Two problems in theories of tone-melody matching

D. Robert Ladd
University of Edinburgh
January 2022

To appear in *Studies in Prosodic Grammar,* volume 9
Beijing and Hong Kong, 2023

Abstract

Recent work on how lexical tones are matched with musical melodies has shown that a satisfactory match generally requires the *pitch direction* between musical notes to be the same as the pitch direction between the corresponding tones. This principle applies in a number of unrelated tone languages in both Asia and Africa. The interaction of this pitch direction principle with other principles is often approached (at least implicitly) in terms of ranked violable constraints, and at least one study has made explicit use of the descriptive framework of Optimality Theory. Nevertheless, some empirical and theoretical issues remain poorly understood. This paper considers two such problems on the basis of recent studies of modern Cantonese popular music ('Cantopop'). First, it discusses so-called 'oblique' matches, where either the tone sequence is unchanging while the musical sequence rises or falls, or the musical pitch stays level across a sequence of different tones. These have often been treated as near-matches, but investigation shows this to be empirically incorrect, with unclear implications for the idea of violable constraints. Second, the paper discusses the treatment of contour tones in determining a good match: contour tones are handled differently in different languages, and there is evidence that they may be intrinsically difficult to accommodate in a well-matched melody.

Keywords: Cantopop, text-setting, tone, contour tones

1. Introduction

The past few decades have seen considerable research on the principles that govern the correspondence between text and melody in songs in tone languages. It is now generally recognised that the tonal sequences in a song text and the note sequences in the accompanying musical melody should *match* in some way. The most important principle of a good match between the tones of a song text and the melody appears to involve the *direction of the pitch movement* from one note to the next: ideally, the musical pitch movement should be in the same direction as the direction of the natural tonal movement between the syllables sung on those notes. More particularly, rising melodic sequences should not accompany syllables in which the tonal sequence is falling (e.g. High tone followed by Mid tone), and rising tonal sequences should not be set to falling musical melodies. Like the artistic constraints involved in meter and rhyme, the prohibition on mismatched pitch direction is not absolute, but the empirical evidence that such a principle exists now comes from many unrelated tone languages, and the basic idea of avoiding such mismatches is now shared across most recent work on tone-melody matching.

One of the earliest studies in this current line of research on tone-melody matching is an exploratory paper by Chan (1987a), which dealt with Cantonese popular songs and film music (see also Chan 1987b). Chan observed that in songs with multiple verses set to the same melody, the *sequence of syllable tones* on any given melodic line was nearly identical in each verse, even though the words were completely different. That is, it appears that a musical melody somehow dictates which Cantonese words can be used at specific places in the melody, just as a rhyme scheme dictates which words are available for use at the end of a line of rhyming verse. Chan proposed the basic pitch-direction rule summarised above and discussed various specific problems of detail. The relevance of pitch direction was confirmed experimentally for Cantonese by Wong and Diehl (2002), and has been demonstrated for other languages by Schellenberg (2009, 2012) and others. McPherson and Ryan (2018) present a formal analysis of the pitch-direction principle expressed in terms of the interacting violable constraints of Optimality Theory (e.g. both a preference for matching pitch direction and a prohibition against non-matching pitch direction); their work was based on folk songs in Tommo So (a Dogon language spoken in Mali). A summary review of the past few decades' work is given by Ladd and Kirby (2020).

By far the best-studied language in this line of research is Cantonese, in part because of its links to an active tradition of westernised pop music, so-called 'Cantopop' (粤语流行). This was the subject of a Hong Kong PhD dissertation by Ho (何詠詩) 2010, which firmly established the systematic nature of Cantopop's tone-melody matching and investigated many of the details. Two more recent Master's dissertation projects at the University of Edinburgh, one (Lo 2013) supervised by me and one (Lin 2018) supervised by my colleague James Kirby, have confirmed and built on Ho's conclusions. My aim in this paper is to use the fairly well understood facts about tone-melody matching in Cantopop to shed light on two problems with current work on tone-melody matching in general.

2. Basic empirical findings on tone-melody matching in Cantopop

2.1. Explanation of terminology and methods

Empirical evaluation of the correspondence between tone and melody in recent work – not just work on Cantopop but all the research briefly summarised in the introduction – has generally been based on the description of **bigrams**: sequential pairs of musical notes and the accompanying syllables. This is illustrated in Fig. 1, a line taken from the 2003 song 'Next station Tin Hau', by Twins:

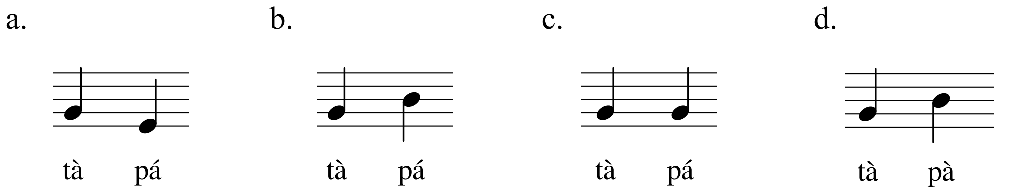
Figure 1: Bigrams in a line from a Cantopop song (adapted from Lin 2018)

The bigrams are labelled with capital letters A-K. For every bigram, the pitch direction of both the musical sequence and the tonal sequence can be described as rising, falling, or unchanging. The definition of these three possibilities, when applied to the musical sequence, is straightforward: for example, bigrams C and F are rising, bigrams D and G are falling, and bigrams A and H are unchanging. For the tonal sequence, in bigrams consisting of level tones, the pitch direction is also easy to define: for example, bigram F (consisting of a mid-level tone followed by a high-level tone) is rising, while bigram G is falling and bigrams A, B, and H are unchanging. For bigrams containing contour tones – like bigram D, with a high-rising tone followed by a low-falling tone, or bigram I, with a mid-level tone followed by a low-rising tone – the situation is somewhat more complex. We return to this question shortly, and the place of contour tones in tone-melody matching more generally is one of the two problems we consider in the second half of the paper (sec. 3.2).

