Proceedings of CLS 58 (2022), 211-230
© Chicago Linguistic Society 2023. All rights reserved.

Comparative tonal text-setting in Mandarin and Cantonese popular song

James P. Kirby
Institute of Phonetics and Speech Processing, LMU Munich

1 Introduction

A common question asked by speakers of European languages with a naïve (and even not-so-naïve) understanding of tone languages is, "how is it possible to sing in a tone language?" This question highlights the fact that in a tone language linguistic and musical uses of pitch appear to conflict: if a Thai singer sings the word 'far' (*klāj*) on a descending musical melody, won't listeners understand 'near' (*klâj*) instead? But clearly, people sing in all languages, including tone languages, so there must be a way that artists and composers meet this challenge.

There are three important components to the answer to this question. The first is what might be called *top-down cues*, which is to say, context. Language is redundant, and song texts are often formulaic, predictable, and/or repetitive. Loss of tone information may not affect comprehension. In many cases, the other words in the lyric, or knowledge of the likely subject material, or other kinds of structural linguistic knowledge, will enable the listener to decode the message. Another important component of the answer is *residual phonetic cues*, that is, pitch modulations laid on top of the melody contour, which may assist the listener in recovering the underlying lexical tone (see Schellenberg & Gick 2020 for a recent review).

The focus of the present study, however, is on a third component: the *text-setting constraints* which govern the productive ability to set texts to songs. There is a long tradition of work on stress-beat text-setting in non-tonal languages, but a body of research over the past few decades suggests that a wide variety of musical traditions in many—possibly all—tone languages enforce some sort of congruence between musical melody and tonal sequences. In other words, given a vocal melody, different sequences of words are judged to "fit" the melody better than others, depending (in part) on the tones of the words in that sequence. One research challenge then becomes to determine which principles determine the goodness of this fit within and across languages.

1.1 Text-setting in non-tonal languages

In languages like French and English, it has long been observed that native speakers familiar with a particular tradition of sung or chanted verse can readily set novel lines to existing rhythms with a significant degree of agreement (Jakobson 1960; Kiparsky 1977; Halle & Lerdahl 1993; Dell & Halle 2009; Hayes 2009 *inter alia*). To take just one example, in English (as well as in other languages) it is important for major stressed syllables to occur on strong musical beats, and perhaps even more important for unstressed syllables *not* to occur on strong beats. To see this preference in action, consider the third line of the "Happy Birthday" song, the

line in which a name needs to be inserted. If the name is a trochee like *Susan* (Figure 1a), this is unproblematic, but with an iamb like *Suzanne* (Figure 1b), the setting which results from the same strategy—1-to-1, left-to-right association of notes and syllables—is generally judged suboptimal by most native speakers of English. The repair, which is to insert an upbeat preceding the final measure and realize the stressed syllable melismatically (Figure 1c), is perhaps still somewhat forced, but nevertheless clearly preferable to setting the unstressed syllable of *Suzanne* on the strong downbeat of the third measure. This intuition reflects what we might call the basic English text-setting constraint: unstressed syllables should not be set to metrically prominent positions in the verse.

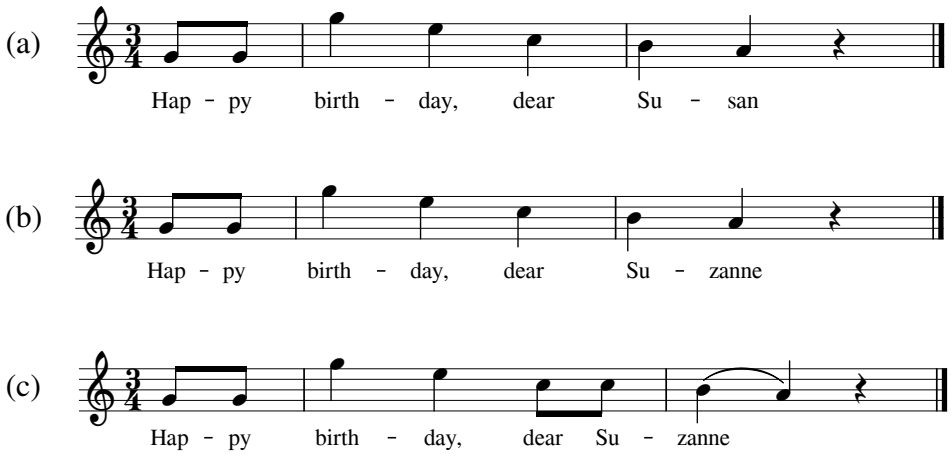

Figure 1: Three examples of setting English texts to the "Happy Birthday" meter. (a): a well-set text. (b): a suboptimal setting, with an unstressed syllable set to a strong beat. (c): a melismatic repair, which ensures that only stressed syllables are associated with strong metrical positions.

1.2 Text-setting in tonal languages

There is now a considerable body of work which establishes that a good match between tone and melody is not simply or even primarily a matter of aesthetics, but that structural-linguistic principles are also involved. While text-setting has been studied in an increasingly broad range of tonal languages, here I will focus on just two, Mandarin Chinese [cmn] and Cantonese [yue]. The tonal inventory of Mandarin consists of four tones: a level tone (T1), a rising tone (T2), a low-falling tone (T3) and a high-falling tone (T4). In addition, some function words, like the adjective suffix 的 /de/ or the perfective verb suffix 了 /le/, are usually realized with a so-called neutral tone, whose surface realization is largely predictable on the basis of the preceding tone (Cao 1992). Cantonese, on the other hand, contrasts 6 tones in sonorant-final syllables (Bauer & Benedict 1997), although some of these are presently undergoing various types of mergers (Mok *et al.* 2013). In this paper, Mandarin is transliterated using *hànyǔ pīnyīn*, with tones represented using diacritics, and Cantonese using the *jyutping* system, with tones represented as Chao

numbers.

Part of the reason for restricting our focus to these two languages is that one of the most intensely studied instances of tonal text-setting is that of 'Cantopop' (粵語流行音樂), an umbrella term for Cantonese-language popular music produced from the late 1970s onward. Text-setting in Cantopop was first studied by Marjorie Chan in a 1987 BLS paper (Chan 1987a) and accompanying UCLA working paper (Chan 1987b), and has subsequently been the focus of several further studies (Wong & Diehl 2002; Ho 2006; Ho 2010; Schellenberg 2011; Lo 2013). Given that this line of work has informed a great deal of subsequent research, it is worth considering the findings in some detail.

Chan made two important observations, both of which have been corroborated many times since. First, when there is divergence between the tones of the syllables in different verses set to the same positions in the melodic line, the difference seems to only be in the tonal onset. This can be seen clearly in the first and last syllables of the texts shown in Figure 2: the first syllable in the first verse has a high tone /55/, while the first syllable in the second verse has a rising tone /25/. Similarly, the syllables set to the last note in the phrase have the opposite tones. What appears to be critical is not that the tones across verses are the same, but rather that they end high.

