

[Pre-print draft of March 2017; to appear in *Language*]

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

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

Laura McPherson
15 N. College St. HB 6220
Dartmouth College
Hanover, NH 03755
laura.emcpherson@gmail.com

Kevin M. Ryan
Boylston Hall 317
Harvard University
Cambridge, MA 02138
kevinryan@fas.harvard.edu

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, while 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 in common with metrics than previously recognized, setting the groundwork for a more general theory of phonological mapping across modalities.*

Keywords

Phonology, tone, textsetting, metrics, music, harmonic grammar

*We would like to thank Megan Crowhurst, Morgan Sonderegger, and an anonymous *Language* reviewer for helpful feedback on this paper, in addition to audiences at UCLA, UC San Diego, UC Berkeley, LSA 2014, and AMP 2016 for valuable discussions in earlier stages of this work. The first author gratefully acknowledges Jeffrey Heath and the financial support of NSF grant BCS-0853364 in carrying out the fieldwork for this paper. Finally, we are indebted to our consultants and singers for sharing their culture, language, and music with us. All music and lyrics remain the intellectual property of Tepama Ouologuem, Roukiatou Djeboukile, and Kounjay Ouologuem.

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. Spechtone and musical tone must be definitely correlated’ (p. 204). However, studies have shown that loss of tonal information has little effect on intelligibility, even in spoken (i.e. nonsung) language. For instance, Schirmer et al. (2005) demonstrate that changing the tone of a word is not overly disruptive to spoken word comprehension 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 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 real lyrics.

Nevertheless, many languages have been shown to match tone and melody at higher than chance levels, as we will discuss in §2. In this paper, we follow the assumption that singers may be motivated to match not by a drive for comprehension but rather by 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 multifaceted 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 the creative adaptation of language. These effects are demonstrated

with statistical analysis of the corpus and then modeled in maximum entropy Harmonic Grammar (e.g. Goldwater and Johnson 2003, Wilson 2006, Hayes and Wilson 2008), using indexed constraints in a stringency relationship. To the best of our knowledge, this represents the first generative model of the tone-tune association, despite the not insignificant number of publications on the topic.

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, among other parallels (§7.2). 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). While these parallels have been hinted at in the past by authors such as Leben (1983) and Ladd (2014), no concrete principles have been discussed. Combined with a lack of generative modeling, this has left tonal textsetting as somewhat of a side note in the linguistics literature and prevented its phonological implications from being fully appreciated.

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 in §8.

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 (Karls-son 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), Duna (Sollis 2010), Ewe (Jones 1959, Schneider 1961), Hausa (Richards 1972, Leben 1983), Shona (Schellenberg 2009), Xhosa (Starke 1930), and Zulu (Rycroft 1959, 1979). These studies are almost exclusively empirical and follow the same general procedure: A song or a corpus of songs is musically transcribed along with the lyrics, and the authors investigate whether the trajectory of the speech melody (tone) is mirrored by the sung melody (tune). The corresponding transitions between notes and tones are tallied, and the degree of tone-tune association is discussed based on these percentages.

In calculating and interpreting results, we need to distinguish between so-called ‘parallel’ and ‘nonopposing’ transitions (see e.g. Schellenberg 2012). 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. Nonopposing transitions, on the other hand, simply avoid direct clashes in direction: rising tone with falling tune, or vice versa. They are a superset of parallel transitions, also including transitions in which a level sequence is mapped to a rise or fall. Thus, nonopposing transitions define a looser metric of tone-tune association. Kirby and Ladd (2016) (following Richards 1972) explicitly define these three states of correspondence, which we follow here: parallel, opposing (cf. Richards' 'contrary'), and 'oblique' (e.g. level tonal transition on a rising melody or a rising tonal transition on a level melody).

The results of these studies are mixed, both across languages and sometimes within a single language depending upon musical genre, definitions of correspondence, or even different analyses of the same data. In terms of genre, Rycroft (1979) suggests for Southern African music that traditional genres like war chants have a greater degree of tone-tune correspondence than contemporary music like church hymns. Herzog (1934) similarly notes low levels of correspondence in Navajo healing songs but high correspondence in gambling songs, though neither is considered 'more traditional' than the other. On the other hand, Schellenberg (2009) finds no significant difference between genres in three Shona songs ranging from traditional to the national anthem. As we will see, the genre examined in this paper is traditional, consisting of women's folk songs; assuming Rycroft's hierarchy, we might expect these songs to display higher rates of tone-tune association than contemporary Tommo So songs, though we are not aware of any such genre to compare them to.

To illustrate how different analyses of the same language result in different rates of tone-tune correspondence, consider Hausa. Richards (1972) reports only 53.4% correspondence in one Hausa song, with correspondence defined as parallel (i.e. an exact match in the direction of the transition from one syllable to the next); 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 (taking into account not just the tonal categories but also the effects of downdrift and the variable pronunciation of L within a register), then correspondence becomes much more exact ('near 100%', p. 153). On the other hand, if tone-tune correspondence is defined as nonopposing rather than parallel, then even Richards' original counts arrive at 96% correspondence, demonstrating that degree of correspondence depends on definitions and 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.

Results can also differ widely even in the same linguistic family or geographic area; in other words, there do not appear to be any areal tendencies in degree of tone-tune association. This is illustrated by comparing Mandarin and Cantonese contemporary music, where the former shows little tone-tune correspondence and the latter a very high degree (Ho 2006, Schellenberg 2013). Similarly, the Shona songs discussed above display only modest rates of correspondence (Schellenberg 2009) while the associations reported for Zulu are very strict (Rycroft 1959, 1979).

Though Leben claims that his Hausa 100% result fits into what '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’ (p. 148), a partial fit may not be so surprising after all. Results from the literature show that many languages display levels of correspondence significantly greater than chance but far from a perfect fit. A selection of these results are summarized in Table 1, with many drawn from Schellenberg’s (2012) survey paper.

Language	Citation	Parallel	Nonopposing
Cantonese	Wong and Diehl (2002)	92%	98%
Duna	Sollis (2010)	66%	92%
Ewe	Jones (1959)	68%	95%
Hausa	Richards (1972)	53%	96%
Kalam Kohistani	Baart (2004)	48%	89%
Shona	Schellenberg (2009)	53%	67%
Vietnamese	Kirby and Ladd (2016)	77%	99%
Xhosa	Starke (1930)	67%	95%
Zulu	Rycroft (1959, 1979)	92%	97%

Table 1: Rates of tone-tune correspondence from a selection of published studies.

Considering these data, the literature may in fact support a middle ground as the more common tendency for tone-tune correspondence rather than Leben’s all-or-nothing dichotomy. Particularly if we look at the parallel metric, most languages match tonal and melodic movement less than 80% of the time. Writing about Navajo, Herzog (1934) states, ‘A slavish following of speech-melody by musical melody is not implied. Rather, the songs illustrate a constant conflict and accommodation between musical tendencies and the curves traced by the speech-tones of the song-text’ (p. 466). Indeed, as our model below will demonstrate, tone-tune association requires a balance between constraints on melodic form, musical interpretation of tone, and lyrical wellformedness; each language (and to some extent each performer) must assign weight to each of these constraints, with some languages favoring musical form, others tonal adherence, but most compromising on the conflicting demands.