Once we have defined both the musical and the tonal pitch direction in a given bigram, we can define three types of correspondence between the tones and the melody. Following Ladd and Kirby, we refer to these three as **similar, contrary**, and **oblique.** In a 'similar' setting[1] both the musical sequence and the tonal sequence are identical – both rising, for example. In a 'contrary' setting, the two sequences go in the opposite direction – one rising and the other falling.

[1] The term *setting* here is based on the notion of 'text-setting', the process of taking a given text (the song lyrics) and fitting them to a melody. This notion is not necessarily very accurate for describing tone-melody matching in many folk traditions (e.g. Lissoir 2016 on Tai Dam, Karlsson 2018 on Kammu), but it applies fairly well to composed music like Cantopop songs.

'Oblique' settings cover the remaining possibilities: either the musical sequence is unchanging and the tonal sequence rises or falls (oblique I), or the tonal sequence is unchanging and the accompanying melody rises or falls (oblique II). The terms are illustrated graphically in Fig. 2.



contrary setting *similar setting* *oblique setting I* *oblique setting II*

Figure 2. Possible bigram types (from Ladd and Kirby 2020)

Ho, Schellenberg, and McPherson and Ryan all use the term **parallel** for Ladd and Kirby's 'similar'. Ho and Schellenberg treat oblique as 'non-opposing'; McPherson and Ryan adopt Ladd and Kirby's term 'oblique' but group it together with 'parallel' as 'non-contrary'. In both cases the terminology conveys the idea that oblique is somehow neutral or intermediate between similar and contrary, a less serious violation of the expectation that the tonal sequence and the melodic sequence should match. This conception of 'oblique' is the other of the two problems considered in the second half of the paper.

Given the categorisation of pitch direction into three types (rising, falling, unchanging) and given the categorisation of the tone-melody correspondence in any given bigram into three types (similar, contrary, oblique), quantitative data from a corpus of songs can be displayed in a 3x3 matrix, with the musical pitch direction on one axis and the tonal pitch direction on the other. The distribution of the bigram types in the nine cells of the 3x3 matrix is illustrated in Table 1.

Table 1. Matrix diagram summarising bigram types

		Musical pitch direction		
		Rising	Falling	Unchanging
Tonal pitch direction	Rising	**similar**	contrary	• oblique
	Falling	contrary	**similar**	• oblique
	Unchanging	• oblique	• oblique	**similar**

With actual data presented in such a matrix, we can evaluate empirical predictions based on various principles governing tone-melody matching. For example, if a prohibition on contrary settings were the only determinant of a good tone-melody match, we would predict that bigrams in the two 'contrary' cells of the matrix should be uncommon, but would have no real basis for predicting the proportion of bigrams in the other cells. On the other hand, if we assume that there is not only a constraint *prohibiting* contrary settings, but also a constraint *preferring* 'similar' settings, we would predict that bigrams in the cells along the diagonal containing similar settings should be more frequent than those in other cells, as shown in Table 2:

Table 2. Matrix diagram displaying frequency predictions about bigram types

		Musical pitch direction		
		Rising	Falling	Unchanging
Tonal pitch direction	Rising	**frequent**	rare	• (unclear)
	Falling	rare	**frequent**	• (unclear)
	Unchanging	• (unclear)	• (unclear)	**frequent**

We would still have no clear prediction about the 'oblique' cells of the matrix, but as we shall see in section 2.3, quantitative data from Cantopop seem to confirm that 'oblique' is not a coherent category and that the two sub-types – those where the musical note on the two syllables is the same and those where the two tones are identical – can be clearly distinguished.

2.2. Cantonese tone and the problem of contour tones

Cantonese is generally analysed (e.g. Matthews and Yip 2013) as having six tones: high level (55), high-rising (35), mid level (33), low-falling (21), low-rising (23), and low-level (22)[2]. The numbers in parentheses give the conventional phonetic description of the tones in terms of the five-level system introduced by Y. R. Chao (赵元任) (1930). The six tones are illustrated in Figure 3, which displays: (1) the traditional Western identifying numbers for each tone (T1 [tone 1], T2, etc.); (2) the traditional Chinese designations; and (3) the conventionally assumed 'Chao numbers' indicating the tone's beginning and ending pitch levels. Throughout the rest of the paper we will use all three methods of referring to specific tones.

