Journal of Phonetics 68 (2018) 50–68

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

# The phonetics of information structure in Yoloxóchitl Mixtec

Christian DiCanio [a,*], Joshua Benn [a], Rey Castillo García [b]

[a] University at Buffalo, United States
[b] Secretaría de educación pública (SEP), Chilpancingo, Guerrero, Mexico

ARTICLE INFO

*Article history:*
Received 26 September 2017
Received in revised form 25 January 2018
Accepted 1 March 2018

*Keywords:*
Tone
Prosody
Information structure
Oto-Manguean
Fieldwork

ABSTRACT

Research on speech prosody has shown that higher-level phonological constituents can be examined directly via their influence on low level phonetic processes (Beckman & Edwards, 1990; Fougeron & Keating, 1997). Despite the strong tradition of research in this area, the existing work has focused mainly on languages which lack lexical tone. This contributes to the view that prosodic structures show little influence on tone, i.e. a language may either have lexical tone or lexical/phrasal stress, the latter of which fits into the prosodic hierarchy. The current paper examines prosodic focus in Yoloxóchitl Mixtec, an endangered Otomanguean language spoken in Mexico. Using experimental data from ten speakers in the field, we investigated how sentence position, stress, and focus type influenced the realization of $F_0$ and duration in different tonal melodies. The findings show that the tonal $F_0$ space was expanded and raised on words produced with contrastive focus, less on words produced with narrow focus, and least on words produced under broad, sentential focus. Focus-related lengthening asymmetrically affected stressed syllables in the language more than unstressed syllables. In stressed syllables, this resulted in an increase in tonal hyperarticulation.



## 1. Introduction

Research throughout the past several decades has shown that lexical tone targets can vary substantially (Andruski, 2006; Chang & Hsieh, 2012; DiCanio, 2012; Gandour, Tumtavitikul, & Satthamnuwong, 1999; Liu & Xu, 2005; Peng, 1997; Scholz, 2012; Xu, 1994; Xu, 1999; Xu & Xu, 2003; Zhang & Liu, 2011). Such variation, either due to coarticulation with adjacent tones and consonants or due to individual speaker differences, has an impact on both the accuracy and timecourse of tone perception (Francis, Ciocca, King Yu Wong, Ho Yin Leung, & Cheuk Yan Chu, 2006; Nixon, Chen, & Schiller, 2014; Peng, Zhang, Zheng, Minett, & Wang, 2012; Xu, 1994). Apart from these local phonological and socio-indexical sources of variation, an additional source of variation in tone production is the informational content of the word or unit to which the lexical tone is assigned. The flow of discourse among speakers requires that certain lexical items be brought to the attention of the listener while others be backgrounded (Baumann, 2006; Lambrecht, 1994). This has an impact on the production of lexical tone (Scholz, 2012; Xu, 1999) as well as the degree to which segmental targets are hyperarticulated (de Jong & Zawaydeh, 2002; de Jong, 1995; Mücke & Grice, 2014).

The current study investigates how information structure influences tone production and the degree to which it is sensitive to stress. These topics are investigated in Yoloxóchitl Mixtec (YM, henceforth; ISO 639 code xty), an indigenous Oto-Manguean language of Southern Mexico (Castillo García, 2007). YM possesses both a complex lexical tone inventory and fixed lexical stress. The relationship between information structure and intonational pitch accents is well-established in non-tonal languages (for an overview, see Baumann (2006), Gussenhoven (2004), Jun (2005), Ladd (2008)), but substantially less is known about how information structure impacts lexical tone production. Moreover, work on non-tonal languages encompasses a typologically-diverse sample of languages, but the existing work on tonal languages is mostly limited to those lacking lexical stress (Kügler & Genzel, 2011; Liu & Xu, 2005; Scholz, 2012; Xu, 1999). YM is different in this regard. If the placement of nuclear pitch accents in non-tonal languages is sensitive to the prosodic hierarchy within the word (Gussenhoven, 2004; Jun, 2005), where do tone languages fit in?

* Corresponding author.
*E-mail address:* cdicanio@buffalo.edu (J. Benn).

https://doi.org/10.1016/j.wocn.2018.03.001
0095-4470/

We investigate the relationship between information structure and tone in YM through a speech production study carried out in the field with a population of ten native speakers. We examined how narrow (argument) focus and contrastive (or corrective) focus are realized via a naturalistic response task (c.f. Clopper & Tonhauser, 2013; Kügler & Genzel, 2011) and separately compare these results to tones produced under broad (sentential) focus via a repetition task. This experiment addresses both an empirical question and a theoretical one. First, how do tones vary in their realization in a complex tone language and with lexical stress? Do unstressed syllables vary more than stressed syllables? Second, how is the prosodic realization of information structure functionally constrained in a lexical tone language? Do tonal type (level, rise, etc) and tonal position within the tonal space (highest tone, lowest tone) matter?

### 1.1. Background: focus

Information structure refers to those components of the linguistic system that interlocutors use to negotiate shared knowledge of entities and states in discourse (Lambrecht, 1994). Languages utilize different strategies for expressing whether an entity is new/old, topical/focal, or recently identifiable (or not) in the discourse. Of primary importance to phonetic and phonological studies of information structure is the realization of *focus*. Focus refers to "the speaker's assessment of the relative predictability or unpredictability of the relations between propositions and their elements in a given discourse situation." (Lambrecht, 1994, p. 6). In utterances produced with broad focus, the entire sentence or predicate conveys pragmatically unpredictable information. In utterances produced with narrow focus, a single argument or state is pragmatically unpredictable.[1] This single constituent may then be linguistically-marked as distinct from others in the utterance. Languages frequently utilize one of three possible tactics for marking focus: morphosyntactic marking, focus particles, and prosodic marking (Ladd, 2008). Additionally, languages like Northern Sotho may mark only pragmatically *predictable* (non-focal) information via backgrounding and pronominalization, leaving focus entirely unmarked (Zerbian, 2007).[2]

In many languages which mark narrow focus with suprasegmentals, the focus domain (c.f. Lambrecht, 1994) may be marked with an intonational pitch accent. This accent is aligned to the most prominent syllable in the phrase via the *focus-to-accent* (FTA) principle (Gussenhoven, 1983a). In this way, utterance-level prosodic distinctions are directly sensitive to stress and the heads of prosodic constituents. Intonational pitch accents "arrange themselves according to the demands of the metrical structure" (Ladd, 2008, p. 268). A simple corollary of this view is the idea that metrically-weak prosodic constituents within words will be less affected by FTA than metrically-strong ones. It is this particular corollary that we investigate in the current study.

There is some debate over the extent to which different focal domains and types are distinguished by speakers/listeners. Bishop (2013) provides an overview of this debate. In terms of focal domain, speakers of English, Dutch, and German distinguish broad and narrow focus in speech production with prosodic features (Baumann, Grice, & Steindamm, 2006; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Gussenhoven, 1983b; Xu & Xu, 2005). However, while listeners may be able to successfully discriminate between focal domains using prosodic cues (Breen, Fedorenko, Wagner, & Gibson, 2010), they are less reliable at using these cues to identify the the context which elicited them (Birch & Clifton, 1995). In terms of focal types (i.e. narrow vs. contrastive focus), there is some debate regarding the extent to which these are phonologically categorical or pragmatically unique (c.f. Büring (2007), Katz & Selkirk (2011) and the references therein). Broad, sentential focus and narrow, nominal focus are distinguished morphosyntactically in YM. In the latter context, the NP is pre-verbal and ex-situ while in the former, the NP is in situ (post-verbal). Yet, both contrastive and non-contrastive (narrow) focus occur ex-situ in the same syntactic position. For the purposes of the current paper, we explore whether this distinction is prosodically marked and compare it with the in situ context.

The general goal of morphosyntactic marking of focus cross-linguistically is to align the constituent with the edge of a prosodic domain (Féry, 2013). This need not involve any particular type of prosodic marking, but in many cases, it does.[3] While a language may be described as marking focus by constituent dislocation, e.g. Italian (Lambrecht, 1994), such dislocation does not preclude prosodic marking at the same time. For instance, speakers of Bilbao Spanish or Central Catalán may front a constituent with narrow focus while simultaneously producing it with a pitch accent (Vanrell & Fernández Soriano, 2013). This same type of "double marking" is found with speakers of Balearic Catalán and Castillian Spanish when producing contrastive focus (ibid). In Zulu, focus is realized both through lengthening of the penultimate vowel of the focused word and via dislocation into the post-verbal position (Cheng & Downing, 2012). Finally, a focused constituent that is pre-posed in English may also be produced with a particular intonational pitch accent, e.g. '**BROCCOLI** *I hate,* **PEARS** *I love.*' (see Prince, 1981). The use of morphosyntax to mark information structure does not preclude prosodic marking. In fact, aligning a constituent to a phrase boundary may be used as a strategy to increase the unit's prosodic prominence (Féry, 2013).

### 1.2. Background: prosodic marking of focus in tone languages

The idea that there are multiple, simultaneous strategies for marking focus is pertinent to understanding prosodic focus marking in tonal languages. From the standpoint of the functional load hypothesis (Berinstein, 1979), one predicts that languages which use tone to mark lexical or morphological contrasts (lexical tone languages) would avoid the use of pitch to mark pragmatic distinctions like focus. However, numerous studies have shown that tone languages can use pitch to mark

[1] The interaction of predictability and focus is specifically explored in Turnbull (2017), to which the reader is referred.

[2] While there is little work on this question, languages appear prima facie to treat constituents with narrow focus as marked (either prosodically or morphosyntactically) and un-focused constituents as un-marked.

[3] See Féry (2013) for additional examples where languages do not mark focus prosodically.

focus, as in Mandarin (Xu, 1999; Liu & Xu, 2005), Wenzhou Chinese (Scholz, 2012), Taiwanese (Pan, 2007), Santa Ana del Valle (SAV) Zapotec (Esposito, 2010), Akan (Kügler & Genzel, 2011), and others.

In Mandarin, focus is marked in situ and modifies the global $F_0$ contour. The effect of this is an expansion of the $F_0$ range for the tone on the focused word along with increased duration (Peng et al., 2005; Xu, 1999; Liu & Xu, 2005). A similar process takes place in Taiwanese (Pan, 2007). In Santa Ana del Valle Zapotec, focus occurs ex-situ (pre-verbal) but also involves a systematic global raising of the $F_0$ contour for all tonal categories (high tone with modal phonation, rising tone with modal phonation, falling tone with breathy phonation, and falling tone with creaky phonation) (Esposito, 2010). In Akan, focus can be marked either in situ or ex-situ. In both syntactic contexts, focus produces register *lowering* for both high and low tones (Kügler & Genzel, 2011).

In Swedish, which possesses a simple tonal distinction between an early and late aligned word accents, prosodic marking of focus varies by dialect (Bruce, 2005). Word accents produced with narrow focus are realized with $F_0$ range expansion and raising of H tones in the Southern and Central dialects. In Eastern and Western dialects, a H phrasal accent is appended after the word accents. In Serbo-Croatian, which possesses a simple tonal distinction between rising and falling word accents, a LH phrase accent replaces the word accent (Godjevac, 2005; Inkelas & Zec, 1988). However, since the accents are aligned differently, the word accentual distinction is maintained despite tonal replacement. The $F_0$ range on the post-focal constituents is also compressed, which Godjevac captures with an abstract phrasal accent Ø. In Curaçao Papiamentu, there is a contrast between final and penultimate stress (Remijsen & van Heuven, 2005). The penult in words with penultimate stress may carry either tone I, realized with a rising-falling contour, or tone II, realized with a level $F_0$. There is no tonal contrast in words with final stress. When tones are placed under focus, the excursion size of lexical tones is increased.

These particular studies highlight a common way in which focus influences lexical tone languages – via changes in $F_0$ scaling. In the case of Mandarin, Taiwanese, and Serbo-Croatian, it is the entire $F_0$ range that is expanded.[4] In the case of SAV Zapotec and Akan, the ranges are raised and lowered, respectively. Within intonational phonology, modifications to pitch range have been alternately argued to be part of the phonological representation within an autosegmental tonal tier (Sosa, 1999) or a paralinguistic effect that is independent of a phonological tonal specification (Beckman & Pierrehumbert, 1986). Ladd (2008) discusses this issue extensively and argues that local changes to $F_0$ range arise from the prosodic structure of the utterance. Higher-level prosodic constituents determine the scaling factor that may be applied to a particular tonal target. If $F_0$ scaling adjustments occur on syllables bearing lexical tone, is this relationship a paralinguistic effect or more properly part of the phonological representation of information structure in the language? This issue remains unresolved.

One way to answer this question is to examine the degree to which $F_0$ range expansion is sensitive to stress. If range expansion is sensitive to the prosodic hierarchy in lexical tone languages as much as intonational pitch accents are argued to be in many non-tonal languages, this would argue in favor of treating $F_0$ scaling differences as phonological in nature and not simply paralinguistic. The implicit assumption here is that paralinguistic effects have a global effect on an utterance or constituent and are not sensitive to internal syntactic or prosodic structure. In other words, it is the entire constituent that is raised/lowered, not the prosodic unit predicted to have greatest metrical strength within the constituent. To date, effects of focus on $F_0$ range in lexical tone languages have mostly excluded a discussion of the prosodic hierarchy. The reasons for this are twofold: (1) there are rather few phonetic studies examining information structure in lexical tone languages and (2) those languages that have been studied more extensively have been East and Southeast Asian.[5] A characteristic of East and SE Asian languages is their tendency for monosyllabic or sesquisyllabic word structure (Brunelle & Pittayaporn, 2012; Matisoff, 1990). Tone in such languages occurs on either a monosyllabic word or on the final syllable of an iamb where the penult lacks a distinct phonological specification for tone. Any effects of focus on $F_0$ scaling here are ambiguous in terms of domain – either focus causes range expansion/movement at the lexical level or it targets the prosodically-prominent syllable; there has been no way to distinguish between these two.

The language investigated in the present study, YM, has polysyllabic word structure with fixed, final stress. Lexical tones contrast on all syllables. As a result, it presents an ideal test case to investigate the effect of focus on tone production. Moreover, with the exception of work on SAV Zapotec (Esposito, 2010), there exists no previous work on the phonetics of information structure in any Otomanguean language. Given the tonal complexity typically found in Mixtecan languages (c.f. DiCanio, 2016b), constraints on functional load predict $F_0$ to play a lesser role in YM than in languages like Akan, Mandarin, or SAV Zapotec. The current study investigates the role of stress on tone production in different focus conditions, serving both the empirical goal of expanding the range of languages for which information structure has been examined and addressing the questions raised above regarding the phonological status of $F_0$ scaling effects.

