is produced via a combination of two distinct control mechanisms, discussed in DiCanio et al. (2018): (1) F0 scaling and range adjustments which target the lexeme and map lexical tones to the articulatory-acoustic space and (2) prosodic lengthening that permits target rescaling and hyperarticulation. While the results here are in line with these production mechanisms, it remains an open question as to how changes in the global F0 range are controlled. Is a single control mechanism responsible or do speakers control several mechanisms in tandem to produce the observed effects?

6.3 F0 scaling reconsidered

The traditional view of F0 scaling within intonational phonology is that, with the exception of downstep, it is *paralinguistic* and not part of the phonological representation of tonal events (for an overview, see Ladd (2008, 304–309)). Whereas pitch accents, boundary tones, and lexical tones are intrinsic parts of the tonal string, adjustments to F0 range are not specified within the tonal string and are, as a result, considered extrinsic to the phonological specification of tone. Within this view, adjustments to tonal scaling are a systematic way that speakers will hyperarticulate prosodic distinctions —for instance, via a universal constraint like the *effort code* (Gussenhoven, 2004), which maximizes articulatory effort in prosodically prominent contexts.

The current article demonstrates that range expansion is systematically used at utterance boundaries in Yoloxóchitl Mixtec in a similar way to how it is used under narrow and contrastive focus contexts (DiCanio et al., 2018). Yet, unlike what has been observed in many non-tonal languages, no specific prosodic category (i.e., no pitch accent) is being targeted by such adjustments. Moreover, adjustments to F0 range occur in precisely those contexts where one typically observes pitch accents and boundary tones. This is not a new finding within the literature on tonal languages. Adjustments to F0 scaling under focus also occur in Mandarin (Chen, 2006; Chen & Gussenhoven, 2008) and in Akan (Kügler & Genzel, 2011) (in the latter case contrastive focus is realized via F0 lowering). Like Yoloxóchitl Mixtec, neither of these lexical tone languages mark focused constituents with pitch accents.

Together, the findings on F0 range adjustment in tonal languages suggest that these adjustments are controlled processes with language-specific implementations. Speakers of tonal languages use range expansion to enhance tonal contrast and distance. While more research on this question is warranted, a likely view is that speakers control the degree of range expansion in prosodically-prominent contexts and this is part of the intonational phonology of tonal languages, but a gradient aspect. This view fits closely with what Ladd identifies as a *metrical* factor in the scaling of tonal targets, defined as “a localisable linguistically meaningful modification of the tonal space controlled by relations in prosodic structure” (Ladd, 2008, 306).

6.4 Methodological constraints

Experiments 1 and 3 involved sentence repetition tasks. One of potential criticisms of this method is that speakers may mimic speech that they are repeating rather than produce the speech naturally. Leaving aside whether or not speakers of Yoloxóchitl Mixtec, in particular, will mimic speech this way,[14] there are a few alternatives to the methods chosen. Experiment 2 attempts to address the methodological issue of Experiment 1, but it also demonstrates a the limit to relying on spontaneous speech corpora for prosodic research on complex tonal languages. Namely, not all tones are equally likely to occur in phrase-final contexts.

A sentence translation task (from Spanish) is another possible alternative to a repetition task chosen, though this type of task is most appropriate where the population is younger and has more balanced bilingualism where they might make more consistent choices for translating particular words. In our experience, translation tasks in elicitation with older speakers (the population here) often involve a shift in word order to Spanish (subject-verb-object) as opposed to the native verb-subject-object order of YM; confusion between the tense system of Spanish and the tense-aspect system of Mixtec; and difficulty ensuring that the information structure of the target sentence is maintained in translation. For instance, the sentence “Él comió piña” (“He ate pineapple”) necessarily requires an overt subject pronoun, “él”(“he”) in Spanish to ensure that the gender of the obligatory YM clitic pronoun is specified in the expected translation /ni^1-ʃa^{3}ʃi^4=ɾa^2 ʃu^{4}ɲu^4/ (perf-eat=3S.Masc pineapple). Yet, this construction in Spanish requires narrow focus on the subject and can elicit the unintended construction in YM /ta^{1}a^3 ni^1-ʃa^{3}ʃi^4=ɾa^{2}ʃu^{4}ɲu^4/ (man perf-eat=3S.Masc pineapple) “The man, HE ate pineapple.” Attempting to train speakers in this nuance, especially older ones with limited experience working with researchers, is quite difficult. As an alternative, a repetition task where the participant is speaking with a native speaker (our third author), avoids the issues that come with translation.