As a kind of artistic adaptation of phonological form, tone-tune association calls to mind metrics and nontonal textsetting, but surprisingly, apart from a few nods in the literature, no studies have explicitly investigated similarities or differences between the two. Equally surprising given the rich history of generative metrics (Halle and Keyser 1966, 1971, et seq.), we are unaware of any generative models of tone-tune association. Richards (1972) develops a Markov model to compare real and simulated Hausa songs based on the expected sequences of tonal transitions and uses this model to conclude that the rate of tone-tune correspondence found in the attested data is significant; apart from this, no paper goes beyond a report of the empirical data. We therefore seek to contribute not only to the empirical study of tone-tune mapping, but also to the formalization of its grammatical properties and its parallels with metrics.

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 (1-a), while verbal morphosyntactic features are expressed by a combination of suffixes and grammatical tone patterns, illustrated in (1-b). The aspectual portmanteau suffix in (1-b) is associated with a {L} grammatical overlay on the verb stem, indicated here with a superscripted L.

- (1) a. gámmá=ge=mbe=j̀
 cat=DEF=PL=OBJ
 ‘the cats (focused object)’
 b. {pò̀-nd-ìyè-m}^L-éélè-ỳ
 fat-FACT-MP-CAUS-NEG.IMPF-1PL
 ‘we will not make (sb) fat’

The language has a seven vowel system /i, e, ε, a, ɔ, 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)	/H/	Gloss	/LH/	Gloss
	náá	‘mother’	nàá	‘cow’
	dámmá	‘village’	dàmmá	‘hoe’
	ísé	‘empty’	ìsé	‘dog’

/HL/ lexical melodies are rare in native vocabulary but fairly common in loanword vocabulary, especially loanwords from Ffulfulde, where /HL/ mimics the language’s initial stress pattern, as in (3).

- (3) a. ámìru
 ‘chief’
 b. hó̀̀l̀̀l̀̀
 ‘honor’

- c. fájìrì
‘predawn Muslim prayer’

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 (1-a) and for the latter see the final [u] in (3-a). In what follows, we label this underspecified tone type the ‘zero’ tone, labeled ‘0’.

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 /jòbó/ ‘run’ illustrates these patterns:

- | | | | |
|-----|------------------------------------|------|---------------------------------|
| (4) | Affirmative imperfective | {HL} | jó b ò ^{HL} -dè |
| | Negative imperfective | {L} | jò b ^L -éélè |
| | Affirmative defocalized perfective | {L} | jò b ^L -è |
| | Negative perfective | {L} | jò b ò ^L -lí |
| | Imperative | {H} | jó b ó ^H |

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 /bàbè/ ‘uncle’:

- (5) a. bàbè^L nó
uncle this
‘this uncle’
b. mí ^Hbábé
1SG.PRO uncle
‘my uncle’
c. bàbè únó
uncle five
‘five uncles’

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

- (6) {bàbè kòmmò ònò}^L nò=mbé
uncle skinny five this=PL
‘these five skinny uncles’

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).

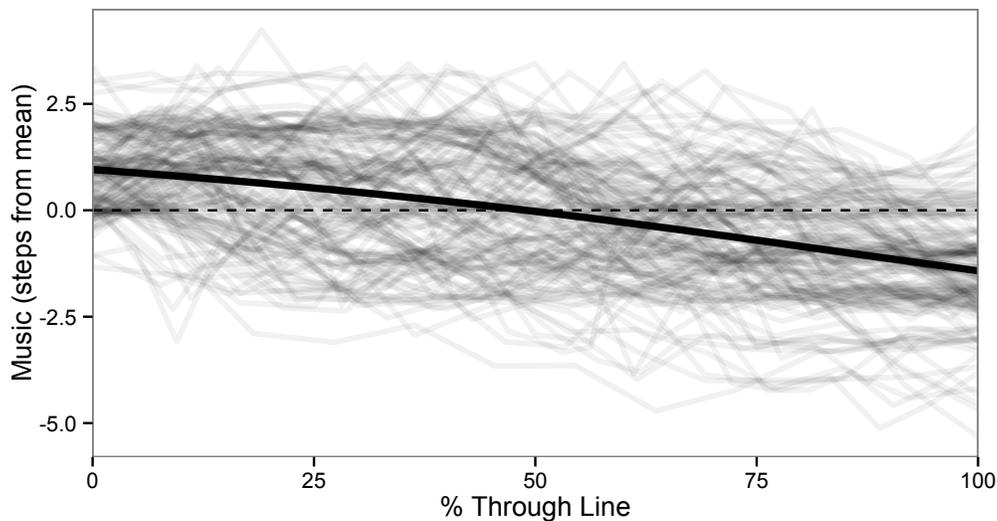


Figure 2: Musical lines tend to decline in tune. The x-axis represents percentage of the way through the line. The y-axis is the height of each note in terms of steps from its line’s mean. Each gray line is a single musical line. The dark band is a GAM smoother over all of the lines in the corpus.

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 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.

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 Djeboukile, 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 musical scores appear in the appendix to illustrate the music; the full corpus of coded bigrams is available online at <URL> and a video recording of a sample of the songs is available at <URL>. 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. The tones that were coded were either lexical tone or grammatical overlays, depending on what would surface in the given context; intonational effects on tone like declination and downstep were not coded. The example in (8) illustrates the encoding of tone in a phrase, where grammatical tones are subscripted with G.

- (8) ì rù gí yé à w-è-w =yo nò nó yè l-ée á rá -mó =ye
 L_G L_G H H L_G L_G 0 L H L H H H H 0
 breast hunger catch-PFV-2SG=if here come-NF suckle-CAUS=VOC
 ‘If you are thirsty for milk, come here and drink.’

The first word, *ìrù-gíyé* is a compound; the compound initial stem is lexically H *íru* ‘breast’ (with 0 on the epenthetic [u]), but as a compound initial it takes a {L} grammatical overlay. It is this surface grammatical tone that is coded in the corpus.

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 the strength of the juncture intervening between the two syllables, with four levels, namely, no boundary (i.e. intraword, including across affix boundaries), clitic boundary, compound boundary, and word boundary. Compound boundaries included both the juncture between two stems in a compound noun and the juncture in a Possessor+N sequence (which is tied together by grammatical tone, see (5-b)). Furthermore, many lines contain the vocable *ee* or *yee*, akin to English ‘oh’; this was treated 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 (nonconsecutive in the song) from ‘An elephant gave birth’ in (9), in which improvisation is bolded.