### 1.3. Background: focal lengthening and word accent

In addition to the $F_0$-related changes mentioned above, focus induces patterns of lengthening on prosodic constituents of different sizes. The degree of lengthening may be asymmetrical within the word and, in certain cases, within the syllable. When a polysyllabic word receives contrastive focus in English (Turk & Sawusch, 1997; Turk & White, 1999), Dutch (Cambier-Langeveld & Turk, 1999), or Swedish (Heldner & Strangert, 2001), the stressed syllable undergoes greater durational expansion than unstressed syllables do. Thus, stressed syllables in non-tonal languages are targets for both intonational pitch accents and accentual lengthening.

[4] Though typically such expansion is asymmetrical where the lower edge of a speaker's range is less mobile than the upper edge (Ladd, 2008, 203).

[5] A notable exception is research on Curaçao Papiamentu (Remijsen & van Heuven, 2005).

Though unstressed syllables are not lengthened as much as stressed syllables when they occur in a focused constituent, they do undergo some lengthening relative to a baseline, unfocused constituent. The literature on this topic has focused specifically on the directionality of this lengthening relative to the stressed syllable. Addressing previous findings in Turk and Sawusch (1997), Turk and White (1999) find that the stressed syllable undergoes 23% lengthening under focus and the following, unstressed syllable undergoes significant, but reduced lengthening (13%). However, very little lengthening is observed in unstressed syllables that precede a stressed syllable, either within the word or across word-boundaries. They conclude that the domain of accentual lengthening in English begins at the stressed syllable and extends rightward to the word boundary. Using similar methods, research on Dutch reached the same conclusions (Cambier-Langeveld & Turk, 1999). Results from a related study on Swedish argued that the domain of lengthening was not the word, but a disyllabic span beginning at the stressed syllable (Heldner & Strangert, 2001). While little accentual lengthening occurred in pre-tonic syllables, the domain of accentual lengthening in Swedish only extended one syllable rightward (not to the word boundary).

Heldner and Strangert's findings on Swedish also differed from previous work with respect to the sub-syllabic domain of lengthening. In English, Dutch, and Mandarin Chinese, vowels and consonants were equally lengthened on a stressed syllable with focal lengthening (Cambier-Langeveld & Turk, 1999; Chen, 2006; Turk & Sawusch, 1997; Turk & White, 1999). In Swedish, if the accented stressed syllable contained a long vowel and short coda consonant (Vː C), the vowel and consonant were also equally lengthened. However, if the accented stressed syllable contained a short vowel and a longer coda consonant (VCː), the adjacent consonants were lengthened to a much greater degree than the vowel was. The authors argue that focus enhances the vowel length distinction in Swedish and the sub-syllabic domain of accentual lengthening is affected because of it. In Northern Finnish, accentual and utterance-final lengthening in disyllabic words is also sensitive to the vowel length contrast (Nakai, Kunnari, Turk, Suomi, & Ylitalo, 2009; Nakai et al., 2012). Short vowels underwent less lengthening when the word is accented than long vowels did. However, lengthening on vowels was inhibited when both syllables contained vowels of the same length, i.e. CVCV and CVVCVV. Nakai et al. (2012) argue that prosodic lengthening is restricted by a *quantity neighbor constraint* whereby the syntagmatic durational relationship across syllables must be maintained. Accentual lengthening is inhibited by the necessity to maintain the quantity contrast within the word.

In a study on Mandarin Chinese, Chen (2006) investigated the domain of focus-related lengthening in four-syllable nonce-word place names. She found greater lengthening when the pragmatic domain of focus was reduced in size. When the syllable was assigned focus, it underwent greater lengthening than when the foot was assigned focus. Support for two types of prosodic units in these phrases was found. At the foot level, there was a trochaic bias, but the final syllable was the locus of greatest lengthening overall. This particular research is relevant as it demonstrates evidence for prosodic lengthening within a tone language which lacks lexical stress.

This work on focal lengthening raises two important questions that we examine in the YM data here. First, is the primary domain of focal lengthening the stressed syllable or the foot? Each of the Germanic languages previously studied have trochaic foot structure and their domains of accentual lengthening begin at the stressed syllable and extend rightward (with limited pretonic lengthening). Languages like YM, with fixed stem-final stress, might show an overall greater degree of pre-tonic lengthening than found for English, Dutch, and Swedish. Second, YM possesses a limited contrast between short and long vowels. Might this constrain the degree of vowel lengthening in a manner similar to the findings for Swedish and Finnish? These questions are addressed in the first study.

### 1.4. Background: Yoloxóchitl Mixtec

The Mixtecan branch of Oto-Manguean contains three daughter language groups: Mixtec, Triqui, and Cuicatec.[6] Yoloxóchitl Mixtec, the focus of the current study, is an endangered language spoken in the towns of Yoloxóchitl, Cuanacaxtitlán, Buena Vista, and Arroyo Cumiapa (Castillo García, 2007), located approximately 20 miles north of the town of Marquelia, Guerrero, along the southeastern coast "la costa chica" of Guerrero, Mexico. There are approximately 4,000 speakers remaining, though many younger speakers are more dominant in Spanish than in YM.

YM possesses a relatively small consonant inventory: /p, t, k, k$^{w}$, tʃ, m, n, m$^{b}$, n$^{d}$, ŋ$^{g}$, s, ʃ/h, β, j, l, ɾ/. There are five contrastive vowels, /i, e, a, o, u/, and nasality is contrastive on all vowels, i.e. /ĩ, ẽ, ã, õ, ũ/ (Castillo García, 2007). Glottalization is contrastive and orthogonal to tone. It occurs in stem-medial position, i.e. between two light syllables, e.g. /na$^{3}$ʔmã$^{3}$/ 'thick', or intervocalically in a monosyllabic word, e.g. /ka$^{3}$ʔa$^{3}$/ 'to ring.' The syllable structure is simple, consisting of only open syllables (V, CV) with (rare) /sC/ clusters.

Content words are minimally bimoraic, consisting of either two light syllables (CVCV) or a one heavy syllable (CVV). Trimoraic stems are attested as well, having the shape CVCV(ʔ)CV or CVCV(ʔ)V, though these are rarer than bimoraic stems and often the result of historical fusion in compounding. As a result of this process, there is a contrast between short and long vowels in word-final syllables, e.g. /n$^{d}$a$^{3}$ʃa$^{2}$/ '*cryer*' vs. /n$^{d}$a$^{3}$ʃa$^{2}$a$^{2}$/ '*to arrive to live*'. Morphological complexity in YM is found mainly in the verbal system. Verbs may possess up to two monosyllabic prefixes marking aspect, negation, or non-productive derivational processes (Palancar, Amith, & Castillo García, 2016). Note that since inflectional prefixes apply to certain monosyllabic verbal stems, the surface contrast between short and long vowels can occur between a prefixed monosyllabic stem and a monomorphemic disyllabic stem, e.g. /ni$^{1}$-tʃi$^{3}$i$^{3}$/ 'PERF-get.wet' vs. /n$^{d}$i$^{3}$tʃi$^{2}$/ 'bean.'

[6] While the latter two groups have limited internal diversification (three and two dialects, respectively, (Anderson & Concepción Roque, 1983; DiCanio, 2008)), Mixtec has extensive internal diversification, possessing roughly sixty distinct varieties spoken in twelve pan-dialectal regions (Josserand, 1983). As a result of this, there are a large number of languages, each of which is labelled "Mixtec", but many of which are as distinct as modern-day Italian and Portuguese. The internal diversification of Mixtec began in the late Preclassical period in Mexico (Josserand, 1983, 458), giving it roughly the same time depth as the diversification of Romance languages, beginning 1800–2000 years ago (Adams, 2007).

The TBU in YM is the mora, but tone is asymmetrical distributed. That is, a greater number of tonal contrasts occur on the final mora of a stem than on the non-final mora. A single mora may be phonologically associated with either a single tone level (/4, 3, 2, 1/, where /4/ is high and /1/ is low) or a contour (/13, 14, 24, 32, 42/). There are two consequences of the association of tones with moras in YM. First, there is a contrast between an underlying rising tone and a derived rising tone on a monosyllabic word (with a long vowel). For instance, /ta$^{13}$a$^{3}$/ '*laid eggs*' is distinct from /ta$^{1}$a$^{3}$/ '*man*' (c.f. DiCanio, Amith, & Castillo García, 2014). Thus, we must distinguish between a rising tone, e.g. /13/, and a rising melody, e.g. /1+3/. Second, since monosyllabic roots and disyllabic roots have an equivalent number of moras, the attested tonal melodies are similar on each root type. There are five possible tones on the initial mora (/1, 3, 4, 13, 14/) and nine possible on the final mora (/1, 2, 3, 4, 14, 13, 24, 32, 42/).[7] Table 1 shows the tonal melodies which surface on both monosyllabic and disyllabic roots. Note that roots with glottalization are excluded here as fewer tonal melody types surface on these words.

Table 1 illustrates how tones may combine on individual moras to create complex word melodies.[8] Up to 27 distinct melodies surface on disyllabic words and up to 20 distinct melodies surface on monosyllabic words. Though in each case the tonal system is simplified if one analyzes each melody as composed of a sequence of two tones, one associated with each mora (see DiCanio et al. (2014) for an in-depth analysis). The table also illustrates that there is an asymmetry in the tonal distribution; while only five tones may be associated with the first mora, nine may be associated with the following mora. This asymmetry in tonal distribution mirrors other phonological asymmetries in polysyllabic words in the language. For instance, nasal vowels contrast only on final syllables, e.g. /ka$^{1}$kã$^{1}$/ 'to ask for' vs. /ka$^{1}$ta$^{1}$/ 'press, mill'; all vowels on non-final syllables are phonologically oral. Moreover, mid vowels are rare in penultimate syllables and the vowel /o/ surfaces in the penult only if it also occurs in the final syllable. Final syllables in YM are also phonetically longer than penultimate syllables (DiCanio et al., 2014). These phonological patterns suggest that stem-final syllables are stressed in YM compared with non-final syllables.[9] There exists no published instrumental work on intonation in YM (though such work is currently in progress). There are no intonational pitch accents in YM, but utterance-level declination may affect tone production. In the current study, this potential effect is ignored.[10]

Cross-linguistically, lexical stress may condition localized hyperarticulation of vowels and articulatory differences in consonant production (Bombien, Mooshammer, Hoole, Rathcke, & Kuhnert, 2007; de Jong, 1995; de Jong & Zawaydeh, 2002; Fougeron & Keating, 1997; Mücke & Grice, 2014) (for an overview, see Gordon (2011)). Though there is an asymmetry in the phonological distribution of tone across syllables in YM, it remains unclear how identical tones are produced in different prosodic contexts. Rather few studies have investigated the degree to which lexical stress conditions localized hyperarticulation of tonal targets. Recent work on Ixpantepec Nieves (IN) Mixtec found that lexical stress was distinguished mainly by vowel duration and vowel quality, with unstressed syllables having both shorter vowel duration and more centralized vowel targets than stressed syllables (Carroll, 2015). $F_0$ was found to play less of a role in distinguishing stressed and unstressed syllables along with other cues like Cepstral Peak Prominence, spectral tilt, and intensity. However, for one of the two speakers that was examined, tones were uniformly raised in the tonic syllable when compared to a pre-tonic and post-tonic syllable (stress in IN Mixtec is penultimate). The current study differs from this recent work on IN Mixtec insofar as it examines how focus influences tone and duration in unstressed and stressed syllables with a larger sample of speakers and in a larger variety of prosodic contexts.

Prior to the current study, there is no published literature on how focus is marked in YM and very little work on other Mixtecan or Oto-Manguean languages. Sentential focus occurs where the entire predicate reflects new information and the individual noun phrase is not specifically under focus; it is a type of broad focus. Narrow, argument focus occurs where the particular noun phrase falls under focus. In YM, noun phrases under sentential focus *always* occur in post-verbal position (YM has VSO word order). However, noun phrases under both argument and corrective focus *always* occur ex-situ; in a sentence-initial or pre-verbal position. Fig. 1 shows examples of the noun /yu$^{3}$βa$^{4}$=õ$^{4}$/ [yu$^{3}$βõ$^{4}$] 'my father' under different focus conditions in YM. Note that argument and corrective focus conditions are identical in YM and include a resumptive 3S clitic pronoun /ɾi$^{4}$/. The argument focus condition was uttered in response to the question /yo$^{3}$o$^{3}$ ni$^{1}$-ta$^{3}$ʃi$^{3}$ kwa$^{4}$-yu$^{2}$ nda$^{3}$ʔa$^{4}$ = $^{2}$//who PERF-give horse hand = 1S/, 'Who gave me the horses?' whereas the contrastive focus condition was uttered in response to the question /ã$^{4}$ si$^{3}$ʔi$^{4}$=$^{2}$ ta$^{13}$ʃi$^{3}$ kwa$^{4}$yu$^{2}$ nda$^{3}$ʔa$^{4}$=$^{2}$//Q mother=1S PERF.give horse hand=1S/ 'Was it my mother who gave me the horses?'

## 2. Speech production study: tone and focus type in YM

To recap, we examine three hypotheses in the current study. First, is the domain of prosodic lengthening the stressed syllable or the word? As a corollary, does the vowel length contrast in YM constrain the degree of prosodic lengthening on short vowels? Second, what is the effect of stress on $F_0$ range expansion and how are register effects associated with information structure? Third, as with the durational effects, is the domain of $F_0$ range expansion and register shift the stressed syllable or the word?

[7] Trimoraic roots are excluded here because they are rare and, as a product of their rarity, they include a restricted set of tonal melody types.

[8] As one observes from the more complex tonal combinations here, tone is heavily used in the language's morphology as well. For an in-depth discussion of YM tonal morphology, see Palancar et al. (2016).

[9] A similar argument is made for lexical stress in Itunyoso Triqui (DiCanio, 2008). Par excellence, evidence for lexical stress in Oto-Manguean languages is primarily based on distributional asymmetries (c.f. DiCanio & Bennett (2018)) as few of the languages possess pitch accents, all of the languages are tonal, and most have polysyllabic roots. Note that this criterion for lexical stress does not contradict one dependent on pitch accents. One may similarly argue that the syllables where all vowels are fully realized in English are stressed.

[10] Preliminary work suggests that the domain most influenced by intonational patterns is utterance-final position. The current study does not examine tone production in this position.