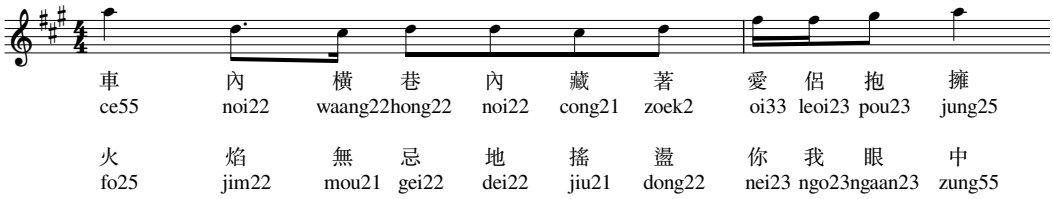

Figure 2: Excerpt from 黑色午夜 by Leslie Cheung. Adapted from Lo (2013).

Second, Chan observed a striking degree of parallelism between the direction that the musical note sequence takes in relation to the direction of the associated tone sequences. It seems that in Cantopop, rather than a particular lexical tone being associated with a particular note (a strategy employed in some classical and religious genres: see e.g. Yung 1983; Tanese-Ito 1988), for any given sequence of two syllables, the *direction* of linguistic pitch change from the first syllable to the second tends to be mirrored in the corresponding musical note sequence. For example, a low tone-high tone sequence (such as 眼中 /ngaan23-zung55/ in Figure 2) will tend to be set to a rising melodic sequence (G# rising to A), while a high tone-low tone sequence (such as 車內 /ce55-noi22/) will tend to be set to a falling sequence (A falling to D). In what follows, I will refer to sequential pairs of lexical tones as *tonal bigrams*, or just *bigrams* for short.

A subsequent study by Wong & Diehl (2002) presented a small quantitative survey of the way tonal bigrams are actually treated melodically in text-setting. For the purposes of defining pitch direction from one tone to the next, they grouped the six tones of Cantonese into three sets of high, mid, and low based on their relative

pitch offsets. Defining pitch direction on this basis, they observed that the direction of the tone sequence matched that of the note sequence in over 90% of cases.

To describe the various possible correspondences between the directions of musical melody and linguistic tone, Ladd & Kirby (2020) repurposed some standard terms from classical Western music theory, exemplified in Figure 3. They use the term *similar setting* to refer to the tendency described by Chan and Wong and Diehl for tone and melody to move together in the same direction, *oblique setting* for when either the tonal or melodic contour stays level while the other rises or falls, and *contrary setting* to refer to instances where one of the sequences rises while the other falls. This space of possibilities defines the three-by-three matrix shown in Table 1.

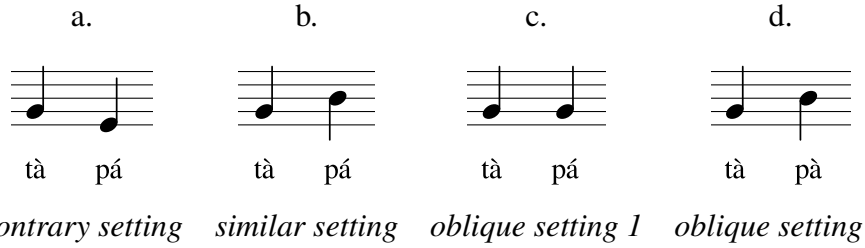

Figure 3: Different possibilities for setting bigrams to musical melodies. Contrary setting (a): a rising tonal sequence set to a falling musical sequence (or vice versa). Similar setting (b): both tonal and musical sequences are rising (or falling). Oblique setting 1 (c): consecutive syllables have different tones but are set to the same note. Oblique setting 2 (d): consecutive syllables have the same tone but are set to different notes. Adapted from Ladd and Kirby (2020).

		Note seq.	
Tone seq.	up	down	level
up	**Similar**	*Contrary*	Oblique
down	*Contrary*	**Similar**	Oblique
level	Oblique	Oblique	**Similar**

Table 1: Similar, contrary, and oblique settings, as defined by the relation between pitch direction in a sequence of two tones and the two corresponding musical notes.

Subsequent work by Ho (2006, 2010) and Lo (2013) further refined Wong and Diehl's classifications, but all studies have observed the same basic constraint at work, which we might describe as "avoid contrary settings". Note that, much as the basic English text-setting constraint formulated in Section 1.1 above referred to avoidance ("unstressed syllables should not be set to metrically strong positions") rather than enforcement ("stressed syllables are preferred on metrically strong positions"), avoidance of contrary settings is subtly but importantly different from a preference for similar settings. This is because there is some indication that oblique settings are not universally avoided. For example, Lo (2013) found that in one specific context—sequences of high-final (/55/ or /35/) tones set to a descending melody line—oblique settings were not uncommon, and Kirby & Ladd

(2016) found a significant number of level tone sequences set to rising or falling melodic lines in Vietnamese.

For whatever reason, rather less attention seems to have been directed at tone-melody correspondence in Mandarin popular music. The general consensus appears to be that tone-tune correspondence is largely ignored in Mandarin popular song, an observation going back at least to Chao (1956). For example, Ho (2006) writes that "many Mandarin songs, if not all, exhibit disagreement between tone and tune. Neither individual lexical tone contour nor tonal target transition is reflected in the melody systematically", while Chan claims that "word tones are often suspended in favour of the overall melody" (Chan 1987b:163). On the other hand, Wee (2007) suggests that tone-tune correspondence in Mandarin is brought about by preserving contrasts in the tonal and melodic structures at positions of metrical prominence. Specifically, if a note sequence is falling, the first syllable must end high (i.e. be tone 1 or 2); if a note sequence is rising, the first syllable must end low (i.e. be tone 3 or 4). Wee argues that these constraints are observed more strictly in structurally prominent positions. However, it should be noted that Wee's study considered Mandarin folk songs, rather than acculturated song, and difference in genre can have a considerable impact on tone-tune correspondence (see e.g. List 1961 on Thai).

