JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Tonal context influences tone-duration interaction: Evidence from Cantonese

Alan C. L. Yu[a]
UChicago Phonology Laboratory, University of Chicago, Chicago, Illinois 60637, USA
aclyu@uchicago.edu

Abstract: Phonetic typological studies suggest that syllable duration is inversely correlated with the accompanying tone's approximate average f_0, and tones with dynamic f_0 movement tend to be in longer syllables rather than shorter ones. Systematic instrumental investigations on tone-duration interaction remain scant, however; existing studies might be confounded as tonal context may impact duration realization due to phonetic constraints on tonal movement. This study investigates the effect of tonal environment on the durational realization of tones in Cantonese, showing that tone-dependent duration variation is governed by the tonal context. Implications of these findings for existing phonetic typology concerning tone-duration interaction are discussed.

[Editor: Tomoki Toda] https://doi.org/10.1121/10.0017517

Received: 15 September 2022 **Accepted:** 16 February 2023 **Published Online:** 9 March 2023

1. Introduction

Beyond differences in fundamental frequency (f_0) and phonation, tonal contrasts in the world's languages are often accompanied by systematic differences in duration. For example, vowels bearing rising tones are longer than those with falling tones (Gordon, 2001; Zhang, 2002). Several typological surveys (e.g., Duanmu, 1994; Gordon, 2001; Zhang, 2002) also reported that contour tones (tones with dynamic f_0 movement) show a more restricted distribution than level tones. In particular, contour tones are often licensed only in phonetically long syllables (e.g., stressed syllables, final syllables, and syllables with long vowels). For example, in Cantonese, when a syllable changes its tonal specification from mid level to low mid-rising as a consequence of morphological zero-derivation, a significant increase in syllable duration is observed (Yu, 2003a,b). Focusing on the relationship between tone height and duration, many have reported that low tone syllables are longer than high tone ones [see, e.g., Faytak and Yu (2011) for a typological survey]. Gandour (1977) suggests that, cross-linguistically, vowel duration tends to be inversely correlated with its accompanying tones' approximate average f_0. For example, Blight and Pike (1976) observed that, in the Oto-Manguean language, Tenango Otomi, which has three tones (high, low, and rising), "when in otherwise analogous environments, a vowel with low tone is longer than a vowel with high tone" (p. 56). In the Indo-Aryan language, Panjabi, which has three contrastive tones (low, mid, high), Gill (1960) observed that "the low tone is longer in duration than the mid tone and the mid tone is longer than the high tone" (p. 11). Most descriptions of tone-duration interaction rely on impressionistic judgments, however. Instrumental studies remain scarce, although there are notable exceptions. Based on an acoustic study of the production of three speakers, Alderete (2005), for example, found that, at least on long vowels, syllables with low tone in Tahltan, a Northern Athabaskan language, are approximately 25% longer than the non-low tone syllables; low stressed syllables typically have a low-level f_0 profile, while non-low tone syllables may have rising and falling f_0 profiles as a function of stress.

The negative correlation between tone height and duration is not without exceptions, however (cf. Faytak and Yu, 2011). Shanghainese is reported to have shorter low tone than high and mid tone syllables (Zee, 1978). A particularly puzzling case is found in Cantonese. Cantonese, a major language within the Chinese family, has six lexical tones. The three level tones are high-level (T55), mid-level (T33), and low-mid-level (T22). There are two rising tones [i.e., a low-mid to high, rising (T25) and a low-mid to mid, rising tone (T23)] and a falling tone [i.e., falling from low-mid to low (T21)]. Kong (1987) investigated the effects of tone on duration and reported that, among the level tones in the language, syllables with mid-level tone (T33) are the longest, syllables with the low-mid-level tone (T22) are intermediate, and syllables with the high level tone (T55) are the shortest; of the six tones in Cantonese, the longest is the low-mid to high rising tone (T25). While tone-duration interaction was not their main focus, many studies report duration measurements in their investigations of Cantonese tones. For example, the duration values reported in Rose (2000) and Mok *et al.* (2013) suggest that T22 and T33 are among the longest, while T21 is the shortest (Rose: T22/T33 > T55 > T25/T23 > T21; Mok *et al.*:

[a] Author to whom correspondence may be addressed.