**Table 1**
Tonal melodies in monosyllabic and disyllabic words. Periods indicate moraic boundaries.

| Tonal Melody | Monosyllable | | Disyllable | |
|---|---|---|---|---|
| on word | Word | Gloss | Word | Gloss |
| 1.1 | nda$^{1}$a$^{1}$ | *flat* | ta$^{1}$ma$^{1}$ | *without appetite* |
| 1.3 | ndo$^{1}$o$^{3}$ | *to stay* | na$^{1}$ma$^{3}$ | *to change (intr)* |
| 1.4 | ndo$^{1}$o$^{4}$ | *sugarcane* | na$^{1}$ma$^{4}$ | *soap* |
| 1.32 | | | na$^{1}$ma$^{32}$ | *I will change myself* |
| 1.42 | ndi$^{1}$i$^{42}$ | *pink* | na$^{1}$ma$^{42}$ | *my soap* |
| 3.2 | nda$^{3}$a$^{2}$ | *steep* | na$^{3}$ma$^{2}$ | *wall* |
| 3.3 | nda$^{3}$a$^{3}$ | *go up* | na$^{3}$ma$^{3}$ | *to change (tr)* |
| 3.4 | nde$^{3}$e$^{4}$ | *strong* | na$^{3}$ma$^{4}$ | *sprout* |
| 3.42 | ɲũ$^{3}$ũ$^{42}$ | *night* | na$^{3}$ma$^{42}$ | *I will pile rocks* |
| 4.1 | | | ka$^{4}$nda$^{1}$ | *is moving (intr)* |
| 4.2 | nda$^{4}$a$^{2}$ | *where (Q)* | na$^{4}$ma$^{2}$ | *I am changing* |
| 4.3 | | | na$^{4}$ma$^{3}$ | *it is changing* |
| 4.4 | nda$^{4}$a$^{4}$ | *black* | na$^{4}$ma$^{4}$ | *is piling rocks* |
| 4.13 | nde$^{4}$e$^{13}$ | *they enter* | na$^{4}$ma$^{13}$ | *is changing* |
| 4.14 | ndi$^{4}$i$^{14}$ | *it is burning* | nda$^{4}$ta$^{14}$ | *is splitting up* |
| 4.24 | ni$^{4}$i$^{24}$ | *skinny* | ya$^{4}$ma$^{24}$ | *Amuzgo person* |
| 4.42 | | | na$^{4}$ma$^{42}$ | *I often pile rocks* |
| 13.2 | ʃi$^{13}$i$^{2}$ | *resistant* | hi$^{13}$ni$^{2}$ | *has seen* |
| 13.3 | nda$^{13}$a$^{3}$ | *went up* | na$^{13}$na$^{3}$ | *has photographed oneself* |
| 13.4 | ka$^{13}$a$^{4}$ | *slipped* | na$^{13}$ma$^{4}$ | *has piled rocks* |
| 14.2 | | | na$^{14}$ma$^{2}$ | *I will not change* |
| 14.3 | nda$^{14}$a$^{3}$ | *to not go up* | na$^{14}$ma$^{3}$ | *to not change* |
| 14.4 | | | na$^{14}$ma$^{4}$ | *to not pile rocks* |
| 14.13 | nde$^{4}$e$^{13}$ | *they do not enter* | na$^{14}$ma$^{13}$ | *to not change oneself* |
| 14.14 | sa$^{14}$a$^{14}$ | *to not heat up* | nda$^{14}$ta$^{14}$ | *to not split up* |
| 14.24 | ka$^{14}$a$^{24}$ | *to not slip* | | |
| 14.42 | | | na$^{14}$ma$^{42}$ | *I will not pile rocks* |

(1) ni$^{1}$-ta$^{3}$ʃi$^{3}$ yu$^{3}$βa$^{4}$=õ$^{4}$ kwa$^{4}$yu$^{2}$ nda$^{3}$ʔa$^{4}$=õ$^{4}$
PERF-give father=2S horse hand=2S
'Your father gave you a horse.'

**Sentential focus**

(2) yu$^{3}$βa$^{4}$=õ$^{4}$ ni$^{1}$-ta$^{3}$ʃi$^{3}$=ɾi$^{4}$ kwa$^{4}$yu$^{2}$ nda$^{3}$ʔa$^{4}$=õ$^{4}$
father=2S PERF-give=3S horse hand=2S
'Your father gave you a horse.'

**Argument focus**

(3) yu$^{3}$βa$^{4}$=õ$^{4}$ ni$^{1}$-ta$^{3}$ʃi$^{3}$=ɾi$^{4}$ kwa$^{4}$yu$^{2}$ nda$^{3}$ʔa$^{4}$=õ$^{4}$
father=2S PERF-give=3S horse hand=2S
'Your father gave you a horse.'

**Contrastive focus**

**Fig. 1.** NPs under different focus conditions in YM.

### 2.1. Methods

#### 2.1.1. Design

Three types of nominal focus were examined in the current study: sentential focus, argument (narrow) focus, and corrective focus. For the purposes of the current study, the argument focus target consisted of the noun phrase produced in an utterance as a response to a WH-question, e.g. '*Who arrived?* **JOHN** arrived.' This is a type of narrow focus. Contrastive focus occured on the particular noun phrase uttered as a correction to a Yes–No question, e.g. '*Did John arrive?* **MARY** arrived.'[11] Two types of designs were used in the current study in order to investigate the different types of focus. For argument and corrective focus, we constructed short narratives of 3–5 sentences where the target nouns were either subjects or objects of a predicate, following similar designs in recent studies (Clopper & Tonhauser, 2013; Kügler & Genzel, 2011). Prior to recording, subjects were asked to answer questions using complete sentences (not single words). For the argument focus condition, a native speaker (Castillo García) read each narrative to the speaker and then asked them a series of WH questions about the narrative. For the corrective focus condition, Castillo García read each narrative to the speaker and then asked them a series of Yes–No questions about the narrative, each of which presupposed the incorrect entity, e.g. '*Did your mother give you a horse?* **YOUR FATHER** gave you a horse.'.

For the sentential focus condition, speakers repeated the target sentence that was uttered by Castillo García. There were two reasons for having a repetition task for this condition. First, when we attempted to include a similar, natural design for the sentential focus condition, speakers had difficulty providing the target stimulus word in their responses.[12] Given that YM is tonal, it was particularly important for the speakers to produce

[11] We use the term *contrastive focus* here instead of *corrective focus* out of simplicity. We make no claim regarding the meaning of this term.

[12] This problem is not unique to fieldwork studies on prosody. Following a picture-naming task eliciting contrastive and non-contrastive focus from native speakers of American English recorded in a laboratory, Breen et al. (2010) discarded 17% of all speakers' productions for this precise reason.

each target word in each condition; a substituted word would likely have a different tone. Second, while an alternative to the current design could have been to ask speakers to read the target sentences, there is currently no native language literacy among speakers of YM (excluding Castillo García and a couple speakers currently being trained). Thus, a repetition-elicitation task was a viable alternative. Given the different designs for eliciting focus types, we present the contrastive and argument focus data first and then separately compare it to the in situ sentential focus data elicited via repetition.

*2.1.2. Stimuli, speakers, and recording*

The target stimuli for the current experiment were much more limited in terms of tonal melodies than the larger set of melodies shown in Table 1. A total of nine tonal melodies on disyllabic words were investigated, comprising tone combinations /1.1, 1.3, 1.4, 3.2, 3.3, 3.4, 4.2, 4.4, 1.42/. There were three reasons for choosing a smaller set of possible melodies for the study. First, this study addressed information structure differences with nouns under sentential, argument, and corrective focus conditions. The set of possible tonal contrasts is much smaller in nouns than in verbs owing to the extensive derivational and inflectional morphology that is marked on verbs via tone in YM (Palancar et al., 2016). Second, in order to examine how stress influences tone production, the tones on each disyllabic word must occur on each syllable. Level tones are freely distributed on YM words. Third, each of the target words possessing these tonal melodies needed to be incorporated both into short texts and into the set of expected responses provided by each speaker. Tonal melodies were excluded if no example words could fit pragmatically within a short text for each of the elicited conditions.

A total of 28 words possessing the nine tonal melodies above were selected. For four of these, only two lexemes were suitable as stimuli (/1.1, 3.2, 4.2, 1.42/). For the remaining five melodies, four lexemes were chosen as suitable. As a result, the number of words containing each tonal melody in this study was not balanced. This issue is addressed in the discussion of statistical methods in Section 2.1.3. The stimuli also differed with respect to onset voicing, e.g. /ju$^{1}$ku$^{1}$/ 'leaf' vs. /k$^{w}$i$^{4}$jo$^{4}$/ 'roadrunner.' It is well-known that onset voicing may influence $F_0$ even in languages already possessing lexical tone contrasts (Chen, 2011; Xu & Xu, 2003; Zee, 1980). However, the current study is principally concerned with the effect of prosodic context on the production of tone in the same words. As far as we know, microprosodic effects like these are insensitive to higher level prosodic distinctions like focus.

Ten native speakers of YM were recorded from the Yoloxóchitl community. Five females and five male speakers participated (mean age 52 years old). No speakers reported any speech or hearing difficulties. All speakers were transported from Yoloxóchitl to the nearby town of San Luis Acatlán for recording purposes. Recording took place in a quiet room. 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.

Two sets of recordings were made for each speaker. The first session contained all three conditions (argument focus, sentential focus, and corrective focus, in this order). The second session was a repetition of the same tasks as the first session. Speakers returned either later in the day or on a different day to record the second session. Within each condition, speakers repeated each response three times. All target words were repeated six times within each condition by each speaker. A total of 234 words (13 words $\times$ 3 conditions $\times$ 3 repetitions $\times$ 2 sessions) were analyzed for each speaker. This totalled 2595 analyzed sentences (and not 2340 sentences) for all ten speakers since certain speakers repeated utterances more than three times and these were included in the analysis. The third author asked speakers to repeat their answers at the time of recording if they produced a disfluency.

*2.1.3. Analytical methods*

Each of the recorded sentences were transcribed by the third author using ELAN (Wittenburg, Brugman, Russel, Klassman, & Sloetjes, 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 duration were measured along with $F_0$ on the vowel. As consonant duration was examined, tokens with utterance-initial voiceless stops were excluded from durational analysis. Average $F_0$ measures for five equal intervals were taken from each vowel with a script written for Praat (DiCanio, 2016a). The $F_0$ range used by the script was adjusted on an individual basis, i.e. higher for female speakers (120–400 Hz) and lower for males (80–300 Hz). For one male speaker, we used a range of 100–300 Hz, as this produced less pitch-halving/doubling. This script also extracted durational information for each segment within the word. Dynamic $F_0$ measures were extracted from only those tokens that were 50 ms or longer in duration. For tokens shorter than 50 ms, no $F_0$ information was extracted. The bases for excluding these tokens were concerns over the reliability of dynamic $F_0$ measurements on short durations and concerns over extracting $F_0$ on vowels which lack voicing, e.g. a word like /ki$^{1}$si$^{3}$/ 'pot' may be produced as [ki̥$^{(1)}$si$^{3}$] when spoken quickly. These shorter vowels comprised 6.2% of the total vowels (321/5195 vowels).

All F0 values were visually examined by the second author for accuracy prior to analysis. $F_0$ data (in Hz) were then converted to $\log_{10}$ values and statistically normalized (*z*-score normalization) to correct for individual speaker differences in $F_0$ range and level prior to statistical analysis. For the duration data, words with each of the different tonal melodies were pooled together. The data were analyzed using a linear mixed effects model with Tonal melody (the combinations listed in 2.1.2), Focus Type (Argument vs. Corrective), and Syllable (non-final/unstressed vs. final/stressed) as fixed effects. Separate statistical models were used for the onset duration and vowel duration data. Speaker and Item were treated as random intercepts and random slopes of Focus by Speaker and Focus by Item were also included. For the consonant duration data, only the Focus by Item random slope was included (the model did not converge with additional random slopes). The fixed effects were evaluated using *lmerTest* (Bates, Maechler, & Bolker, 2011; Kuznetsova, Brockhoff, & Christensen, 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 of freedom. There is currently no way to approximate the upper bound on degrees of freedom for linear mixed effects models (Baayen, 2008). All statistics were calculated using R (R Development Core Team, 2017).

For the $F_0$ data, statistical analyses were done individually for each tonal melody; e.g. analyses for melody 1.1, analyses for melody 1.3, etc. Separating out each of the tonal melodies here allows one to control for possible differences in the direction of $F_0$ movement. For instance, 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 $F_0$ range is expanding. Two types of statistical models were examined for each tonal melody: a dynamic model treating $F_0$ as the dependent variable with time as a fixed effect and a static model treating the $F_0$ midpoint (the average $F_0$ from 40–60% of the vowel's duration) as the dependent variable but without time included in the model. For the dynamic model, *Time* was an ordered predictor with five values (1–5) corresponding to the five points at which $F_0$ was extracted in each vowel. The dynamic models are useful for specifically examining changes in $F_0$ slope and the static ones for examining changes in $F_0$ height. In both types of statistical models, random intercepts for Speaker and Word were included alongside random slopes of Focus Type by Speaker. This was the maximal random effects structure that converged for all of the tonal melodies. We attempted to include random slopes for Focus Type by Word, but these models did not converge in many cases since many tonal categories were represented by a small group of words (one or two exemplars, c.f. Section 2.1.2). For those tonal melodies' models which converged with more complex random effects structure, we obtained a pattern of significant effects virtually identical to the patterns described for the random intercept-only models. Focus Type (Argument, Contrastive) and Syllable (non-final and unstressed, final and stressed) were treated as fixed effects for $F_0$ models. Identical to the durational model described above, lmerTest (Kuznetsova et al., 2017) was used to evaluate the significance of the fixed effects.

### 2.2. Results – duration

Fig. 2 shows the durational differences by focus type and stress position for onset consonants and vowels in disyllabic YM words. With respect to onset duration, there were significant fixed effects of focus type; where onsets produced in words under contrastive (corrective) focus significantly differed in duration from those produced under argument focus ($t$[13.4] = −3.1, $p$ < .01). Onsets on words with argument focus were slightly longer than those on words with contrastive focus (mean durations 114 ms vs. 104 ms). There was a significant main effect of word position on onset duration ($t[1634] = 25.2, p < .001$). Onsets in final, stressed syllables were significantly longer than those produced in non-final, unstressed syllables (126 ms vs. 80 ms, a ratio of 1.58:1). In addition to these main effects, there was a significant interaction between Focus and Position ($t[484] = -5.6, p < .001$). The average duration of unstressed syllable onsets was identical across focus types (80 ms), but the onset duration in stressed syllables was longer under argument focus than under contrastive focus (133 vs. 118 ms).