1.3 The present study

If it is true that modern Mandarin music shows little evidence of tone-tune correspondence, why might this be the case? One possibility is that this simply represents a point of cross-linguistic divergence: there is no *a priori* reason that avoidance of contrary settings is necessarily a universal constraint active to the same extent in all tonal languages. However, there are a number of facts about the realization of tone in Mandarin that may also play a role. The goal of the present study was to take into account three aspects of Mandarin in order to assess their effect (or lack thereof) on tone-melody correspondence:

1. *Influence of (word and phrase) boundaries*. Virtually all off the instances of contrary settings observed in Ho (2010) and Lo (2013) for Cantonese were found to occur across syntactic and/or musical phrase boundaries. More critical still for Mandarin might be *word* boundaries, given that a large proportion of the lexicon is minimally disyllabic—at least 50% according to Wu *et al.* (2020) and perhaps as much as 67% according to Li & Thompson (1981). To the best of my knowledge, no previous work on tonal text-setting in either Mandarin or Cantonese has explicitly coded for word boundaries or compared correspondence counts when word-internal transitions were included versus excluded.
2. *Neutral tones*. Unlike Cantonese, in Mandarin there are certain grammatical morphemes, including some suffixes, particles and kinship reference terms, which lack tone at both underlying and surface levels (Cao 1992). For example, the nominalizer 子 /zi/ surfaces with a different tone depending on what it is affixed to: compare 钉子 /dīng zi/ [ding^{55} zi^{2}] 'nail' vs. 脑子 /nǎo zi/ [nao^{21} zi^{4}] 'brain'. Whether or not lyricists take the surface tone of neutral-toned syllables into account is an open question. Chan (1987b) explicitly disregarded neutral toned syllables, but did not compare her results with an

analysis that included them, so it is not yet clear to what extent they do or do not impact tone-melody correspondence.

3. *Tone sandhi*. In Mandarin, there is a well-known phonological alternation in which (low-falling) T3 changes to (high-rising) T2 when followed by another T3. Impressionistically, this phenomenon is not uncommon, although at the time of writing I have been unable to find any source which provides frequency counts of tonal bigrams in running text. An obvious example is the greeting 你好 /ní hǎo/ [ni^{35} hao^{214}], pronounced with a rising tone on the first syllable: in isolation, both syllables (你 nǐ and 好 hǎo) would be produced with the low or low-rising tone. As with the neutral tone, it is unclear whether bigrams consisting of two underlying T3s are treated with respect to tone-melody correspondence (i.e. as level or falling tonal transitions).

In addition to exploring the role of these features, the present study also attempts to control for differences in melodic transitions by holding the melodic component constant. One problem when attempting to compare rates of tone-melody correspondence between any two languages is that not only do the tone systems differ, but the songs do as well. This means that it is unclear whether the differences observed have to do with the languages themselves, or with (incidental or systematic) properties of the melodies of the song corpora. To attempt to mitigate the potentially confounding effects of changing both melody and text, the present study focuses on a corpus of 8 song pairs for which versions exist with both Mandarin and Cantonese lyrics. The practice of releasing bilingual versions of popular songs is one that enjoys a long tradition in the Chinese-speaking popular music world, due perhaps in part to the fact that song lyric writing is generally considered as a separate skill from musical song composition (Mitchell 2006). This tradition has produced a rich body of song texts in both languages which are set to the same musical melodies, providing a natural laboratory in which to isolate the influence of linguistic, as opposed to musical, factors on rates of tone-melody correspondence.

Alongside the three factors mentioned above, the present work considers the role of interval step size, as some authors (e.g. Ho 2010; McPherson & Ryan 2017) have suggested that contrary mappings are avoided more stringently for larger musical steps than smaller ones. Holding the melodic corpus constant makes it possible to see if the same differences in step size have different effects when setting texts in different languages, i.e. if the step size constraint is more highly ranked by Cantonese or Mandarin lyricists.

2 Materials and methods

2.1 Corpus

The corpus was chosen opportunistically for a master's thesis (Lin 2018). As seen in the Appendix, the Cantonese lyricists were primarily from Hong Kong; the Mandarin lyricists were primarily Mandarin native speakers, with a few Hong Kong-based bilinguals. Many of the Cantonese songs feature lyrics written by Lam Chik (Albert Leung), although there is greater diversity among the composers. The songs are typically written with particular performers in mind, so it is potentially also relevant that Eason Chan appears three times and Faye Wong twice. The Can-

tonese versions of the songs in this corpus were typically released first, although in many cases the same artist recorded both versions. The exception is song 8《痴心绝对》'Absolutely Infatuated', which was a Mandarin-language hit by 李圣杰 Sam Lee in 2002 and only later appeared as a Cantonese-language cover version《残酷游戏》'Cruel Game' by 卫兰 Wai Lan (Janice Vidal) in 2009. Figure 4 shows the first two phrases of the song 'Red Rose' (in Mandarin, examples 1-2) or 'White Rose' (in Cantonese, examples 3-4), popularized by Eason Chan, who sings both of the versions analyzed here. Glosses and translations for the first phrase of each language are given in (1-4).

(1) 红 是 朱砂 痣 烙印 心口
hóng shì zhū shā zhì lào yìn xīn kǒu
red COP scarlet birthmark brand.print heart.mouth
'Red, the scarlet beauty mark over your heart[1] …'

(2) 红 是 蚊子 血 般 平庸
hóng shì wén zi xuè bān píng yōng
red COP mosquito blood kind common
'Red is as common as mosquito's blood …'

(3) 白 如 白忙 莫名 被 摧毁
baak^{2} jyu^{21} baak^{2} mong^{21} mok^{2} ming^{21} bei^{22} ceoi^{55} wai^{35}
white as IDIOM inexplicably PASS destroy
'White like working in vain[2] as I'm inexplicably cut down …'

(4) 得 到 的 竟 已 非 那 位
dak^{5} dou^{33} dik^{5} ging^{35} ji^{23} fei^{55} naa^{23} wai^{25}
get arrive POSS actually already not DEM PASS
'Who I catch is not actually "the one" ...'

2.2 Transcription and annotation

For each text/melody pair, song melodies were transcribed as fixed do solfège values (e.g. A, B, C...) along with accidentals and octave range, which were then automatically converted to MIDI note numbers. The corpus was then enriched with word and phrase boundary annotations, to enable comparison of the counts of similar, contrary, and oblique settings depending on whether or not transitions which crossed a boundary were ignored. Annotation of word boundaries was fairly conservative; anything that occurred as a dictionary headword was marked as a compound, so this encompassed examples such as 领悟 *lǐngwù* 'understand' (understand+comprehend), 伤心 *shāngxīn* 'sad, grieved' (hurt+heart), 蝴蝶 *húdié* 'butterfly' (non-decomposible), or 号码牌 *hàomǎpái* 'number plate' (number+code+plate).

[1]烙印 can also be 'indelible mark, lasting mark, stigma'; 心口, literally 'heart and mouth' also has a figurative reading 'words and thoughts'.

[2]While awkward in translation, this idiom works because the other verses are 'white as white teeth', 'white sugar' (the Cantonese pronunciation of 'red' is /hung^{21}/) and because "white" has a connotation like "wiped clean" (Alan Yu, p.c.).