- (9) a. Verse 1 (Tepama)
gwéⁿ nàl-è^l èm^l kémíjé sàm-è^l, gwéⁿ nàl-è^l
‘An elephant gave birth; there was so much colostrum; an elephant gave birth’
Yààndó gíné úwɔ=ne gwéⁿ nàl-è^l, èm^l kémíjé sàm-è^l
‘**Yààndó, in your house** an elephant gave birth; there was so much colostrum’
ámíru=ge ^lgìnè=**ge=ne** gwéⁿ nàl-è^l, èm^l kémíjé sàm-è^l
‘**In the chief’s house** an elephant gave birth; there was so much colostrum’
- b. Verse 2 (Roukiatou)
gwéⁿ nàl-è^l èm^l kémíjé sàm-è^l, gwéⁿ nàl-è^l
‘An elephant gave birth; there was so much colostrum; an elephant gave birth’
àbáá=mbe ^ldàmmà gwéⁿ nàl-è^l èm^l kémíjé sàm-è^l
‘**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 nonfinal, allowing us to easily identify the ends of lines. 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 replaceive 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. If a lexical tone was overwritten with the same grammatical tone, it was still coded as grammatical.

5 Results

5.1 Basic results

We first discuss the basic results of tone-tune alignment in the corpus before analyzing how various factors affect these rates of alignment in §5.2. A summary of empirical results can be found in §5.3. 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* if 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*.

		musical sequence				
		<i>2+ up</i>	<i>1 up</i>	<i>same</i>	<i>1 down</i>	<i>2+ down</i>
tone	<i>up</i> (LH)	40 (8%)	178 (35%)	243 (48%)	46 (9%)	0 (0%)
sequence	<i>same</i> (LL or HH)	7 (1%)	97 (13%)	404 (52%)	238 (31%)	27 (3%)
	<i>down</i> (HL)	3 (1%)	10 (3%)	193 (58%)	112 (33%)	17 (5%)

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

Tommo So singers show a significant tendency to avoid contrary mappings between tonal and musical contours. For simplicity, we put ‘zero’ tones (see §3.1) aside for the moment to consider only the four fully specified tonal transitions, namely, LH, LL, HH, and HL. Of 1,615 such transitions in our corpus, 751 (46.5%) are parallel, 805 (49.8%) oblique, and 59 (3.7%) contrary. Full counts are given in Table 2 and plotted in Figure 3. These results show that while contrary mappings are strongly (though not categorically) avoided, oblique mappings are relatively tolerated, the latter being approximately as frequent as parallel mappings.

Moreover, contrary mappings are avoided more stringently for larger musical steps than smaller ones. 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 Table 2) is 5.2 (290 to 56). For 2+ steps (the most peripheral columns), the ratio is significantly greater, at 19.0 (57 to 3; Fisher’s exact test: odds ratio 3.7, $p = .03$).

Next, we turn to the grouping of tonal bigrams. Inspection of Figure 3 suggests that flat tones (middle column) are treated more like falling tones (rightmost column) than rising tones (leftmost column). This asymmetry is significant. For flat vs. rising, $\chi^2(4) = 197$ (a larger χ^2 indicates a greater discrepancy), while for flat vs. falling, $\chi^2(4) = 25$, and the difference between the two is significant (Cochran-Mantel-Haenszel test for two 2×5 tables: $M^2(4) = 90.9$, $p < .0001$; Agresti 2002). In this sense, then, the most important

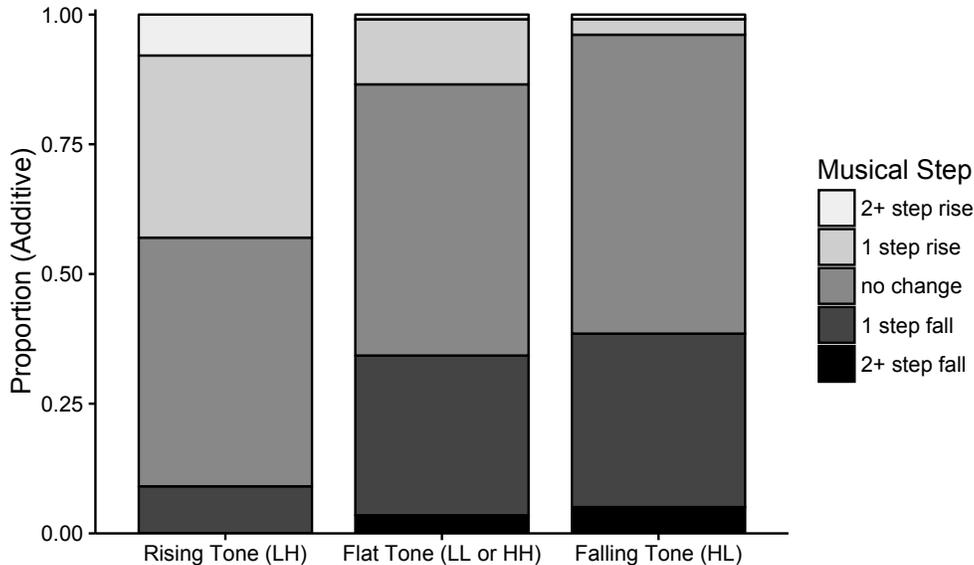


Figure 3: Mapping of fully specified two-tone sequences to their respective musical notes. Musical sequences are coded for both polarity (rising or falling) and degree (one step or more than one step, rounding half steps to the nearest more peripheral whole).

tonal opposition in Tommo So textsetting is not falling vs. rising, but nonrising vs. rising, as reinforced further below by cluster analysis of all transitions.

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

		musical sequence				
		<i>2+ up</i>	<i>1 up</i>	<i>same</i>	<i>1 down</i>	<i>2+ down</i>
tone	LH	40 (8%)	178 (35%)	243 (48%)	46 (9%)	0 (0%)
	0H	4 (9%)	23 (50%)	10 (22%)	8 (17%)	1 (2%)
	00	0 (0%)	2 (4%)	21 (41%)	24 (47%)	4 (8%)
	0L	2 (1%)	23 (14%)	81 (50%)	47 (29%)	10 (6%)
	LL	5 (1%)	70 (16%)	197 (46%)	141 (33%)	19 (4%)
	HH	2 (1%)	27 (8%)	207 (61%)	97 (28%)	8 (2%)
	L0	0 (0%)	1 (1%)	63 (85%)	9 (12%)	1 (1%)
	H0	1 (0%)	6 (3%)	124 (53%)	96 (41%)	7 (3%)
	HL	3 (1%)	10 (3%)	193 (58%)	112 (33%)	17 (5%)

Table 3: Count data upon which Figure 4 is based. Percentages are by row.