The results from the spontaneous speech analysis in Experiment 2 mostly replicated those from Experiment 1. We believe that this discovery lends some support to our choice of a repetition task as a possible method for investigating prosody with the population examined here. Though, it is nonetheless a constraint to the current research. We see no alternative for the repetition task in Experiment 3, as all tones in the utterance were fixed, following similar methods used in past research (see Connell 2011, 2017). At the present time, alternative methods do not seem to exist for carefully controlling the prosodic context while eliciting speech from illiterate populations. However, we also question why the alternative available for literate populations, read speech, need be standard for prosodic research. Even exemplary research on speech prosody often involves a reading task (e.g., in Mandarin (Chen & Gussenhoven, 2008; Xu, 1999), Guaraní (Clopper & Tonhauser, 2013), Arabic (de Jong & Zawaydeh, 2002), German (Mücke & Grice, 2014), and Dutch (Peters et al., 2014)). Though, just like repetition, read speech is phonetically distinct from elicited and spontaneous speech (Keating & Huffman, 1984; Koopmans van Beinum, 1980). A comparison of methods for eliciting prosodic distinctions across different populations of speakers is likely to be a fruitful avenue for future research.

7 Conclusions

Through a series of experiments, the current article examined the realization of tonal melodies in Yoloxóchitl Mixtec, an endangered languages spoken in Mexico, in utterance-final and non-final contexts. We found evidence for all three types of tonal-prosodic interactions highlighted in Section 2. First, phrase-final F0 shift occurs on tones produced in utterance-final position. Speakers lengthened utterance-final syllables and expanded their F0 range here. The effect of this range expansion was the raising of the highest level tone in the tonal space, a lowering of the lowest level tone, and a lowering of final falling tonal contours. Second, we found evidence for positional allotony for rising tones, which underwent substantial tonal leveling when produced in a non-utterance-final context. Third, we found distinct evidence for utterance-level, global F0 shift. Controlling for level tone across the utterance, we found a pattern of declination for sentences consisting of low tone /1/ and mid tone /3/, but no effect for high tone /4/. Tonal effects occurring at utterance boundaries were not extensions of more global patterns of declination or raising since local patterns of F0 range expansion which affected tones /4/ and /1/ were independent from the broader patterns across the utterances for these tones.

An additional noteworthy finding in the current article was the relation between tonal reduction and prosody. In tonal languages possessing no intonational pitch accents and boundary tones, it may be difficult to find evidence for prosodic constituency and structure. Examining the degree to which lexical tones are hypo- and hyper-articulated (see Cheng & Xu, 2015) in a range of contexts may provide such evidence in much the same way that patterns of vowel reduction co-occur with degrees of prosodic prominence (Chen, 2008; de Jong, 1995; de Jong & Zawaydeh, 2002). The articulatory control mechanisms responsible for the patterns tonal reduction and F0 range expansion closely match those argued for in work on segmental lenition, namely prosodically-conditioned undershoot and hyperarticulation (Parrell, 2014).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by NSF Grant #1603323 (DiCanio, PI) at the University at Buffalo. This manuscript also benefitted from the commentary provided by Wei-Rong Chen and audiences at the University of Toronto and PhonFest at the University of Indiana.

ORCID iD

Christian DiCanio https://orcid.org/0000-0001-9164-188X

Supplemental material

Supplemental material for this article is available online.