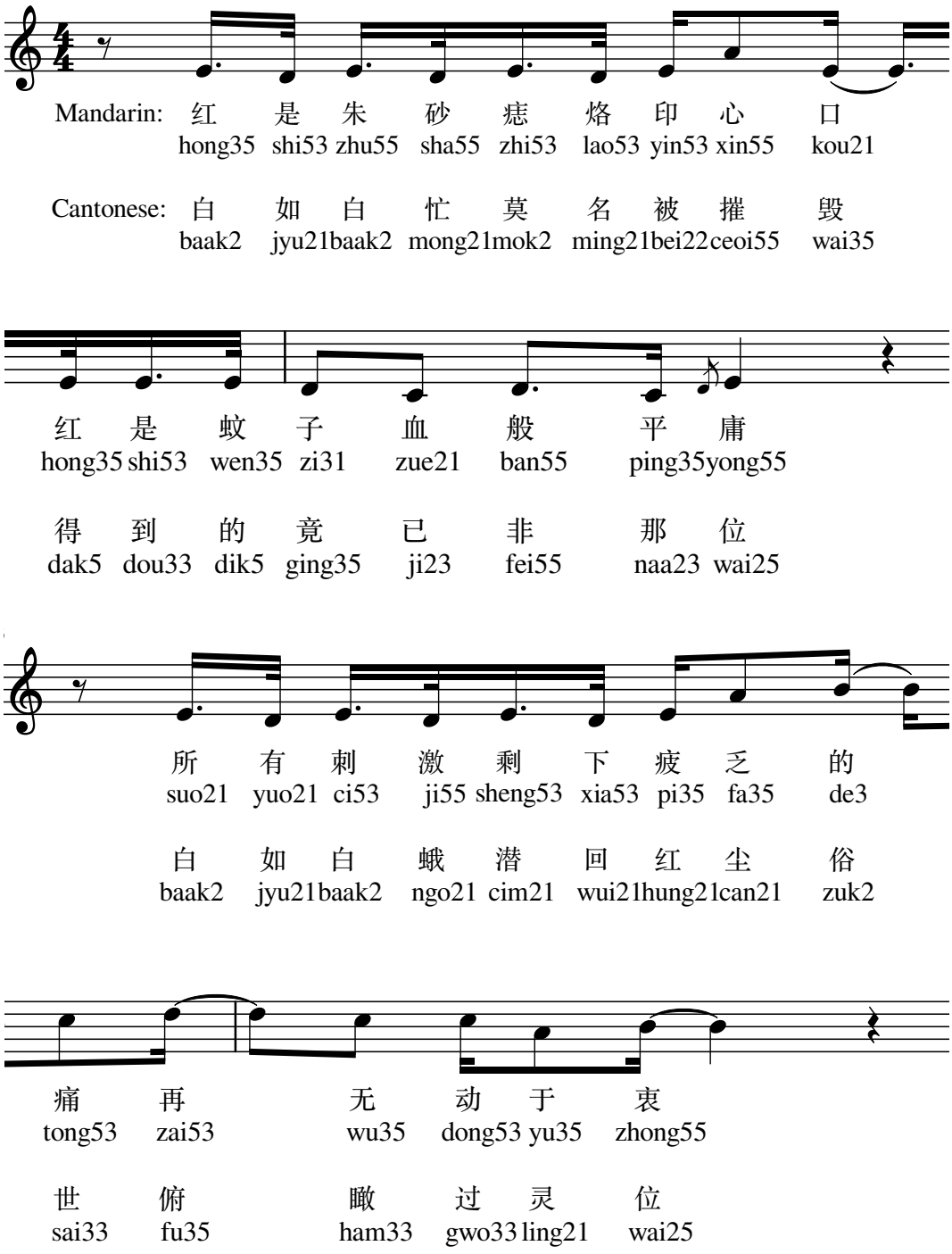

Figure 4: First two phrases of 白玫瑰/红玫瑰 performed by 陳奕迅 Eason Chan. A musical phrase boundary occurs at the ends of the 2nd and 4th lines.

Next, musical phrase boundaries were added by musical proficient annotators. Again, the aim was to be maximally conservative, so in cases of annotator uncertainty or disagreement, a boundary was annotated. This made it possible to compute transitions over the entire corpus, or just within musical phrases, since previous work makes clear that constraints on tonal text-setting are often ignored or relaxed at phrase edges.

Finally, each corresponding syllable of the lyric was first annotated with its dictionary tone value. In Cantonese, 变音 *pinyām* tone modifications (Bauer & Benedict 1997; Alderete *et al.* 2022), which are not easily predicted by general rule, were hand-annotated by native speakers; for Mandarin, tone changes were applied by rule. T3 was treated as low-falling /21/ unless it was phrase-final (which occurred very rarely, less than 10% of all occurrences of T3 and only 2% of all total transitions) or if it occurred before another T3, in which case it was treated as high-rising /35/ (81 instances representing about 20% of all lexical T3s, but less than 5% of the total corpus). This modification was applied both within and across word boundaries within utterances (Shih 1997). Neutral tones (of which there 96, or again around 5% of the total corpus) were assigned surface values based on those of the preceding syllable (Cao 1992), and these were hand-checked by native speaker annotators.

All corpus materials and data analysis scripts are available at this paper's accompanying OSF archive (https://osf.io/k6uma).

2.3 Analysis

One methodological issue that arises in the analysis of tone-melody correspondence when contour tones are involved is that it is not immediately obvious whether a given tonal bigram should be encoded as rising, level, or falling (Ladd & Kirby 2020:679-680). For example, the tonal sequence /35-53/ (as in 红 是 *hóng shì*, the first two lyrics of the Mandarin text given in Figure 4) might be classified as level (the transition from offset 5 to onset 5), rising (the transition from onset 3 to onset 5) or falling (the transition from offset 5 to offset 3).

In the present study, transitions were computed in two ways: first, in the manner of Chan (1987a) and Wong & Diehl (2002), i.e. by considering the offsets of any two syllables in a musical bigram; and second, by considering the offset of the first syllable and the onset of the second. Using this latter method, a tonal sequence like /35-53/ would count as a level tonal transition, meaning the setting would be marked as oblique. As offset-offset mappings always resulted in higher rates of similar (or at leat non-opposing) settings in both languages, the remainder of the paper considers the first method only.

3 Results

3.1 Tone-melody correspondence

In Cantonese, we find the expected high occurrence of similar settings: out of around 1700 transitions, nearly 80% were similar (Table 2); contrary settings comprised just 5% of the total. The remaining transitions are oblique, although as seen in other studies of Cantopop, sequences of identical tones far outnumber sequences

of identical notes: that is, there are many more oblique level *tone*-sequences than oblique level *note*-sequences.

In Mandarin, there are slightly fewer total transitions, but similar and oblique settings each make up around 38% of the total, with the remaining 25% contrary settings (Table 3).

	Note seq.		
Tone seq.	up	down	level
up	**530**	*39*	30
down	*45*	**486**	26
level	85	165	**314**

Table 2: Frequencies of similar (bold), contrary (italic), and oblique (plain) settings in the Cantonese song corpus. Similar settings: 77% (1330); contrary settings: 5% (84); oblique settings: 18% (306) out of 1720 total transitions.