Returning to the question of subgrouping based on this full set of transitions, as is appar-

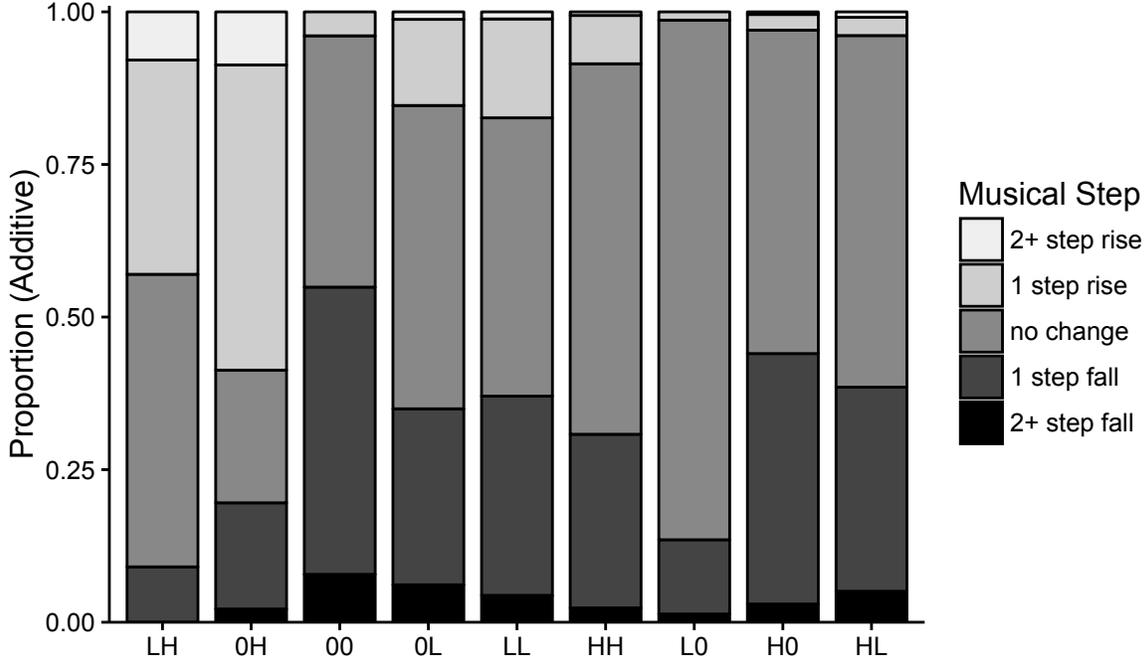


Figure 4: Mapping of all two-tone sequences to their respective musical notes (cf. Figure 3).

ent from Figure 4, musical rises are strongly skewed towards two tone sequences, namely, LH and OH, which together pattern as RISING tones. All other tone sequences, including falling and flat tones, comprise the NONRISING condition. This bifurcation is further supported by cluster analysis of the raw data in Table 3. Four dendrograms are shown in Figure 5, corresponding to four distance metrics (Rousseeuw and Kaufman 1990), namely, Fisher’s exact test (using $\log(1/p)$), Euclidean distance (i.e. $\sqrt{\sum(q_i - p_i)^2}$), Pearson’s correlation r^2 , and Spearman’s correlation ρ . Euclidean distance is based on the percentages in Table 3; the other metrics use counts. For each metric, a matrix containing the distance measure of every pairing of rows in Figure 5 was analyzed by AGNES (with default settings), an R method for agglomerative clustering (*ibid.*, §6).³ In all four diagrams, the first division separates LH and OH — the rising tones — from all other tones. Furthermore, except for the isolated case of L0 on the Euclidean metric, the distance between the rising and nonrising groups exceeds the heterogeneity within those groups, as indicated by height on the y-axis.

Cluster analysis therefore supports grouping LH and OH as RISING, to the exclusion of all other tone sequences, which are NONRISING. That said, however, given that 0 is filled in by interpolation in the spoken language, one might expect the treatment of OH (and other

³Divisive clustering (DIANA) was also tested, but did not differ qualitatively from the results for AGNES for any metric. Both parametric and nonparametric correlations are shown in Figure 5. On the one hand, the underlying distributions are borderline nonnormal, favoring a nonparametric test. On the other hand, because there are only five levels of musical change and two are sparsely populated, a rank-order correlation has little power compared to the other methods, as can be seen in the ties for Spearman’s ρ . Since the methods agree on the point at hand, however, there is no need to consider parametricity further.

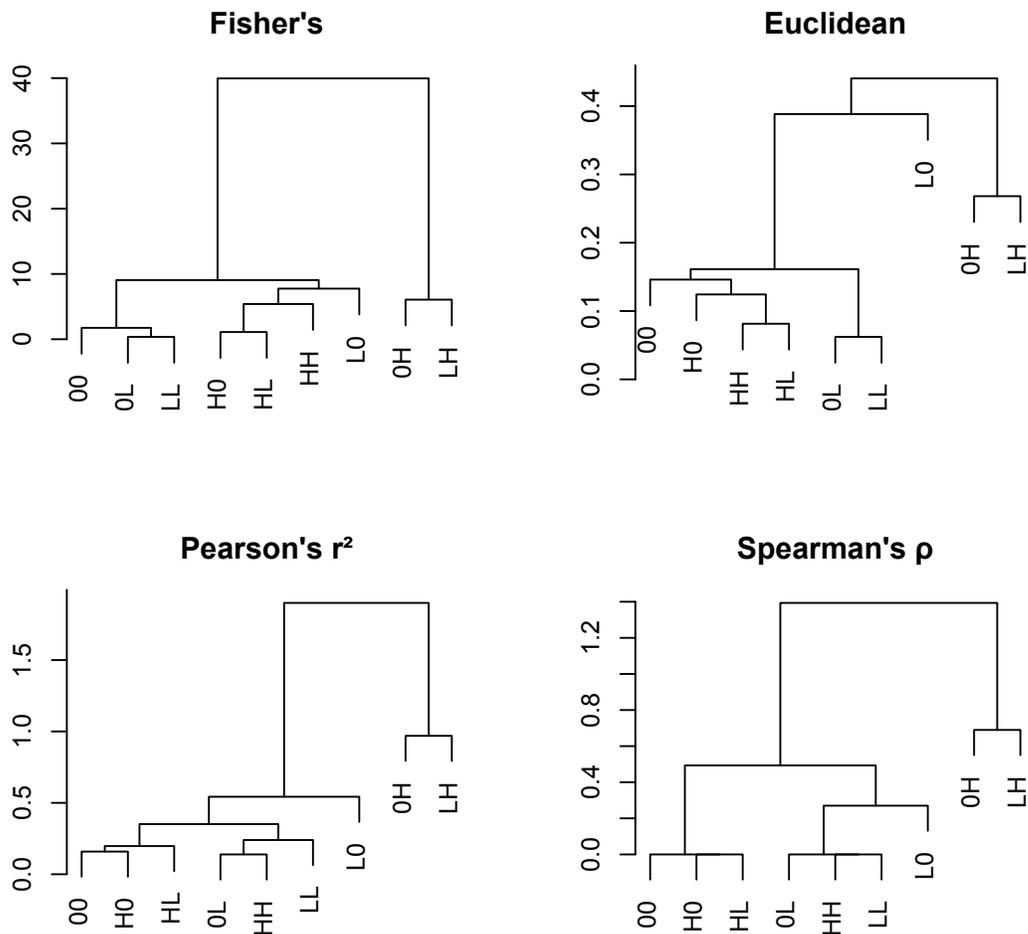


Figure 5: Dendrograms showing the clustering of the nine tonal sequences with respect to their treatment in textsetting.

sequences containing 0) to be affected by tonal context. Indeed, a more detailed, context-sensitive cluster analysis reveals that OH behaves like LH only when the former follows L (notated L-OH). H-OH, by contrast, is treated as nonrising. Dendrograms are provided in Figure 6. Whenever a zero tone occurs, it is given with its context, including both sides for the bigram 00. ‘B’ stands for line-final position, for which there is no following tone. Only sequences attested ten or more times are shown, as rare sequences can be less reliable for grouping analysis (e.g. 0-00-0 is attested only once, and therefore excluded). Other possible sequences, such as 0-00-H, are unattested altogether.

