Tone-tune association in Tommo So (Dogon) folk songs

Laura McPherson (Dartmouth College) and Kevin M. Ryan (Harvard University)

Abstract

This article explores the relationship between linguistic tone and musical melody in Tommo So, a Dogon language of Mali. Most fundamentally, contrary mappings (rising tone on falling music, or vice versa) are strongly penalized, though oblique mappings (flat tone on changing music, or vice versa) are largely tolerated. Strictness of mapping is further modulated by several factors, including whether the tones straddle a word boundary, whether their source is lexical or grammatical, position in the line, and so forth. We model these conditions using weighted, stringent constraints, and conclude that tone-tune setting bears more commonalities with metrics than previously recognized, setting the groundwork for a more general theory of phonological mapping across modalities.

1 Introduction

In a tone language, where pitch is employed to create linguistic contrasts, what happens when a speaker wants to sing? Are musical melodies constrained by lexical tone, or words chosen carefully to fit the contours of the musical line? Or are the tonal demands of the language set aside in favor of creative musical expression?

This question of “tone-tune association” has been a long-standing one in the literature, often with a central concern of intelligibility or loss of linguistic contrasts in musical settings. As Schneider (1961) puts it, “if a word is to be grammatically intelligible, the individual syllables cannot be sung arbitrarily high or low. Speechtone and musical tone must be definitely correlated” (204). However, studies have shown that loss of tonal information has little effect on intelligibility, even in spoken language. For instance, Schirmer et al. (2005) demonstrate that a change in tone was no more disruptive to spoken word comprehension than a change in a segment in Cantonese, and that speakers are able to retrieve intended meanings based on context. Cutler and Chen (1997) find that spoken word recognition is more constrained by segmental information than tonal information in Cantonese, suggesting that musically-motivated divergences from linguistic tone should not pose a significant challenge for understanding, given that words are not sung in isolation but are part of a broader poetic context. Wong and Diehl (2002) specifically studied the question of tonal perception in sung music, showing that speakers did use musical melody to disambiguate words in a short sung phrase; however, each word was embedded in the same carrier phrase, thus removing crucial contextual information likely to be present in actual lyrics.

Nevertheless, many languages have been shown to match tone and melody at higher than chance levels. In this paper, we follow the assumption that singers may be motivated to match not out of a drive for comprehension but rather due to aesthetic principles. Consider the case of English stress. While intelligibility has been shown to be greater when prosodic stress matches musical stress (Johnson et al. 2013), English speakers still have little
problem understanding in instances of mismatch. But such mismatches are avoided, and often consciously jarring, because they violate wellformedness principles. These principles, rather than intelligibility, are the focus of this article; as such, we join the broader research program of generative metrics and textsetting (e.g. Halle and Keyser 1966, 1971, Kiparsky 1977, Hayes 1988, Halle and Lerdahl 1993, Hanson and Kiparsky 1996, Hayes and Kaun 1996, et seq.).

This paper focuses on women’s folk songs in Tommo So, a Dogon language spoken by about 60,000 people in Mali (McPherson 2013). As we demonstrate, tone and tune are definitely correlated, but the degree of correlation is sensitive to a number of linguistic and musical factors, most of which have not been documented previously for textsetting, including whether tone is lexical or grammatical, whether a sequence of notes is contained within a word or straddles a word boundary, whether the lyrics are rote or improvised, and the position in the musical line. In other words, the question of tone-tune association is complex and multi-faceted even within a single language, and we show how a nuanced study of the question can shed light on questions of mental representations of tone, processing, and effects of prosodic structure on creative adaptation of language. These effects are demonstrated with statistical analysis of the corpus and then modeled in maximum entropy Harmonic Grammar (e.g. Hayes and Wilson 2008), using a combination of scalar constraints and indexed constraints in a stringency relationship.

In setting up this model, we show that tone-tune association, or tonal textsetting, is hardly exotic, as sometimes implied by previous treatments, but rather subject to many of the same constraints invoked in metrics research, such as final strictness (e.g. Hayes 1983), boundary effects (e.g. Kiparsky 1975), and even basic principles of mapping. For example, our constraint penalizing mapping a falling tone onto rising music is akin to a constraint in Hayes et al. (2012: 701) forbidding falling stress in a rising (iambic) metrical context. We take such parallels as further evidence that the set of constraints on artistic adaptation can be employed for different phonological parameters (e.g. stress, weight, tone) mapped onto different artistic structures (e.g. meter, music).

In what follows, we first review the findings of previous work on tone-tune association in §2 and then provide an introduction to Tommo So linguistic structure and music in §3. The corpus and coding are described in §4, followed by a discussion of the results in §5, which illustrates the basic principles of Tommo So tone-tune association and the significance of several factors influencing it. These results are then implemented as an omnibus constraint-based model in §6. We discuss remaining issues in §7 and conclude.

2 Previous work

Tone-tune association has received a fair amount of scholarly attention, spanning back to at least the early 20th century. A diverse group of tonal languages is represented in this literature, including Navajo (Herzog 1934), Kalam Kohistani (Baart 2004), Kammu (Karlsson et al. 2014), Tai Phake (Morey 2010), Thai (List 1961, Saurman 1999), Vietnamese (Kirby and Ladd 2016), Cantonese (Wee 2007), Mandarin (Chao 1956, Wong and Diehl 2002),
Lushai (Bright 1957), Ewe (Jones 1959, Schneider 1961), Hausa (Richards 1972, Leben 1983), Shona (Schellenberg 2009), Xhosa (Starke 1930), and Zulu (Rycroft 1959, Rycroft 1979). The results of these studies are mixed, both across languages and sometimes within a single language. Even languages in the same linguistic family or geographic area can show drastically differing behavior, as is the case for Mandarin versus Cantonese contemporary music, where the former shows little tone-tune correspondence and the latter a very high degree (Ho 2006, Schellenberg 2013).

To illustrate how a single language produces different results for textsetting studies, consider Hausa. Richards (1972) reports only 53.4% tone-tune correspondence in one Hausa song, with correspondence defined as an exact match in the direction of the transition from one syllable to the next (rising, falling, level); though greater than chance, this is hardly a result that shows musical melody to “slavishly adhere” (Schellenberg 2009) to linguistic tone (or vice versa). However, Leben (1983) reanalyzes the data, showing that if correspondence is calculated between the melody and the intonational realization of tone, then correspondence becomes much more exact. While this fits into what Leben claims “we know from other studies of the correspondence between tunes and tones, where in general the result is that the correspondence is either extremely close or practically nonexistent” (148), a partial fit may not be so surprising after all. Schellenberg’s (2012) survey paper shows that many studies have found levels of correspondence significantly greater than chance but far from a perfect fit. If anything, the literature may support this as the natural state of tone-tune correspondence rather than Leben’s all-or-nothing dichotomy.