With this background, let us return to the issue of how to define the pitch direction across bigrams involving contour tones (in Cantonese, T2, T5, and T4). Consider for example a bigram in which both syllables bear T4 (阳平, 21). Taking the Chao pitch levels as a rough phonetic description of the pitch direction on each tone, we can describe the tone sequence across the bigram is something like the following:

```
2 1   2 1
 \/    \/
 σ     σ
```

We can imagine a variety of principles that might define a good musical match to this tone sequence. For example, perhaps a good match would prescribe rising

[2] There are also three 'checked tones' (入声), High, Mid, and Low, that occur only in syllables closed by a stop, but these are readily identified with the three level tones of the main six-tone set (i.e. T1 (阴平), T3 (阴去), and T6 (阳去)). Note also that in present-day Hong Kong Cantonese there appears to be a merger in progress between T5 (阳上, 23) and T2 (阴上, 35) (e.g. Bauer et al. 2003).

musical pitch across the bigram, corresponding to the rise from level 1 to level 2 across the syllable boundary. Or perhaps the musical pitch should be falling, to express the overall fall in pitch from level 2 at the beginning of the bigram to level 1 at the end. Or perhaps the musical pitch should remain unchanging, because the tones on each syllable are phonologically the same.

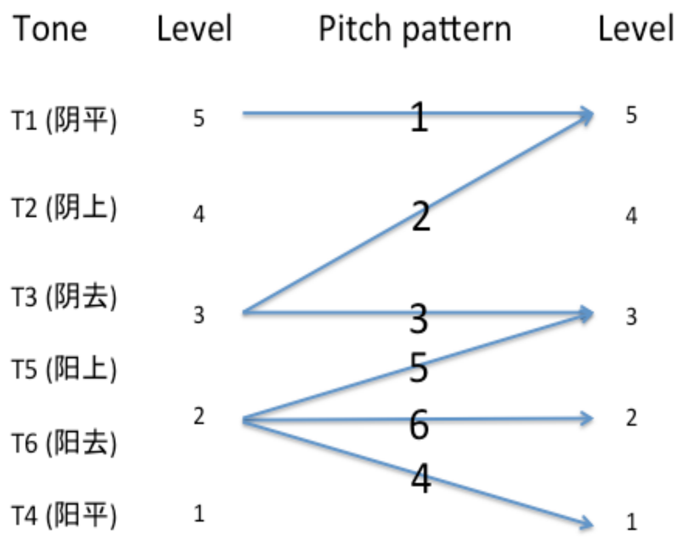


Figure 3. Cantonese tones

As it happens, Chan's original description of tone-melody matching addressed this problem and suggested that, in Cantonese, what counts in evaluating the pitch direction from one tone to the next is the **endpoint** – the final pitch level – of each tone. Roughly speaking, a falling tone that ends low counts as low, and a rising tone that ends high counts as high. This also seems to be true of some other Asian tone languages as well (e.g. Kirby and Ladd 2016 on Vietnamese), though in some languages (e.g. Thai; see Ketkaew and Pittayaporn 2014) the situation may be considerably more complex. Even within Chinese, recent work by Zhang and Cross (2021) on folksongs in Chaozhou (Teochew, 潮州) shows that for text-setting purposes tones must be grouped into three categories (High, Mid and Low), but that the endpoint principle fails to group them correctly. For Cantonese, however, Ho's thesis amply confirms that the end pitch (Ho sometimes uses the term *tonal target*) is the key to dealing with contour tones.

Concretely, the endpoint principle means that the tonal pitch direction in a tonal bigram consisting of T1 (阴平, 55) and tone T2 (阴上, 35) counts as 'unchanging', because both tones end at level 5; the same is true of a bigram consisting of T3 (阴去, 33) and T5 (阳上, 23), where both tones end at level 3. By the same logic, a bigram consisting of T1 (阴平, 55) followed by T3 (阴去, 33) and a bigram consisting of T2 (阴上, 35) followed by T5 (阳上, 23) both involve falling pitch direction, from level 5 (at the end of the first tone of the bigram) to level 3 (at the end of the second).

For purposes of their experiment Wong and Diehl settled on a three-way distinction between High, Mid and Low endpoints, grouping both T6 (22) and T4 (21) together as Low. Ho's subsequent work showed clearly that we get a more accurate picture of tone-melody matching in Cantopop if we assume that T6 (阳去, 22) counts as Low and T4 (阳平, 21) counts as Extra-Low. I return to this issue in section 3.2; to avoid the potentially confusing use of both Chao's pitch level numbers and Cantonese tone numbers, I will adopt the terms High, Mid, Low, and Extra-Low throughout the paper for referring to tone endpoint levels.

2.3. Quantitative studies of tone-melody matching in Cantopop

In this section I summarise the data from two relatively recent studies of Cantopop songs. Both are based on Master's dissertation projects at the University of Edinburgh, one (Lo 2013) supervised by me and one (Lin 2018) supervised by my colleague James Kirby. (Kirby has subsequently done further work on Lin's data and throughout the paper I refer to the 'Kirby/Lin corpus'.) Lo's corpus was based on 11 songs and a total of 2665 bigrams; the Kirby/Lin corpus was based on 8 songs and a total of 1720 bigrams. Lin's study focused on a comparison between Cantonese and Mandarin (普通话) versions of the same songs, but with respect to Cantonese very similar methodology was used in both studies. Songs were transcribed and the bigrams classified in the ways described in section 2.1; the classification of tonal bigrams treats T4 as Extra-Low, as just described. Both studies reveal a very similar picture that is in line with Ho's conclusions, but I focus on the two Edinburgh studies in order to illustrate my points with quantitative data. In this short section I treat the two datasets together, whereas the discussion in section 3 is based primarily on the Kirby/Lin corpus, as it is the only one for which I have access to a complete file of raw data.