	Note seq.		
Tone seq.	up	down	level
up	**244**	*218*	126
down	*204*	**272**	121
level	173	179	**107**

Table 3: Frequencies of similar (bold), contrary (italic), and oblique (plain) settings in the Mandarin song corpus. Similar settings: 38% (623); contrary settings: 25% (416); oblique settings: 37% (605) out of 1644 total transitions.

	Similar	Contrary	Oblique 1	Oblique 2
Cantonese	77%	5%	4%	15%
no internal trans.	77%	5%	4%	15%
Mandarin	38%	25%	15%	22%
no internal trans.	36%	28%	16%	20%
no neutral tones	38%	24%	15%	23%

Table 4: Comparison of similar, contrary, and oblique settings in the two song corpora. Oblique 1 is level melody; Oblique 2 is level tone (see Figure 3).

Table 4 compares the rates of similar, contrary, and oblique settings in the two corpora, as well as how the percentages change when disregarding word-internal transitions and, in Mandarin, neutral tones. In effect, nothing changes; indeed, in Mandarin the number of similar settings *decreases* slightly when word-internal transitions are disregarded. So from this we may conclude that contrary and oblique settings truly seem to be more common in Mandarin than in Cantonese, and that this cannot be trivially accounted for by the existence of a primarily disyllabic lexicon or neutral tones.

3.2 Interval size effects

One suggestion that has been made by a number of other researchers (e.g. Ho 2006; McPherson & Ryan 2017) is that oblique or even contrary settings are more likely when the difference between melody steps is small. While step size was not systematically manipulated in the present study, we can nonetheless make some preliminary observations.

Figure 5 shows a simple histogram of the distribution of melodic interval step sizes (which of course are the same for the two corpora). Generally speaking, the empirical preference appears to be to go up or down by single whole steps. Figures 6 and 7 give the distribution of interval sizes by direction of tonal transitions. In Cantonese, we can see that there is some indication of a slightly stronger preference to avoid contrary settings with larger melodic steps—as the interval step size increases, the number of contrary settings decrease (i.e., the distributions for the falling and rising tonal transitions are clearly skewed). In Mandarin, however, this tendency does not appear to be as pronounced. The tendency for rising tone bigrams to be set to rising note sequences and falling tone bigrams to be set to falling note sequences is still discernible, but the differences between 1-2 up/down and larger step sizes is closer the general distribution of interval sizes seen in Figure 5. So while melodic interval step size may well be an active constraint in Cantonese, it does not appear that it will help to explain the large discrepancy between the setting percentages in Cantonese and those in Mandarin, either.

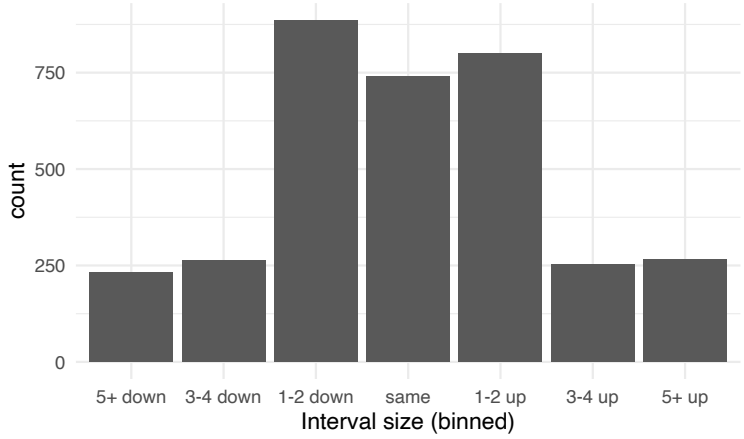

Figure 5: Distribution of musical interval step sizes between melodic bigrams.

4 Discussion

4.1 Frequency effects and the problem of contours

The foregoing investigation indicates that, despite the fact that the Mandarin lexicon contains a rather larger percentage of polysyllabic words than does the Cantonese lexicon, rates of tone-melody correspondence do not change appreciably regardless of whether or not word-internal transitions and/or neutral tones are disregarded. In this section, I will suggest a different possibility for the divergence between languages, namely that spoken Mandarin contains a rather larger percentage of contour tones and that these are inherently more difficult to set.

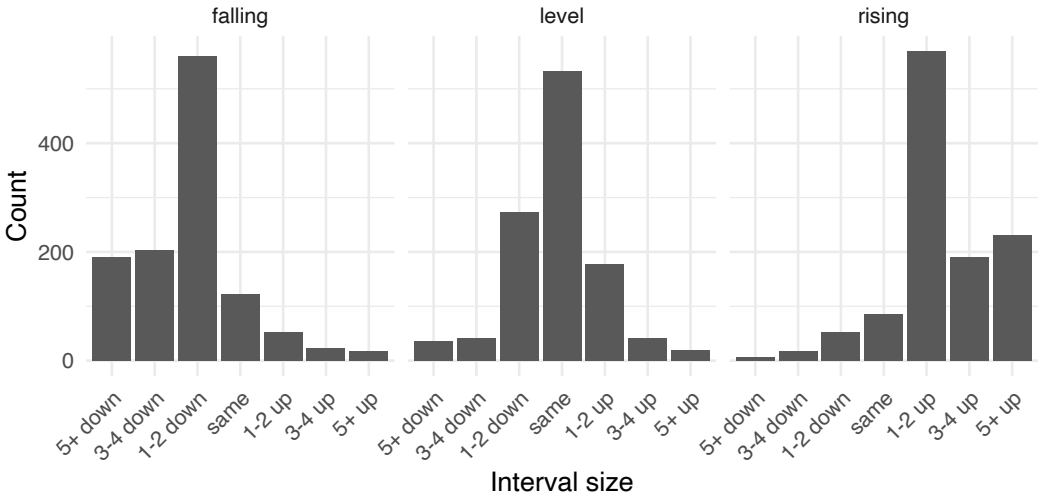

Figure 6: Distribution of musical interval step sizes between melodic bigrams by tonal bigram type, Cantonese texts.