Notes

1. An anonymous reviewer points out that utterance-final position is rarely the position of tonal contrast neutralization. While we agree with this sentiment more generally, tone languages with boundary tones which replace a final lexical tone would seem to fall into this category, for example, Serbo-Croatian (Inkelas & Zec, 1988). Lexical tonal contrast can be neutralized by an intonational boundary tone.
2. See Clements (1979) for an earlier discussion of upstep as well.
3. See Connell (2011) for an overview of the literature on F0 declination and downstep in tone languages.
4. We distinguish here between what is traditionally called *tone sandhi*, the phonological alternation of a particular tone in the context of an adjacent tone, and *positional allotony*, which is not triggered by a

particular tone, but by phrase position. A third type of change, morphophonological sandhi, is ubiquitous within Otomanguean languages but often undiscussed in the literature on tone sandhi.

5. An alternative description of this process is that tone /21/ is realized as /213/ in phrase-final contexts.
6. All these past statements on Mixtec languages reflect impressionistic research.
7. See Hyman (2007) and Zhang (2001) for additional examples.
8. Tone /4/ is high and /1/ is low.
9. This distribution is rather odd, cross-linguistically, since falling tones are restricted to root-final position and rising tones have an unrestricted distribution. The expected distribution is the opposite (see Zhang, 2001).
10. Why does focus target the vowel in these languages? It might relate to the fact that Mixtec languages (and Otomanguean languages more generally) are prefixing languages—stem-final syllables would not be reinterpreted as suffixal.
11. This particular effect is not due to pre-low raising (see Lee & Xu, 2016) since both the preceding tone /4/ and tone /1/ were lowered.
12. Note that utterance-final rising tones were not particularly longer than other utterance-final tones. It is for this reason that we discuss a process of contraction here and not utterance-final lengthening.
13. Three speakers' data was excluded here since the duration of the utterance non-final vowels for these speakers was less than 50 ms and no dynamic F0 data were collected from vowels that were this short (see Section 3.1.3 for a discussion).
14. For prosodic mimicry to take place, we must assume that (a) speakers will pay attention to careful phonetic detail when asked to repeat sentences and (b) there is prosodic content to mimic, in particular pitch accents or boundary tones. Within Yoloxóchitl Mixtec, (b) is mostly lacking, so the basis for this criticism is mostly (a). While (a) is assumed in laboratory settings with western, educated, industrialized, rich, and democratic (WEIRD) participants (see Henrich et al., 2010) it is unclear how much this extends to non-WEIRD populations.

References

Akumbu, P. W. (2015). Babanki verb tone. *Studies in African Linguistics*, *44*(1), 1–26.

Amith, J. D., & Castillo García, R. (2019). *Documentation of Yoloxóchitl Mixtec (Glottocode: Yolo1241; ISO 639-3: xty): Corpus, Lexicon, and Grammar*. Linguistic Data Consortium, Philadelphia.

Amith, J. D., & Castillo García, R. (n.d.). *A dictionary of Yoloxóchitl Mixtec*.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to Statistics using R*. Cambridge University Press.

Barnes, J. (2006). *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Mouton de Gruyter.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and the physics of speech* (pp. 152–178). Cambridge University Press.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [computer program]. www.praat.org.

Brunelle, M. (2017). Stress and phrasal prominence in tone languages: The case of Southern Vietnamese. *Journal of the International Phonetic Association*, *47*(3), 283–320.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*(1): 3–16.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*, *120*(3), 1589–1599.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2):149–180.

Cambier-Langeveld, T., & Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, *27*(3), 255–280.