This context-sensitive analysis of zeroes reveals that interpolation affects tone-tune mapping. As hypothesized, L-OH behaves like LH. Interestingly, L-OL (but not e.g. H-OH) also patterns as rising, rather than flat, though this may be due to a confound with boundary

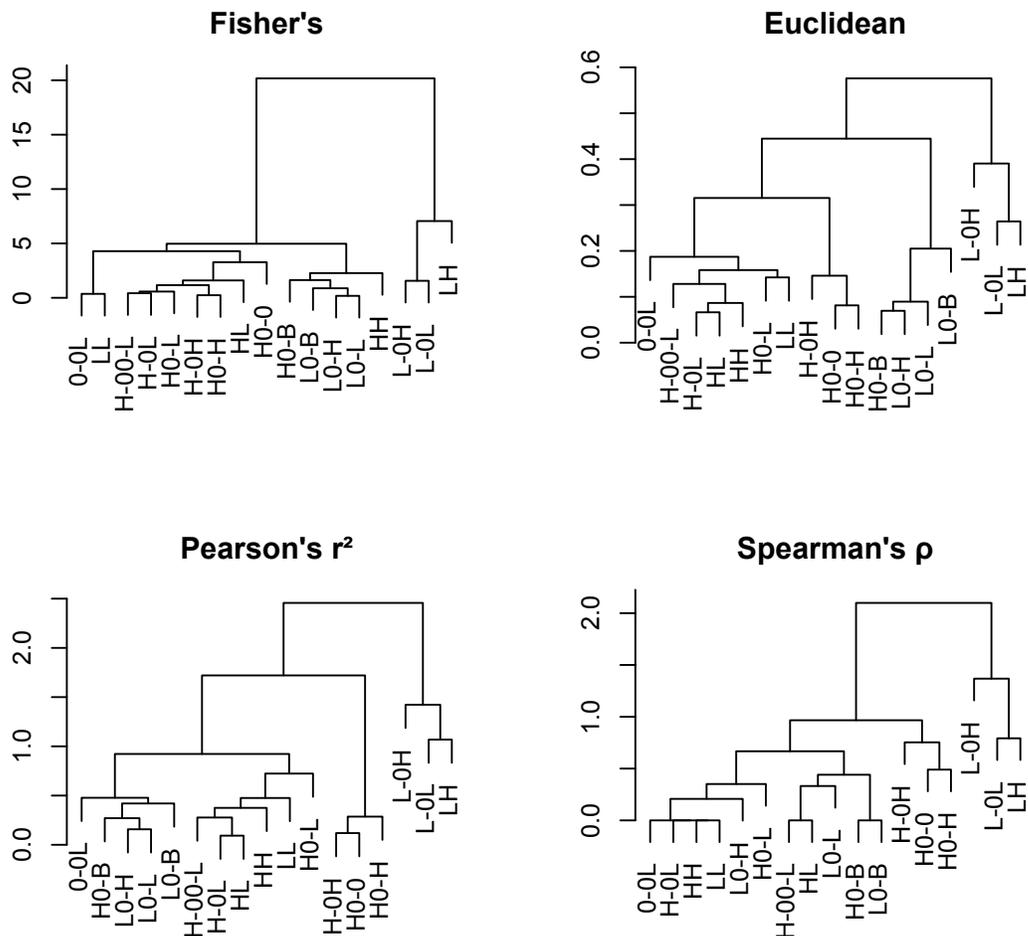


Figure 6: Dendrograms for bigrams including immediate tonal contexts for any zero tones. For example, L-0H refers to a 0H sequence preceded by an L.

distribution.⁴ In what follows, we treat LH and L-0H as RISING and all of the tone sequences specified in the left clusters in Figure 6 as NONRISING. L-0L and sequences that were excluded from Figure 6 due to rarity (e.g. 0-00-0) are put aside as indeterminate. These exclusions affect only 67 bigram tokens, or about 3% of the original data. Figure 7 illustrates the aggregate alignments of the two categories as just defined.

⁴Visual inspection of the data reveals that a disproportionate number of the 28 L-0L tokens involve a line-internal phrase boundary between the 0 and the L, which we hypothesize to allow some degree of pitch reset. As the present analysis does not code for phrase boundaries, we leave deeper investigation of this behavior to future work, and put L-0L aside in the subsequent analyses.

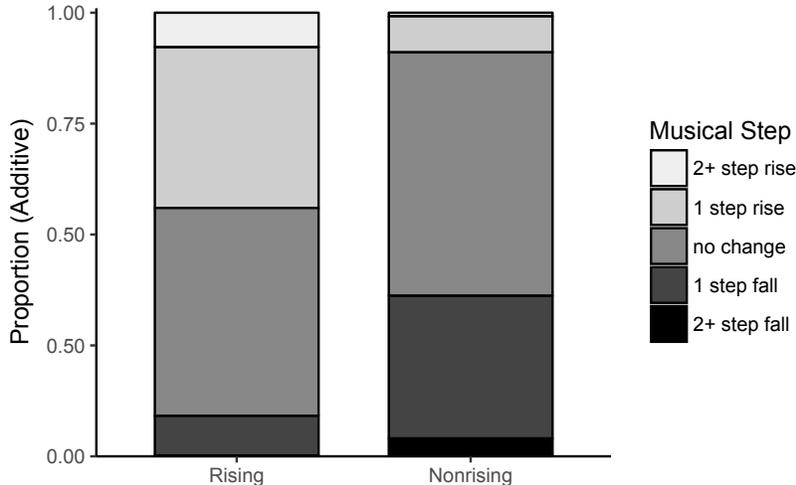


Figure 7: Figure 4 collapsed into two overarching groups.

5.2 Factors modulating strictness

We now turn to factors modulating the strictness of mapping, namely, position in the line, juncture level, compositional status (improvised vs. rote), morphosyntactic source of the tone (lexical vs. grammatical), and singer identity. Binomial logistic regression was employed to gauge the contributions of these predictors. The dependent variable is taken to be whether each transition is parallel (coded as 1) or contrary (coded as 0), such that factors favoring strictness have positive coefficients. The definitions of parallel and contrary reflect the discussion in §5.1. Specifically, we divide tonal transitions into two groups, rising and nonrising. Rising transitions include LH and L-0H (i.e. 0H immediately preceded by L). All other transitions are nonrising, except for certain infrequent transitions that were excluded altogether because their classification was indeterminate.⁵ The motivation for taking nonrising as opposed to falling is that falling and flat tonal transitions behave almost indistinguishably in textsetting (§5.1). Parallel and contrary are then defined as in Table 4, which also gives the frequency of each cell. In total, 997 transitions are evaluated by the model, 807 parallel and 190 contrary.