T22/T33 > T23 > T55/T25/T21). Counterexamples such as Shanghainese and Cantonese, thus, cast doubt on the validity of Gandour's typological claim concerning the inverse relationship between tone height and duration. To be sure, the inconsistencies between reports regarding Cantonese tone-duration interaction point to a need to scrutinize the nature of this case further, especially in light of their methodological differences. For example, the duration data reported in Mok *et al.* are based on one female speaker. Rose's duration data came from ten speakers, gender balanced, but the segmental composition of the target syllables was not controlled. Kong's production study was most controlled, but the results were based on three subjects, one of whom was the author, who spoke a different variety of Cantonese than the other two subjects. There are also significant inconsistencies between patterns across subjects within that study. While Kong (1987) found that his speakers, as a group, exhibit the following hierarchy of rime duration based on tone: T25 (272 ms) > T33 (264 ms) > T23 (260 ms) > T22 (257 ms) > T55 (236 ms) > T21 (229 ms), the exact hierarchy actually differs across the three speakers he examined. While T25 is always longer than the other non-T55 tones and T33 is longer than the lower tones (i.e., T23, T22, and T21), T25 is longest for speakers 1 and 3, and T55 is among the shortest; T55 is longest for speaker 2. Also, the relative positions between the lower tones in the duration hierarchy vary across speakers. A larger sample of speakers is needed to ascertain the validity of the tone-duration hierarchy observed in Kong (1987) as a general claim about Cantonese. Another complication to the picture is the fact that the target syllables were elicited in different tonal contexts. While the target syllables were elicited in isolation in Rose's study, the target syllables were embedded within a fixed carrier sentence in the studies of Kong and Mok *et al.* Specifically, the target syllables were sandwiched between two low-mid (T22) tone syllables (Kong: ni˥ ko˧ hɐi˨ ___ tsi˨ "This word is ___"; Mok *et al.*: ŋɔ˨˧ tʊk˨ ___ tsi˨ "I say ___ word"). Given the significant contextual effects tones have on one another in Cantonese (Flynn, 2003; Wong, 2006), it is not clear if the reported durational differences observed in Kong (1987) and Mok *et al.* (2013) are due to genuine tonal differences or if they are phonetic artefacts induced by the effects of tonal context along the f_0 dimension. Wong (2006), for example, reported assimilatory carryover and dissimilatory anticipatory tonal effects in Cantonese. Thus, target tones preceded by T22 generally begin with a low onset on par with the f_0 at the terminus of the preceding T22. There is also a dissimilatory anticipatory tonal effect such that a high f_0 offset target of a preceding tone (i.e., T55 and T25) is higher when the following tone begins with a low f_0 onset (e.g., T25, T23, T21, and T22). In a fixed tonal context design, the effect of tonal context on a target tone could vary drastically depending on the nature of the target tone. For example, if the preceding and following tones are T22, like in the studies of Kong and Mok *et al.*, a high level tone such as T55 is expected to have a rising f_0 onset following T22 and a falling f_0 offset before the following T22. The following T22 might also cause the terminus of T55 to be higher due to the abovementioned dissimilation effect. On the other hand, little f_0 movement is expected for a T22 target tone in the same context. Given that dynamic tones take longer to realize than non-dynamic ones (Ohala and Ewan, 1973; Xu and Sun, 2002), the greater degree of f_0 movements in a target tone in the fixed-context condition, like in the case of a T55 flanked by two T22 tones, might require the speaker to take more time to realize the target tone than in a tonal context where little context-induced f_0 movement is expected (e.g., a T55 flanked by two T55 tones). Such durational adjustments induced by the tonal context might obfuscate the underlying relationship between tone and duration.

This study reexamines the nature of tone-duration interaction in Cantonese, focusing in particular on the effect of tonal context on duration realization. Specifically, we ask if and how the nature of tone-duration interaction would differ in tonal environments that minimize or encourage contextual effects on tonal realization. This paper begins with a presentation of the methodology of the production experiment (Sec. 2). The experimental results appear in Sec. 3, focusing first on the durational profile of Cantonese tones in different tonal contexts in Sec. 3.1, followed by an examination on how the tonal specification of neighboring syllables affect the f_0 realization of the target tone in Sec. 3.2. The discussion and conclusion appear in Sec. 4.

2. Experiment

2.1 Subject

Eleven native speakers of Cantonese (six female) participated in the experiment for course credit or a nominal fee. The subjects, all undergraduates at a university in the United States, were originally from Hong Kong and had been in the United States for less than four years at the time of recording.