Table 3 presents the numbers from both studies combined into a single dataset. The total number of bigrams in each cell is given in section (a), while section (b) presents the same numbers as percentages of the total of 4385 bigrams. Section (c) shows the proportions row-by-row (so that each row sums to 100%), while section (d) shows the proportions column-by-column (so that each column sums to 100%). This allows us to consider the treatment of the three types of tonal pitch direction (section (c)) and the three types of musical pitch direction (section (d)) separately.

Table 3. Distribution of bigram types in the combined Lo/Lin Cantopop data

(a)

		Musical pitch direction		
		Rising	Falling	Unchanging
Tonal pitch direction	Rising	**1410**	98	• 65
	Falling	154	**1279**	• 80
	Unchanging	·· 190	·· 375	**734**

(b)

		Musical pitch direction		
		Rising	Falling	Unchanging
Tonal pitch direction	Rising	***32.2%***	*2.2%*	• *1.5%*
	Falling	*3.5%*	***29.2%***	• *1.8%*
	Unchanging	·· *4.3%*	·· *8.6%*	***16.7%***

(c)

		Musical pitch direction			
		Rise	Fall	Unchg	
Tonal pitch direction	Rise	**89.6%**	6.2%	• 4.1%	100%
	Fall	10.2%	**84.5%**	• 5.3%	100%
	Unchg	·· 14.6%	·· 28.9%	**56.5%**	100%

(d)

		Musical pitch direction		
		Rise	Fall	Unchg
Tonal pitch direction	Rise	**80.4%**	5.6%	• 7.4%
	Fall	8.8%	**73.0%**	• 9.1%
	Unchg	·· 10.8%	·· 21.4%	**83.5%**
		100%	100%	100%

It can be seen immediately from Table 3(b) that 'contrary' settings (shown in the two shaded cells of the matrix) are very uncommon, while at the same time Table 3(d) shows that 'similar' settings (those in the three cells along the diagonal, with heavy black borders) constitute more than three-quarters of the whole corpus. This clearly suggests that tone-melody matching in Cantopop involves both a prohibition on contrary settings and a preference for similar settings, and that these constraints may interact in some way. In the context of the current research reviewed in the introduction, this is hardly a surprising conclusion, but it is striking to see just how large the quantitative differences are between prohibited and preferred bigrams. Moreover, this finding appears to confirm the view (e.g. Ho 2010: 36ff.; Schellenberg 2013) that tone-melody matching in Cantonese is quite strict compared to many languages. As just noted, the larger purpose of Lin's 2018 study was to make an explicit comparison between Cantonese and Mandarin, and her results show clearly that Cantonese is much stricter in this respect than standard Mandarin.

Somewhat more surprisingly, it can also be seen that there is a clear difference between the two different types of 'oblique' settings illustrated in Fig. 1 above. These are shown in Table 3 in the cells marked with the symbols • (I) and ·· (II).

From sections (a) and (b) we can see that 'Oblique II' bigrams, where the tone remains the same but the musical pitch is either rising or falling, are far more common in the corpus than 'Oblique I', where the tonal pitch direction is rising or falling but the musical pitch stays the same. The bottom row of Table 3(c) shows that 'Oblique II' settings are relatively common with unchanging tonal sequences, although it is clear that 'similar' settings (where unchanging musical pitch mirrors the unchanging tone) are preferred in these cases as well. By contrast, the upper two rows of Table 3(c) show that in rising or falling tonal sequences, oblique I settings are less common even than contrary settings. This suggests that the category 'oblique' may be somewhat incoherent.

3. Two problems

The remainder of the paper is devoted to more detailed discussion of two specific problems that the approach sketched so far brings to light. These are the apparent incoherence of the category 'oblique', just mentioned, and the question of how to determine pitch direction across bigrams that include contour tones.

3.1. Oblique settings

The Cantopop evidence just reviewed seems to show clearly that the two subtypes of oblique are not equivalent. It can be seen from the first two rows of Table 3(b) that oblique I bigrams – where the tonal sequence is rising or falling but the musical sequence remains unchanging – are less frequent than 'contrary' settings, where the tonal and musical sequences go in opposite directions. Ho 2010 reports exactly the same finding, and states (p. 60) that what we are calling an oblique I bigram 'always represents a perceived mismatch'. If the only constraints on tone-melody matching involve the closeness of the match between the music and the text, this finding is unexpected.

The difference can be easily explained if we assume the existence of a strictly musical constraint that disfavours same-note sequences (or, looked at the other way, a constraint that prefers musical melodies that keep moving). This constraint is 'strictly musical' in the sense that it applies only to a feature of the melody, not to how closely tone and melody match; it is thus comparable to McPherson and Ryan's proposal that in Tommo So songs rising musical bigrams are disfavoured. In terms of a ranked-constraint approach like Optimality Theory, this musical constraint would be outranked only by the preference for 'similar' settings: unchanging musical sequences are allowed in the case of unchanging tonal sequences, where they yield a 'similar' bigram, but not otherwise. Looking only at the rightmost column of Table 3(d) above, we see that roughly five-sixths of all repeated-note musical sequences in the dataset involve unchanging tonal sequences and therefore 'similar' bigrams.