Carroll, L. S. (2015). *Ixpantepec Nieves Mixtec word prosody*. PhD thesis, University of California, San Diego.
Castillo García, R. (2007). Descripción fonológica, segmental, y tonal del Mixteco de Yoloxóchitl, Guerrero. Master's thesis, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), México, D.F.
Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics*, *34*(2), 176–201.
Chen, Y. (2008). The acoustic realization of vowels in Shanghai Chinese. *Journal of Phonetics*, *36*(4), 629–648.
Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, *36*(4):724–746.
Cheng, C., & Xu, Y. (2015). Mechanism of disyllabic tonal reduction in Taiwan Mandarin. *Language and Speech*, *58*(3), 281–314.
Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8: Varieties of Phonological Competence* (pp. 519–548). Mouton de Gruyter.
Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyperarticulation in Korean. *Journal of Phonetics*, *39*(3), 344–361.
Chung, Y., Piccinini, P. E., & Rose, S. (2016). The interaction of polar question and declarative intonation with lexical tone in Moro. In *Proceedings of Speech Prosody 8, Boston, MA* (pp. 615–619). ISCA Archive. http://doi.org/10.21437/SpeechProsody.2016-126.
Clements, G. N. (1979). The description of terraced-level tone languages. *Language*, *55*(3):536–558.
Clopper, C. G., & Tonhauser, J. (2013). The prosody of focus in Paraguayan Guaraní. *International Journal of American Linguistics*, *79*(2), 219–251.
Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? In *Proceedings from Interspeech 201* (pp. 969–972). ISCA.
Connell, B. (2011). Downstep. In M. Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 824–847). Wiley-Blackwell.
Connell, B. (2017). Tone and intonation in Mambila. In L. J. Downing & A. Rialland (Eds.), *Intonation in African Tone Languages* (pp. 132–166). De Gruyter.
de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics*, *30*(1), 53–75.
de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*(1), 491–504.
DiCanio, C., Amith, J. D., & Castillo García, R. (2014). The phonetics of moraic alignment in Yoloxóchitl Mixtec. In *Proceedings of the 4th Tonal Aspects of Language Symposium*. *Nijmegen, the Netherlands*.
DiCanio, C., Benn, J., & Castillo García, R. (2018). The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics*, *68*, 50–68.
DiCanio, C., & Bennett, R. (Forthcoming). Prosody in Mesoamerican languages. In C. Gussenhoven & A. Chen (Eds.), *The Oxford prosody handbook*, Chapter 28. Oxford University Press.
DiCanio, C., Nam, H., Amith, J. D., Castillo García, R., & Whalen, D. H. (2015). Vowel variability in elicited versus spontaneous speech: evidence from Mixtec. *Journal of Phonetics*, *48*, 45–59.
DiCanio, C., & Whalen, D. H. (2015). The interaction of vowel length and speech style in an Arapaho speech corpus. In *Proceedings of the 18th International Congress of the Phonetic Sciences*, *Glasgow, Scotland* (pp. 513–517).
DiCanio, C., Zhang, C., Whalen, D. H., & Castillo García, R. (2019). Phonetic structure in Yoloxóchitl Mixtec consonants. *Journal of the International Phonetic Association*, https://doi.org/10.1017/S0025100318000294.
DiCanio, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, *40*(1), 162–176.
DiCanio, C. T. (2016). *Pitch dynamics 5.0*. http://www.acsu.buffalo.edu/~cdicanio/scripts/Pitch_Dynamics_5.praat.