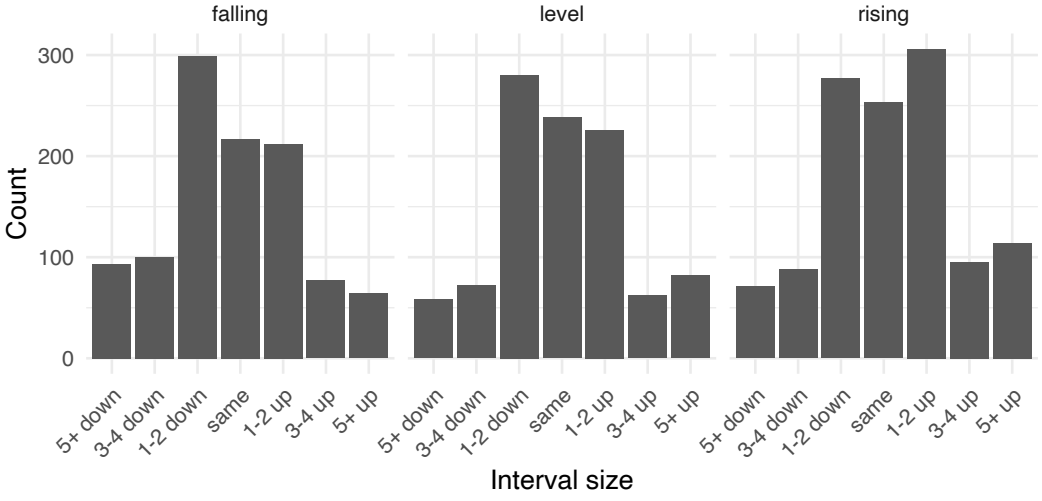

Figure 7: Distribution of musical interval step sizes between melodic bigrams by tonal bigram type, Mandarin texts.

The intuition which underlies this suggestion begins with the observation (originally due to Bob Ladd: see Ladd in press) that there appear to be far fewer Cantonese syllables with high-rising tones in Cantopop songs than one would expect, given their frequency in running text. To explore this a bit more quantitatively, Table 5 gives the (unigram) tone frequencies for the six Cantonese tones computed from the Hong Kong Cantonese Adult Language Corpus (Leung & Law 2001; Leung *et al.* 2004), compared to the frequencies of the same tones in the current song corpus. While most of the tones have similar frequencies in both the song and text corpora, the high-rising tone (tone 2) does indeed occur rather less often

Tone	Description	Value	Example			Corpus	Song
1	High level	55	/si1/	詩	‘poem’	21%	22%
2	High rising	35	/si2/	史	‘history’	17%	10%
3	Mid level	33	/si3/	嗜	‘fond of’	19%	17%
4	Mid-low falling	21	/si4/	時	‘time’	11%	15%
5	Mid-low rising	23	/si5/	市	‘market, city’	11%	13%
6	Mid-low level	22	/si6/	視	‘to view’	21%	23%

Table 5: Frequency distributions in text vs. song: Cantonese. Corpus frequencies from Leung and Law (2004).

Tone	Description	Value	Example			Corpus	Song
1	High level	55	/fū/	孵	‘to hatch’	25%	16%
2	High rising	35	/fú/	福	‘good fortune’	24%	28%
3	Low	21(4)	/fǔ/	腐	‘to ferment’	16%	20%
4	High fall	53(51)	/fù/	父	‘father’	35%	37%

Table 6: Frequency distributions in text vs. song: Mandarin. Corpus frequencies from Wu et al. (2020).

in the song corpus. This suggests the possibility that text-setters may be trying to avoid this tone when possible, perhaps because it is the most dynamic tone in the Cantonese repertoire.[3]

What is interesting if we compare with Mandarin (Table 6) is not so much the extent to which the tones occur in the song corpus relative to running text (frequencies from the corpus study of Wu *et al.* 2020), but instead the frequency distribution of the tones in the speech corpus. As shown by Wu *et al.* (2020), high-falling tone 4 is the most common tone in Mandarin, followed by tones 1 and 2. But together, tones 2 and 4—the high-rising and high-falling contour tones, whose dynamic changes in F0 are the critical cues for their perception (Jongman *et al.* 2012)—make up nearly 60% of the lexicon. Avoiding contour tones in Mandarin would make much of the lexicon off-limits, and presumably make for rather awkward and unnatural sounding lyrics. In Cantonese, however, the nature of the tone system makes it much easier to avoid potentially problematic contours, since there is really just one potentially troublesome tone.

This is not to imply that Mandarin artistic traditions ignore tone *tout court*; this is clearly not the case, as shown by e.g. Chao (1956) and Stock (1999). Instead, it seems that the fact that the Cantonese tone system consists of (a) more and (b) primarily less dynamic tones gives it an advantage at this particular task.

4.2 Observed vs. expected ratios of Mandarin bigram sequences

If we consider the observed/expected (O/E) ratios for the tonal bigrams set to rising, falling, and level melodic transitions (Tables 7a-7d), we see something interesting. Recall that the O/E ratio is a simple way of calibrating our expectations

[3] Anecdotally, a number of Vietnamese songwriters have told me in informal discussions that they consciously try to avoid the high-rising “broken” tone (*ngã*) because they feel it is especially difficult to set.

(a)

	Rise	Fall	Level
T1-T1	0.58	1.07	1.62
T1-T2	0.80	1.55	0.30
T1-T3	0.49	1.41	1,12
T1-T4	0.80	1.21	0.95

(b)

	Rise	Fall	Level
T2-T1	1.40	0.51	1.13
T2-T2	0.88	0.88	1.43
T2-T3	0.82	1.35	0.63
T2-T4	0.89	1.04	1.13

(c)

	Rise	Fall	Level
T3-T1	1.41	0.66	0.93
T3-T2	1.39	0.76	0.78
T3-T3	1.21	0.86	0.90
T3-T4	1.43	0.61	0.99

(d)

	Rise	Fall	Level
T4-T1	0.95	1.00	1.08
T4-T2	0.84	1.13	1.03
T4-T3	0.84	1.11	1.07
T4-T4	1.02	1.05	0.87

Table 7: O/E ratios for Mandarin melodic bigrams where corresponding texts are headed by (a) T1 /55/, (b) T2 /35/, (c) T3 /21/, (d) T4 /53/.

about co-occurrences—-here, how often we should expect to see a particular tone sequence would be set to a particular melodic transition. When O/E is about 1, then a particular tonal bigram is set to a particular melodic transition about as often as one would expect if bigrams are independent of melodic transitions; values greater than or less than 1 thus indicate over- or underrepresentation respectively, under this assumption. O/E ratios for the Mandarin data were computed by tallying the number of tone bigrams (T1-T1, T1-T2...) that occurred with each type melodic transition (level, rising, or falling) and using these counts to generate the number of times that bigram would be expected to occur with a given transition. For example, there are 83 instances of T1-T4 sequences in the Mandarin corpus out of a total of 1447 total bigrams (ignoring transitions involving neutral tones or spanning phrase boundaries), of which 547 are rising note transitions; thus T1-T4 is expected to occur with a rising note transition $(83 \times 547)/1447 \approx 31$ times. Since this bigram is only set to rising transitions 25 times in the corpus, the O/E ratio is slightly less than 1 (0.8), meaning that this bigram occurs somewhat less often than we might expect with rising transitions, were the distributions of tonal bigrams and melodic sequences independent.