The suggestion that moving melodies are preferred is also consistent with the distribution of oblique II bigrams. From the bottom row of Table 3(c) above, we see that scarcely more than half of the unchanging tone sequences are set to unchanging musical sequences, even though such a setting would be 'similar' (and therefore, on a simple account, preferred). Here again, the preference for similar settings seems to interact with an independent preference for moving melodies. In fact, a more detailed breakdown of the oblique II data in the Kirby/Lin corpus suggests that we are dealing with the special treatment of a particular sequence. Specifically, any 'unchanging' tonal sequence involving either of the two 'High' tones (i.e. the sequences T1+T1, T2+T2, T1+T2, and T2+T1) shows a moderate preference for a *falling* musical setting over the expected unchanging or same-note sequence. All other unchanging tonal sequences, as expected by the general preference for 'similar' settings, show a preference for unchanging or same-note musical sequences. This difference is presented graphically in Fig. 4[3]. The explanation for this special treatment of High tonal sequences is hardly clear, but one might speculate that the tendency to 'declination' of F0 during spoken phrases and utterances may make it more natural to set a High-High sequence to a slightly falling musical melody. The same phenomenon is noted by Ho (2010: 65f.), who also suggests the same speculative explanation.

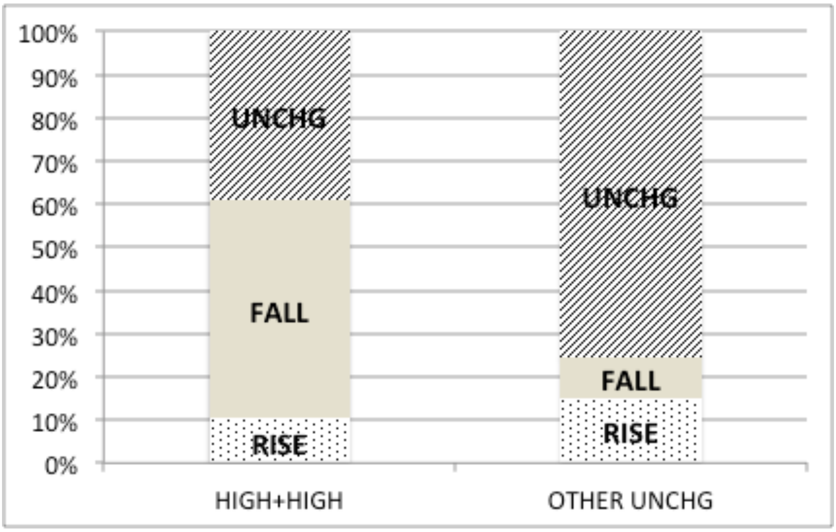


Figure 4. Special treatment of High+High tonal sequences

In any case, the evidence just reviewed suggests that there are problems with any descriptive framework for tone-melody matching that takes the category 'oblique' to be on a par with the categories 'contrary' and 'similar'. The clear differences between what we have called 'oblique I' and 'oblique II' certainly argue for distinguishing those two types from one another. More importantly, the fact that contrary settings are more frequent than oblique I settings seems to

[3] Ho (2010:63ff.) finds a complementary tendency for unchanging Extra-Low tonal sequences (sequences consisting of T4 (阳平, 21) syllables) to be set to a *rising* musical melody. The Kirby/Lin data hint at the same conclusion, but there are too few cases to be confident, which is why I have grouped T4+T4 sequences with Mid+Mid and Low+Low bigrams in Figure 4.

point to a more general problem with the way the notion of constraint violation is used to account for the details of tone-melody matching.

Specifically, as Proto (2016) has pointed out, 'oblique' settings are treated in most work as if they were neutral in some way, neither a complete match nor a complete mismatch; we have already mentioned this in connection with the implications of Schellenberg's term 'non-opposing' (which Proto also uses) and McPherson and Ryan's term 'non-contrary'. This idea is a consequence of conceptualising the principles of tone-melody matching in terms of interacting but contradictory violable constraints: if the two key constraints are a prohibition on contrary settings and a requirement for similar settings, then oblique settings should be preferable to contrary settings, because a 'contrary' bigram violates *both* the prohibition on contrary settings and the requirement for similar settings, whereas an 'oblique' bigram (either oblique I or oblique II) violates only the latters but not the former. But since the Cantopop data shows that oblique I settings are actually *less* frequent than contrary settings and that oblique I and oblique II are not comparable, it is plainly misleading to regard 'oblique' as simply a less serious mismatch than 'contrary'. At present it is hardly clear how to refine the approach so that it better fits actual song data, but Proto's critique directs our attention to the need for some sort of rethinking.

3.2. Setting contour tones

So far I have simply adopted Chan's 'endpoint principle' for categorising pitch direction across tonal sequences involving contour tones, and have simply stated that Ho's application of this principle to Cantonese (distinguishing between 'Low' and 'Extra-Low' endpoints) is superior to Wong and Diehl's version that lumps together tones 6 (阳去, 22) and 4 (阳平, 21) as 'Low'. In this section I first present the evidence for Ho's analysis and then discuss further evidence for the general correctness of the endpoint principle itself. Finally, I discuss a few specific cases that suggest that the endpoint principle may nevertheless coexist with special treatment of certain specific tones or combinations of tones. All the data are taken from the Kirby/Lin corpus.

The evidence for Ho's analysis can be found in a quantitative comparison of T4+T6 bigrams and T6+T4 bigrams. If the endpoint principle is correct and if T6 (阳去, 22) ends Low but T4 (阳平, 21) ends Extra-Low, we should expect that T4+T6 bigrams should preferentially be set to rising musical melodies while T6+T4 bigrams should prefer falling melodies. This is exactly what we see in the Kirby/Lin corpus. Altogether, the corpus contains 90 T4+T6 bigrams and 68 T6+T4 bigrams. The vast majority of the T4+T6 tonal sequences are set to a rising musical sequence, whereas the majority of the T6+T4 tonal sequences are set to a falling musical sequence. The comparison is summarised graphically in Fig. 5. It is clear that the two types of bigrams are treated differently.