2.2 Stimuli and procedure

To test the effects of tonal context on tone-duration interaction, each participant recited a list of target words embedded in two types of carrier sentences. In the first carrier sentence condition (the *matched-context* condition), the target syllable is flanked by syllables of similar or matching pitch levels. Thus, a high tone syllable (T55) is flanked by two high tone (T55) syllables. A low-mid to high rising (T25) syllable is flanked by a preceding low-mid tone (T22) and a following high tone (T55). A complete listing of the target syllables and their corresponding carrier phrases in the *matched-context* condition is shown in Table 1. This *matched-context* method is designed to reduce potential tonal context effects on syllable duration realization. As the control condition, we followed Kong's original fixed carrier phrase design, thus, providing a *fixed-context* condition to contrast with the *matched-context* condition. However, to ensure that the target words in two

Table 1. Target syllables and corresponding carrier phrases in the *matched-context* conditions.

Tone	Target syllable	Carrier phrase	Translation
T55	ji˥ “clothes”	ŋɔ˨˧ tʊk˺˨ tsʰyn˥ ___ sɑːm˥ tsʰi˧	“I say ‘wear ___’ three times.”
T25	ji˨˥ “chair”	ŋɔ˨˧ tʊk˺˨ muk˺˨ ___ sɑːm˥ tsʰi˧	“I say ‘wood ___’ three times.”
T33	ji˧ “wish”	ŋɔ˨˧ tʊk˺˨ kʰyt˺˨ ___ sei˧ tsʰi˧	“I say ‘determined ___’ four times.”
T21	ji˨˩ “doubt”	ŋɔ˨˧ tʊk˺˨ jim˨˩ ___ liŋ˨˩ tsʰi˧	“I say ‘suspect ___’ zero times.”
T23	ji˨˧ “ear”	ŋɔ˨˧ tʊk˺˨ muk˺˨ ___ sei˧ tsʰi˧	“I say ‘wood ___’ four times.”
T22	ji˨ “two”	ŋɔ˨˧ tʊk˺˨ luk˺˨ ___ luk˺˨ tsʰi˧	“I say ‘six ___’ six times.”

conditions are flanked by the same number of syllables in the carrier phrase, the carrier sentence in the control (*fixed--context*) condition is /ŋɔ˨˧ jiu˧ tʊk˺˨ ___ sɑːm˥ tsʰi˧/ “I need to read ___ three times”, rather than Kong’s /ni˥ ko˧ hɐi˨ ___ tsi˨/ “This word is ___.” Subjects were recorded in a soundproofed booth with a Marantz (Kawasaki, Japan) PMD670 solid-state recorder and a Shure (Niles, IL) SM10A head-mounted microphone at a sampling rate of 44 kHz, reading ten times the target syllables ([ji] in six tones) embedded in the two types of carrier sentences. The sentences from the *matched-context* condition were presented first. Items within each condition were randomized. A total of 120 utterances were recorded per subject [6 tones × 2 conditions (*matched-context* vs *fixed-context*) × 10 repetitions].

2.3 Measurements

The target syllables and the respective sentences were first manually segmented in PRAAT (Boersma and Weenink, 2019). Syllable duration was measured from the onset of the onset glide of the target syllable [ji] to the cessation of voicing of the target vowel [i]. f_0 extraction was done using ProsodyPro (Xu, 2013). The script extracts an f_0 track from pre-labeled.wav files and transforms the intervals between successive labels into f_0 values using a method that combines automatic vocal pulse marking by PRAAT (Boersma and Weenink, 2019) and manual correction by the investigator. The same program smooths the f_0 curves obtained using a trimming algorithm that eliminates discontinuities in the curves. The f_0 measurements were Z-transformed by participant prior to further analysis. To evaluate the temporal dynamics of the f_0 contours of the target syllables and for proper averaging, the Z-transformed f_0 curves were time-normalized; measurements were obtained at ten equidistant temporal locations throughout the target syllable. Both the duration of the target syllable and the duration of the respective sentence were measured.

3. Analysis and results

3.1 Duration analysis

Syllable duration was analyzed in terms of a mixed-effects linear regression, which included trial within each list (1–60); speaking rate, tone (T21, T22, T23, T25, T33, T55); condition (*matched-context* vs *fixed-context*); and the interaction between tone and condition as fixed factors. Speaking rate refers to the average syllable duration within each target phrase, which is the ratio of the acoustic duration of the utterance, measured from the onset of nasal murmur of [ŋ] in /ŋɔ˨˧/ to the cessation of vocalic voicing of [i] in /tsʰi˧/, to the number of syllables in the utterance (i.e., six). The model also included by-subject random intercepts and by-subject random slopes for trial, rate, tone, and condition to allow for by-subject variability in their effects on the syllable duration. All categorical variables were sum-coded, and the continuous variables were scaled and centered. The residuals of the initial fit were examined and found to deviate strongly from normality. As a result, residuals that were more than 2.5 standard deviations from the mean were trimmed, which amounted to no more than 2.5% of the data, and the model was refitted to the trimmed data set. The new model had a residual distribution much closer to normality, and it is the refitted models that are reported below. All data sets and the analysis scripts can be found at https://osf.io/98etp/?view_only=35a3ba419b7d4961b4d066961dad88fa.