First, consider the melodic bigrams headed by the high-level tone T1 (Table 7a). What we observe is that bigrams headed by T1 occur rather less than would be expected with rising note sequences, and rather more than would be expected with falling note sequences, some more than others. In other words, if the melodic transition is rising, tonal bigrams headed by T1 are disprefered, presumably because these would always involve some kind of fall on the tonal tier. Conversely, when the melodic transition is falling, sequences headed by T1 are overrepresented, the exception being oblique T1-T1 sequences. These same sequences are overrepresented on level musical transitions (O/E = 1.62). If we then consider the melodic bigrams headed by T3 (Table 7c)—that's the low or low-falling tone, which only rises phrase-finally—we see the opposite pattern as we observed for bigrams headed by T1: namely, rises are overrepresented, while falls are underrepresented. (Note that T3-T3 sequences here are based on the underlying/lexical tone values of the

syllables, not their surface sandhi values.)

For comparison, we can also look at the bigrams headed by the dynamic contour tones, rising T2 and falling T4 (Tables 7b and 7d). Here, the pattern, if there is any, is far less obvious. It appears that T2 may be treated as a low tone when followed by T1 (overrepresentation of T2-T1 sequences on rising melodies) but high when followed by T3 (overrepresentation of T2-T3 sequences on falling melodies). However, bigrams headed by falling T4 are set to falling melodic transitions no more than would be expected if bigrams and melodic transitions were independent, nor do they appear to be especially disprefered with rising transitions.

These differences between T1/T3 on the one hand and T2/T4 on the other suggest that, when Mandarin lyricists can avoid contrary motion and prefer similar motion, they do; it's just that the facts about the tone distribution in this language make it difficult for them to do so very often, leading to correspondingly lower rates of tone-tune correspondence.

4.3 General constraints on tone-tune correspondence

Recall that in Section 1.2, the fundamental tonal text-setting constraint was formulated as something like "avoid contrary settings" (*NONPARALLEL[4] in the parlance of McPherson & Ryan 2017). At least in Cantonese, this may also interact with a dispreference for large musical step size transitions, what McPherson & Ryan (2017) characterize as *STEP. To this, we might add a third constraint penalizing the use of dynamic contour tones in texts, something like "avoid contour tones" (*CONTOUR).

As we are still at what we might call the constraint-discovery stage in the study of tone-melody correspondences, a formal treatment of how these constraints interact would be premature. Nevertheless, the Mandarin data indicate that a general dispreference for setting dynamic contour tones may well be important for the broader understanding of tonal text-setting and should be investigated in more detail in other languages, such as Thai, where the general rates of similar motion have been found to be low (List 1961; Saurman 1999; Ketkaew & Pittayaporn 2014; Kirby 2021). Moreover, the observed/expected analysis of Mandarin suggests that in at least some languages, even an optimal ranking of the relevant constraints may result in a sizeable number of contrary and oblique settings.

5 Conclusion and perspectives

This paper has considered the extent to which three characteristic properties of Mandarin Chinese—a primarily polysyllabic lexicon, the existence of neutral tones, and the frequent application of tone sandhi—might account for its seemingly lax adherence to tone-melody correspondence constraints. Even when discounting transitions between syllables with neutral tones, as well as those which spanned word and musical phrase boundaries, the relative rates of tone-melody correspondence remained constant. This was also true of texts set in Cantonese to the exact

[4]Implicitly or explicitly, terms such as "nonparallel" or "non-opposing" suggest that the notion of "oblique" may be reasonably treated as a kind of neutral, which as mentioned briefly in Section 1.2, seems unlikely to be empirically valid. For more discussion on this point, see Ladd (in press).

same melodies, underscoring that the fact that the movements of tones and tunes are more closely aligned in Cantonese than they are in Mandarin is not due to differences in melody.

Instead, I have suggested that the explanation may actually be primarily an accident of the languages' tone systems, the relevance of which was also remarked on by Chan (1987b). However, while Chan drew a parallel between the effects of intonation and the "dominance" of melody over word tone in Mandarin, the current study has highlighted how the canonical phonetic realizations of the tones themselves may also play an important role. Cantonese contains (a) more tones which (b) mostly occupy single registers, with just one tone, the mid-high rising tone, involving a considerable pitch excursion, and this tone occurs less in the song corpus than one would expect based on its frequency in the lexicon (or in running text). Mandarin, on the other hand, has a smaller tone inventory overall, but well over half of the most commonly occurring tones are contour tones. To the extent that texts involving contour tones will more frequently result in non-similar settings, Cantonese lyricists are at a distinct advantage, given that syllables bearing contour tones make up a rather smaller proportion of the lexicon. In Mandarin, on the other hand, it is much more difficult to compose a text of any sort without making extensive use of contour tones.

As emphasized by Ho (2010) (and more recently from an ethnomusicological perspective by Li 2021), low rates of tone-melody correspondence do not necessarily mean that non-corresponding sequences are judged as "ill-formed" or otherwise unacceptable to listeners. Other, potentially more subtle considerations are often at play. Nevertheless, text-setting constraints are clearly taken into account by lyricists and are salient to listeners, so it is worth studying them in detail. It is hoped that this paper has made a small contribution to exploring the space of possibilities, and may serve as an inspiration for more extensive future investigations.

6 Acknowledgments

This study revises and extends an analysis based on data originally gathered by Ruoqi Lin for her 2018 University of Edinburgh MSc thesis "A comparison of tonal text-setting in Mandarin and Cantonese popular songs". Thanks to Bob Ladd, as well as audiences at CLS58, the Chinese University of Hong Kong, and the 41st DGfS workgroup "Prosody from a cross-domain perspective: How language speaks to music (and vice versa)" for discussion, comments, and suggestions. Special thanks to Matthew Sung and Grace Wenling Cao for assistance with transliteration, data annotation and proofreading. The author remains solely responsible for any errors of fact or interpretation. This project was funded in part by AHRC grant AH/M005240/1 "Singing in tone: text-setting constraints in Southeast Asia".

References

Alderete, J., Q. Chan, & S.-i. Tanaka. 2022. The morphology of Cantonese "changed tone": Extensions and limitations. 言語研究（*Gengo Kenkyu)* 161. 139–169.

Bauer, R. S., & P. K. Benedict. 1997. *Modern Cantonese Phonology*. De Gruyter Mouton.

Cao, J. 1992. On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics* 20. 49–50.

Chan, M. K. M. 1987a. Tone and melody in Cantonese. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society (1987)*, p. 26–37.

Chan, M. K. M. 1987b. Tone and melody interaction in Cantonese and Mandarin songs. *UCLA Working Papers in Phonetics* 68. 132–169.

Chao, Y. R. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*, ed. by M. Halle, H. Lunt, H. McLean, & C. H. van Schooneveld, p. 52 – 59. The Hauge: Mouton.