Studies also point to the need to distinguish between what Schellenberg (2012) calls “parallel” versus “non-opposing” transitions. Parallel transitions are a perfect match in direction: falling tone with falling tune, level tone with level tune, rising tone with rising tune. This is tone-tune correspondence in the strictest sense. Non-opposition transitions, on the other hand, simply avoid direct clashes in direction: rising tone with falling tune, or vice versa; level transitions may neutrally combine with all transitions. In Kirby and Ladd’s (2016) terminology, which we follow here, we can define three states of correspondence: parallel, opposing, and “oblique”, wherein a level sequence is mapped to a rise or fall (e.g. level tonal transition on a rising melody or a rising tonal transition on a level melody). If tone-tune correspondence is defined using just the non-opposing metric (including both parallel and oblique transitions), many languages reach near-perfect levels of correspondence. For instance, Richards’ statement of 53.4% correspondence for Hausa was based on parallel transitions; Schellenberg’s (2012) calculations show a 96% rate of non-opposition, suggesting that the artist may still accommodate the language’s tonal system to achieve an aesthetically pleasing match between lyrics and melody but without rigid determinism.

3 Tommo So

Tommo So is one of around twenty Dogon languages spoken in east central Mali. The genetic affiliation of the Dogon family has been the subject of much debate, but it is currently thought to form its own branch of Niger-Congo (Blench 2005). All data for this paper were
gathered by the first author in Mali in 2012, and they represent the northernmost dialect of the language spoken in the commune of Tédié.

### 3.1 Linguistic structure

Tommo So is an SOV language with isolating nominal morphology and agglutinating verbal morphology; nominal morphosyntactic features are expressed with typically toneless enclitics, shown in (1a), while verbal morphosyntactic features are expressed by a combination of suffixes and grammatical tone patterns, illustrated in (1b):

(1)  

a.  
gámá=mbe=ñ
  cat=DEF=PL=OBJ
  ‘the cats (focused object)’

b.  
{pò-nd-iyè-m}^{L}-éélè-ý
  fat-FACT-MP-CAUS-NEG.IMPF-1PL
  ‘we will not make (sb) fat’

The aspectual portmanteau suffix in (1b) is associated with a \{L\} grammatical overlay on the verb stem, indicated here with a superscripted L.

The language has a seven vowel system (/i, e, ə, a, o, u/) for which length is contrastive. A strict system of vowel harmony greatly restricts possible vowel combinations on roots and stems (McPherson and Hayes 2016).

In terms of tone, Tommo So has two tonal primitives H and L, which can combine to form the contour tones LH and HL. Lexical tone melodies are very constrained, with nearly all native vocabulary falling into one of two melodies, /H/ and /LH/. Minimal pairs are provided in (2):

(2)  

<table>
<thead>
<tr>
<th></th>
<th>/H/</th>
<th>Gloss</th>
<th>/LH/</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>náá</td>
<td>‘mother’</td>
<td>náá</td>
<td>‘cow’</td>
<td></td>
</tr>
<tr>
<td>dámáa</td>
<td>‘village’</td>
<td>dámáa</td>
<td>‘hoe’</td>
<td></td>
</tr>
<tr>
<td>ísé</td>
<td>‘empty’</td>
<td>ísé</td>
<td>‘dog’</td>
<td></td>
</tr>
</tbody>
</table>

/LH/ lexical melodies are rare in native vocabulary but fairly common in loanword vocabulary, especially loanwords from Fulfulde, where /HL/ mimics the language’s initial stress pattern:

(3)  

a.  
ámíru
  ‘chief’

b.  
hóláâl
  ‘honor’
Regardless of the overall melody, all lexical stems must bear at least one H tone.

In addition to H and L, Tommo So has surface-underspecified syllables that receive their f0 contour by interpolation from surrounding tones (McPherson 2011). Underspecification is confined to functional elements like clitics and epenthetic vowels; for the former, see (1a) and for the latter see (3a).

Tommo So also displays complicated grammatical tone patterns, both at the word level and at the phrase level (McPherson 2014, McPherson and Heath 2016). Both kinds of grammatical tone are replacive in Tommo So, with lexical tone overwritten by grammatically-controlled melodies (“tonal overlays”). Word-level grammatical tone, triggered by inflectional affixes or morphosyntactic features, is found on verbs. The following partial paradigm for the verb /jɔbɔ/ ‘run’ illustrates these patterns:

(4)  
<table>
<thead>
<tr>
<th>Tone Pattern</th>
<th>Affirmative imperfective</th>
<th>Negative imperfective</th>
<th>Affirmative defocalized perfective</th>
<th>Negative perfective</th>
<th>Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>{HL} jɔbɔ̃H̱L-dɛ</td>
<td>{L} jɔbL-ééɛɛ</td>
<td>{L} jɔbL-ɛɛɛɛ</td>
<td>{L} jɔbL-ɛɛɛɛ</td>
<td>{H} jɔb3H</td>
</tr>
</tbody>
</table>

Phrase-level grammatical tone is found in the noun phrase, where tonal overlays are triggered by modifiers of certain syntactic categories on c-commanded words. For instance, demonstratives and possessors impose tonal overlays while numerals do not, illustrated in (5) with the lexically LH-toned noun /bàbɛ/ ‘uncle’:

(5)  
<table>
<thead>
<tr>
<th>Tone Pattern</th>
<th>bàbɛ̃</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Hbàbɛ</td>
</tr>
<tr>
<td>uncle</td>
<td>nɔ</td>
</tr>
<tr>
<td>this</td>
<td>‘this uncle’</td>
</tr>
<tr>
<td>b.</td>
<td>1sg.pro bàbɛ</td>
</tr>
<tr>
<td>‘my uncle’</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>bàbɛ</td>
</tr>
<tr>
<td>uncle</td>
<td>ɛnɔ</td>
</tr>
<tr>
<td>five</td>
<td>‘five uncles’</td>
</tr>
</tbody>
</table>

If the modifier c-commands multiple words, the tonal overlay will affect this whole c-command domain:¹

¹The L on the demonstrative is due to redistribution of an underlying /LH/ sequence onto the toneless enclitic; in the absence of a host for the second tone, /LH/ simplifies to H, as seen in (5).
For more detailed information on Tommo So tone, see McPherson (2013, 2014) and McPherson and Heath (2016). As we will see below, lexical tone and grammatical tone are treated differently for the purposes of tone-tune association.

### 3.2 Musical structure

The structure of a Tommo So folk song consists of a solo verse combining fixed, rote lyrics with bits of improvisation (such as adding in people’s names, repeating different lines, or altering the melody), followed by a chorus. The verses are often elaborated versions of the chorus. The scale used in the current data sample is the Eb major scale, though most songs use only five of seven notes, making them largely pentatonic:

(7) **Dogon scale**

Eb F (G) Ab Bb C ((D)) 1 2 (3) 4 5 6 ((7))

The third note of the scale, G, is frequently employed in one of the songs in the corpus, *Simila*, but is omitted in most other songs. Even in this case, however, G typically takes the place of Ab, resulting once again in a pentatonic scale, albeit a different one compared to most of the corpus. Similarly, the leading tone, D, is seldom used, but a handful of cases exist. Notes \{1 2 4 5 6\} of the Western scale constitute the heart of the Dogon scale.