Table 2 summarizes the estimated coefficients of the fixed effects predictors. The model shows that there is a significant effect of rate, suggesting the slower the speaking rate, the longer the target syllable duration ($\beta = 26.71$, $t\text{-value} = 5.54$, $p < 0.001$). There is a significant effect of condition on syllable duration [Fig. 2(a)]: Syllables in the *matched-context* condition are significantly shorter than syllables in the *fixed-context* condition ($\beta = -21.38$, $t\text{-value} = -4.25$, $p < 0.001$). This is noteworthy as it suggests that the duration difference observed across conditions is not due to a practice effect although the subjects read the *fixed-context* sentences last; practice effects generally lead to duration reduction rather than lengthening.

There are significant effects of tone on syllable duration. On average, shown in black in Fig. 1, T23 is longest ($\beta = 8.58$, $t\text{-value} = 2.71$, $p < 0.01$), while T55 is the shortest ($\beta = -12.70$, $t\text{-value} = -3.64$, $p < 0.001$). Crucially, there are significant interactions between condition and tone. As illustrated in Fig. 1, in the *fixed-context* condition, the syllable duration of T25 is much longer than it is in the *matched-context* condition ($\beta = -6.03$, $t\text{-value} = -3.51$, $p < 0.01$). When

Table 2. Regression analysis of the syllable duration measurements. ***, $p < 0.001$; **, $p < 0.01$. The p-values were obtained using normal approximation, which assumes that the t distribution converges to the z distribution as degrees of freedom (df) increase [see Mirman (2014) for details].

	Coefficient (s.e.[a])	t-value
Intercept	220.63 (8.92)	24.73***
Trial order	2.16 (2.20)	0.98
Rate	26.71 (4.82)	5.54***
Condition	−21.38 (5.04)	−4.25***
T55	−12.70 (3.49)	−3.64***
T25	4.29 (2.50)	1.71
T21	−0.88 (2.24)	−0.39
T23	8.58 (3.17)	2.71**
T22	4.23 (2.97)	1.42
Condition: T55	−0.05 (1.71)	−0.03
Condition: T25	−6.03 (1.72)	−3.51**
Condition: T21	−1.22 (1.68)	−0.72
Condition: T23	−3.03 (1.72)	−1.76
Condition: T22	8.61 (1.69)	5.08***

[a] Standard error (s.e.).

the target tone is T22, the effect of condition is significantly smaller, which is driven by the longer duration of T22 in the *matched-context* condition and the shorter duration of T22 in the *fixed-context* condition ($\beta = 8.61$, t-value = 5.08, $p < 0.001$). Figure 1 also suggests a marked difference in syllable duration when the tone is T23; the regression analysis confirms that there is a marginal interaction between condition and T23 ($\beta = -3.03$, t-value $= -1.76$, $p = 0.078$).

To further investigate the nature of the condition × tone interaction, a series of pairwise *post hoc* comparisons was conducted using the `emmeans` package in R to determine significant differences ($\alpha = 0.05$) in each tonal environment condition, with p-values corrected using the Tukey HSD method; results are provided in Table 3. In the *fixed-context* condition, T25 is significantly longer than T55 ($\beta = 22.97$, t-value = 4.44, $p = 0.003$) and T33 ($\beta = 15.57$, t-value = 3.63, $p = 0.013$); T23 is significantly longer than T55 ($\beta = 24.25$, t-value = 3.66, $p = 0.025$), T33 ($\beta = 16.85$, t-value = 3.50, $p = 0.023$), and T22 ($\beta = 15.98$, t-value = 3.35, $p = 0.033$). In the *matched-context* condition, T22 is significantly longer than T55 ($\beta = 25.58$, t-value = 4.00, $p = 0.012$) and T33 ($\beta = 14.65$, t-value = 3.17, $p = 0.044$) and marginally so relative to T21 ($\beta = 14.93$, t-value = 3.07, $p = 0.056$).