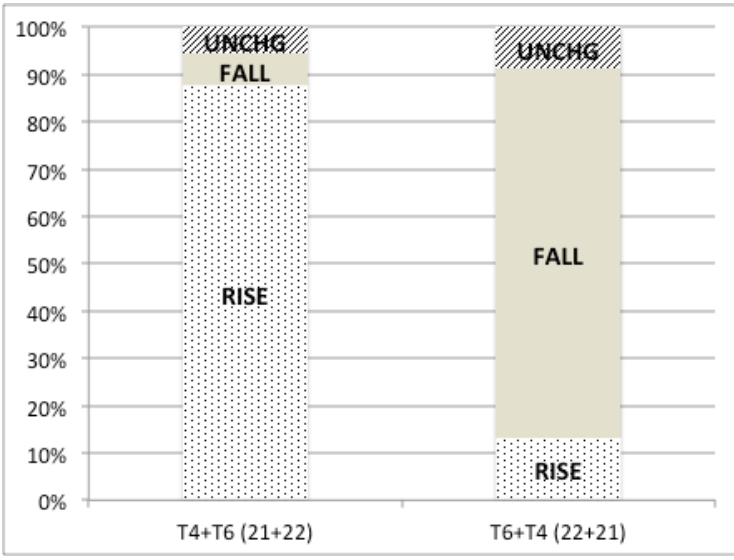


Figure 5. Different patterns of setting tonal sequences T4+T6 and T6+T4 in the Kirby/Lin corpus.

A more basic test of the validity of the endpoint principle is to compare bigrams consisting only of level tones with bigrams involving contour tones. If the endpoint principle is correct, there should be no quantitative difference between the two types of bigram. For example, we can compare the level-tone bigram T1(阴平, 55) + T3 (阴去, 33), which involves only High level and Mid level, with bigrams that also include T2 (阴上, 35) and T5 (阳上, 23). According to the endpoint principle, T2 and T5 should be treated as High and Mid, respectively. We should therefore expect that the musical settings of bigrams with T2 and T5 should be indistinguishable from those involving T1 and T3. (Ho 2010: 47 makes this same point.)

Three sets of such comparisons[4] are displayed in Table 4. In each section of Table 4, the data from the bigram type that consists only of level tones is shaded; the other three types of bigrams (level+contour, contour+level, and contour+contour) follow in the three immediately following rows. Clear evidence for the validity of the endpoint principle is seen in the fact that – with a few exceptions that I return to below – there is no obvious overall difference between bigrams involving only the High and Mid level tones (the shaded top rows of each section of Table 4) and those involving contour tones. The percentages of bigrams in the boldface columns of the table – where the musical sequence matches the tonal sequence – appear roughly comparable.

[4] The fourth possible comparison, namely unchanging High+High sequences, is not included, because of the fact (shown graphically in Fig. 4) that these sequences are treated as a special case for tone-melody matching.

Table 4. Comparison of bigrams composed only of level tone (shaded rows) with bigrams including at least one contour tone

Tonal pitch direction	**Musical pitch direction**			
a. Falling tonal sequences (High-Mid)	Falling	Rising	Unchanging	Total
T1 (High level) + T3 (Mid level)	**76 *(84%)***	11	3	90
T1 (High level) + T5 (Rising to Mid)	**41 *(77%)***	9	3	53
T2 (Rising to High) + T3 (Mid level)	**31 *(91%)***	1	2	34
T2 (Rising to High) + T5 (Rising to Mid)	**29 *(81%)***	3	4	36
b. Rising tonal sequences (Mid-High)	Falling	Rising	Unchanging	Total
T3 (Mid level) + T1 (High level)	7	**54 *(76%)***	10	71
T5 (Rising to Mid) + T1 (High level)	11	**43 *(70%)***	7	61
T3 (Mid level) + T2 (Rising to High)	7	**22 *(71%)***	2	31
T5 (Rising to Mid) + T2 (Rising to High)	2	**24 *(92%)***	0	26
c. Unchanging tonal sequences (Mid)	Falling	Rising	Unchanging	Total
T3 (Mid level) + T3 (Mid level)	4	6	**46 *(82%)***	56
T5 (Rising to Mid) + T3 (Mid level)	3	12	**19 *(56%)***	34
T3 (Mid level) + T5 (Rising to Mid)	6	9	**27 *(64%)***	42
T5 (Rising to Mid) + T5 (Rising to Mid)	3	1	**14 *(78%)***	18

More generally, Table 4 displays the overwhelming evidence – already seen in Table 3 – that Cantopop prefers 'similar' settings, not merely avoiding tone-melody mismatches, but actively seeking matches. The pattern is seen clearly in the difference between the columns in boldface and those in normal font: falling tonal sequences tend to be set to falling musical sequences, rising tonal sequences to rising musical sequences, and unchanging tonal sequences to unchanging musical sequences. There appear to be minor differences depending on the pitch direction, but the overall preference for matching is not in doubt.