	Rising Music	Falling Music
Rising Tone	Parallel (232)	Contrary (48)
Nonrising Tone	Contrary (142)	Parallel (575)

Table 4

The resulting regression table is given in Table 5. The column labeled $\hat{\beta}$ shows the coefficient estimate of each predictor. If positive, the predictor favors greater strictness. If

⁵The following bigram types were excluded, amounting to 67 tokens in total: L-0L, 0-0H, H-00-B, H-00-H, L-00-H, L0-0, 0-00-0, 0-00-B, B-00-H, B-0H, H-00-0. Nonrising transitions include LL, HH, HL, H-0L, H0-L, H0-B, H0-0, 0-0L, H-00-L, L0-L, L0-H, L0-B, H0-H, and H-0H, amounting to 1,589 tokens in total.

the p -value (rightmost column) is less than 0.05, the effect is significant. More generally, the model makes predictions according to the formula in (10).

$$(10) \quad Pr(y_i \text{ is parallel}) = \text{logit}^{-1}(\beta^0 + X_i^1 \cdot \beta^1 + X_i^2 \cdot \beta^2 + \dots + X_i^n \cdot \beta^n)$$

For any transition y_i , the probability that y_i is parallel as opposed to contrary is given by the inverse logit (i.e. $1/(1 + e^{-x})$) of the weighted sum of the intercept and applicable predictors. X_i^j is the value of transition y_i for predictor j . For example, if y_i occurs in an improvised context, X_i is 1 for ‘Composition = Improvised’ and 0 for ‘Composition = Mixed’. The only numerical predictor is ‘Position in the Line’, which is the proportion of the way through the line that y_i occurs. For example, if the token occurs 75% of the way into its line, X_i is 0.75 for position.

	N	$\hat{\beta}$	SE	z	p
(Intercept)	—	1.264	0.182	6.95	< .0001 *
Position in Line	—	1.241	0.313	3.97	< .0001 *
Juncture = Word	286	-0.725	0.218	-3.33	.001 *
Juncture = Clitic	72	0.316	0.407	0.78	.437
Juncture = Compound	80	0.642	0.488	1.32	.188
Composition = Improvised	149	-0.582	0.224	-2.60	.009 *
Composition = Mixed	39	-0.353	0.383	-0.92	.357
LexGram = Grammatical	184	-0.457	0.223	-2.05	.040 *
LexGram = Mixed	188	0.347	0.285	1.22	.223

Table 5: Regression table for tone-tune alignment. For factors, N gives the number of transitions in each condition. $\hat{\beta}$ is the coefficient estimate (positive $\hat{\beta} \Rightarrow$ greater strictness); SE is $\hat{\beta}$ ’s standard error; z is $\hat{\beta}/\text{SE}$; and p indicates significance (* if $p < 0.05$).

The four predictors are now considered in turn (followed by the effect of speaker identity at the end of this section).⁶ First, position in the line is significant with a positive coefficient, meaning that mapping is stricter towards the end of the line. This accords with a putative universal of metrics, whereby endings are (if anything) stricter than beginnings (see Hayes 1983:373 for an overview and Chen 1979 for a case from tone mapping in Chinese verse), though it has not been previously documented for tone-tune mapping to our knowledge. The empirical increase in strictness over the course of the line is depicted in Figure 8. The dark line is a LOESS smoother indicating the proportion of transitions that are parallel as opposed to contrary at any point in the line, and the gray band is its 95% confidence interval. Strictness increases from approximately 72% to 88% from beginning to end: Contrary mappings are more than halved.

⁶Given that bigrams overlap (e.g. a tone sequence HLH comprises the bigrams HL and LH), tonal bigrams are not independently distributed. To ensure that this lack of independence was not substantively affecting any of the results in Table 5, the regression was also tested using only even-numbered bigrams as data. Word juncture was the only factor that was significant in Table 5 but lost significance in this $\sim 50\%$ smaller corpus; however, it retained its significance in the odd-parity half of the data.

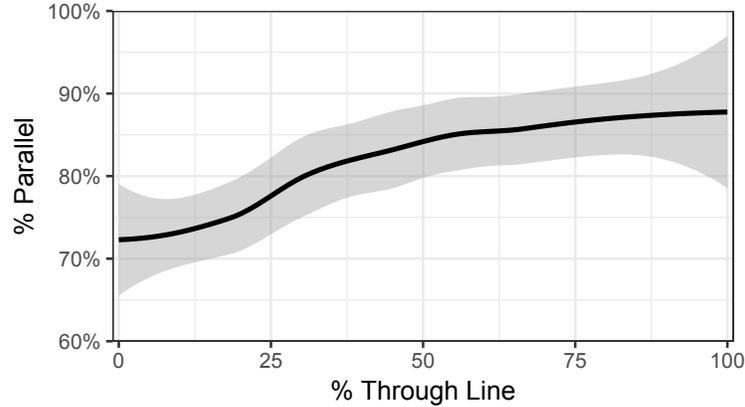


Figure 8: Increasing strictness across the line.

Independent of increasing strictness, lines tend to become musically flatter towards the end. As illustrated by Figure 9, the proportion of transitions that are nonflat steadily descends in the second half of the line. However, this musical trend is not driving the increase in strictness observed in Figure 8. Figure 8 and the regression analysis (Table 5) take only transitions in which the music changes, excluding musically flat sequences.

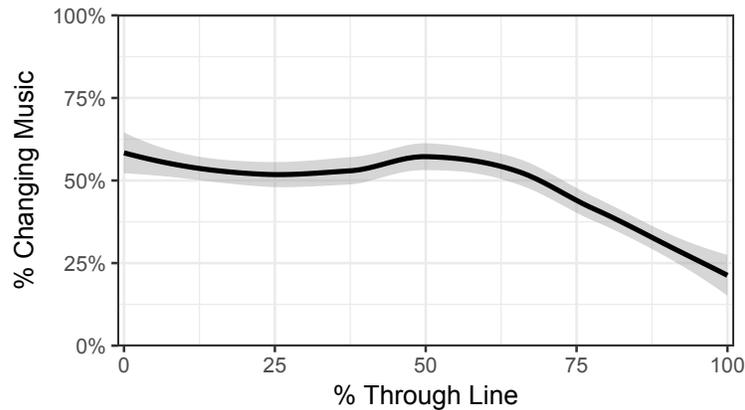


Figure 9: Changes in the music become less likely towards the end of the line.