The duration analysis reveals that there are significant durational differences depending on what tone a syllable carries, but the tone-based differences differ across tonal environment conditions. Consistent with previous reports, the rising tones (T25 and T23) are longest when the carrier sentence is fixed, that is, when the syllables, and therefore the tones,

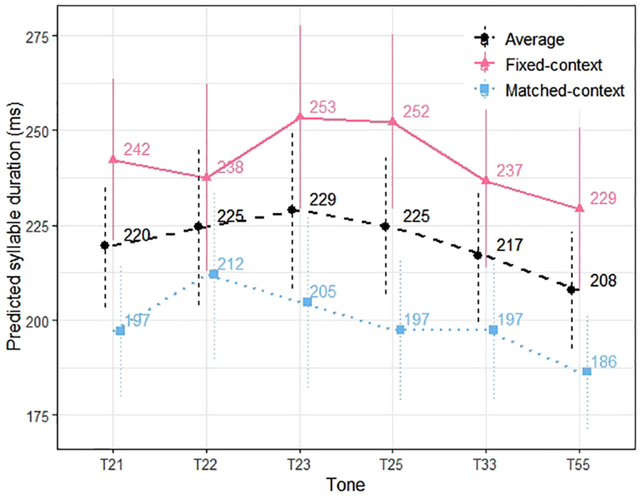

Fig. 1. Syllable duration across Cantonese tones in the *matched-context* (blue) and *fixed-context* (red) conditions. The average syllable duration measures are shown in black. The model-predicted duration values for the target tones in different conditions are provided. The error bars indicate 95% confidence intervals.

Table 3. Pairwise *post hoc* tests were applied to determine which pair of tones significantly differs in duration (α = 0.05) in each tonal environment condition, with *p*-values corrected using the Tukey honestly significant difference (HSD) method.

Matched-context		Fixed-context								
Contrast	Estimate	s.e.	df	*t*-ratio	*p*-value	Estimate	s.e.	df	*t*-ratio	*p*-value
T55–T25	−11.011	5.123	18.346	−2.149	0.307	−22.968	5.167	18.914	−4.445	0.003
T55–T21	−10.652	4.659	21.526	−2.286	0.242	−12.986	4.628	21.221	−2.806	0.095
T55–T23	−18.300	6.656	14.239	−2.750	0.125	−24.254	6.631	14.108	−3.657	0.025
T55–T22	−25.581	6.389	15.098	−4.004	0.012	−8.272	6.270	14.301	−1.319	0.770
T55–T33	−10.931	4.777	21.358	−2.288	0.242	−7.401	4.749	20.865	−1.558	0.633
T25–T21	0.358	4.785	19.718	0.075	1.000	9.982	4.858	20.834	2.055	0.347
T25–T23	−7.290	4.905	19.370	−1.486	0.676	−1.286	4.977	20.410	−0.258	1.000
T25–T22	−14.571	5.232	18.105	−2.785	0.106	14.696	5.220	18.101	2.815	0.100
T25–T33	0.080	4.203	26.624	0.019	1.000	15.567	4.295	28.277	3.625	0.013
T21–T23	−7.648	5.221	18.007	−1.465	0.689	−11.268	5.207	18.018	−2.164	0.301
T21–T22	−14.929	4.859	21.015	−3.073	0.056	4.714	4.739	19.744	0.995	0.914
T21–T33	−0.279	4.047	30.668	−0.069	1.000	5.585	4.033	30.345	1.385	0.735
T23–T22	−7.281	4.844	20.947	−1.503	0.666	15.982	4.772	20.041	3.349	0.033
T23–T33	7.370	4.802	20.809	1.535	0.647	16.853	4.819	20.875	3.497	0.023
T22–T33	14.650	4.625	22.994	3.168	0.044	0.871	4.566	21.788	0.191	1.000

flanking the target syllable do not vary across target syllables. Between level tones, T55 is shortest, while T22 and T33 are comparable in duration. However, when the tonal environments vary to minimize contextual influences between the neighboring tones and the target tone, the longest tones are no longer the rising tones. Rather, T22, the low level tone, is the longest. Moreover, consistent with Gandour's prediction, T22, a low-mid tone, is longer than T33, a mid tone, and T55 is the shortest. It is worth noting that, even in the *matched-context* condition, the rising tones also trended higher than the high tone, T55, but those differences were not statistically significant. What factors could lead to the marked differences in tone-duration interaction in the two conditions? We hypothesized at the outset that tonal context effects along the f_0 dimension might influence tonal realization and complicate the relationship between tone and duration. To ascertain the effect of tonal context on tonal realization, Sec. 3.2 examines the f_0 contours of the Cantonese tones in detail.