With respect to vowel duration, there was a near significant main effect of focus type (t[15] = −1.9, p =.08). Vowels in words under argument focus were only slightly longer than vowels in words under contrastive focus (108 ms vs. 101 ms). The main effect of position was not significant; no effect of stress on vowel duration was found. A small, but significant interaction between Focus and Position was found ($t$[3176] = −2.2, $p$ < .05). Final and non-final vowels were of roughly equal duration under contrastive focus (102 vs. 100 ms) but final vowels were slightly longer than non-final vowels under argument focus (110 vs. 104 ms). The strongest effects observed for duration were those associated with consonants and (stress) position.

### 2.3. Results – $F_0$

The left panel of Fig. 3 shows the $F_0$ trajectories for words with level tonal melodies /1.1/,/3.3/, and /4.4/. For each of these tonal melodies, a significant main effect of Focus was found: /1.1/ ($t$[8.5] = 4.7, $p$ < .01), /3.3/ ($t$[8.8] = 2.5, $p$ < .05), /4.4/ ($t$[8.9] = 3.8, $p$ < .01). In each case, the tonal melody produced with contrastive focus was raised relative to the same word produced under argument focus. No main effects of (stress) position and no interactions of focus and position were found. The magnitude of the effect of focus varied by tone level: 0.27 s.d. for /1.1/, 0.37 s.d. for /3.3/, and 0.49 s.d. for /4.4/. This difference by tone height is observed despite the normalization used here.

For the dynamic model of melody /1.1/, a significant main effect of Time ($t[943] = -5.3, p < .001$) and a small but significant interaction of Time and Focus was found ($t[943] = -2.4, p < .05$).[13] Tone /1/ was realized with a falling $F_0$ contour, though this fall was steeper under contrastive focus (falling 0.41 s.d.) than under argument focus (falling 0.25 s.d.). No main effects or interactions with stress/position were observed. For the dynamic model of melody /3.3/, main effects of Time ($t[2080] = -6.3, p < .001$) and Position were found ($t[2080] = 4.7, p < .001$). A significant interaction between Time and Position was also observed ($t[2080] = -2.7, p < .01$). Tone /3/ was realized with a falling trajectory in both unstressed and stressed syllables, but the fall was greater in the stressed syllable (0.39 s.d.) than in the unstressed syllable (0.19 s.d.). For the dynamic model of melody /4.4/, no main effect of Time was observed, but a significant effect of Position was observed ($t[2261] = 5.4, p < .001$). A significant interaction between Time and Position was also observed ($t[2261] = -6.2, p < .001$). The peak $F_0$ for tone /4/ is reached in the latter half of the vowel (at about 70% of the vowel duration) in the unstressed syllable, but early in the vowel in the stressed syllable (at about 30% of the vowel duration).

The right panel of Fig. 3 shows the $F_0$ trajectories for words with rising tonal melodies /1.3/,/1.4/, and /3.4/. For the statistical tests examined here, we exclude significant effects of Position, as these reflect differences between distinct tones, not between identical tones. Rather, we are interested in interac-

[13] For all static models where a significant effect of Focus was found, a significant effect of Focus was found in the dynamic model. These latter effects are excluded here.

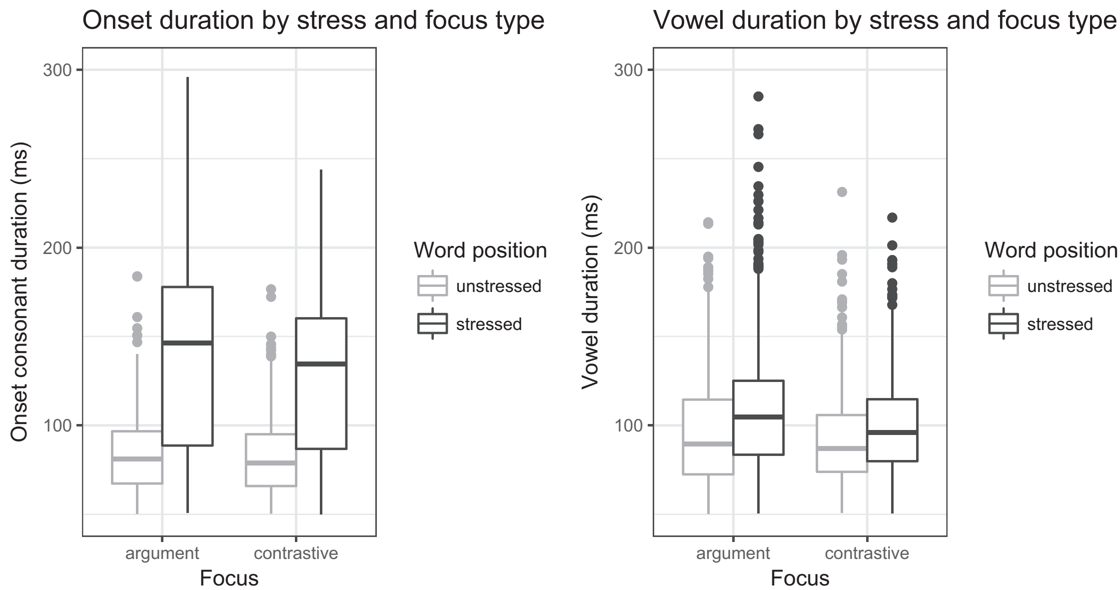


**Fig. 2.** Duration results for onsets and vowels in unstressed and stressed syllables across focus conditions.

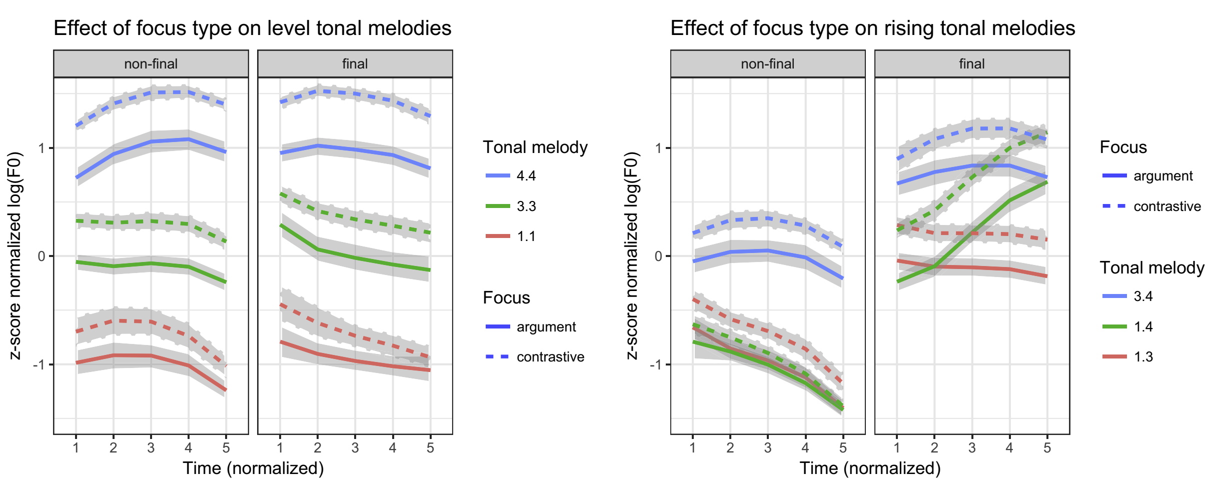


**Fig. 3.** Level and rising tonal melody data. The grey bars indicate a 95% confidence interval around the mean values, assuming a normal distribution. The panels are ordered sequentially for each melody, e.g. the $F_0$ trajectory for the unstressed syllable of melody followed by the trajectory for the stressed syllable.

tions between Focus Condition and Position and those between Time and Focus. Similar to the findings for level tones, small, but significant main effects of Focus were found for each of these melodies: /1.3/ ($t[8.4] = 2.3, p < .05$), /1.4/ ($t[8.5] = 2.7, p < .05$), /3.4/ ($t[8.7] = 2.9, p < .05$). All tones were raised under contrastive focus compared with argument focus (each by approximately 0.30 s.d.). A significant Position by Focus interaction was found only for melody /1.4/ ($t[410] = 4.1, p < .001$) but not for either of the other melodies. The initial tone /1/ of this melody was raised to a smaller degree (0.1 s.d.) under contrastive focus than the final tone /4/ was (0.5 s.d.).

For the dynamic model of melody /1.3/, a significant main effect of Time ($t[1936] = -10.7, p < .001$) was found along with a significant interaction of Time by Position ($t[1936] = 7.7, p < .001$). This latter result reflects the difference in slope between the two tones (where tone /1/ is realized with a falling trajectory). No other interactions were significant. For the dynamic model of melody /1.4/, a significant main effect of Time ($t[2067] = 5.4, p < .001$) was found along with significant interactions between Time and Position ($t[2067] = 19.5, p < .001$) and Focus by Position ($t[2067] = 2.6, p < .01$). The first interaction indicates that tones /1/ and /4/ were realized with different $F_0$ trajectories. The second indicates the same pattern observed for the static model – tone /1/ undergoes less $F_0$ raising with focus than tone /4/ does. For the dynamic model of melody /3.4/, the main effect of Time was not significant, but a significant interaction between Time and Position was observed ($t[2175] = 2.8, p < .01$). This pattern reflected tone-specific slope differences; tone /3/ has a relatively level trajectory here but tone /4/ is realized with a rising contour.

Fig. 4 shows the $F_0$ trajectories for words with falling and complex tonal melodies /3.2/, /4.2/, and /1.42/. A significant main effect of focus was found for melodies /3.2/ ($t[9] = 3.2, p < .05$) and /4.2/ ($t[9.3] = 2.7, p < .05$), but not

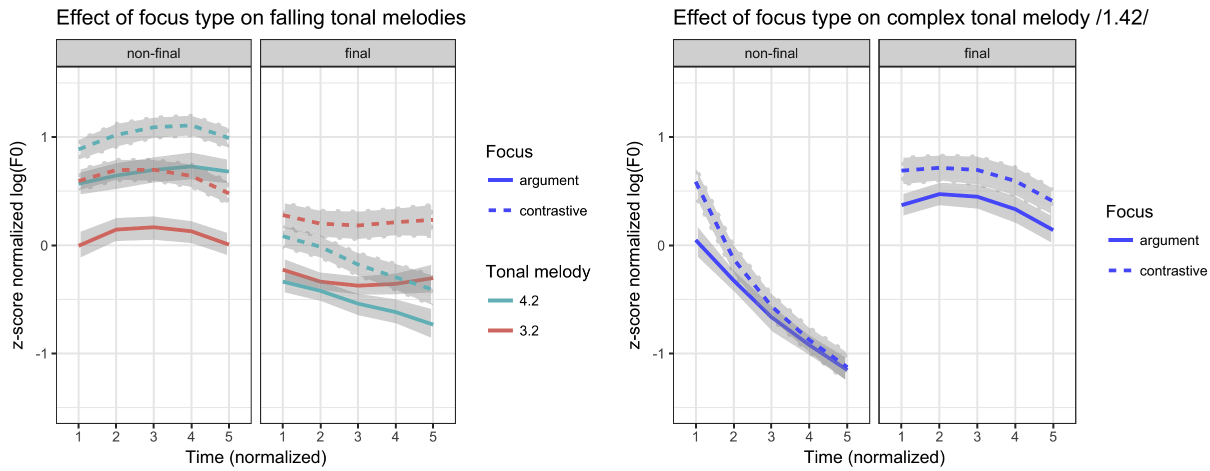


**Fig. 4.** Falling and complex tonal melody data. The grey bars indicate a 95% confidence interval around the mean values, assuming a normal distribution. The panels are ordered sequentially for each melody, e.g. the $F_0$ trajectory for the unstressed syllable of melody followed by the trajectory for the stressed syllable.

for /14.2/. Each melody underwent $F_0$ raising under contrastive focus when compared with argument focus. The effect of focus on $F_0$ was stronger for melodies /3.2/ (0.52 s.d.) and /4.2/ (0.40 s.d.) than it was for tonal melody /1.42/ (0.21 s.d.). No significant Position by Focus interactions were observed.

For the dynamic model of melody /3.2/, the main effect of time was not significant and no interactions with time or focus reached significance. For the dynamic model of melody /4.2/, a significant main effect of Time was observed ($t[1181] = -2.9, p < .01$) along with a significant interaction of Time by Position ($t[1181] = -6.2, p < .001$). These observations correspond with tone-specific slope differences across the word; tone /4/ is realized with a rising $F_0$ trajectory while tone /2/ has a falling $F_0$ trajectory. No Time by Focus interaction was found. For the dynamic model of melody /1.42/, a significant main effect of focus was observed ($t[19.1] = 2.8, p < .05$). A significant interaction of Time by Position ($t[863] = 7.6, p < .001$) was also found, which reflects tone-specific slope differences across the word.

In summary, the static models demonstrate a clear main effect of focus type on $F_0$ level for all tonal melodies. Contrastive focus raises $F_0$ for all tonal melodies relative to non-contrastive argument focus. The effect of focus varied by tone, with the lowest tone (/1/) undergoing less $F_0$ raising than higher tones and/or contour tones. The dynamic models demonstrate three patterns: (1) tone /4/ is realized with later $F_0$ peak alignment in unstressed syllables than in stressed syllables, (2) tone /3/ is realized with a steeper falling $F_0$ trajectory in stressed syllables than in unstressed syllables, and (3) tone /1/ is realized with a steeper fall under contrastive focus than under argument focus.

### 3. Comparison with sentential focus via a repetition task

Sentential focus was elicited via a repetition task and it is compared with the narrow focus conditions (argument/contrastive focus) above. Fig. 5 shows both durational differences among consonants and vowels. For all syllables, onsets produced in words under sentential focus significantly differed slightly in duration from those produced under argument focus ($t[16] = 2.5, p < .05$) and contrastive focus ($t[16] = 2.0, p = .06$). The target noun in these sentences was of shorter duration than the same noun in the narrow focus conditions. Similar to the narrow focus conditions, onset duration in stressed syllables was longer than in unstressed syllables in the sentential focus condition, though the magnitude of the stress-related difference was significantly smaller (unstressed onset duration = 70 ms; stressed onset duration = 95 ms; a ratio of 1:1.36). This magnitude difference was significant; there was a strong interaction between Position and Argument focus when compared to sentential focus ($t[2779] = 10.8, p < .001$) and between Position and Contrastive focus when compared with sentential focus ($t[1857] = 5.1, p < .001$). The onset of the stressed syllable was significantly longer in the argument and contrastive focus conditions in comparison to the sentential focus condition.