Musical lines have a falling melodic contour; a small amount of rising movement can be found in the first part of the line, but the majority of the line descends towards either the supertonic, for non-final lines in a verse or chorus, or the tonic, for the final line. The melody typically levels out at this final note for the final four or five syllables of the line, as shown in the following verse-final line of *An Elephant Gave Birth*.

(8) **An Elephant Gave Birth: Line 23**

\[
\text{nɔm-báá yɛl-ɛɛ } \text{i-rù gi-yɛ ɲ-dɛ́ } \text{á-wi-nɛ } \text{nɔ-nú yɛl-ɛɛ } \text{á-rá-má ye}
\]

Translation: “Come here, whoever is thirsty for milk, come and suckle.”

To confirm this impression of a falling melodic contour in each line, we averaged the melodic contour of each line in our data corpus, normalized for length; figure 1 displays the result.

The folk songs are polyrhythmic, with the drum line (performed in the corpus on a calabash) in a different time signature from the singing. We will generally abstract away from questions of time signature, as the rhythmic or metrical aspects of textsetting are not
Figure 1: Musical lines tend to decline in tune. The x-axis represents percentage of the way through the line. The y-axis is the height of each note in terms of steps from the mean. The dark band is a GAM smoother over all of the lines in the corpus.

The focus of this work. The musical transcriptions in the appendix are intended to capture the general rhythm of the singing rather than an exact representation.

A final point raises interesting questions for future research: For the Dogon people, Tommo So is the language of song. Nearly all Dogon people sing in Tommo So regardless of their native language (Hochstetler et al. 2004, personal experience of the first author). The songs in our corpus are sung by native speakers, but future work could examine the lyrical structure and degree of tone-tune association when singers do not speak the language.

4 Data corpus

The analyses in this paper are based on a corpus of women’s folk songs recorded in Tédié, Mali in 2012. They were performed by three older women noted for their singing talent: Tepama Ouologuem, Roukiatou Djeoukile, and Kounjay Ouologuem. The women sang continuously for an hour and a half, from which the first author transcribed roughly thirty minutes of lyrics with the help of a native speaker consultant, Sana “M. le Maire” Ouologuem. Of these thirty minutes, we musically transcribed and coded eleven minutes containing eight different songs. This resulted in a data corpus of 172 musical lines, consisting of 2,223 musical bigrams (two-note sequences). Lines range in length from three syllables to thirty-four syllables. Three illustrative songs can be found in the appendix; the full corpus is available online at <URL suppressed>.

For the purpose of analysis, each bigram was coded for a number of other factors, briefly described in the following subsections.
4.1 Tone

In the majority of bigrams, two notes correspond to two syllables, and the tone of each syllable was then coded. With three possible tones (H, L, 0), this yielded nine possible two-tone sequences: HH, LL, HL, LH, 00, 0L, L0, 0H, H0. In a few cases of melisma, a single syllable spans multiple notes, in which case, both notes correspond to the same tone.

4.2 Change in note

As shown in (7), the most typical scale used in the folk songs is a pentatonic one consisting of the notes Eb, F, Ab, Bb, and C. This was taken as the core scale and coded as notes 1, 2, 3, 4, and 5, respectively. Change in note is thus calculated as Note2 - Note1, yielding negative integers for falling sequences (e.g. -1 is a one note musical fall), positive integers for rising sequences (e.g. 2 is a two note musical rise), and zero for a level sequence of notes. For the songs that employ G rather than Ab, G was treated as a half step (2.5), but as discussed below, there were not enough data of this type and for the purpose of analysis, it was rounded up to 3; since songs typically do not use both G and Ab, we see no conceptual problem with this simplification.

4.3 Juncture strength

To test whether tone-tune association is sensitive to word boundaries, we coded each bigram for juncture strength: 0 for two notes within a word (including across affix boundaries), 1 across a clitic boundary, and 2 across a word boundary. The boundary between two stems in a compound noun or Possessor+N sequence (tied together by grammatical tone, see (5b)) was coded as 1, i.e. as a clitic boundary. Many lines contain the vocable ee or yee, analogous to something like English oh, and this was treated grammatically as an enclitic.

4.4 Improvised vs. rote

Each Tommo So folk song consists of a few rote lines, which are repeated several times in verses and choruses, passing between singers. Interspersed with these rote lyrics are bits of improvisation, consisting mainly of people’s names and slight changes in wording. Consider the following two verses (non-consecutive in the song) from “An elephant gave birth”, in which improvisation is bolded:

(9) a. Verse 1 (Tepama)

\[
gwéⁿ nálëᵗ emⁿ kémínjè såmëᵗ, gwéⁿ nálëᵗ
\]

An elephant gave birth, there was so much colostrum, an elephant gave birth

\[
Yàândó giné úwó=ne gwéⁿ nálëᵗ, èmⁿ kémínjè såmëᵗ
\]

Yàândó, an elephant gave birth in your house, there was so much colostrum

\[
ámíru=ge 'ginè=ge=ne gwéⁿ nálëᵗ', èmⁿ kémínjè såmëᵗ
\]

An elephant gave birth in the chief’s house, there was so much colostrum
b. Verse 2 (Roukiatou)

\[
gwe^n \text{nålê}^\circ \text{ème}^\circ \text{kémînjé} \text{sâmê}^\circ, \ gwe^n \text{nålê}^\circ
\]

An elephant gave birth, there was so much colostrum, an elephant gave birth

\[
\text{âbáá}=\text{mbe} \ 'dàmmà \ gwe^n \text{nålê}^\circ \text{ème}^\circ \text{kémînjé} \text{sâmê}^\circ
\]

In the fathers’ village an elephant gave birth, there was so much colostrum

Here we see that rote lyrics consist of the phrases an elephant gave birth, there was so much colostrum, while the place where the elephant gave birth varies and can include names of those present (such as Yààndó, the first author’s Tommo So name). In the data coding, we treat all of these interchangeable phrases as improvised, though certain elements of them (in the house of, in the village of) may be considered rote as well.

4.5 Singer

The data corpus contained data from three singers, and each bigram was coded for singer. Verses are sung by a soloist, and so were tagged for only a single singer (K, T, R), while choruses are sung by the remaining two, yielding the codings KT, KR, and TR. Coding for singer allows us to test for interspeaker differences in tone-tune association rates.

4.6 Position in the line

Each bigram was coded for its position in the line. On a broad scale, bigrams were coded as final and non-final, allowing us to easily identify the ends of lines (which were also uniquely numbered and coded). On a finer scale, bigrams were also coded for their location in terms of percentage through the line of the first element.

4.7 Lexical vs. grammatical tone

In §3.1, we described the extensive use of replacive grammatical tone in Tommo So. Each tone in the data corpus was coded for being either Lexical (L) or grammatical (G), to test the hypothesis that tone-tune association is stricter for lexical tone sequences than for grammatical tone. This yielded two-note sequences LL, GG, LG, and GL.