3.2 f_0 analysis

f_0 values were analyzed using smoothing spline analysis of variance (SSANOVA), a curve-comparing technique that has been used previously for the comparisons of tongue shapes (e.g., Davidson, 2006), formant trajectories (e.g., Decker and Nycz, 2006), and f_0 contours (e.g., Gao and Arai, 2018). SSANOVA estimates a mean contour and a s.e. based on distribution of the

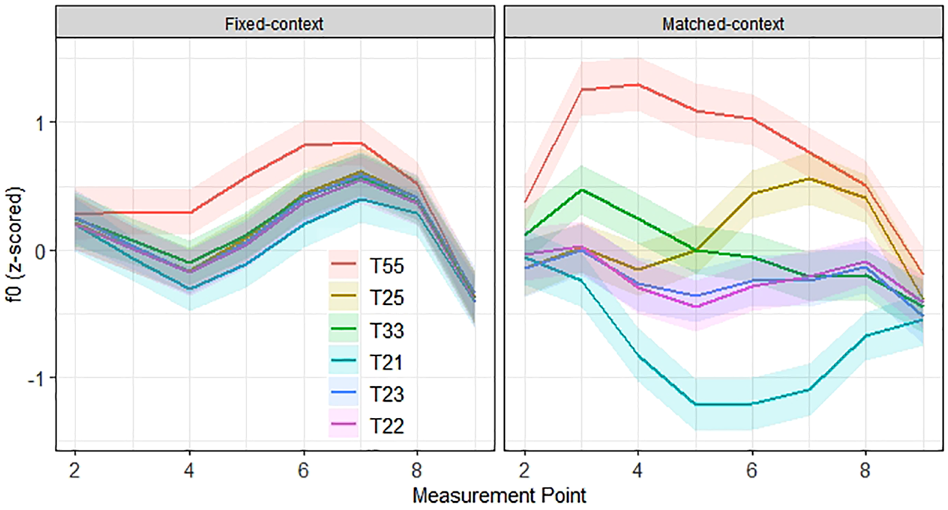

Fig. 2. SSANOVA predictions of the f_0 contours of Cantonese tones in the *matched-context* (left panel) and *fixed-context* (right panel) conditions.

input. When plotted with 95% confidence interval, all the non-overlapping regions signify significant differences in the contours. In comparison with analysis of variance (ANOVA) of selected points along the tonal contour, SSANOVA modeling compares the whole trajectory and is, therefore, capable of offering more fine-grained comparisons between tonal contours. To avoid potential word boundary effects, such as f_0 perturbation from the coda consonant of the preceding syllable and the initial consonant of the following syllable, measurements at measurement points 1 and 10 were excluded from the f_0 analysis.

Figure 2 illustrates the SSANOVA predictions for all six tones in Cantonese across the two pitch conditions. The variability within each of the Bayesian confidence intervals shows inter-subject variability across speakers. For each subject, the f_0 values were averaged across repetitions for each token before SSANOVA modeling. In f_0 regions where the confidence intervals overlap greatly, the f_0 values between tones are not distinct. Conversely, if the confidence intervals do not overlap, the f_0 values are phonetically distinct. Several observations are particularly noteworthy. First, with the exception between the contours of T23 and T22, the six tones have clear and distinct contours under the *matched-context* condition, while, in the *fixed-context* condition, all tonal contours, except for T55, are quite overlapping. Indeed, all tones have a rising f_0 contour in the *fixed-context* condition, likely due to the fact that the target syllable is preceded by a low-mid tone (T22) and followed by a high tone (T55). Recall that syllable duration is, on average, significantly longer in the *fixed-context* condition compared to the *matched-context* condition, yet tonal distinctions are much more dispersed in the *matched-context* condition than in the fixed condition. These findings clearly establish that tonal context plays an important role in the tonal realization of the target syllable, independent of, if not despite, the overall durational profile of the syllable itself.

4. Discussion and conclusion

The results of the present study show that syllable duration in Cantonese varies as a function of tone, although the nature of this tone-dependent duration variation is modulated by the tonal context in which the target syllable is situated. Specifically, the rime duration hierarchies (from longest to shortest), according to the model-predicted duration values (see Fig. 1), are as follows: *fixed-context* condition: T23, T25 > T21 > T22 > T33 > T55; *matched-context* condition: T22 > T23 > T25, T21, T33 > 55. Thus, contour tone syllables (T25 and T23) are longer than level tone syllables, but only in the *fixed-context* condition, i.e., when the tonal environment flanking the target tones does not vary. The results from the *fixed-context* condition are also broadly consistent with those reported in Kong (1987) as he found that the three speakers in his study, as a group, exhibit the following hierarchy of rime duration based on tone: T25 (272 ms) > T33 (264 ms) > T23 (260 ms) > T22 (257 ms) > T55 (236 ms) > T21 (229 ms). That is, T25 and T23 are among the longest tones, and T33 is longer than T55. The duration pattern observed in the *matched-context* condition, however, shows a drastically different picture. That is, when the tonal context varies to match the tonal profile of the target tone, the longest tone is T22, a low-mid tone, not the rising tones. Of particular interest is the finding of the f_0 analysis, which reveals significant neutralization in f_0 contour distinctions in the *fixed-context* condition; tonal distinctions are well-maintained in the *matched-context* condition. The neutralization of all tones in the *fixed-context* condition toward an f_0 rising profile with varying degrees of curvature and steepness might have obscured underlying durational differences due to constraints on dynamic f_0 realization. Further research is needed to elucidate the mechanism behind the duration adjustments, however. For example, if T22 is underlyingly the longest tone, as seen in the *matched-context* condition, did it come to be shorter than T25 and T23 in the *fixed-context* condition due to the lengthening of the two rising tones in the *fixed-context* condition, the shortening of T22, or both?