D'Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, *5*(1226), 1–10.
Downing, L. J. (2017). Tone and intonation in Chichewa and Tumbuka. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 365–391). De Gruyter.
Downing, L. J., & Rialland, A. (2017). *Intonation in African tone languages*. De Gruyter.
Edwards, J., Beckman, M., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*, *89*(1), 369–382.
Everett, K. M. (1998). The acoustic correlates of stress in Pirahã. *Journal of Amazonian Linguistics*, *1*(2), 104–162.
Féry, C. (2013). Focus as prosodic alignment. *Natural Language & Linguistic Theory*, *31*(3), 683–734.
Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 521–602).
Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*(6), 3728–3740.
Garellek, M., Aguilar, A., Caballero, G., & Carroll, L. (2015). Lexical and post-lexical tone in Choguita Rarámuri. In *Proceedings of the 18th International Congress of the Phonetic Sciences*, (pp. 254–258). University of Edinburgh.
Gerfen, C., & Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics*, *33*(3), 311–334.
German, J., Pierrehumbert, J. B., & Kaufmann, S. (2006). Evidence for phonological constraints on nuclear accent placement. *Language*, *82*(1), 151–168.
Grønnum, N. (1991). Prosodic parameters in a variety of regional Danish standard languages, with a view towards Swedish and German. *Phonetica*, *47*, 182–214.
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Research Surveys in Linguistics. Cambridge University Press.
Heldner, M., & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics*, *29*(3), 329–361.
Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61–83.
Herman, R. (1996). Final lowering in Kipare. *Phonology*, *13*(2), 171–196.
Himmelmann, N. P., & Ladd, D. R. (2008). Prosodic description: An introduction for fieldworkers. *Language Documentation and Conservation*, *2*(2), 244–274.
Huang, Y. (2018). *Tones in Zhangzhou: Pitch and beyond*. PhD thesis, Australia National University.
Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics*, *33*(3), 335–362.
Hyman, L. M. (2007). Universals of tone rules: 30 years later. In T. Riad & C. Gussenhoven (Eds.), *Tones and tunes, volume 1: Typological studies in word and sentence prosody* (pp. 1–34). Mouton de Gruyter.
Inkelas, S., & Zec, D. (1988). Serbo-Croatian pitch accent: The interaction of tone, stress, and intonation. *Language*, *64*(2), 227–248.
Josserand, J. K. (1983). *Mixtec dialect history*. PhD thesis, Tulane University.
Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, *44*, 62–82.
Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology 6* (pp. 145–163). Cambridge University Press.
Keating, P. A., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*, *41*:191–207.
Kisseberth, C. (2017). Chimiini intonation. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 225–284). De Gruyter.
Koopmans van Beinum, F. (1980). *Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions*. PhD thesis, University of Amsterdam, The Netherlands., Academische Pers B. V., Amsterdam.

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*(3):430–442.
Kügler, F. (2017). Tone and intonation in Akan. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 89–129). De Gruyter.
Kügler, F., & Genzel, S. (2011). On the prosodic expression of pragmatic prominence: The case of pitch register lowering in Akan. *Language and Speech*, *55*(3), 331–359.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13),1–26.
Ladd, D. R. (2008). *Intonational phonology* (2nd Edition). Cambridge Studies in Linguistics 119. Cambridge University Press.
Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics*, *31*(2), 203–250.
Lee, A., & Xu, Y. (2016). Effect of speech rate on pre-low raising in Cantonese. In C. DiCanio, J. G. Malins, J. Good, K. Michelson, J. J. Jaeger, & H. Keily (Eds.), *Proceedings of the 5th Symposium of the Tonal Aspects of Language* (pp. 75–79). *Buffalo, NY*.
Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff, & R. Oehrle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students* (pp. 157–233). MIT Press.
Lindau, M. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America*, *80*(3):757–764.
Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439. Kluwer.
Lunden, A. (2006). *Weight, final lengthening, and stress: A phonetic and phonological case study of Norwegian*. PhD thesis, University of California, Santa Cruz. Available online as ROA-833, Rutgers Optimality Archive, http://roa.rutgers.edu/.
Maddieson, I. (2013). Tone. In M. Dryer & M. Haspelmath (Eds.), *The World Atlas of language structures online*, chapter 13. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/13.
Makasso, E.-M., Hamlaoui, F., & Lee, S. J. (2017). Aspects of the intonational phonology of Bàsàá. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 89–129). De Gruyter.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*:305–315.
Mayer, M. (1969). *Frog, where are you?* Dial Books.
McKendry, I. (2013). *Tonal association, prominence, and prosodic structure in South-Eastern Nochixtlán Mixtec*. PhD thesis, University of Edinburgh.
Morén, B., & Zsiga, E. C. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory*, *24*, 113–178.
Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, *44*, 47–61.
Myers, S. (1996). Boundary tones and the phonetic implementation of tone in Chichewa. *Studies in African linguistics*, *25*(1), 29–60.
Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, *37*(1), 29–45.
Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica*, *41*, 1–16.
Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 325–347). Cambridge University Press.
Padgett, J., & Myers, S. (2014). Domain generalization in artificial language learning. *Phonology*, *31*(3), 399–433.
Palancar, E. L., Amith, J. D., & Castillo García, R. (2016). Verbal inflection in Yoloxóchitl Mixtec. In E. L. Palancar & J.-L. Léonard (Eds.), *Tone and inflection: New facts and new perspectives* (pp. 295–336). Mouton de Gruyter.