Dell, F., & J. Halle. 2009. Comparing musical textsetting in French and in English songs. In *Towards a typology of poetic forms: from language to metrics and beyond*, ed. by J.-L. Aroui & A. Arleo, Language Faculty and Beyond (LFAB): Internal and External Variation in Linguistics, p. 63–78. Amsterdam: John Benjamins.

Halle, J., & F. Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55. 3–23.

Hayes, B. 2009. Textsetting as constraint conflict. In *Towards a typology of poetic forms: from language to metrics and beyond*, ed. by J.-L. Aroui & A. Arleo, p. 43–62. Amsterdam: John Benjamins.

Ho, W.-S. V. . 2006. The tone-melody interface of popular songs written in tone languages. In *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC 2006)*, ed. by M. Baroni, A. R. Addessi, R. Caterina, & M. Costa, p. 1414–1422.

Ho, W.-S. V. . 2010. *A phonological study of the tone-melody correspondence in Cantonese pop music*. Hong Kong: University of Hong Kong dissertation.

Jakobson, R. 1960. Closing statement: Linguistics and poetics. In *Style in language*, ed. by T. A. Sebok, p. 350–377. New York; London: The Technological Press; John Wiley & Sons.

Jongman, A., Y. Wang, & C. B. Moore. 2012. Perception and production of Mandarin Chinese tones. In *Handbook of East Asian psycholinguistics, Volume 1: Chinese*, ed. by J. A. Sereno, P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng, p. 209–217. Cambridge: Cambridge University Press.

Ketkaew, C., & P. Pittayaporn. 2014. Mapping between lexical tones and musical notes in Thai pop songs. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation (PACLIC 28)*, p. 160–169.

Kiparsky, P. 1977. The rhythmic structure of English verse. *Linguistic Inquiry* 8. 189–247.

Kirby, J. 2021. Towards a comparative history of tonal text-setting practices in Southeast Asia. In *Transcultural music history*, ed. by R. Strohm, 291–312. Berlin: Berliner Wissenschafts-Verlag.

Kirby, J., & D. R. Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. In *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL-2016)*, p. 48–51, Buffalo.

Ladd, D. R. In press. Two problems in theories of tone-melody matching. *Studies in Prosodic Grammar (*韵律语法研究*)* 8.

Ladd, D. R., & J. Kirby. 2020. Tone-melody matching in tone language singing. In *The Oxford Handbook of Language Prosody*, ed. by C. Gussenhoven & A. Chen, p. 676 – 687. Oxford: Oxford University Press.

Leung, M.-T., & S.-P. Law. 2001. HKCAC: the Hong Kong Cantonese adult language corpus. *International Journal of Corpus Linguistics* 6. 305–326.

Leung, M.-T., S.-P. Law, & S.-Y. Fung. 2004. Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments, & Computers* 36. 500–505.

Li, C., & S. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

Li, E. K. C. 2021. Cantopop and speech-melody complex. *Music Theory Online* 27.
Lin, R. 2018. A comparison of tonal text-setting in Mandarin and Cantonese popular songs. University of Edinburgh MSc thesis.
List, G. 1961. Speech melody and song melody in central Thailand. *Ethnomusicology* 5. 16–32.
Lo, T. C. 2013. Correspondences between lexical tone and music transitions in Cantonese pop songs: a quantitative and analytic approach. Hons dissertation, The University of Edinburgh.
McPherson, L., & K. M. Ryan. 2017. Tone-tune association in Tommo So (Dogon) folk songs. *Language* .
Mitchell, T. 2006. Tian ci - Faye Wong and English songs in the Cantopop and Mandopop repertoire. In *Access All Eras: Tribute Bands and Global Pop Culture*, ed. by S. Homan, p. 215–228. Open University Press.
Mok, P. P. K., D. Zuo, & P. W. Y. Wong. 2013. Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change* 25. 341–370.
Saurman, M. E. 1999. The agreement of Thai speech tones and melodic pitches. *Notes on Anthropology* 3. 15–24.
Schellenberg, M. 2011. Tone contour realization in sung Cantonese. In *Proceedings of the 17th International Congress of the Phonetic Sciences*, p. 1754–1757.
Schellenberg, M., & B. Gick. 2020. Microtonal variation in sung Cantonese. *Phonetica* 77. 83–106.
Shih, C. 1997. Mandarin third tone sandhi and prosodic structure. In *Studies in Chinese Phonology*, ed. by J. Wang & N. Smith, 81–123. Berlin, Boston: De Gruyter.
Stock, J. P. J. 1999. A reassessment of the relationship between text, speech tone, melody, and aria structure in Beijing Opera. *Journal of Musicological Research* 18. 183–206.
Tanese-Ito, Y. 1988. The relationship between speech-tones and vocal melody in Thai court song. *Musica Asiatica* 5. 109–39.
Wee, L. H. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics* 35. 128–144.
Wong, P. C. M., & R. L. Diehl. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30. 202–209.
Wu, Y., M. Adda-Decker, & L. Lamel. 2020. Mandarin lexical tones: A corpus-based study of word length, syllable position and prosodic position on duration. In *Interspeech 2020*, p. 1908–1912.
Yung, B. 1983. Creative process in Cantonese opera I: the role of linguistic tones. *Ethnomusicology* 27. 29–47.

Appendix: Song data

Annotated song data and R scripts for analysis are available at https://osf.io/k6uma.

Cantonese

Pair	Title	Year	Performance	Lyrics	Composition	Arrangement
1	白玫瑰	2006	陈奕迅	李焯雄	梁翘柏	梁翘柏
2	回旋木马的终端	2003	梁咏琪	林夕	林一峰	张人杰
4	给自己的情书	2000	王菲	林夕	江志仁	江志仁
5	下一站天后	2003	Twins	黄伟文	伍乐城	伍乐城
6	明年今日	2002	陈奕迅	林夕	陈小霞	陈辉阳
7	K 歌之王	2000	陈奕迅	林夕	陈辉阳	陈辉阳
8	残酷游戏	2009	卫兰	林夕	蔡伯南	雷颂德
9	暗涌	1997	王菲	林夕	陈辉阳	陈辉阳

Mandarin

Pair	Title	Year	Performance	Lyrics	Composition	Arrangement
1	红玫瑰	2007	陈奕迅	李焯雄	梁翘柏	梁翘柏
2	遇见	2003	孙燕姿	易家扬	林一峰	Terrence Teo
4	笑忘书	2001	王菲	林夕	江志仁	江志仁
5	莫斯科没有眼泪	2005	Twins	许常德	伍乐城	Mac Chew
6	十年	2003	陈奕迅	林夕	陈小霞	陈辉阳
7	K 歌之王	2001	陈奕迅	林夕	陈辉阳	陈辉阳
8	痴心绝对	2002	李圣杰	蔡伯南	蔡伯南	陈飞午
9	暗涌	2013	杨丞琳	林夕	陈辉阳	陈辉阳