Vowels produced in words under sentential focus significantly differed in duration from those produced under argument focus ($t[15] = 2.7, p < .05$) and those produced under contrastive focus ($t[15] = 2.3, p < .05$). Vowels were shorter in words produced in the sentential focus position than in words produced under the narrow focus conditions. In summary, words were of shorter duration in the sentential focus condition than under the narrow focus conditions and displayed similar stress-related lengthening patterns.

Fig. 6 shows the tonal data from the narrow focus conditions compared with the sentential focus condition. We evaluate the sentential focus data in relation to the argument and contrastive focus data using only the static mixed effects models where the sentential focus condition is treated as the model intercept. The midpoint $F_0$ in the sentential focus condition significantly differed from the contrastive focus condition for tonal melody /3.3/ ($t[7.7] = 4.9, p < .01$), but not for other tonal melodies. The midpoint $F_0$ in the sentential focus condition significantly differed from the argument focus condition for tonal melodies /1.1/ ($t[10.9] = -3.2, p < .01$) and /3.3/ ($t[9.1] = 2.5, p < .05$). For melody /4.4/, $F_0$ under sentential focus was intermediate between contrastive and argument focus, the latter which was realized with a lower $F_0$ than sentential focus. For melodies /1.1/ and /3.3/, the $F_0$ under sentential focus was lower than that found for both narrow focus conditions. For melody /1.1/, a more dramatic falling contour on the final

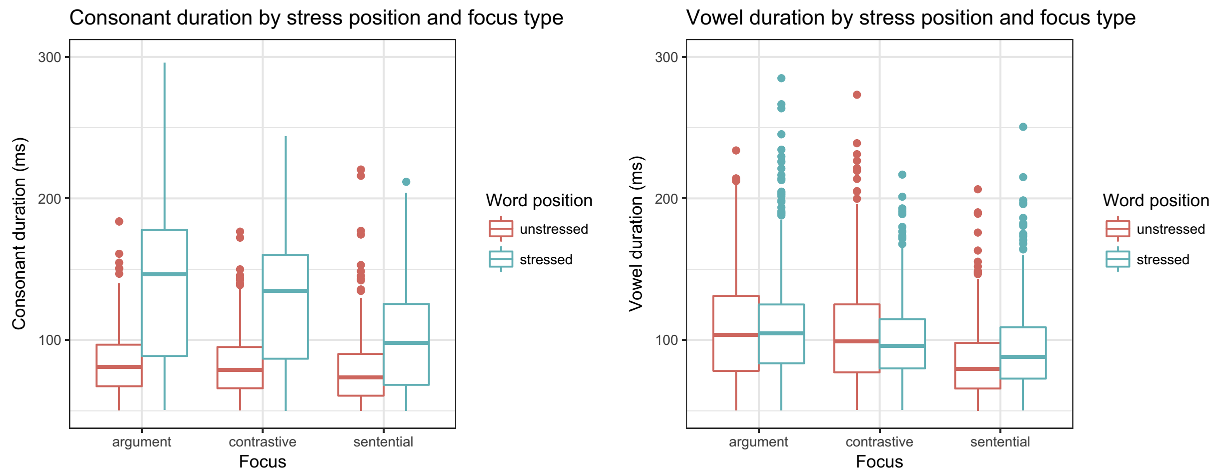


**Fig. 5.** Comparison of durational data across focus conditions.

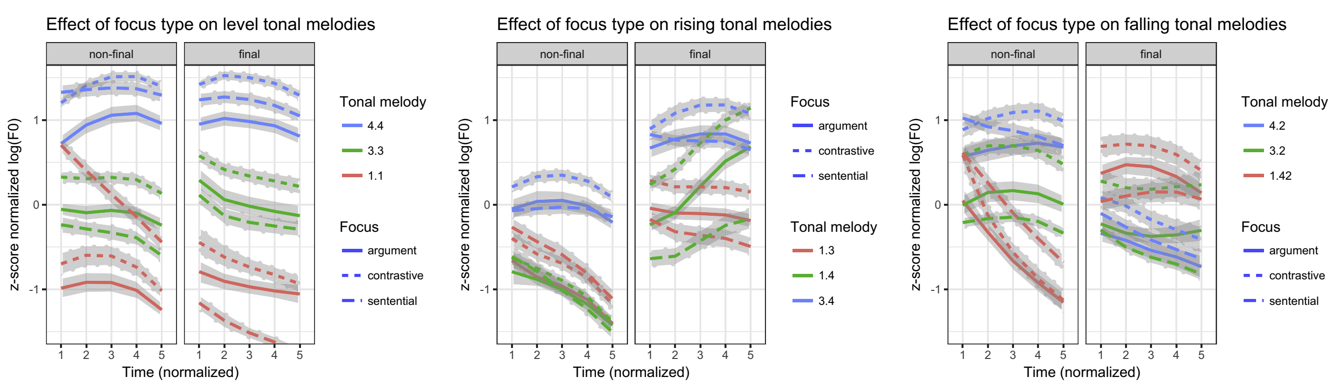


**Fig. 6.** Comparative tonal data. The grey bars indicate a 95% confidence interval around the mean values, assuming a normal distribution. The panels are ordered sequentially for each melody, e.g. the $F_0$ trajectory for the unstressed syllable of melody followed by the trajectory for the stressed syllable.

syllable was also observed under sentential focus than under the narrow focus conditions. For tonal melody /1.1/, a strong interaction of Position by Argument Focus was found ($t[431] = 16.5, p < .001$) along with a strong Position by Contrastive Focus interaction ($t[431] = 15.3, p < .001$). Under sentential focus, melody /1.1/ was realized with a falling $F_0$ trajectory. The effect of this was that tone /1/ in the penultimate syllable was significantly higher than the same tone under contrastive focus, but significantly lower than the same tone under argument or contrastive focus in the final syllable. No other significant Focus x Position interactions were found for the other level tonal melodies.

The midpoint $F_0$ in the sentential focus condition significantly differed from the contrastive focus condition for rising tonal melodies /1.3/ ($t[9.3] = 3.7, p < .01$), /1.4/ ($t[8.5] = 4.0, p < .01$), and /3.4/ ($t[9.4] = 2.8, p < .05$). Higher $F_0$ was always observed in the contrastive focus condition relative to the other two conditions. A significant interaction between contrastive focus and position was observed for melodies /1.3/ ($t[550] = 7.0, p < .001$) and /1.4/ ($t[560] = 10.7, p < .001$). While the penultimate tone /1/ produced under contrastive focus did not differ much from tone /1/ produced under sentential focus, the final syllable tones (/3/ or /4/) were significantly higher when produced under contrastive focus. The midpoint $F_0$ in the sentential focus condition did not significantly differ from the argument focus condition for any of the rising tonal melodies. However, as above, there were significant interactions between argument focus and position for tonal melodies /1.3/ ($t[550] = 6.4, p < .001$) and /1.4/ ($t[560] = 7.4, p < .001$). While the penultimate tone /1/ produced under argument focus did not differ much from tone /1/ produced under sentential focus, the final syllable tones (/3/ or /4/) were significantly higher when produced under argument focus.

The midpoint $F_0$ in the sentential focus condition significantly differed from the contrastive focus condition for falling tonal melody /3.2/ ($t[8.1] = 7.1, p < .001$) but not for either melody /4.2/ nor /1.42/. The midpoint $F_0$ of the sentential focus condition also differed from the argument focus condition for tonal melody /3.2/ ($t[8.5] = 2.4, p < .05$). The midpoint $F_0$ in the sentential focus condition was signficantly lowered for melody /3.2/ relative to the narrow focus conditions, specifically in the initial syllable. The $F_0$ of /4.2/ fell between that of the contrastive and argument focus $F_0$ values, similar to melody /4.4/ above, and did not significantly differ from the other focus conditions. Though, the lack of a general effect for tonal melody /1.42/ can be attributed to a strong interaction between focus and position here, e.g. ($t[231] = 7.2, p < .001$ for argument focus, $t[265] = 7.6, p < .001$ for contrastive focus). Under sentential focus, tone /1/ was raised relative to the narrow focus conditions, but tone /42/ was lowered. The net effect of this pattern is tonal levelling under sentential focus relative to the other focus conditions.

## 4. Discussion

### 4.1. Duration

Using the data from Section 2.2, Table 2 summarizes the mean duration values with percentages and ratios for individual segments and syllables. Under contrastive or argument focus, the entire word was lengthened. However, the lengthening was asymmetric. Examining the entire syllable duration, unstressed syllables underwent slightly less durational expansion when the target word was realized with argument or contrastive focus (21% vs. 31%). Examining sub-syllabic units, a divergent pattern emerges: consonants in stressed syllables underwent greater lengthening than consonants in unstressed syllables but vowels in unstressed syllables underwent roughly equivalent lengthening as those in stressed syllables. Taken as a whole, these results demonstrate that stressed syllables in YM are realized with longer overall duration than unstressed syllables and consonant duration is the main contributor to these differences. Yet, why might focal lengthening be stronger with onsets than with vowels?

While all segments were lengthened under focus, greater focal lengthening occurs word-medially than at word edges. This result is compatible with two interpretations, one of which maintains a paradigmatic distinction (vowel length) and one of which maintains a syntagmatic one (stress). With respect to the first, recall that vowel length is only contrastive on stem-final syllables in YM. If the goal of argument or contrastive focus is to enhance or preserve phonological contrast, there may be a resistance to focal lengthening on final short vowels. In this scenario, focal lengthening would preserve a vocalic contrast in the language. This particular interpretation matches that of Heldner and Strangert (2001) for Swedish and Nakai et al. (2012) for Northern Finnish and is in agreement with the *quantity neighbor constraint*. Though the linguistic details are different - vowel length is not contrastive in unstressed syllables in YM, so any focal lengthening in this position does not result in the potential for contrast neutralization. As a result of how vowel length is distributed with respect to stress, focus in YM induces greater lengthening in onsets than on vowels. One way to test this pattern might be to examine the degree to which focus-related lengthening occurs on vowels in monosyllabic words with long vowels. Since the current study focused on the interaction of lexical prosody and focus, only polysyllabic words were considered.

With respect to the second perspective, note that speakers may also prefer to maintain a durational difference between unstressed and stressed syllables. Table 2 provides the unstressed to stressed syllable ratio, which is approximately 1:1.3 across all conditions. Lengthening may occur in the vowels of unstressed syllables for the purpose of maintaining this ratio. Such a perspective is not incompatible with the length preservation hypothesis above, as speakers may attempt to preserve both syntagmatic and paradigmatic contrasts simultaneously. The preservation of the vowel length contrast and stress-based durational asymmetries between syllables both serve the general aim of enhancing phonological contrasts in narrow focus contexts.

Apart from the distributional evidence for stress in YM, the findings here suggest that the acoustic duration of the onset consonant is a stable cue for marking stress on the final syllable. Consonant duration is not a common acoustic cue for stress contrasts in languages of the world (Fletcher, 2010), but it is the main cue used in Pirahã (Everett, 1998). The work here suggests that medial consonant duration cues stress in YM.

In sum, words realized with contrastive or argument focus were lengthened relative to the same word under sentential focus. While the entire word was lengthened, such lengthening was asymmetrical; it affected stressed syllables more than unstressed syllables. This finding parallels research on prosodic lengthening in English (Turk & Sawusch, 1997; Turk & White, 1999), Dutch (Cambier-Langeveld & Turk, 1999), and Swedish (Heldner & Strangert, 2001). While focal lengthening affects the entire word in these languages, it induces greater lengthening on stressed syllables.

### 4.2. Focus and $F_0$ range

Globally, contrastive focus induced a raising of $F_0$ in comparison with argument and sentential focus. This effect was asymmetrical across the different tonal melodies; while syllables with the highest tones were uniformly raised under contrastive focus, tone /1/ was resistant to $F_0$ raising. The tonal differences between argument and sentential focus conditions were less consistent than the effects observed with the contrastive focus condition. $F_0$ raising was observed with argument focus for certain tonal melodies (/1.3, 1.4, 3.3, 1.42, 3.2/). However, relative to sentential focus, $F_0$ was lowered under argument focus for melody /4.4/ and no $F_0$ difference between these focus conditions was observed for the /3.4/ melody. On the whole, argument focus involved $F_0$ raising relative to sentential focus, but the differences were smaller and less consistent than those observed for the contrastive focus condition.

In addition to $F_0$ raising, $F_0$ range expansion occurred in the production of contrastive and argument focus relative to the sentential focus condition. This pattern was observed most

**Table 2**
Durational patterns across focus types. Except for ratios and percentages, all numbers are in milliseconds.

| | $C_1$ | $V_1$ | $C_2$ | $V_2$ | $\sigma_1$ | $\sigma_2$ | $\sigma$-duration ratio |
|---|---|---|---|---|---|---|---|
| **Baseline (sentential focus)** | 70 | 77 | 95 | 90 | 141 | 185 | 1:1.31 |
| **Contrastive focus** | 77 | 92 | 120 | 99 | 169 | 219 | 1:1.30 |
| (comparison to baseline) | 10% | 19% | 26% | 10% | 20% | 18% | |
| **Argument focus** | 76 | 94 | 136 | 107 | 170 | 242 | 1:1.42 |
| (comparison to baseline) | 9% | 22% | 43% | 19% | 21% | 31% | |

clearly in the production of the rising tonal melodies. Here, no raising effect was observed with argument/contrastive focus in the production of the initial tone /1/ in melodies /1.3, 1.4, 1.42/, but significant $F_0$ raising was observed on the tones in the stressed syllables. The resistance of tone /1/ to raising here may, at first glance, appear to be positive evidence for a stress-based difference with respect to $F_0$ raising. That is, tone raising was restricted to the stressed syllable in these words. However, note that this tone resists raising only in unstressed syllables in melodies which rise across the word, not with tonal melody /1.1/. A more plausible explanation for this effect is that, by maintaining a lower initial tone and raising the final tone, speakers increase the acoustic/perceptual distance across syllables on words with rising melodies. In this way, the resistance of tone /1/ to raising under the narrow focus conditions is an active strategy for the expanding the $F_0$ range and producing hyperarticulated tonal melodies on words.

Taken together, the results show that contrastive focus is realized via the greatest $F_0$ raising and range expansion while argument focus is realized with slightly less $F_0$ raising and range expansion. NPs with sentential focus are realized with the smallest $F_0$ range and a lower $F_0$ register. These results are comparable to findings with Zapotec (Esposito, 2010), Mandarin (Xu, 1999; Liu & Xu, 2005), and Taiwanese (Pan, 2007). Like Zapotec, focus in YM is realized via $F_0$ raising. Like Mandarin and Taiwanese, the $F_0$ range also appears to be expanded. The effect of both processes is the hyperarticulation of tonal melodies with distinct tones on each syllable within the word.