Pankratz, L., & Pike, E. V. (1967). Phonology and morphotonemics of Ayutla Mixtec. *International Journal of American Linguistics*, *33*(4), 287–299.

Parrell, B. (2014). *Dynamics of consonant reduction*. PhD thesis, University of Southern California.

Parrell, B., & Narayanan, S. (2018). Explaining coronal reduction: Prosodic structure and articulatory posture. *Phonetica*, *75*, 151–181.

Patin, C. (2017). Tone and intonation in Shingazidja. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 285–320). De Gruyter.

Peng, S.-h. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, *25*(3), 371–400.

Peters, J., Hanssen, J., & Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics*, *4*, 185–209.

Pike, E. V., & Cowan, J. H. (1967). Mixtec phonology and morphophonemics. *Anthropological Linguistics*, *9*(5), 1–15.

Pike, E. V., & Ibach, T. (1978). The phonology of the Mixtepec dialect of Mixtec. *Linguistic and literary studies in honor of Archibald A. Hill*, *2*, 271–285.

Pike, E. V., & Oram, J. (1976). Stress and tone in the phonology of Diuxi Mixtec. *Phonetica*, *33*, 321–333.

Pike, E. V., & Wistrand, K. (1974). Step-up terrace tone in Acatlán Mixtec. *Advances in Tagmemics*, 81–104.

Pitt, M. A., Johnson, K., Hume, E. V., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*(1), 89–95.

R Development Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 3.3, retrieved from http://www.R-project.org/.

Rakerd, B., Sennett, W., & Fowler, C. (1987). Domain-final lengthening and foot-level shortening in spoken English. *Phonetica*, *44*, 147–155.

Rialland, A., & Embanga Aborobongui, M. (2017). How intonations interact with tones in Embosi (Bantu C25), a two-tone language without downdrift. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 195–222). De Gruyter.

Rose, S., & Piccinini, P. (2017). Intonation in the Thetogovela dialect of Moro. In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 19–50). De Gruyter.

Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333–382.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th conference on speech prosody, Campinas, Brazil*.

Scholz, F., & Chen, Y. (2014a). The independent effects of prosodic structure and information status on tonal coarticulation: Evidence from Wenzhou Chinese. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, & E. van Zanten (Eds.), *Above and beyond the segments: Experimental linguistics and phonetics* (pp. 275–287). John Benjamins Publishing Company.

Scholz, F., & Chen, Y. (2014b). Sentence planning and f0 scaling in Wenzhou Chinese. *Journal of Phonetics*, *47*, 81–91.

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis, A. (Ed.) *Intonation: Analysis, modelling, and technology*. Springer Netherlands.

Sinclair, D. E., & Pike, K. L. (1948). The tonemes of Mesquital Otomi. *International Journal of American Linguistics*, *14*(2), 91–98.

Snider, K. L. (1990). Tonal upstep in Krachi: Evidence for a register tier. *Language*, *66*(3), 453–474.

Sun, X. (2017). *Gradience in contextual tonal variations in Nanjing Chinese*. PhD thesis, University at Buffalo.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, *7*, 71–79.