An aspect of our findings that might seem surprising at first glance concerns the main effect of condition. That is, the target syllables are generally longer in the *fixed-context* condition than in the *matched-context* condition. The situation is especially puzzling in the case of the T25 target syllable, where a durational difference is observed across conditions although the flanking tonal contexts are identical across conditions (i.e., T25 is preceded by T22 and followed by T55). This difference might stem from the difference in the nature of the quoted constituent within each carrier phrase condition. The target syllable in the *fixed-context* condition, as the sole syllable in the constituent under focus, might enjoy stronger focus-induced duration expansion than the target syllable in the *matched-context* condition, which is part of a disyllabic phrase under focus. The disyllabic constituents in the *matched-context* condition also vary in phrase structure; some are verb-object-structured phrases (e.g., wear clothes), while others are compounds (e.g., wood ear). The tone-dependent duration difference observed here is not likely to be attributable solely to the difference in focus constituent size, however, given that the f_0 range is more expanded and tonal contrast more pronounced in the *matched-context* condition than in the *fixed-context* condition. The opposite effect is otherwise expected given the longer syllable duration in the *fixed-context* conditions. More research is warranted to examine any potential effects of focus, constituent size, and phrase structure on tone-duration interaction.

Another intriguing aspect of the results concerns the duration patterning of T21. Many studies have found T21 to be the shortest tone in Cantonese (e.g., Kong, 1987; Mok *et al.*, 2013; Rose, 2000), yet, in both tonal context conditions, the duration of T21 hovers around the mean within each condition. Given that this tone is simultaneously the lowest tone and a falling tone, we might predict *a priori* that it should be longer than the level tones since low tone tends to be longer than higher tones, and contour tones tend to be longer than level tones. The nature of T21 is complex, however, since, unlike other tones in Cantonese, phonation is also a significant cue to the realization of this contrast (Yu and Lam, 2014; Zhang and Kirby, 2020). Further investigation is needed to elucidate the typological relationship between f0, phonation, and duration in the realization of tonal contrasts as there is a general lacuna in the current tonal literature.

Our findings have significant implications for various typological claims about the relationship between tone and duration. While the findings from the *fixed-context* condition are consistent with typological claims about contour tone distribution and partly consistent with the typological claim about the inverse relationship between tone height and duration, the duration patterns revealed in the *matched-context* condition paint a different picture. Contour tones are not always longer than level tones. In fact, the low tone, T22, is longer than the two rising tones in terms of their mean values, even if the differences are not statistically significant. However, consistent with the hypothesized inverse relationship between tone height and duration, T22 is longest and T55 is shortest among the level tones. Given that most acoustic studies of tone-duration either employed a fixed carrier phrase design (Kong, 1987; Zee, 1978) or did not control for tonal contexts (e.g., Alderete, 2005; Zhang, 2002), this raises questions about how results from such studies should be evaluated against typological claims. That is, are these findings representative of the effects of phonological tonal specification on duration realization, or are they illustrative of the effects of phonetic constraints on tonal movement on duration? To be sure, the use of a fixed carrier phrase is common practice in most phonetic studies as it is generally believed to minimize contextual effects on research outcomes. However, when the target stimuli interact with the context in different ways, contextual effects could be amplified, rather than the opposite. In the present case, the fixed tonal context imposes strong tonal context influence on the target tones such that most tones end up with a rising contour, even if the tones might still be differentiated by f_0 height. While more cross-linguistic instrumental studies are needed to ascertain whether or not the influence of tonal context on tone-duration interaction is language-specific, the present study shows that the *matched-context* design can offer a way to minimize the potential influence of phonetic constraints on tonal movements, which is shown here to confound the nature of the tone-duration interaction.