The findings here differ slightly from those found for Mandarin Chinese focus constructions. In work investigating the degree of tonal hyperarticulation in three discourse contexts (no emphasis, emphasis, more emphasis), Chen and Gussenhoven (2008) found strong differences between no emphasis (backgrounded) contexts and those involving emphasis (contrastive focus) in terms of both duration and $F_0$ scaling. However, there were non-significant differences observed between the two emphatic conditions (contrastive focus and clarificational contrastive focus) in terms of $F_0$ scaling. While Chen and Gussenhoven analyze the difference between the non-emphatic and emphatic contexts as discrete, they analyze the differences between the emphatic contexts as gradient. In YM, tones with contrastive focus were produced with significant $F_0$ raising relative to the same tones in the argument focus condition. The argument focus condition also involved significant, but diminished $F_0$ raising relative to the sentential focus condition. For both the contrastive and argument focus conditions, tonal hyperarticulation was observed. In comparison with Chen and Gussenhoven's findings for Mandarin, the YM patterns are not easily categorized as either gradient or discrete.

### *4.3. Tonal hyperarticulation and the focal domain in YM*

In addition to the global effects of focus on tone, additional, tone-specific differences were observed that were unrelated to $F_0$ scaling and height. For both the contrastive and argument focus conditions, the stressed syllable of the rising tonal melodies (/1.3, 1.4, 3.4/) was produced with a greater phonetic rise (greater $F_0$ excursion) in comparison with the sentential focus condition. Excursion degree was directly related to the expansion in $F_0$ range which accompanied tone production under narrow focus. Similar findings for rising tones have been reported in research on Mandarin focus (Xu, 1999).

However, unlike the findings for Mandarin, greater $F_0$ excursion was not observed with the falling tonal melodies in YM. Tonal melodies with a phonetic fall across syllables (/4.2, 3.2, 1.1/) were produced as two sequences of level tonal contours under contrastive or argument focus conditions with a raised $F_0$ level. In the sentential focus condition, these melodies were produced with $F_0$ movement on one or both of the syllables of the word. In particular, the unstressed syllable of melodies /4.2, 1.1/ was realized as a phonetically falling contour; an anticipatory coarticulatory effect. This finding suggests that $F_0$ movement is an important acoustic cue for the production of rising tonal melodies in YM but a less important cue in the production of falling tonal melodies.[14] Under focus, this cue is enhanced for rising tonal melodies but not for falling tonal melodies.

Are these patterns of hyperarticulation the result of global processes which expand the $F_0$ range and increase the register for higher tones? Like the durational data, there are two possible interpretations. One possibility is that the prosodic lengthening observed in Section 4.1 permits a longer window over which tonal contours may be realized. Tones may be hyperarticulated under argument or contrastive focus because the word duration is greater (c.f. Xu (1999) for Mandarin). The second possibility is that distinct phonological processes apply to YM tones when they occur on focused constituents. That is, just as intonational pitch accents are distinct $F_0$ contours realized on focused constituents, argument or contrastive focus may condition phonetic allotones of existing tonal melodies.

While tones on words with narrow focus and in stressed syllables were produced with more canonical phonetic shapes, $F_0$ raising and range expansion did not directly interact with word stress. The effect of stress on tone was instead indirect: focus induced greater lengthening on stressed syllables than unstressed syllables and this greater lengthening permitted the hyperarticulation of tonal targets. Thus, we understand these additional tonal changes as particular instances of tonal reduction and hyperarticulation that result from changes in articulatory timing under focus rather than targeted phonological patterns associated with stressed syllables via the prosodic hierarchy.

Since prosodic lengthening is distinct from the focus-related effects on tone, the results here argue in favor of two distinct phonetic mechanisms which influence lexical tone production: (1) $F_0$ 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. These two mechanisms correspond, respectively, to the articulatory control parameters of *target modification* and *movement rescaling* within models of prosody production (Beckman, Edwards, & Fletcher, 1992; Cho, 2006; Mücke & Grice, 2014). Along these lines, focus in YM

[14] Absent from the current work are contour tones on individual syllables (c.f. Table 1). We assume that changes in $F_0$ are relevant for the perception/production of these tones (c. f. DiCanio et al. (2014)). The current paper notes that $F_0$ movement is a cue that is enhanced under focus specifically in the context where level tones are rising across a word.

**Table 3**
Fixed effects for consonant duration linear mixed effects model: cdur ~ Focus.2 * Position + (1 + Focus|Word) + (1|Speaker). All models were estimated using lmertest (Kuznetsova et al., 2017), which examines the predictor variance using the *t* distribution; the *t* values here are not the result of a standardized *t* test.

| Effect | Estimate | St.Error | df | *t* value | Pr(> \|*t*\|) |
|---|---|---|---|---|---|
| (Intercept) | 1.067e−01 | 1.145e−02 | 1.400e+01 | 9.319 | 2.22e−07*** |
| Focuscontrastive | −7.465e−03 | 2.460e−03 | 1.340e+01 | −3.035 | 0.0093** |
| Position.L | 4.234e−02 | 1.678e−03 | 1.634e+03 | 25.233 | <2e−16*** |
| Focuscontrastive:Position.L | −1.283e−02 | 2.305e−03 | 4.838e+02 | −5.568 | 4.27e−08*** |

**Table 4**
Fixed effects for vowel duration linear mixed effects model: dur ~ Focus * Position + (1 + Focus — Word) + (1 + Focus — Speaker). All models were estimated using lmertest (Kuznetsova et al., 2017), which examines the predictor variance using the *t* distribution; the *t* values here are not the result of a standardized *t* test.

| Effect | Estimate | St.Error | df | *t* value | Pr(> \|*t*\|) |
|---|---|---|---|---|---|
| (Intercept) | 1.077e−01 | 7.495e−03 | 1.800e+01 | 14.373 | 1.88e−11*** |
| Focuscontrastive | −5.105e−03 | 2.710e−03 | 1.500e+01 | −1.884 | 0.0793 |
| Position.L | 1.163e−03 | 9.384e−04 | 3.176e+03 | 1.239 | 0.2154 |
| Focuscontrastive:Position.L | −3.007e−03 | 1.371e−03 | 3.176e+03 | −2.193 | 0.0283* |

**Table 5**
Fixed effects for static linear mixed effects models for each tonal melody. Most models have the structure: zlogF0 ~ Focus * Position + (1 + Focus|Speaker) + (1|Word), but models for melodies /1.1/,/3.2/,/4.2/, and /1.42/ exclude a random intercept by item since only one target word with each melody served as the target in the experiment. All models were estimated using lmertest (Kuznetsova et al., 2017), which examines the predictor variance using the *t* distribution; the *t* values here are not the result of a standardized *t* test.

| Melody | Effect | Estimate | St.Error | df | *t* value | Pr(> \|*t*\|) |
|---|---|---|---|---|---|---|
| /1.1/ | (Intercept) | −0.93865 | 0.08980 | 9.22 | −10.453 | 2.04e−06*** |
| | Focus.2contrastive | 0.31376 | 0.06620 | 8.52 | 4.739 | 0.00123** |
| | Position.L | −0.02823 | 0.04710 | 182.44 | −0.599 | 0.54968 |
| | Focus.2contrastive:Position.L | −0.08165 | 0.06626 | 181.77 | −1.232 | 0.21939 |
| /3.3/ | (Intercept) | −0.03442 | 0.18835 | 2.00 | −0.183 | 0.8720 |
| | Focus.2contrastive | 0.35335 | 0.13968 | 8.80 | 2.530 | 0.0329* |
| | Position.L | 0.05415 | 0.04298 | 406.70 | 1.260 | 0.2085 |
| | Focus.2contrastive:Position.L | −0.03166 | 0.06188 | 406.70 | −0.512 | 0.6092 |
| /4.4/ | (Intercept) | 1.02010 | 0.16564 | 8.30 | 6.159 | 0.000233*** |
| | Focus.2contrastive | 0.49116 | 0.12963 | 8.90 | 3.789 | 0.004349** |
| | Position.L | −0.05037 | 0.03439 | 436.10 | −1.465 | 0.143707 |
| | Focus.2contrastive:Position.L | 0.03921 | 0.05092 | 436.10 | 0.770 | 0.441639 |
| /1.3/ | (Intercept) | −0.55380 | 0.19268 | 2.20 | −2.874 | 0.0926 |
| | Focus.2contrastive | 0.28549 | 0.12132 | 8.40 | 2.353 | 0.0450* |
| | Position.L | 0.61590 | 0.04102 | 376.80 | 15.014 | <2e−16*** |
| | Focus.2contrastive:Position.L | 0.05301 | 0.06023 | 376.90 | 0.880 | 0.3793 |
| /1.4/ | (Intercept) | −0.43936 | 0.19805 | 2.00 | −2.218 | 0.1540 |
| | Focus.2contrastive | 0.33993 | 0.12586 | 8.50 | 2.701 | 0.0255* |
| | Position.L | 0.90432 | 0.04353 | 408.90 | 20.773 | <2e−16*** |
| | Focus.2contrastive:Position.L | 0.25540 | 0.06304 | 409.50 | 4.052 | 6.09e−05*** |
| /3.4/ | (Intercept) | 0.44691 | 0.22440 | 2.90 | 1.992 | 0.1434 |
| | Focus.2contrastive | 0.33982 | 0.11714 | 8.70 | 2.901 | 0.0181* |
| | Position.L | 0.55152 | 0.04248 | 419.20 | 12.982 | <2e−16*** |
| | Focus.2contrastive:Position.L | 0.04418 | 0.06203 | 419.30 | 0.712 | 0.4767 |
| /3.2/ | (Intercept) | −0.10225 | 0.09463 | 9.01 | −1.081 | 0.3080 |
| | Focus.2contrastive | 0.53362 | 0.16943 | 9.00 | 3.150 | 0.0118* |
| | Position.L | −0.37415 | 0.04713 | 205.28 | −7.939 | 1.31e−13*** |
| | Focus.2contrastive:Position.L | 0.01116 | 0.06845 | 205.15 | 0.163 | 0.8706 |
| /4.2/ | (Intercept) | 0.05693 | 0.11689 | 8.99 | 0.487 | 0.6379 |
| | Focus.2contrastive | 0.35933 | 0.13293 | 9.27 | 2.703 | 0.0237* |
| | Position.L | −0.85770 | 0.04950 | 221.39 | −17.328 | <2e−16*** |
| | Focus.2contrastive:Position.L | −0.03536 | 0.07269 | 221.42 | −0.486 | 0.6272 |
| /1.42/ | (Intercept) | −0.13624 | 0.07399 | 7.91 | −1.841 | 0.103 |
| | Focus.2contrastive | 0.18483 | 0.10288 | 9.48 | 1.797 | 0.104 |
| | Position.L | 0.83609 | 0.07006 | 166.61 | 11.933 | <2e−16*** |
| | Focus.2contrastive:Position.L | 0.07594 | 0.10093 | 169.53 | 0.752 | 0.453 |

would induce changes to tonal targets on a word via $F_0$ rescaling, but adjustments to a movement rescaling parameter would result in processes of tonal enhancement/reduction in accordance with the overall duration of the tone-bearing units. The role of these particular parameters and their relationship to tone production and focus in YM necessitates further research, but the current results suggest that different articulatory parameters may be responsible for the different tonal patterns observed here.

## 5. Conclusions

Differences in the information structure in Yoloxóchitl Mixtec are conveyed via parallel processes of NP dislocation and

**Table 6**
Fixed effects for dynamic linear mixed effects models for level and rising tonal melodies. Most models have the structure: zlogF0 ∼ Time * Focus * Position + (1 + Focus|Speaker) + (1|Word), but the model for melody /1.1/ excludes a random intercept by item since only one target word for this melody served as the target in the experiment. All models were estimated using lmertest (Kuznetsova et al., 2017), which examines the predictor variance using the *t* distribution; the *t* values here are not the result of a standardized *t* test.