Sybesma, R., & Li, B. (2007). The dissection and structural mapping of Cantonese sentence final particles. *Lingua*, *117*(10), 1739–1783.

Turk, A., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*(4), 397–440.

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, *25*(1), 25–41.

Wagner, M. (2002). The role of prosody in laryngeal neutralization. In A. Csirmaz, Z. Li, A. Nevins, O. Vaysman, & M. Wagner (Eds.), *Phonological answers* (pp. 357–376). MIT Working Papers in Linguistics.

Wagner, P., Trouvain, J., & Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics*, *48*:1–12. The Impact of Stylistic Diversity on Phonetic and Phonological Evidence and Modeling.

Winter, B. (2015). The Other N: The role of repetitions and items in the design of phonetic experiments. In *Proceedings of the 18th International Congress of the Phonetic Sciences, Glasgow, Scotland* (pp. 181–185). ISCA Archive.

Wittenburg, P., Brugman, H., Russel, A., Klassman, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research*. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. http://tla.mpi.nl/tools/tla-tools/elan/.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *96*(4), 2240–2253.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*(1), 55–105.

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure, and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, *18*(1), 125–159.

Xu, Y., & Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication*, *57*,181–208.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America*, *111*(3), 1399–1413.

Zerbian, S. (2017). Sentence intonation in Tswana (Sotho-Tswana group). In L. J. Downing & A. Rialland (Eds.), *Intonation in African tone languages* (pp. 393–433). De Gruyter.

Zhang, J. (2001). *The effects of duration and sonority on contour tone distribution – Typological survey and formal analysis*. PhD thesis, UCLA.

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In B. Hayes, R. Kirchner, R. M. Kirchner, & D. Steriade (Eds.), *Phonetically based Phonology*. Cambridge University Press.

Zhang, J. (2006). *The phonology of Shaoxing Chinese*. PhD thesis, Leiden University.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone Sandhi. *Phonology*, *27*(1), 153–201.

Appendix A. Declination stimuli

Sentence	Pattern	Yoloxóchitl Mixtec	Spanish/English Gloss
1a	T4	An4 ki4xi(3)=un4 sa4bi4? Q PERF.come=2S rainy.season	"Vienes en tiempo de lluvia?" "Did you come in the rainy season? "
1b	T4	Ta4 xu4ñu4 ku4u4=ra4. agua.CL pineapple be=3.liquid	"Es de piña (el agua)." "It's pineapple (the drink)."
2a	T3	Bi3tin3 ku3xi3 ndi'3i3=ndu3. now POT.eat finish=1P.INCL	"Ahora todos vamos a comer." "Now all of us are going to eat."
2b	3	Bi3tin3 kwa'3a3 be'3e3=ndu3. now POT.make house=1P.INCL	"Ahora se construirá nuestra casa." "Now our house is going to get built."
2c	T3	Le3ka3 ti3in3=na3. palm.leaf.bag trap=3P	"Morral de palma van a hacer (ellos)." "They are going to make a palmleaf bag."
2d	T3	Chi3ñu3 ko3o3 a3sa3=ndu3. work exist do=1P.INCL	"Vamos a tener trabajo." "We are going to do (some) work."
3a	T1	Ni1nu1 ni1-ki1xin1 su1tu1. below PERF-sleep priest	"El sacerdote durmió abajo." "The priest slept down below."
3b	T1	Tio1to1 ni1-ko1so1=ndu1. clothing PERF-take.apart=1P.INCL	"Aventamos ropa." "We took apart the clothing."
3c	T1	Kwi1ya1 ni1-ko1so1=ndu1. year PERF-spill=1P.INCL	"Regamos (agua) por muchos años." "We spilled water for many years. "
3d	T1	Bi1xin1 ni1-ki1xin1=ndu1. hot PERF-sleep=1P.INCL	"Nosotros dormimos calientito." "We slept warmly."