5 Results

We first discuss the basic results of tone-tune alignment in the corpus before discussing how various factors affect these rates of alignment.

Tommo So singers show a significant tendency to align tonal contours with musical contours and to avoid aligning conflicting contours. Following Kirby and Ladd (2016), we distinguish among three types of alignment. First, alignment is PARALLEL if a rising tone (LH) is mapped onto rising music, a falling tone (HL) onto falling music, or a flat tone (HH or LL) onto flat music. Second, alignment is CONTRARY is a rise is mapped onto a fall, or
vice versa. Otherwise — when a contour tone is mapped onto flat music, or a flat tone onto changing music — the alignment is said to be **OBLIQUE**.

<table>
<thead>
<tr>
<th>tone</th>
<th>sequence</th>
<th>2+ up</th>
<th>1 up</th>
<th>same</th>
<th>1 down</th>
<th>2+ down</th>
</tr>
</thead>
<tbody>
<tr>
<td>up (LH)</td>
<td>same (LL or HH)</td>
<td>40 (80%)</td>
<td>178 (62%)</td>
<td>243 (29%)</td>
<td>46 (12%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>down (HL)</td>
<td></td>
<td>3 (6%)</td>
<td>10 (4%)</td>
<td>193 (23%)</td>
<td>112 (28%)</td>
<td>17 (39%)</td>
</tr>
</tbody>
</table>

Figure 2: Alignment between tonally specified syllabic bigrams and musical changes. Bolded cells are parallel, blank cells oblique, and shaded cells contrary. Percentages are by column.

We first consider only specified tones, putting aside for the moment zero tones (see §3.1). Our corpus contains 1,615 transitions between two tonally specified syllables. Among these transitions, 751 (46.5%) are parallel, 805 (50%) oblique, and the remaining 59 (3.5%) contrary. Full counts are given in Figure 2 and plotted in Figure 3. In short, 96.5% of specified sequences are non-contrary. These percentages also highlight that Tommo So textsetting is highly tolerant of oblique alignment, which is approximately as frequent as parallel alignment.

Figure 3: Mapping of fully specified two-tone sequences to their respective musical notes. Musical sequences are coded for both polarity (rising or falling) and degree (one step or more than one step, rounding half steps to the nearest more peripheral whole).

Parallel alignment is evident not only in polarity but also in degree. Taking only contour (LH or HL) tones, the ratio of parallel to contrary alignment for one step up or down (second and fourth columns of Figure 2) is 5.2 (290 to 56). For two or more steps (peripheral columns
of Figure 2), the ratio is significantly greater (Fisher’s $p = .03$, odds ratio 3.7), at 19.0 (57 to 3).

Inspection of the plot (Figure 3) further reveals that flat tones are treated more like falling tones ($r^2 = .95$) than rising tones ($r^2 = .52$). A Mantel-Haenszel test supports this finding of asymmetric similarity ($M^2(4) = 86, p < .0001$). In this sense, then, the most important tonal opposition in Tommo So textsetting is not falling vs. rising, but non-rising vs. rising.

Until this point, only fully specified tone sequences (viz. HH, LL, LH, and HL) have been considered. Figure 4 repeats the specified data but also introduces the alignment data for tone sequences containing one or more zero (i.e. unspecified) tones. As in Figure 3, the peripheral columns are LH and HL, respectively. The nonperipheral columns now comprise the flat tones as well as zero-containing sequences. Because there are three tone types (L, H, and 0), there are $3 \times 3 = 9$ columns in the plot.

![Figure 4](image)

Figure 4: Mapping of all two-tone sequences to their respective musical notes (cf. Figure 3).

Musical rises are strongly skewed towards two particular tone sequences, namely, LH and 0H, which together pattern as rising tones (the slightly higher rate of falls on 0H notwithstanding). All other tone sequences, including falling and flat tones, comprise the ‘nonrising’ condition. This bifurcation is supported by cluster analysis. The dendrogram in Figure 5 depicts a divisive hierarchical clustering of the correlation matrix for the nine tonal sequences.
(Rousseeuw and Kaufman 1990: §6). The first division separates LH and 0H — the rising tones — from all other tones. Within these two groups, rising and nonrising, there is not as much disparity (height on the y-axis) as there is between them. Figure 6 collapses all nine tone combinations into rising vs. nonrising to illustrate their aggregate alignments.

Figure 5: Dendrogram showing the clustering of the nine tonal sequences with respect to their treatment in textsetting.

Figure 6: Figure 4 collapsed into two overarching groups.
5.1 Improvised vs. rote

This section considers whether tone-tune mapping is stricter in the improvised or rote sub-corpus. For the present purposes, only bigrams in which both syllables have the same compositional status (improvised or rote) are taken as data.

A metric for the strictness of alignment is taken to be the ratio of parallel and oblique mappings to contrary mappings. Moreover, the bifurcation between ‘rising’ and ‘nonrising’ established in §5 is adopted here, such that the conditions for alignment for the rising group are the same as those for LH and the conditions for alignment for the nonrising group are the same as those for HL. For example, 00 being mapped to a rise is counted as contrary, since 00, as a nonrising tone, is treated the same as HL.

Most (88%) of our corpus is rote. In the rote material, the ratio of non-contrary (parallel or oblique) to contrary mappings is 11.2 (1,702 to 152). In the improvised material, by contrast, the ratio is 5.0 (218 to 44), significantly lower (Fisher’s $p < .0001$, odds ratio 2.3). Thus, as one might expect, tone-tune alignment is not as strict for extemporaneous lyrics as it is for established songs. See §5.4 for further support of this conclusion based on a logistic model containing several factors.

5.2 Singer specificity

We now consider whether singers differ significantly in their strictness of tone-tune alignment. The present corpus comprises data from three speakers, labeled R (25% of the corpus), K (20%) and T (20%). The remaining 35% of the corpus is sung by pairs of these three singers. For this section, only data sung by individuals are considered.

Figure 7 shows the percentage of contrary as opposed to parallel/oblique mappings for each singer. Within each speaker, the data are further separated according to improvisational status, putting aside bigrams with mixed statuses. As discussed in §5.1, contrary mappings are more frequent under improvisation. As figure 7 reveals, this generalization holds for every speaker individually, though K’s alignment under improvisation is numerically stricter than R’s and T’s. Nevertheless, speaker identity is not a significant factor for either the rote data ($\chi^2(2) = 1.5, p = .47$) or improvised data ($\chi^2(2) = 2.3, p = .32$). That said, improvised data are underrepresented in the corpus, and if K’s better alignment persisted in a somewhat larger corpus, singer identity would readily reach significance.