Second, juncture level affects strictness of mapping. As mentioned in §4.3, four levels of juncture are operationalized, namely, none (i.e. the two syllables are contained within a word), clitic, compound, and word. The coefficient for word-level juncture in Table 5 is negative and significant, meaning that mapping is significantly stricter within words (the baseline level) than between them. The same tendency is observed in metrics, as in English, where stress-meter mapping is stricter within tighter phrasal constituents (Magnuson and Ryder 1970, Kiparsky 1975, Hayes et al. 2012). Meanwhile, clitic and compound boundaries are not significantly different from no boundary. A Tukey’s HSD test reveals that of six possible pairwise combinations of these four levels, three are significant: Intraword, clitic, and compound are all significantly stricter than cross-word. No other pair (e.g. clitic vs. compound)

was significant. Thus, mapping is stricter within clitic groups (including compounds) than it is across them.

Third, rote material is stricter than improvised material (on this distinction, see §4.4). In Table 5, ‘Improved’ refers only to bigrams in which both syllables are part of an improvised section, and ‘Mixed’ refers to bigrams in which one syllable is part of an improvised section and the other is part of a rote section. As one might expect, bigrams of mixed status are intermediate in strictness between fully improvised and fully rote bigrams, though not significantly different from either. But completely rote bigrams are significantly stricter than completely improvised ones.

Fourth, mapping is stricter for lexical tone than it is for grammatical tone. Recall that grammatical tone occurs when the surface tone of a syllable is determined by a grammatical overlay rather than by the syllable’s underlying tone (§4.1). Specifically, bigrams in which both tones are grammatical (‘GG’) are significantly worse aligned than bigrams in which both tones are lexical (‘LL’) and bigrams of mixed status (‘LG’ and ‘GL’) (the latter confirmed by Tukey’s HSD test). Meanwhile, mixed and purely lexical bigrams are not significantly different from each other. In short, {LL, GL, LG} is stricter than GG.

An anonymous reviewer asks whether the underlying tone might still affect mapping even when it is overwritten by a grammatical overlay. In other words, perhaps the reason that GG appears to be less strict is that there is some tendency to match the underlying as opposed to surface tones in such bigrams. To test this, we take all bigrams in which (1) both tones are grammatical, (2) the music is nonflat (as before), (3) the underlying tone sequence is one of {LL, LH, HL, HH}, and (4) the surface tone sequence is one of {LL, LH, HL, HH}. 175 bigrams meet these criteria. As before, tone sequences are classified as rising (LH) and nonrising (LL, HH, HL). The question is then whether the musical changes in these bigrams are determined by surface tone, underlying tone, both, or neither. Logistic regression suggests that both matter. The binary outcome is whether the music is rising (1) or falling (0). Two binary predictors are employed, one for surface tone (rising or nonrising) and one for underlying tone (rising or nonrising). As revealed in Table 6, both predictors are significant.⁷

	N	$\hat{\beta}$	SE	z	p
(Intercept)	—	-0.982	0.194	-5.05	< .0001 *
scale(Underlying Tone = Rising)	48	0.465	0.177	2.62	.009 *
scale(Surface Tone = Rising)	17	0.988	0.234	4.22	< .0001 *

Table 6: Regression table showing that both underlying and surface tones affect mapping in grammatical overlay contexts.

The predictors in Table 6 are scaled (i.e. transformed to standard deviations from the mean) so that their effect sizes can be meaningfully compared. This comparison suggests that surface tone is aggregately about twice as important as underlying tone. For example,

⁷The interaction of these two predictors was also tested. It was not significant, but only six tokens were rising on both predictors (Table 7).

when surface and underlying tone sequences make conflicting demands (i.e. one is rising and the other is nonrising), the music is over twice as likely to agree with the former (of 53 such bigrams, 36 musical changes agree with the surface tone and the remaining 17 agree with the underlying tone). But the regression also suggests that underlying tone is not inert. This can be seen when surface tone is controlled, as in Table 7. For any given surface tone, the likelihood of rising music increases if the underlying tone is rising. We will discuss these findings further in §7.

Surface tone sequence	Underlying tone sequence	% of musical changes that are rises	Out of N total bigrams
Nonrising	Nonrising	17%	116
Nonrising	Rising	36%	42
Rising	Nonrising	82%	11
Rising	Rising	100%	6

Table 7: Effects of underlying tone on the music when surface tone is controlled.

Finally, we tested whether singers differ significantly in their strictness of tone-tune alignment. The present corpus comprises data from three singers, labeled R (25% of the corpus), K (20%), and T (20%), as well as pairs of these three, KT (13%), RT (12%), and KR (11%). Figure 10 shows the percentage of contrary as opposed to parallel mappings for each individual singer (putting aside pairs), based on the subset of data used for the regression above in Table 5. Within each speaker, the data are further separated into rote and improvised, putting aside mixed rote-improvised bigrams.

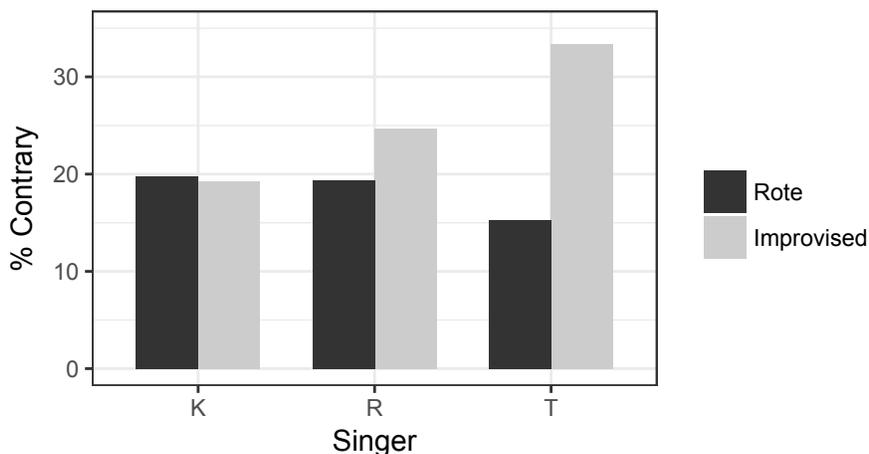


Figure 10: Percentages of contrary alignments for three singers, separated by rote vs. improvised for each singer.

As discussed above, contrary mappings are more frequent under improvisation. As Figure 10 reveals, this generalization holds for two of three speakers individually, though K’s alignment is approximately the same for rote and improvised material, and she is likewise

the strictest singer under improvisation.⁸ Nevertheless, singer identity is not significant in our data. When ‘Singer’ (with all six levels) is added as a predictor to the logistic model in Table 5, the AIC (Akaike information criterion) becomes worse (that is, higher): 941.6 with singer vs. 936.6 without. The same is true if ‘Singer’ is included as both a main effect and as an interaction with ‘Composition’ (AIC = 949.7). In either case, none of the levels of singer (or the singer/composition interaction) reaches significance, and none of the other predictors is qualitatively affected. The conclusion is the same if singer identity is included in the model as random intercepts. However, improvised material is underrepresented in the corpus, being 12% of our data, and therefore speaker differences (which emerge mainly in improvised material) have little opportunity to reach significance.