At the same time, though, in sections (b) and (c) of Table 4 we find a few cells (highlighted by heavy black borders) that may involve special cases. Perhaps the most notable of these involves the rising sequence T5+T2. This sequence is matched by a musical rise in more than 90% of cases, compared to only 70-75% for the other Mid+High tonal sequences. Why might this be special? Phonetically, T5+T2 is the only one of the four Mid+High sequences in which the <u>*tonal*</u> pitch rise is continuous across the two syllables: perhaps this phonetic property makes the rising musical setting especially appropriate. Similar factors may explain the less consistent patterns seen with the unchanging tonal sequences in the third section of the table: sequences of completely identical tones (level tones T3+T3 or contour tones T5+T5) are matched by unchanging musical sequences about 80% of the time, but the two types of mixed sequences in the table are much more variable and are musically matched in only about 60% of bigrams. It therefore seems important to emphasise that the broad validity of the matching principles does not preclude the existence of special cases. As McPherson and Ryan put it, 'a single aggregate rate of tone-tune association calculated across a whole corpus ... obscures a great deal of complexity.' In the same vein, Ho (2010: 55f.) suggests sardonically that finding 80% of 'similar' bigrams in matrix-style diagrams like those in Table 3 above 'is significant enough to satisfy most linguists' curiosity' and states clearly that it is a mistake to assume that the remaining 20% are simply irregular.

Finally, despite the existence of the endpoint principle, there is reason to suspect that contour tones may be inherently difficult to accommodate to a musical melody in an aesthetically satisfying way. We have just seen that the tonal sequence T5+T2 is set to a rising musical melody in over 90% of cases, unlike all other instances of Mid+High sequences where proportions are closer to 70%. A different kind of evidence suggests another reason to think that contour tones may pose a more general problem for songwriters: it appears that words with the high rising tone T2 (阴上, 35) may be actively avoided.

Table 5. Token frequency of the tones in the Kirby/Lin corpus compared to baseline Cantonese data from Leung et al. 2004

	Token frequency (Leung et al.)		Token frequency (Lin/Kirby)	
Tone	number	percent	number	percent
T1 阴平	30145	21.36%	401	23.31%
T2 阴上	23346	16.54%	178	10.35%
T3 阴去	27002	19.13%	319	18.55%
T4 阳平	15898	11.26%	231	13.43%
T5 阳上	15098	10.70%	215	12.50%
T6 阳去	29660	21.01%	376	21.86%
Total	141149	100%	1720	100%

This can be seen by examining the relative frequency of the individual tones in the corpus as a whole. Previous quantitative work by Leung et al. 2004, which reported both type frequency and token frequency of the six tones in a large (>141,000 syllables) corpus of Cantonese text, gives us a baseline for estimating their expected relative frequency in the Kirby/Lin corpus. This is shown in Table 5. It can be seen from the two shaded cells that the proportion of syllables with T2 (阴上, 35) in the Kirby/Lin corpus is less than two-thirds of what would be expected on the basis of the Leung et al. frequency data, while the proportions of all the other tones are roughly the same or greater. This suggests that it is somehow awkward or unsatisfying to set the high-rising contour of T2 to a single note in a musical melody[5]. Such a situation is entirely comparable to the constraints on lexical choice posed by rhyme and meter: for example, the English word *vacuum* cannot be used at the end of a line of rhyming verse because no other word rhymes with it, and words like *automobile* or *seventy-six* are difficult to accommodate in the iambic meter of much traditional English poetry. Although the Kirby/Lin corpus is too small for serious statistical evaluation,

---

[5] This speculation is consistent with the findings of Schellenberg and Gick (2020), who show that Cantonese amateur singers generally show 'microtonal' variation in singing syllables with a rising tone on a single musical note.

Table 5 shows how quantitative studies may at least allow systematic observations on topics that seem at first glance to be purely a matter of aesthetic judgement. A qualitative study exploring the judgements of Cantopop songwriters might confirm or disprove the speculation that they actively avoid using words with Tone 2.

4. Prospect

The discussion in this paper has illustrated in some detail how the simple quantitative approach to data reduction developed in recent work succeeds in capturing important aspects of the principles of tone-melody matching. Whether or not these principles are then expressed explicitly in terms of an Optimality Theory grammar or in some other way, it seems clear that text-setting practice in tone languages is based in some way on the interaction of violable constraints. However, my goal in this paper has been to draw attention to two problems that may require refinements to our way of thinking about the problem. The actual way oblique settings are used in Cantopop seems to make clear that it is inaccurate to think of 'oblique' as simply a neutral way of avoiding a more serious constraint violation. Similarly, the treatment of contour tones in text-setting seems to show that a simple principle such as the endpoint principle will cover a large proportion of the data but that the residue of exceptions and problems should encourage us to take a closer look for special cases and sub-regularities, and perhaps even to accept that some tones and tone sequences are simply awkward to set to music and may be avoided for that reason.

I am not suggesting, of course, that contour tones and oblique settings are the only problems in current work; there are other foundational issues that this paper could have discussed in the same way. In particular, it seems likely that the **interval** between the two notes in a musical bigram is important in evaluating its effect on the match between tone and melody. The role of interval is discussed by Carter-Ényì 2016 in his analysis of tone-melody matching in Yoruba music, and McPherson and Ryan's Optimality Theory analysis incorporates different constraints based on the size of the interval between the two notes of a bigram. Specifically, they suggest that a large musical rise or fall makes a tone-melody mismatch worse than a mismatch involving a small interval. By contrast, Ho's work (2010: 76-80) emphasises the possibility that specific tonal sequences actually favour specific musical intervals, and shows that mismatches can be made worse both by intervals that are too large and by intervals that are too small: roughly speaking, the phonetic interval between the tone endpoints in a bigram must somehow be matched by the musical interval between the corresponding notes. It is not immediately obvious how such cases might be accommodated in a strict Optimality Theory analysis. In any case, interval needs to be more thoroughly taken into account in future work.