5.3 Position in the line: final strictness

Crosslinguistically, mapping of all sorts tends to become stricter towards the ends of poetic constituents (see Hayes 1983: 373 for an overview; Chen 1979 for a case from tone mapping in Chinese verse). Figure 8 shows that this tendency is also observed in Tommo So textsetting, in which contrary alignments decline relatively linearly from 17% in the first decile of the line to 3% in the final decile ($p < .0001$ in a logistic model). This decline in contrary mappings

\[2\text{One consultant fervently asserted that K is considered a very good singer. We speculate that this perceived talent may be partially based on the ability to align tone and tune when singing extemporaneously.}\]
across the line cannot be attributed to an increase in flat music towards the end of the line. Although there is such an increase (from 46% flat in the first decile to 76% flat in the final decile; $r^2 = .50$), it is not enough to explain a nearly sixfold decline in contrary mappings, a much stronger trend ($r^2 = .93$). Moreover, most of the increase in flatness occurs in the final 20% of the line. In the first 80% of the line, flatness and position in the line are not significantly correlated ($p = .50$), while in the same portion of the line, mapping strictness significantly increases ($p = 0.0004$).

### 5.4 Juncture level

Every transition in the corpus is coded for its boundary level: none, clitic, or word. As this section will show, mapping between tones and tunes in Tommo So is more stringent when the musical bigram is contained within a word than when it is divided between words. This tendency accords with the crosslinguistic tendency for greater strictness of metrical mapping within tighter phrasal constituents, especially words (on English, see Magnuson and Ryder 1970, Kiparsky 1975, and Hayes et al. 2012).

To demonstrate this effect of juncture while controlling for the other factors under investigation, a mixed effects logistic regression model is employed, as summarized in figure 9. The dependent variable is taken to be whether each mapping is contrary (0) or non-contrary (1), as in §5.1–5.3. Fixed effects include juncture level, improvisational status (§5.1), position through the line in deciles (§5.3), and the statuses of the two tones as lexical vs. grammatical (to be discussed in §5.5). The model also includes random intercepts for singer identity (six levels, including combinations), not shown. The $p$-values reported in the table are inflated by a Holm-Bonferroni adjustment, which penalizes them according to the total number of
Figure 8: Increasing strictness in mapping from the beginning (left) to the end (right) of the line. The y-axis is the percentage of contrary mappings in each part of the line divided into deciles, each decile given with its 95%-confidence Clopper-Pearson interval.

The factors tested (Holm 1979). Factors with non-significant $p$-values are grayed out.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>$z$-value</th>
<th>Holm-adj. $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.733</td>
<td>0.178</td>
<td>9.73</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Position in Line</td>
<td>0.166</td>
<td>0.027</td>
<td>6.06</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Juncture=Clitic</td>
<td>0.610</td>
<td>0.294</td>
<td>2.08</td>
<td>.114</td>
</tr>
<tr>
<td>Juncture=Word</td>
<td>-0.587</td>
<td>0.183</td>
<td>-3.21</td>
<td>.008</td>
</tr>
<tr>
<td>Composition=Improvised</td>
<td>-0.625</td>
<td>0.205</td>
<td>-3.04</td>
<td>.012</td>
</tr>
<tr>
<td>Composition=Mixed</td>
<td>-0.329</td>
<td>0.353</td>
<td>-0.93</td>
<td>.353</td>
</tr>
<tr>
<td>LexGram=Grammatical</td>
<td>-0.544</td>
<td>0.208</td>
<td>-2.62</td>
<td>.036</td>
</tr>
<tr>
<td>LexGram=Mixed</td>
<td>0.352</td>
<td>0.225</td>
<td>1.57</td>
<td>.234</td>
</tr>
</tbody>
</table>

Figure 9: Regression table for tone-tune alignment. Reported $p$-values are penalized by a Holm adjustment. Factors with Holm-adjusted $p$-values of greater than 0.05 are grayed out.

The factors discussed in previous sections, namely, position in the line and compositional status, remain significant in this more complex model. Juncture is now also observed to be significant: Alignment is significantly worse across words than within them, the latter being the baseline in this model. As confirmed by a Tukey’s HSD test, mapping across word boundaries is also significantly worse than mapping across clitic boundaries ($p < .0001$), while clitic boundary vs. no boundary is non-significant ($p = .114$).
5.5 Grammatical vs. lexical tone

One final linguistic factor affecting textsetting is considered here. In particular, a tone’s status as lexical vs. grammatical has a significant effect on the strictness of its alignment with the tune. Lexical tone refers to tone determined by a lexical representation while grammatical tone is determined by a grammatical overlay (McPherson 2014). As the regression table in figure 9 reveals, bigrams in which both tones are grammatical (“LexGram=Grammatical”) are significantly less well aligned than bigrams in which both tones are lexical (the baseline level, not shown). Bigrams of mixed grammatical-lexical status are not significantly different from either the lexical-lexical or grammatical-grammatical types.

5.6 Summary of the empirical results

Here we summarize the empirical results on tone-tune alignment in Tommo So folk songs.

- Contrary mappings are strongly avoided, but oblique mappings are only weakly avoided (§5).
- Contrary mappings are avoided more stringently for larger musical steps (§5).
- The most significant bifurcation in the behavior of tonal sequences is that between rising sequences (LH and 0H) and all other, ‘nonrising’ sequences (§5).
- Mapping is stricter in rote material than improvised material (§5.1).
- Our corpus does not exhibit significant differences between singers, though this non-result may be driven in part by the small amount of improvised material in the corpus (§5.2).
- Mapping is stricter towards the end of the line (§5.3).
- Mapping is stricter within words than across them (§5.4).
- Mapping is stricter for lexical tone than grammatical tone (§5.5).

The first result aligns with the literature on other languages, such as Vietnamese (Kirby and Ladd 2016), where contrary mappings are avoided more strongly than oblique mappings. However, the degree of difference between the two is stronger for Tommo So, with oblique mappings treated as almost equally optimal as parallel mappings. In this way, it more closely mirrors results from other West African languages like Ewe (von Hornbostel 1928, Jones 1959) or Hausa (Richards 1972) in which only about 50% of mappings are parallel but 90% or more of the corpus is non-opposing (i.e. parallel or oblique). There are currently not enough studies of West African tonal textsetting to determine whether this is an areal feature or a common result for languages of this tonal structure.

The division in Tommo So between ‘rising’ vs. ‘nonrising’ has not been reported in the literature on tone-tune association, but it is not clear that it has ever been explicitly tested.
for. Given the tendency for musical (and speech) declination, it would not surprise us to find this division in other languages.

Two results probe the question of human performance, namely the division between improvised and rote lyrics and the effect of singer. That mapping would be stricter in rote lyrics than in improvised ones is unsurprising; we may attribute it to evolutionary selection involving not just the original composer but subsequent singers who chose to embrace certain phrases more than others, akin to the emergence of formulas in oral poetics (e.g. Bakker 1997, 2005). The fact that singers are capable of tone-tune matching to some degree even when improvising, however, shows that tonal textsetting is also a synchronic effect and that people are constrained by a musicolinguistic production grammar.

The next two results are consistent with the literature on metrics: mapping is stricter at the ends of lines and stricter within words than across them. As discussed in the next section, this suggests that the constraint set responsible for tone-tune alignment may be the same constraints, or the same general constraint templates, responsible for meter.