Acknowledgments

This work was supported in part by National Science Foundation Grant No. BCS1827409.

References and Links

Alderete, J. (**2005**). "On tone and length in Tahltan (Northern Athabaskan)," in *Athabaskan Prosody*, edited by S. Hargus and K. Rice (Benjamins, Amsterdam), pp. 185–207.

Blight, R. C., and Pike, E. V. (**1976**). "The phonology of Tenango Otomi," Int. J. Am. Linguist. **42**, 52–57, available at https://www.jstor.org/stable/1264808.

Boersma, P., and Weenink, D. (**2019**). "Praat (version 6.1) [computer program]," http://www.praat.org/.

Davidson, L. (**2006**). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," J. Acoust. Soc. Am. **120**(1), 407–415.

Decker, P. D., and Nycz, J. (**2006**). "A new way of analyzing vowels: Comparing formant contours using smoothing spline ANOVA," in *Proceedings of New Ways of Analyzing Variation 35*, November 9–12, Columbus, OH.

Duanmu, S. (**1994**). "Syllabic weight and syllable duration: A correlation between phonology and phonetics," Phonology **11**, 1–24.

Faytak, M., and Yu, A. C. L. (**2011**). "A typological study of the interaction between level tones and duration," in *Proceedings of the International Congress of the Phonetic Sciences XVII*, August 17–21, Hong Kong, pp. 659–662.

Flynn, C. Y. C. (**2003**). *Intonation in Cantonese* (LincomEuropa, Munich, Germany).

Gandour, J. (**1977**). "On the interaction between tone and vowel length: Evidence from Thai dialects," Phonetica **34**, 54–65.

Gao, J., and Arai, T. (**2018**). "F0 perturbation in a 'pitch-accent' language," in *Proceedings of the 6th International Symposium of the Tonal Aspects of Languages*, June 18–20, Berlin, Germany, pp. 56–60.

Gill, H. S. (**1960**). "Panjabi tonemics," Anthropol. Linguist. **2**, 11–18, available at https://www.jstor.org/stable/30022266.

Gordon, M. (**2001**). "A typology of contour tone restrictions," Stud. Lang. **25**, 405–444.

Kong, Q.-M. (**1987**). "Influence of tones upon vowel duration in Cantonese," Lang. Speech **30**(4), 387–399.

Mirman, D. (**2014**). *Growth Curve Analysis and Visualization Using R* (Chapman and Hall/CRC, Boca Raton, FL).

Mok, P., Zuo, D., and Wong, P. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Var. Change **25**, 341–370.

Ohala, J. J., and Ewan, W. G. (**1973**). "Speech of pitch change," J. Acoust. Soc. Am. **53**, 345.

Rose, P. (**2000**). "Hong Kong Cantonese citation tone acoustics: A linguistic-tonetic study," in *Proceedings of the 8th Australian International Conference on Speech Science and Technology*, December 4–7, Canberra, Australia, pp. 198–203.

Wong, Y. W. (**2006**). "Contextual tonal variations and pitch targets in Cantonese," in *Proceedings of the 3rd International Conference on Speech Prosody*, May 2–5, Dresden, Germany, pp. 317–320.

Xu, Y. (**2013**). "ProsodyPro a tool for large-scale systematic prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, August 30, Aix-en-Provence, France, pp. 7–10.

Xu, Y., and Sun, X. (**2002**). "Maximum speed of pitch change and how it may relate to speech," J. Acoust. Soc. Am. **111**, 1399–1413.

Yu, A. C. L. (**2003a**). "Contour tone induced lengthening in Cantonese," in *Proceedings of the 15th International Congress of Phonetic Sciences*, August 3–9, Barcelona, Spain, pp. 2381–2384.

Yu, A. C. L. (**2003b**). "Some methodological issues in phonetic typology research: Cantonese contour tone revisited," in *Proceedings of the 29th Annual Meeting of the Berkeley Linguistics Society*, February 14–17, Berkeley, CA.

Yu, K. M., and Lam, H. W. (**2014**). "The role of creaky voice in Cantonese tonal perception," J. Acoust. Soc. Am. **136**(3), 1320–1333.

Zee, E. (**1978**). "Duration and intensity as correlates of F_0," J. Phon. **6**, 213–220.

Zhang, J. (**2002**). *The Effects of Duration and Sonority on Contour Tone Distribution—A Typological Survey and Formal Analysis* (Routledge, New York).

Zhang, Y., and Kirby, J. (**2020**). "The role of F_0 and phonation cues in Cantonese low tone perception," J. Acoust. Soc. Am. **148**, EL40–EL45.