Nevertheless, the basic approach reviewed here seems to offer a wealth of tractable research questions for the future. Among other things, this represents a real opportunity for Chinese scholars. For example, the recent paper by Zhang and Cross (2021) on Chaozhou folksongs, mentioned earlier, reports clear findings within the general framework discussed here; one of the specific questions that they are able to address is the role of complex tone sandhi (变调) in tone-melody matching. As a Southern Min variety (闽南语), Chaozhou syllables exhibit large differences between citation forms and words spoken in phrasal context. Zhang and Cross show clearly that the tonal sequence in a bigram needs to be defined in terms of the surface sandhi tones, not the underlying citation tones. Although this finding is intuitively not surprising, Zhang and Cross's work gives us clear empirical validation of the intuition, based on a clear theoretical and methodological approach. The rich variety of tonal systems and of musical traditions in Chinese thus provides a natural laboratory for comparing different systems and a unique opportunity to deepen our understanding of tone-melody matching.

Acknowledgments

I thank James Kirby for providing me with access to the data on which Lin 2018 is based. A preliminary version of this paper was presented at a one-day workshop on music and language at the University of Massachusetts, Amherst, in May 2019. Thanks to James Kirby and Jiayin Gao (高佳音) for much useful discussion.

References

Bauer, Robert S., Cheung, Kwan Hin, and Cheung, Pak Man. 2003. Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change* 15:211–225.

Carter-Ényì, Aaron 2016. Contour levels: An abstraction of pitch space based on African tone systems. PhD dissertation, The Ohio State University.

Chan, Marjorie K. M. 1987a. Tone and melody in Cantonese. *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society (1987)*: 26–37.

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

Chao, Y. R. 1930. A system of 'tone letters'. *Le Maître Phonétique* 45:24-27.

Ho, W. S. Vincie (何詠詩). 2006. The tone-melody interface of popular songs written in tone languages. *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC 2006)*: 1414–1422.

Ho, W. S. Vincie (何詠詩). 2010. *A phonological study of the tone-melody correspondence in Cantonese pop music*. PhD dissertation, University of Hong Kong.

Karlsson, A. 2018. Coordination between lexical tones and melody in traditional Kammu singing. *Journal of the Phonetic Society of Japan* 22: 30-41.

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

Ketkaew, C. and Pittayaporn, P. 2015. Do note values affect parallelism between lexical tones and musical notes in Thai pop songs? *Proceedings of the 18th International Congress of Phonetic Sciences*.

Kirby, James and Ladd, D. Robert. 2016. Tone-melody correspondence in Vietnamese popular song. *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL-2016)*: 48–51.

Ladd, D. Robert and Kirby, James. 2020. Tone-melody matching in tone language singing**.** In A. Chen and C. Gussenhoven, eds. *The Oxford Handbook of Linguistic Prosody*, Oxford University Press, pp. 676-687.

Leung, Man-Tak, Law, Sam-Po, and Fung, Suk-Yee. 2004. Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments, & Computers* 36: 500–505

Lin, Ruoqi. 2018. A comparison of tonal text-setting in Mandarin and Cantonese popular songs. MSc dissertation, University of Edinburgh.

Lissoir, M.-P. 2016. *Le* khap tai dam*, catégorisation et modèles musicaux.* PhD dissertation, Université Libre de Bruxelles and Université Sorbonne Nouvelle.

Lo, Albert T. C. 2013. Correspondences between lexical tone and music transitions in Cantonese pop songs: a quantitative and analytic approach. MA dissertation, University of Edinburgh.

Matthews, Stephen and Yip, Virginia. 2013. *Cantonese: A comprehensive grammar* (2nd edition). Routledge.

McPherson, Laura and Ryan, Kevin M. 2018. Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94: 119-156.

Proto, Teresa. 2016. Methods of analysis for tonal text-setting. The case study of Fe'Fe' Bamileke. *Tonal Aspects of Language (TAL) 2016*, pp. 162-166. (https://www.isca-speech.org/archive/TAL_2016/pdfs/36-Proto.pdf).

Schellenberg, Murray. 2009. Singing in a tone language: Shona. *In*: Ojo, A. and Moshi, L. (eds.), *Selected Proceedings of the 39th Annual Conference on African Linguistics: Linguistic Research and Languages in Africa*, 137–144. Somerville, MA: Cascadilla Proceedings Project.

Schellenberg, Murray. 2012. Does language determine music in tone languages? *Ethnomusicology* 56: 266–278.

Schellenberg, Murray. 2013. The realization of tone in singing in Cantonese and Mandarin. PhD dissertation, University of British Columbia.

Schellenberg, Murray and Gick, Bryan 2020. Microtonal variation in sung Cantonese. *Phonetica* 77: 83-106. DOI:10.1159/000493755.

Wong, Patrick C. M. (黃俊文) and Diehl, Randy L. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30: 202–209.

Zhang, Xi (张曦) and Cross, Ian. 2021. Analysing the relationship between tone and melody in Chaozhou songs. *Journal of New Music Research* 50: 299-311. DOI:10.1080/09298215.2021.1974490.