Finally, the different behavior of lexical and grammatical tone has not, to our knowledge, been investigated before in studies of tonal textsetting. It is a natural question to ask in studies of African languages, where grammatical tone is far more widely attested than in East Asian tone languages. We return to the question of whether lexical tone should be expected to be more strictly adhered to than grammatical tone in the next section.

6 Constraint-based analysis

This section formalizes a grammatical model for tone-tune association in Tommo So folk songs. The model is set in maximum entropy (maxent) Harmonic Grammar (HG), a weighted-constraints framework in which candidate mappings are assigned probabilities, permitting the modeling of the variation found in the corpus. In maxent HG, as in HG more generally, each constraint is assigned a nonnegative weight, and a candidate’s harmony score is the sum of its weighted constraint violations. In classical HG, the candidate with the highest harmony wins. Maxent HG takes the process one step further: Harmonies are converted to probabilities by $e^{H_0}/\sum_i e^{H_i}$, where $H_0$ is the harmony of the candidate and the sum $\sum_i$ ranges over all candidates (e.g. Hayes and Wilson 2008 and references therein). Maxent HG has been employed for metrics by Ryan (2011), Hayes and Moore-Cantwell (2011), and Hayes et al. (2012).

Given the focus of this article, we operationalize the grammar as evaluating mappings between tonal bigrams and musical transitions. A tonal bigram is $<\{L, H, 0\}, \{L, H, 0\}>$, annotated with any relevant additional linguistic information (boundary level, lexical vs. grammatical tone, etc.). Musical transitions are taken to be integers in the range $[-2, 2]$, where negative is falling. Since only 2.8% of transitions in our corpus involve half-steps, we simplify the candidate set by rounding these outwards to the nearest full step. Furthermore, since only 0.4% of transitions are more than two steps, we truncate the range to $[-2, 2]$ to keep the candidate set more manageable.

Two sample tableaux follow in (10) for inputs LH and HL, respectively. Two illustra-
tive constraints are shown, to be described further below: *CONTRARY penalizes mapping a tonal sequence onto an opposing musical sequence, and *STEP penalizes a change in the music according to its degree. The constraint weights in (10) were optimized only for the shown data; they will be adjusted in the full model below. Weight optimization was carried out by maxent learning software by Wilson and George (2008), in every case with the default settings of $\mu = 0$ and $\sigma^2 = 100,000$.

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{step:} & \text{obs'd} N & \text{obs'd p} & \text{gen'd p} & \text{weighted penalty} & \text{CONTRARY weight = 2.8} & \text{STEP weight = 0.8} \\
\hline
\text{a. } -2 & 0 & .000 & .007 & -4.4 & -1 & -2 \\
\text{b. } -1 & 6 & .012 & .016 & -3.6 & -1 & -1 \\
\text{c. } 0 & 243 & .479 & .594 & 0 & 0 & 0 \\
\text{d. } 1 & 178 & .351 & .265 & -0.8 & 0 & -1 \\
\text{e. } 2 & 40 & .079 & .118 & -1.6 & 0 & -2 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{<H, L>} & \text{obs'd p} & \text{gen'd p} & \text{weighted penalty} & \text{CONTRARY weight = 2.8} & \text{STEP weight = 0.8} \\
\hline
\text{a. } -2 & 17 & .051 & .118 & -1.6 & 0 & -2 \\
\text{b. } -1 & 112 & .334 & .265 & -0.8 & 0 & -1 \\
\text{c. } 0 & 193 & .576 & .594 & 0 & 0 & 0 \\
\text{d. } 1 & 10 & .030 & .016 & -3.6 & -1 & -1 \\
\text{e. } 2 & 3 & .009 & .007 & -4.4 & -1 & -2 \\
\hline
\end{array}
\]

As stated, this formalism takes the linguistic representation as the input, which is then mapped onto music. But it is also possible to conceive of the grammar in a bidirectional sense, such that neither music nor tone is prior. Consider again the data represented by the two tableaux in (10), now reinterpreted as a single tableau in (11), such that each candidate is an unordered tone-tune pair. The optimization program was rerun on this new tableau, in this case learning the same weights (though in general, weights are not always preserved under such a transformation). Thus, the constraints posited here can just as easily be interpreted in a parallel, bidirectional sense.
A further conceptual issue with this approach concerns our choice to operationalize the input as a bigram as opposed to a longer sequence such as a phrase or line. The question is largely moot; the same grammar, with the same weights, can be applied to longer sequences. If the input were increased to a three-syllable sequence, for instance, there would be two transitions to represent in each candidate. Because we consider 5 possible transitions $[-2, 2]$, this makes for $5^2$ candidates for a trigram. More generally, for $n$ syllables in the input, there are $n - 1$ ordered transitions and $5^{n-1}$ candidates. Median line length in the corpus is 15 syllables, corresponding to 6.1 billion candidates. Thus, optimizing the weights based on full lines is infeasible, and at any rate also unnecessary, as none of the constraints that we discuss evaluates a sequence of more than two syllables.

We turn now to the composition of the constraint set (CON), beginning with two purely musical constraints, namely, *Step and *Up. *Step evaluates every pair of adjacent notes, assigning a penalty equal to the number of musical steps separating them. This constraint reflects the strong tendency on the part of Tommo So singers to avoid larger musical intervals. As illustrated in Figure 10 (left), no change is the most frequent outcome, with a roughly logarithmic decay for every additional step. The plot on the right shows that larger steps are similarly avoided even when only parallel tone-tune transitions are considered, meaning that tone-mapping is unlikely to be driving large-step avoidance.

*Up penalizes any rise in the music, reflecting the tendency of musical lines to fall (an average of 2.2 steps per line for our corpus), as discussed in §3.2. Typologically, it may be that falling music is more preponderant than rising music, especially at the ends of lines, perhaps with some connection to the universality of declination in natural language (Huron 1996, 2006). At any rate, this is not a tendency that can be explained by the distribution of high and low tones in Tommo So, which are roughly evenly distributed across lines ($r = .03$ for high tone incidence vs. percentage through line).