| Melody | Effect | Estimate | St.Error | df | *t* value | Pr(> \|*t*\|) |
|---|---|---|---|---|---|---|
| /1.1/ | (Intercept) | −0.799570 | 0.098926 | 11.600 | −8.083 | 4.14e−06*** |
| | Time | −0.059091 | 0.011137 | 942.600 | −5.306 | 1.40e−07*** |
| | Focus.2contrastive | 0.414239 | 0.085859 | 17.300 | 4.825 | 0.000152*** |
| | Position.L | 0.040948 | 0.053060 | 945.100 | 0.772 | 0.440460 |
| | Time:Focus.2contrastive | −0.038332 | 0.015770 | 942.700 | −2.431 | 0.015256* |
| | Time:Position.L | 0.001891 | 0.015750 | 942.600 | 0.120 | 0.904475 |
| | Focus.2contrastive:Position.L | 0.038234 | 0.074606 | 944.700 | 0.512 | 0.608439 |
| | Time:Focus.2contrastive:Position.L | −0.031925 | 0.022303 | 942.500 | −1.431 | 0.152649 |
| /3.3/ | (Intercept) | 1.469e−01 | 1.770e−01 | 2.300e+00 | 0.830 | 0.48355 |
| | Time | −6.178e−02 | 9.743e−03 | 2.080e+03 | −6.342 | 2.78e−10*** |
| | Focus.2contrastive | 3.394e−01 | 1.445e−01 | 1.050e+01 | 2.349 | 0.03952* |
| | Position.L | 2.151e−01 | 4.606e−02 | 2.080e+03 | 4.669 | 3.21e−06*** |
| | Time:Focus.2contrastive | 1.393e−03 | 1.398e−02 | 2.080e+03 | 0.100 | 0.92062 |
| | Time:Position.L | −3.655e−02 | 1.377e−02 | 2.080e+03 | −2.654 | 0.00802** |
| | Focus.2contrastive:Position.L | −6.272e−02 | 6.595e−02 | 2.080e+03 | −0.951 | 0.34172 |
| | Time:Focus.2contrastive:Position.L | 8.609e−03 | 1.977e−02 | 2.080e+03 | 0.436 | 0.66321 |
| /4.4/ | (Intercept) | 9.115e−01 | 1.524e−01 | 1.010e+01 | 5.980 | 0.000128*** |
| | Time | 1.154e−02 | 7.962e−03 | 2.261e+03 | 1.450 | 0.147334 |
| | Focus.2contrastive | 4.928e−01 | 1.320e−01 | 1.040e+01 | 3.734 | 0.003650** |
| | Position.L | 2.020e−01 | 3.730e−02 | 2.261e+03 | 5.415 | 6.78e−08*** |
| | Time:Focus.2contrastive | −4.172e−03 | 1.177e−02 | 2.261e+03 | −0.354 | 0.723140 |
| | Time:Position.L | −6.980e−02 | 1.126e−02 | 2.261e+03 | −6.199 | 6.72e−10*** |
| | Focus.2contrastive:Position.L | −7.201e−03 | 5.520e−02 | 2.261e+03 | −0.130 | 0.896210 |
| | Time:Focus.2contrastive:Position.L | 1.005e−02 | 1.665e−02 | 2.261e+03 | 0.604 | 0.546204 |
| /1.3/ | (Intercept) | −2.770e−01 | 1.806e−01 | 2.300e+00 | −1.534 | 0.2508 |
| | Time | −9.978e−02 | 9.298e−03 | 1.936e+03 | −10.731 | <2e−16*** |
| | Focus.2contrastive | 3.005e−01 | 1.210e−01 | 1.040e+01 | 2.484 | 0.0314* |
| | Position.L | 3.324e−01 | 4.411e−02 | 1.936e+03 | 7.537 | 7.37e−14*** |
| | Time:Focus.2contrastive | −5.103e−03 | 1.362e−02 | 1.936e+03 | −0.375 | 0.7080 |
| | Time:Position.L | 1.016e−01 | 1.314e−02 | 1.936e+03 | 7.728 | 1.73e−14*** |
| | Focus.2contrastive:Position.L | 4.449e−02 | 6.444e−02 | 1.936e+03 | 0.690 | 0.4900 |
| | Time:Focus.2contrastive:Position.L | 7.329e−03 | 1.926e−02 | 1.936e+03 | 0.381 | 0.7036 |
| /1.4/ | (Intercept) | −0.64051 | 0.15857 | 3.60 | −4.039 | 0.01926* |
| | Time | 0.05552 | 0.01032 | 2067.20 | 5.380 | 8.3e−08*** |
| | Focus.2contrastive | 0.41836 | 0.13236 | 11.10 | 3.161 | 0.00893** |
| | Position.L | 0.08802 | 0.04882 | 2067.10 | 1.803 | 0.07152 |
| | Time:Focus.2contrastive | −0.02735 | 0.01502 | 2067.70 | −1.821 | 0.06871 |
| | Time:Position.L | 0.28447 | 0.01459 | 2067.20 | 19.500 | <2e−16*** |
| | Focus.2contrastive:Position.L | 0.18328 | 0.07103 | 2067.10 | 2.580 | 0.00994** |
| | Time:Focus.2contrastive:Position.L | 0.01828 | 0.02123 | 2067.40 | 0.861 | 0.38939 |
| /3.4/ | (Intercept) | 4.025e−01 | 2.283e−01 | 2.800e+00 | 1.763 | 0.18242 |
| | Time | −1.063e−02 | 9.763e−03 | 2.175e+03 | −1.089 | 0.27646 |
| | Focus.2contrastive | 2.618e−01 | 1.246e−01 | 1.110e+01 | 2.102 | 0.05914 |
| | Position.L | 4.468e−01 | 4.580e−02 | 2.175e+03 | 9.755 | <2e−16*** |
| | Time:Focus.2contrastive | 1.838e−02 | 1.425e−02 | 2.175e+03 | 1.290 | 0.19711 |
| | Time:Position.L | 3.908e−02 | 1.381e−02 | 2.175e+03 | 2.830 | 0.00469** |
| | Focus.2contrastive:Position.L | −9.591e−03 | 6.684e−02 | 2.175e+03 | −0.143 | 0.88592 |
| | Time:Focus.2contrastive:Position.L | 1.462e−02 | 2.015e−02 | 2.175e+03 | 0.726 | 0.46806 |

phonetic enhancement. Broad, sentential focus is marked in situ with shorter target nouns and tones produced within a lower, contracted $F_0$ range. Contrastive and argument focus are marked ex-situ (NP dislocation) with longer target nouns and tones produced within a higher, expanded $F_0$ range. This lengthening and range expansion permitted the production of more hyperarticulated tonal melodies. Moreover, words with contrastive focus were realized with a raised $F_0$ range relative to argument focus. Though we observed phonetic modifications accompanying differences in information structure, stress-based durational asymmetries were retained on words regardless of its information structural content. While prosodic lengthening targeted the stressed syllable more than the unstressed syllable, tonal enhancement was not directly sensitive to lexical stress.

One surprising finding of the current study was the lack of correspondence between the degree of $F_0$ raising and the overall duration in YM words. Words with argument focus and contrastive focus were lengthened relative to sentential focus, but argument focus induced greater lengthening than contrastive focus. Why might the ordering of the durational effects differ from the ordering of the $F_0$ raising? There are two possible explanations for this pattern. One explanation may lie in the experimental condition itself. Consider that words are generally reduced in duration with repeated mentions (Fowler & Housum, 1987). In English, this effect is robust even when speech style and prosodic context are controlled (Baker & Bradlow, 2009; Turnbull, 2017). In the current study, the argument focus condition was the first block of the experiment. Thus, responses provided within this block reflected initial productions of these words by each of the speakers. Responses provided in the contrastive focus block reflected second mentions of the target stimuli. While this may appear to account for the durational differences, note that second-mention reduction

**Table 7**
Fixed effects for dynamic linear mixed effects models for falling tonal melodies. Most models have the structure: zlogF0 ~ Time * Focus * Position + (1 + Focus|Speaker). These models exclude a random intercept by item since only one target word for these melodies served as the target in the experiment. All models were estimated using lmertest (Kuznetsova et al., 2017), which examines the predictor variance using the *t* distribution; the *t* values here are not the result of a standardized *t* test.

| Melody | Effect | Estimate | St.Error | df | t value | Pr(> $\lvert t \rvert$) |
|---|---|---|---|---|---|---|
| /3.2/ | (Intercept) | −9.258e−02 | 9.750e−02 | 1.160e+01 | −0.950 | 0.36163 |
| | Time | −8.206e−03 | 1.132e−02 | 1.103e+03 | −0.725 | 0.46857 |
| | Focus.2contrastive | 5.581e−01 | 1.764e−01 | 1.060e+01 | 3.164 | 0.00945$^{**}$ |
| | Position.L | −2.486e−01 | 5.309e−02 | 1.103e+03 | −4.682 | 3.19e−06$^{***}$ |
| | Time:Focus.2contrastive | −9.688e−03 | 1.642e−02 | 1.103e+03 | −0.590 | 0.55527 |
| | Time:Position.L | −1.057e−02 | 1.601e−02 | 1.103e+03 | −0.660 | 0.50923 |
| | Focus.2contrastive:Position.L | −7.675e−02 | 7.702e−02 | 1.103e+03 | −0.997 | 0.31922 |
| | Time:Focus.2contrastive:Position.L | 2.547e−02 | 2.322e−02 | 1.103e+03 | 1.097 | 0.27297 |
| /4.2/ | (Intercept) | 0.13768 | 0.11791 | 10.60 | 1.168 | 0.26846 |
| | Time | −0.03290 | 0.01108 | 1181.00 | −2.969 | 0.00304$^{**}$ |
| | Focus.2contrastive | 0.39833 | 0.14572 | 11.60 | 2.734 | 0.01865$^{*}$ |
| | Position.L | −0.53594 | 0.05217 | 1181.20 | −10.273 | <2e−16$^{***}$ |
| | Time:Focus.2contrastive | −0.01794 | 0.01631 | 1181.00 | −1.101 | 0.27132 |
| | Time:Position.L | −0.09722 | 0.01567 | 1181.10 | −6.203 | 7.67e−10$^{***}$ |
| | Focus.2contrastive:Position.L | 0.03117 | 0.07663 | 1181.10 | 0.407 | 0.68428 |
| | Time:Focus.2contrastive:Position.L | −0.01299 | 0.02306 | 1181.10 | −0.563 | 0.57335 |
| /1.42/ | (Intercept) | 0.35481 | 0.08501 | 19.10 | 4.174 | 0.000508$^{***}$ |
| | Time | −0.17639 | 0.01548 | 863.00 | −11.393 | <2e−16$^{***}$ |
| | Focus.2contrastive | 0.35310 | 0.12835 | 19.10 | 2.751 | 0.012672$^{*}$ |
| | Position.L | 0.25771 | 0.07503 | 868.60 | 3.435 | 0.000621$^{***}$ |
| | Time:Focus.2contrastive | −0.04069 | 0.02318 | 863.20 | −1.755 | 0.079533 |
| | Time:Position.L | 0.16603 | 0.02189 | 863.00 | 7.583 | 8.73e−14$^{***}$ |
| | Focus.2contrastive:Position.L | −0.09696 | 0.11397 | 868.00 | −0.851 | 0.395149 |
| | Time:Focus.2contrastive:Position.L | 0.04344 | 0.03278 | 863.20 | 1.325 | 0.185433 |

**Table 8**
Fixed effects for duration linear mixed effects models in Section 3: dur ~ Focus * Position + (1 + Focus|Word) + (1 + Focus|Speaker). All models were estimated using lmertest (Kuznetsova et al., 2017), which examines the predictor variance using the *t* distribution; the *t* values here are not the result of a standardized *t* test.

| Model | Effect | Estimate | St.Error | df | *t* value | Pr(> $\lvert t \rvert$) |
|---|---|---|---|---|---|---|
| Cons. duration | (Intercept) | 8.132e−02 | 5.524e−03 | 2.400e+01 | 14.721 | 1.89e−13$^{***}$ |
| | Focus.2argument | 2.524e−02 | 9.862e−03 | 1.600e+01 | 2.559 | 0.0211$^{*}$ |
| | Focus.2contrastive | 1.757e−02 | 8.931e−03 | 1.600e+01 | 1.967 | 0.0669 |
| | Position.L | 1.825e−02 | 1.278e−03 | 4.229e+03 | 14.280 | <2e−16$^{***}$ |
| | Focus.2argument:Position.L | 2.380e−02 | 2.196e−03 | 2.779e+03 | 10.835 | <2e−16$^{***}$ |
| | Focus.2contrastive:Position.L | 1.149e−02 | 2.247e−03 | 1.857e+03 | 5.112 | 3.51e−07$^{***}$ |
| Vowel duration | (Intercept) | 8.411e−02 | 4.487e−03 | 2.300e+01 | 18.746 | 1.55e−15$^{***}$ |
| | Focus.2argument | 1.348e−02 | 5.024e−03 | 1.500e+01 | 2.684 | 0.0171$^{*}$ |
| | Focus.2contrastive | 9.095e−03 | 3.929e−03 | 1.400e+01 | 2.315 | 0.0361$^{*}$ |
| | Position.L | 9.281e−03 | 7.721e−04 | 4.304e+03 | 12.021 | <2e−16$^{***}$ |
| | Focus.2argument:Position.L | 6.759e−04 | 1.373e−03 | 3.523e+03 | 0.492 | 0.6225 |
| | Focus.2contrastive:Position.L | −3.157e−03 | 1.404e−03 | 2.407e+03 | −2.248 | 0.0247$^{*}$ |

is inhibited when the target word occurs in a different narrative context (as it would here) (Fowler, Levy, & Brown, 1997; Vajrabhaya & Kapatsinksi, 2011).[15]

The durational changes associated with final syllable onset consonants in the data broadly suggest that the locus of prosodic strengthening is the stem-final syllable in YM. Since final syllable onsets undergo greater lengthening under focus, this finding contrasts with the general observation that intervocalic positions are the locus of processes of lenition (Kaplan, 2010). One characteristic of running speech in YM is a tendency for obstruent spirantization and debuccalization of word onsets in polysyllabic words but less reduction in the onsets of final syllables (DiCanio, Amith, Castillo García, & Lilley, 2016). This asymmetry in reduction mirrors the stress-based pattern in prosodic lengthening found in the current work, but the specific relation between these processes remains an open question.

The word-prosodic systems of Otomanguean languages demonstrate some of the greatest complexity among languages of the world (Gordon, 2016; Maddieson, 2010). Despite this complexity, there is rather little research investigating how prosodic structure above the word interacts with tone and stress in these languages.[16] By combining fieldwork with laboratory methods in phonetic research, the current paper sheds light on this connection in Yoloxóchitl Mixtec and contributes to the growing laboratory phonology literature on the indigenous languages of the Americas (c.f. Gordon, 2017; Whalen & McDonough, 2015).

### Acknowledgements

This work was supported by NSF Grant #1603323 (DiCanio, PI) at the University at Buffalo. The authors wish to thank Jonathan Amith, Douglas Whalen, and the Yoloxóchitl community for their feedback and participation on this project. This manuscript benefitted from the helpful commentary of Karin Michelson, John Pate, Jonathan Barnes, and D. Robert Ladd.

[15] We thank an anonymous reviewer for this point.

[16] For a recent survey, see DiCanio and Bennett (2018).