The remaining constraints regulate tone-tune mapping. *Contrary penalizes mapping a rising tone sequence (LH) onto a fall in the music or a falling tone sequence (HL) onto a rise in the music. Furthermore, given that contrary mappings to larger steps are especially avoided (§5), we take the penalty of *Contrary to scale with the degree of the step. For

<table>
<thead>
<tr>
<th></th>
<th>obs’d</th>
<th>obs’d</th>
<th>gen’d</th>
<th>weighted penalty</th>
<th>*Contrary</th>
<th>*Step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>p</td>
<td>p</td>
<td></td>
<td>weight = 2.8</td>
<td>weight = 0.8</td>
</tr>
<tr>
<td>a.</td>
<td>LH ↔ −2</td>
<td>0</td>
<td>.000</td>
<td>.004</td>
<td>−4.4</td>
<td>−1</td>
</tr>
<tr>
<td>b.</td>
<td>LH ↔ −1</td>
<td>6</td>
<td>.007</td>
<td>.008</td>
<td>−3.6</td>
<td>−1</td>
</tr>
<tr>
<td>c.</td>
<td>LH ↔ 0</td>
<td>243</td>
<td>.303</td>
<td>.297</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>d.</td>
<td>LH ↔ 1</td>
<td>178</td>
<td>.222</td>
<td>.132</td>
<td>−0.8</td>
<td>0</td>
</tr>
<tr>
<td>e.</td>
<td>LH ↔ 2</td>
<td>40</td>
<td>.050</td>
<td>.059</td>
<td>−1.6</td>
<td>0</td>
</tr>
<tr>
<td>f.</td>
<td>HL ↔ −2</td>
<td>17</td>
<td>.021</td>
<td>.059</td>
<td>−1.6</td>
<td>0</td>
</tr>
<tr>
<td>g.</td>
<td>HL ↔ −1</td>
<td>112</td>
<td>.140</td>
<td>.132</td>
<td>−0.8</td>
<td>0</td>
</tr>
<tr>
<td>h.</td>
<td>HL ↔ 0</td>
<td>193</td>
<td>.241</td>
<td>.297</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i.</td>
<td>HL ↔ 1</td>
<td>10</td>
<td>.012</td>
<td>.008</td>
<td>−3.6</td>
<td>−1</td>
</tr>
<tr>
<td>j.</td>
<td>HL ↔ 2</td>
<td>3</td>
<td>.003</td>
<td>.004</td>
<td>−4.4</td>
<td>−1</td>
</tr>
</tbody>
</table>
example, LH being mapped onto a two-step fall in the music would receive a penalty of two. A likelihood-ratio test (Hayes et al. 2012: 712) confirms that the subset grammar with binary *Contrary alone significantly underperforms the superset grammar with binary *Contrary plus a version of *Contrary penalizing only contrary mappings of two or more musical steps ($D = 23.5, p < .0001$). Thus, scalar evaluation of *Contrary (or, alternatively, splitting up *Contrary into multiple constraints for different step sizes) is justified. One could also imagine that *Contrary might be stronger in one direction (e.g. rising tone on falling music) than the other (e.g. falling tone on rising music). A likelihood-ratio test fails to support splitting up *Contrary in this manner for Tommo So ($D = 0.9, p = .34$). We therefore retain only the unified, symmetric *Contrary.

Beyond *Contrary, which penalizes opposing mappings, we posit *Oblique, which penalizes oblique mappings as defined in §5, that is, a flat tone on rising or falling music or a non-flat tone on flat music. While oblique mappings are evidently not strongly avoided in Tommo So, they are in other languages (e.g. Cantonese, Ho 2006; Vietnamese, Kirby and Ladd 2016), justifying the constraint, and indeed the constraint receives some weight in Tommo So as well, though not nearly as much as the *Contrary family.

Finally, as discussed in §§5.1–5.5, mapping is stricter under certain conditions. Four such contexts were established, namely; (1) within words as opposed to across them, (2) for lexical tones as opposed to grammatical tones, (3) towards the end of the line, and (4) for rote as opposed to improvised material. The first two cases are dichotomous (as is roteness, but see below), and can be encoded simply as two-constraint stringency hierarchies (cf. Prince 1999, de Lacy 2004). In such a hierarchy, the stricter condition is given an index, leaving only the more general constraint to cover the remaining cases. In this case, the general constraint is *Contrary. *Contrary_{Lex} penalizes a contrary mapping iff at least one of the tones is lexical. When both tones are grammatical, only generic *Contrary applies. Otherwise, both *Contrary and *Contrary_{Lex} gang up to assign a greater penalty. Defining the stringency hierarchy this way, such that *Contrary_{Lex} but not *Contrary_{Gram} exists, implies that tone-tune mapping is universally stricter, if anything, for lexical tone than for

Figure 10: Distribution of step sizes between notes in the music, ignoring direction. Half steps, which are infrequent, are not shown.
grammatical tone. While this is the case for Tommo So, it is unclear whether it might be a universal, given that it has not been examined elsewhere. Treating *Contrary_{Lex} as a stringency hierarchy is therefore tentative.

Similarly, *Contrary_{Word} assigns a penalty only for a contrary mapping that is contained within a word or clitic group. In this case, the universality of the asymmetry between boundaries and non-boundaries, such that the latter are stricter, is more clearly warranted typologically, albeit for metrics rather than tone-tune mapping per se (see §5.4).

Positional strictness is a rather different case because strictness increases gradiently over the course of the line (§5.3). For the present purposes, we take *Contrary_{Pos} to be a scalar constraint, violated by the fraction of the way through the line in which the contrary mapping occurs. For example, if a contrary mapping starts at the middle syllable of a line, *Contrary_{Pos} receives a violation of 0.5. Other operationalizations of final strictness are conceivable, but this is not the place to explore them.

The final weights of the full model are given in (12). *Oblique was tested with the same three indices as *Contrary, but all three constraints received zero weight, and are therefore omitted. A sample tableau for <L, H> (both lexical, no boundary intervening, 10% of the way into the line) follows in (13).

\begin{verbatim}
(12)
*Contrary  1.151
*Step      0.971
*Up        0.391
*Oblique   0.306
*Contrary_{Word}  0.275
*Contrary_{Pos}   0.180
*Contrary_{Lex}   0.176
\end{verbatim}

\begin{verbatim}
(13)
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\<L, H\> & \<lex, word, 1\> & \obs'd\ & \obs'd\ & \gen'd\ & \weighted\ & \penalty\ & \*Contrary\ & \*Step\ & \*Up\ & \*Oblique\ & \*Contrary_{Word}\ & \*Contrary_{Pos}\ & \*Contrary_{Lex}\ \\
\hline
& & \N\ & \p\ & \p\ & \penalty\ & 1.151 & .971 & .391 & .306 & .275 & .180 & .176 \\
\hline
a. & \-2 & 0 & .000 & .050 & \-3.562 & \-1 & \-2 & 0 & 0 & \-1 & \-0.1 & \-1 \\
b. & \-1 & 0 & .000 & .114 & \-2.591 & \-1 & \-1 & 0 & 0 & \-1 & \-0.1 & \-1 \\
c. & 0 & 14 & .609 & .774 & \-0.306 & 0 & 0 & 0 & \-1 & 0 & 0 & 0 \\
d. & 1 & 8 & .348 & .319 & \-1.362 & 0 & \-1 & \-1 & 0 & 0 & 0 & 0 \\
e. & 2 & 1 & .043 & .141 & \-2.333 & 0 & \-2 & \-1 & 0 & 0 & 0 & 0 \\
\hline
\end{tabular}
\end{verbatim}

As this tableau shows, the model achieves a fairly good fit to the data, matching the majority of instances of lexical LH within a word in the first decile being mapped to level melody or a one-step rise. Because the constraint weights are optimized for the entire data

\textsuperscript{3}Each bigram was annotated with its position in the line in the input. As defined, *Contrary_{Pos} assigns fractional penalties. If one wanted to avoid fractions, one could just as easily take the multiplier to be, say, decile through the line rounded to the nearest integer, introducing negligible noise through rounding.
corpus, rather than this one case, the model assigns non-zero probability to falling melodies as well, though not to same the degree as their level or rising counterparts. The size of the corpus entails that each specific set of conditions will have only a small handful of tokens to represent it (in the case of the last tableau, 25); a larger corpus would be expected to more closely match the generated probabilities of the model.

While §5.1 also established that rote material is slightly stricter than improvised material, this tendency is not implemented as a constraint in the present model, which is intrinsically a model of production. As discussed in §5.1, the strictness of rote material may be due to evolutionary selection of lines. Since the present model is synchronic rather than diachronic, such community-level effects are beyond its scope.

7 Conclusion

To return to the question posed in the introduction, we have demonstrated that when speakers of Tommo So, a tonal language, go to sing, the musical melodies are constrained by linguistic tone, but with certain points of flexibility. Given that oblique mappings are only weakly penalized, singers can always sing a word on a level melodic transition, just as they can sing a level tone sequence on any melody. The dispreference for rising melodies is driven by *Up, a constraint on musical form.

We have also shown that the degree of tone-tune association in a language is not monolithic. Textsetting is influenced by a variety of intersecting grammatical and pragmatic factors, which adjust the strictness of mapping. Grammatically, we saw that mapping is stricter within words than across them, lexical tone is stricter than grammatical tone, and the ends of lines are stricter than their beginnings. Pragmatically, we find that rote lyrics display stricter mapping than improvised lyrics, and that singers differ in their rates of tone-tune mapping, though the latter factor did not reach significance in our corpus.

Though we investigated several factors here, there are likely still others at play. For example, we did not specifically test the intonational realization of tone, as advocated by Leben (1983); it may be that LH sequences in downdrift contexts are mapped more frequently to level melodies, since a downdrifted H in Tommo So is pronounced at essentially the same level as the preceding L (McPherson 2013). Nor did we test effects of higher-level phonological phrasing, which has been shown to affect strictness in metrics (e.g. Kiparsky 1975, 1977, Hayes 1989, Hayes et al. 2012). We leave these factors to future work.

Finally, we have argued that tone-tune association exhibits tighter parallels with metrics than previously realized, pointing towards a more general theory of the artistic adaptation of language. Though textsetting, and more specifically tune-setting here, is not a metrical phenomenon, we have shown that at least certain universals of metrics, such as within-word strictness and final strictness, apply also to tune-setting. Further study is needed to determine whether they are well supported as universals in this domain as well. Phonological theory has long enjoyed a close relationship with metrics, with insights traveling in both directions between them. We suggest that the aesthetic manipulation of tone holds similar promise, still largely untapped, for tonology and for a more crossmodal theory of phonological mapping.
References


**Appendix**
An elephant gave birth
There was so much colostrum
Tommo So folk song

ha koy gwë^n nà-lë^ë ém^ë ké-mïñ-jë sàm-ë-ë gwë^n nà-lë-ë

nàâ't yàâ-ndô gi-në ú-wå^ë ne gwë^n nà-lë-ë ém^ë ké-mïñ-jë sà-më-ë

á-mï-ru ge gi-në ge ne gwë^n nà-lë-mo ém^ë ké-mïñ-jë sà-më-ë

nôm-bââ yël-ëe i-rû gi-yë h-në-ë á-wë-në nô-nû yël-ëe á-rà-màâ ye

gwë^n nà-lë^ë ké-mïñ-jë sà-më-ë gwë^n nà-lë-ë á-më-ru gi-nââ
gwë^n nà-lë^ë ém^ë ké-mïñ-jë sà-mëë m-bââ yël-ëe i-rû gi-yë á-

wëë yo m-bââ yël-ëe á-rà-màâ ye gwë^n nà-lë-ë ém^ë ké-mïñ-jë sà-
mëë gwë^n nà-lë-ë tô-gë bí lu dâm-màâ é-wå gwë^n nà-
lää ᛹ kē-miṅ-jé sā-mē-yo ᛸ Yāą̣ ndō gi-nē ū-wo ne gwē

nā-lē mōn ᛹ kē-miṅ-jé sā-mē bāā Jē-bu tā-nā

ū-wo ne gwē nā-lē ᛹ kē-miṅ-jé sā-mē Min-kā-yu gi-nē

ū-wo ne gwē nā-lē ᛹ kē-miṅ-jé sā-mē mē-bāā yē-lēe i-rū gi-yē

nā-dē á-wi-nē ᛸ nō-nū yēlēe á-rā-mā ye gwē nā-lē-e

ēm kē-miṅ-jē sā-mē gwē nā-lē-e á-nām-gē dām-māi ne gwē nā-lē-e

ēm kē-miṅ-jē sā-mē-ē mē-bāā yēlēe i-rū gi-yā wē yō mē-bāā

yēlēe á-rā-mā yē
Don't Forget Us

Tommo So folk song

Bass

nàậḷ-gù ye Ám-báá Sàạ́ lú jàm le dòëẉ̊ yo yò-gó é-m-mëñ

4

nàậḷ-gù ye bàậ gî-në jàm le dòëẉ yo yò-gó é-m-mëñ nàậḷ-gù ye ú

8

bàậḷ gî-në jàm le dòëẉ yo yò-gó é-m-mëñ nàậḷ-gù ye

11

nàậḷ-gù ú bàậḷ gî-në jàm le dòëẉ yo yò-gó é-m-mëñ nàậḷ-gù ye

15

nàậḷ-gù ye Yàậ-ndó yò-gó é-m-mëñ nàậḷ-gù ye nàậḷ-gù bàậḷ gî-në

19

jam le dòëẉ... nàậḷ-gù ye sàáb Ál-là yò-gó ém-mëñ nàậḷ-gù ye

24

yò-gó ém-mëñ bàậḷ gî-në jàm le dòëẉ yo yò-gó ém-mëñ nàậḷ-gù ye

28

gù ye nàậḷ-gù ú bàậḷ gî-në jàm le dòëẉ yo yò-
The Crane
Who hears her?

Tommo So folk song

kùwⁿá á ëg - wo
á ëg - wo
kùwⁿá á ëg - wo

kùwⁿá
kùwⁿá á ëg - wo
kùwⁿá

á ëg - wo
kùwⁿá jùú le pi-yëlll - dë
á ëg - wo

wⁿá
kùwⁿá á ëg - wo
kùwⁿá
kùwⁿá

kùwⁿá

kùwⁿá
kùwⁿá kùwⁿá á ëg - wo

wⁿá á ëg - wo
kùwⁿá
kùwⁿá á ëg - wo

mòlù-gí - yé miñi bò - dë - lë à - báá mbe som - ye
á ëg - wo - c

kùwⁿá
kùwⁿá á ëg - wo
á ég - wo - c

wə́wə́ kù - wə́wə́ á ég - wo