**Table 9**
Fixed effects for static linear mixed effects models for each tonal melody in Section 3. Most models have the structure: zlogF0 ∼ Focus * Position + (1 + Focus|Speaker) + (1|Word), but models for melodies /3.2/,/4.2/, and /1.42/ exclude a random intercept by item since only one target word with each melody served as the target in the experiment. For the factor of *Focus*, *sentential focus* is the intercept. All models were estimated using lmertest (Kuznetsova et al., 2017), which examines predictor variance using the *t* distribution; the values here are not the result of a standardized *t* test.

| Melody | Effect | Estimate | St.Error | df | *t* value | Pr(> \|*t*\|) |
|---|---|---|---|---|---|---|
| /1.1/ | (Intercept) | −0.57996 | 0.22646 | 3.40 | −2.561 | 0.07419 |
| | Focus.2argument | −0.44843 | 0.13715 | 10.90 | −3.270 | 0.00756** |
| | Focus.2contrastive | −0.13337 | 0.14163 | 10.20 | −0.942 | 0.36813 |
| | Position.L | −1.15422 | 0.03560 | 432.70 | −32.420 | <2e−16*** |
| | Focus.2argument:Position.L | 1.12073 | 0.06775 | 431.40 | 16.541 | <2e−16*** |
| | Focus.2contrastive:Position.L | 1.04499 | 0.06809 | 430.70 | 15.347 | <2e−16*** |
| /3.3/ | (Intercept) | −0.29772 | 0.22668 | 1.10 | −1.313 | 0.39586 |
| | Focus.2argument | 0.26213 | 0.10536 | 9.10 | 2.488 | 0.03435* |
| | Focus.2contrastive | 0.60096 | 0.12314 | 7.70 | 4.880 | 0.00137** |
| | Position.L | 0.07505 | 0.04520 | 586.00 | 1.660 | 0.09739 |
| | Focus.2argument:Position.L | −0.01801 | 0.06085 | 584.10 | −0.296 | 0.76741 |
| | Focus.2contrastive:Position.L | −0.04942 | 0.06186 | 583.80 | −0.799 | 0.42468 |
| /4.4/ | (Intercept) | 1.30539 | 0.11383 | 6.00 | 11.468 | 2.71e−05*** |
| | Focus.2argument | −0.28560 | 0.17427 | 9.00 | −1.639 | 0.1357 |
| | Focus.2contrastive | 0.21347 | 0.15461 | 9.10 | 1.381 | 0.2004 |
| | Position.L | −0.08639 | 0.03713 | 659.70 | −2.327 | 0.0203* |
| | Focus.2argument:Position.L | 0.03596 | 0.05157 | 659.00 | 0.697 | 0.4858 |
| | Focus.2contrastive:Position.L | 0.07646 | 0.05390 | 659.10 | 1.419 | 0.1565 |
| /1.3/ | (Intercept) | −0.53913 | 0.19593 | 1.10 | −2.752 | 0.20009 |
| | Focus.2argument | −0.02328 | 0.14179 | 9.00 | −0.164 | 0.87322 |
| | Focus.2contrastive | 0.26486 | 0.07124 | 9.30 | 3.718 | 0.00455** |
| | Position.L | 0.19729 | 0.04885 | 552.30 | 4.038 | 6.15e−05*** |
| | Focus.2argument:Position.L | 0.42148 | 0.06619 | 549.50 | 6.367 | 4.06e−10*** |
| | Focus.2contrastive:Position.L | 0.47917 | 0.06844 | 549.80 | 7.001 | 7.45e−12*** |
| /1.4/ | (Intercept) | −0.69731 | 0.19861 | 2.20 | −3.511 | 0.0637 |
| | Focus.2argument | 0.25766 | 0.17353 | 8.50 | 1.485 | 0.1739 |
| | Focus.2contrastive | 0.59404 | 0.14882 | 8.50 | 3.992 | 0.0035** |
| | Position.L | 0.40838 | 0.05290 | 561.00 | 7.719 | 5.40e−14*** |
| | Focus.2argument:Position.L | 0.49644 | 0.06886 | 560.40 | 7.210 | 1.82e−12*** |
| | Focus.2contrastive:Position.L | 0.75281 | 0.07017 | 560.40 | 10.729 | <2e−16*** |
| /3.4/ | (Intercept) | 0.30315 | 0.19060 | 2.40 | 1.591 | 0.2333 |
| | Focus.2argument | −0.02753 | 0.13556 | 9.30 | −0.203 | 0.8435 |
| | Focus.2contrastive | 0.31637 | 0.11406 | 9.40 | 2.774 | 0.0207* |
| | Position.L | 0.61565 | 0.04237 | 699.10 | 14.532 | <2e−16*** |
| | Focus.2argument:Position.L | −0.06405 | 0.06159 | 695.60 | −1.040 | 0.2987 |
| | Focus.2contrastive:Position.L | −0.01977 | 0.06369 | 695.30 | −0.310 | 0.7564 |
| /3.2/ | (Intercept) | −0.37594 | 0.04740 | 7.70 | −7.930 | 5.75e−05*** |
| | Focus.2argument | 0.27367 | 0.11271 | 8.45 | 2.428 | 0.0398* |
| | Focus.2contrastive | 0.79693 | 0.11151 | 8.12 | 7.147 | 9.03e−05*** |
| | Position.L | −0.33836 | 0.04646 | 308.73 | −7.282 | 2.75e−12*** |
| | Focus.2argument:Position.L | −0.03574 | 0.06475 | 307.75 | −0.552 | 0.5814 |
| | Focus.2contrastive:Position.L | 0.02457 | 0.06643 | 307.57 | −0.370 | 0.7118 |
| /4.2/ | (Intercept) | 0.223887 | 0.062130 | 8.500 | 3.604 | 0.00632** |
| | Focus.2argument | −0.167013 | 0.132439 | 9.100 | −1.261 | 0.23867 |
| | Focus.2contrastive | 0.190898 | 0.122925 | 8.900 | 1.553 | 0.15529 |
| | Position.L | −0.902341 | 0.053213 | 322.900 | −16.957 | <2e−16*** |
| | Focus.2argument:Position.L | 0.044750 | 0.072073 | 321.400 | 0.621 | 0.53511 |
| | Focus.2contrastive:Position.L | 0.009149 | 0.074609 | 321.300 | 0.123 | 0.90248 |
| /1.42/ | (Intercept) | 0.01366 | 0.09563 | 8.96 | 0.143 | 0.88955 |
| | Focus.2argument | −0.14839 | 0.08020 | 10.15 | −1.850 | 0.09358 |
| | Focus.2contrastive | 0.05506 | 0.13580 | 4.34 | 0.405 | 0.70434 |
| | Position.L | 0.19574 | 0.06111 | 262.92 | 3.203 | 0.00153** |
| | Focus.2argument:Position.L | 0.64386 | 0.08951 | 231.17 | 7.193 | 8.73e−12*** |
| | Focus.2contrastive:Position.L | 0.71497 | 0.09449 | 265.10 | 7.566 | 6.34e−13*** |

## Appendix A. Statistical models

See Tables 3–9.

## Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.wocn.2018.03.001.

## References


Adams, J. N. (2007). *The regional diversification of Latin 200 BC–AD 600*. Cambridge University Press.

Anderson, E. R., & Concepción Roque, H. (1983). *Diccionario Cuicateco*. Number 26 in Serie de Vocabularios y Diccionarios Indígenas Mariano Silva y Aceves. Instituto Lingüístico de Verano: Mexico, D.F.

Andruski, J. E. (2006). Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics, 34*, 388–404.

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

Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech, 52*(4), 391–413.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. Version 0.999375-41, retrieved from http://CRAN.R-project.org/package=lme4.

Baumann, S. (2006). Information structure and prosody: Linguistic categories for spoken language annotation. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, J. Schlieer (Eds.), *Methods in empirical prosody research, language context, and cognition* (Vol. 3, pp. 153–180). Mouton de Gruyter.

Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic marking of focus domains: categorical or gradient? In *Proceedings of speech prosody, 2006* (pp. 301–304).

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*. Cambridge: Cambridge University Press.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology 2: Gesture, segment, prosody* (pp. 68–86). Cambridge: Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255–309.

Berinstein, A. E. (1979). *A cross-linguistic study on the perception and production of stress* (Master's thesis). Los Angeles: University of California. Available online at http://escholarship.org/uc/item/0t0699hc.

Birch, S., & Clifton, C. (1995). Focus, accent and argument structure: effects of language comprehension. *Language and Speech, 38*, 365–391.

Bishop, J. B. (2013). *Prenuclear accentuation in English: Phonetics, phonology, information structure* (Ph.D. thesis). Los Angeles: University of California.

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

Bombien, L., Mooshammer, C., Hoole, P., Rathcke, T., & Kuhnert, B. (2007). Articulatory strengthening in initial German /kl/ clusters under prosodic variation. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 457–460). Germany: Saarbrücken.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7/8/9), 1044–1098.

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. chapter 15.

Brunelle, M., & Pittayaporn, P. (2012). Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia. *Diachronica, 29*(4), 411–433.

Büring, D. (2007). Intonation, semantics, and information structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces*. Oxford: Oxford University Press.

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

Carroll, L. S. (2015). *Ixpantepec nieves mixtec word prosody* (Ph.D. thesis). San Diego: University of California.

Castillo García, R. (2007). *Descripción fonológica, segmental, y tonal del Mixteco de Yoloxóchitl, Guerrero* Master's thesis. México, D.F.: Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS).

Chang, Y.-C., & Hsieh, F.-F. (2012). Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review, 29*, 37–73.

Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics, 34*(176–201).

Chen, Y. (2011). How does phonology guide phonetics in segment-f0 interaction? *Journal of Phonetics, 39*(4), 612–625.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36*(4), 724–746.

Cheng, L. L.-S., & Downing, L. J. (2012). Against FocusP: Arguments from Zulu. In I. Kučerová & A. Neeleman (Eds.), *Contrasts and positions in information structure* (pp. 247–266). Cambridge University Press. chapter 10.

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*. Berlin, New York: Mouton de Gruyter.

Clopper, C. G., & Tonhauser, J. (2013). The prosody of focus in Paraguayan Guaraní. *International Journal of American Linguistics, 79*(2), 219–251.

de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*, 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., Castillo García, R., & Lilley, J. (2016). Obstruent lenition and voicing in a Yoloxóchitl Mixtec corpus (talk). In B. Tucker, M. Ernestus, & N. Warner (Eds.), *Satellite meeting on reduction, 15th conference in laboratory phonology.*

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., & Bennett, R. (2018). Prosody in Mesoamerican languages. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Prosody Handbook*. Oxford University Press. chapter 28.

DiCanio, C. T. (2008). *The phonetics and phonology of San Martín Itunyoso Trique* (Ph. D. thesis). Berkeley: University of California.

DiCanio, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics, 40*, 162–176.

DiCanio, C. T. (2016a). Pitch dynamics 5.0.http://www.acsu.buffalo.edu/~cdicanio/scripts/Pitch_Dynamics_5.praat.

DiCanio, C. T. (2016b). Tonal classes in Itunyoso Trique person morphology. In E. Palancar & J.-L. Léonard (Eds.), *Tone and inflection: New facts and new perspectives*. *Trends in linguistics studies and monographs* (vol. 296, pp. 225–266). Mouton de Gruyter. chapter 10.

Eady, S., & Cooper, W. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America, 80*, 402–415.

Eady, S., Cooper, W., Klouda, G., Mueller, P., & Lotts, D. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech, 29*, 233–251.

Esposito, C. (2010). Variation in contrastive phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association, 40*, 181–198.

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 *The handbook of phonetic sciences*, pp. 521–602). Wiley-Blackwell.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*(6), 3728–3740.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.

Fowler, C. A., Levy, E. T., & Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language, 37*, 24–40.

Francis, A. L., Ciocca, V., King Yu Wong, N., Ho Yin Leung, W., & Cheuk Yan Chu, P. (2006). Extrinsic context affects perceptual normalization of lexical tone. *Journal of the Acoustical Society of America, 119*(3), 1712–1726.

Gandour, J., Tumtavitikul, A., & Satthamnuwong, N. (1999). Effects of speaking rate on Thai tones. *Phonetica, 56*, 123–134.

Godjevac, S. (2005). Transcribing Serbo-Croatian intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. chapter 6.

Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The blackwell companion to phonology* (pp. 924–948). Wiley-Blackwell.

Gordon, M. (2017). Phonetic and phonological research on Native American languages: Past, present, and future. *International Journal of American Linguistics, 83*(1), 79–110.

Gordon, M. K. (2016). *Phonological typology.* Oxford University Press.

Gussenhoven, C. (1983a). Focus, mode, and the nucleus. *Journal of Linguistics, 19*(2), 377–417.

Gussenhoven, C. (1983b). Testing the reality of focus domains. *Language and Speech, 26*, 61–80.

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*(329–361).

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* (Ph.D. thesis). Tulane University.

Jun, S.-A. (Ed.). (2005). *Prosodic typology.* Oxford University Press.

Kaplan, A. (2010). *Phonology shaped by phonetics: The case of intervocalic lenition* (Ph. D. thesis). Santa Cruz: University of California.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language, 87*(4), 771–816.

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.* Cambridge Studies in Linguistics 119. (2nd ed.). Cambridge University Press.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*, 70–87.

Maddieson, I. (2010). Tone. In M. Haspelmat, M. Dryer, S. Matthew, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. chapter 13. Accessed on 10/27/2010.

Matisoff, J. A. (1990). Bulging monosyllables: Areal tendencies in southeast Asian diachrony. In *Proceedings of the sixteenth annual meeting of the Berkeley linguistics society* (pp. 543–559). Berkeley: University of California.

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics, 44*, 47–61.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics, 37*, 29–45.

Nakai, S., Turk, A., Suomi, K., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics, 40*(6), 796–807.

Nixon, J. S., Chen, Y., & Schiller, N. O. (2014). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition, and Neuroscience, 30*(5), 491–505.

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. chapter 12.

Pan, H.-H. (2007). Focus and Taiwanese unchecked tones. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation. Studies in linguistics and philosophy* (vol. 82, pp. 195–213). Dordrecht: Springer.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., & Wang, W. S.-Y. (2012). The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research, 55*, 579–595.

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

Peng, S.-H., Chan, M. K. M., Tseng, C.-Y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology* (pp. 230–270). Oxford University Press. chapter 9.

Prince, E. F. (1981). Topicalization, Focus-Movement, and Yiddish-Movement. In *Proceedings of the seventh annual meeting of the Berkeley linguistics society* (pp. 249–264). Alford: Daniel.

R Development Core Team, Vienna, A. (2017). R: A language and environment for statistical computing [computer program], version 3.3.3. http://www.R-project.org. R Foundation for Statistical Computing.

Remijsen, B., & van Heuven, V. (2005). Stress, tone, and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology, 22*, 205–235.

Scholz, F. (2012). *Tone sandhi, prosodic phrasing, and focus marking in Wenzhou Chinese* (Ph.D. thesis). Leiden University.

Sosa, J. M. (1999). *La entonación del español: su estructura fónica, variabilidad y dialectología*. Madrid: Ediciones Cátedra.

Turk, A., & White, L. (1999). Structural influences on accentual lengthening in english. *Journal of Phonetics, 27*(2), 171–206.

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

Turnbull, R. (2017). The role of predictability in intonational variability. *Language and Speech, 60*(1), 123–153.

Vajrabhaya, P., Kapatsinksi, V. (2011). There is more to the story: First-mention lengthening in Thai interactive discourse. In *Proceedings of the 17th congress of the phonetic sciences, Hong Kong* (pp. 2050–2053).

Vanrell, M. D. M., & Fernández Soriano, O. (2013). Variation at the interfaces in Ibero-Romance Catalan and Spanish prosody and word order. *Catalan Journal of Linguistics, 12*, 253–282.

Whalen, D. H., & McDonough, J. (2015). Taking the laboratory into the field. *Annual Review of Linguistics, 1*, 395–415.

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

Xu, C., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association, 33*(2), 165–181.

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*, 55–105.

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

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

Zerbian, S. (2007). Investigating prosodic focus marking in Northern Sotho. In E. O. Aboh, K. Hartmann, & M. Zimmermann (Eds.), *Focus strategies in African languages: The interaction of focus and grammar in Niger-Congo and Afro-Asiatic* (pp. 55–79). Mouton de Gruyter. chapter 3.

Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica, 68*, 161–191.