5.3 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.
- Contrary mappings are avoided more stringently for larger musical steps.
- The most significant bifurcation in the behavior of tonal sequences is that between rising sequences (LH and 0H preceded by L) and all other, nonrising sequences.
- Mapping is stricter in rote material than improvised material.
- 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.
- Mapping is stricter towards the end of the line.
- Mapping is stricter within words (and clitic groups) than across them.
- Mapping is stricter for lexical tone than grammatical tone.
- Even when the surface tone reflects a grammatical overlay, the underlying tone exerts some influence on mapping preferences.

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 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

⁸One consultant fervently asserted that K is considered an excellent singer. We speculate that this perceived talent may be partially based on the ability to align tone and tune when singing extemporaneously.

West African tonal textsetting to determine whether we could consider this to be 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 been explicitly tested. 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, though this convergence has not been previously observed to our knowledge: Mapping is stricter at the ends of lines and stricter within words than across them. As discussed in the next two sections, 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 §7, where we discuss further implications of our findings.

6 Constraint-based analysis

This section formalizes a grammatical model for tone-tune association. 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 (\mathcal{H}) is the sum of its weighted constraint violations. (Here, violations are taken to be negative, so \mathcal{H} is nonpositive.) In classical HG, the candidate with the highest harmony wins. Maxent HG takes the process one step further: Harmonies are converted to probabilities by $e^{\mathcal{H}_0} / \sum_i e^{\mathcal{H}_i}$, where \mathcal{H}_0 is a candidate’s nonpositive harmony and the sum \sum_i ranges over all candidates (e.g. Goldwater and Johnson 2003, Hayes and Wilson 2008), as will be exemplified in the tableaux below. Maxent HG has been employed for metrics by Ryan (2011), Hayes and Moore-Cantwell (2011), and Hayes et al. (2012). As will be further discussed at the end of this section, maxent HG is closely related to regression models such as the one presented in §5.2.

Given the focus of this article, we operationalize the grammar as evaluating mappings

between tonal bigrams and musical transitions. The question of directionality (tone \Rightarrow tune, tune \Rightarrow tone, or tone \Leftrightarrow tune) will be discussed further in §7.4, where it will be shown that the mapping constraints defined here are compatible with any directionality. For now, we take the input to be linguistic material — a tonal bigram — and the output to be a musical transition. A tonal bigram is $\{L, H, 0\}$, $\{L, H, 0\}$, annotated with any relevant linguistic information (boundary level, 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.

We define the following constraints. First, *STEP penalizes a change in music according to its degree in steps. For example, flat music receives zero violations, while a musical change of two steps up or down receives two violations. As Figure 11 shows, about half of musical transitions are flat, and fewer than 10% are multiple steps.

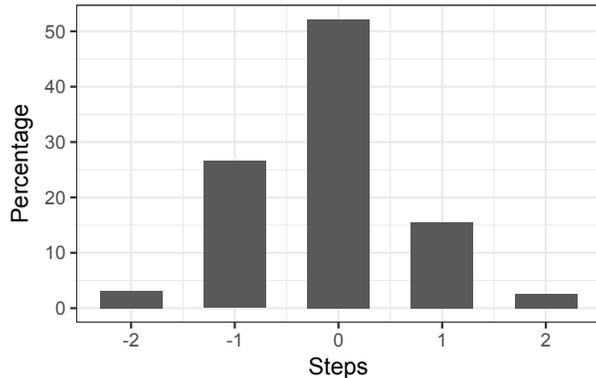


Figure 11: Distribution of musical step sizes in the corpus.

Second, *UP penalizes any rise in the music, reflecting the tendency of musical lines to fall (an average of 2.2 steps per line in our corpus), as discussed in §3.2 and also evident from Figure 11. Typologically, falling music may be 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, whose ratio remains roughly constant across the line; see Figure 12. Therefore, *UP, a constraint on the music, is justified.

These two constraints are the only purely musical constraints in the present analysis. A more sophisticated model of Tommo So music (as independent from textsetting) is a desideratum, but beyond the scope of the present article, which treats only tone-tune mapping. Furthermore, a musical model would require constraints on sequences larger than bigrams, which we do not analyze here.

The remaining constraints concern the mapping between linguistic tone and musical tune. *CONTRARY penalizes falling music on a rising tone sequence (which in Tommo So includes 0H after L; see §5.1) and rising music on a falling tone (including H0 before L). A contrary

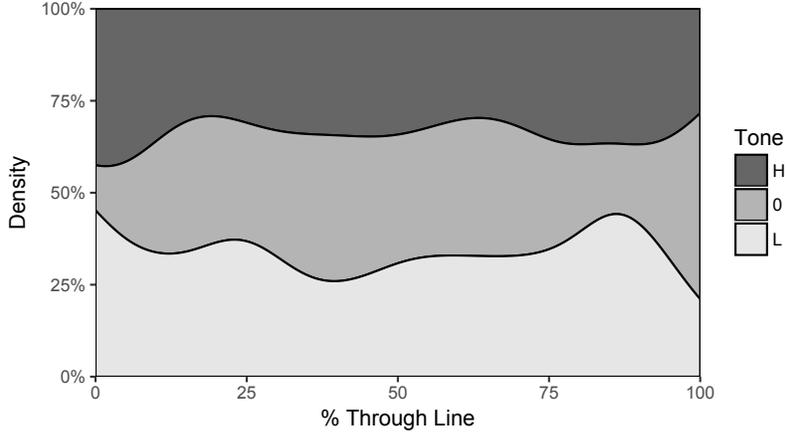


Figure 12: Distribution of high, zero, and low tones across the musical line.

mapping is penalized according to its degree in musical steps (recall from §5.1 that mapping is stricter for larger steps). For example, if a rising tone accompanies a musical change of two notes down, *CONTRARY receives two violations, one for each step. We also tried splitting *CONTRARY into two constraints, one for single steps and one for multiple steps, but this two-constraint model was not a significant improvement over the model with one scalar constraint. Specifically, AIC_C (the corrected Akaike information criterion)⁹ is 5,127.5 for the one-constraint model and 5,128.8, or 1.3 higher (worse), for the two-constraint model. As a rule of thumb, a more complex model is justified only if it reduces AIC_C by at least two (Burnham and Anderson 2004). One could also imagine that *CONTRARY might be stronger in one direction (rising tone on falling music) than the other (falling tone on rising music). But this split was also not a significant improvement, with $AIC_C = 5,128.7$. We therefore retain the monolithic version of *CONTRARY, as in (11).

- (11) *CONTRARY: Penalize each contrary mapping by the absolute size of the interval separating the two notes. A contrary mapping is defined as falling music on a rising tonal transition or rising music on a falling tonal transition.

In addition to generic *CONTRARY, mapping is stricter under certain conditions. Four strictness-favoring contexts were established in §5.2, namely, (1) within words and clitic groups as opposed to across them, (2) lexical as opposed to grammatical tones, (3) rote as opposed to improvised compositions, and (4) endings of lines. We consider each in turn. First, *CONTRARY_{CG} in (12) has the same definition as *CONTRARY in (11), except that it applies only when the two syllables are contained within a clitic group (CG), which we take to subsume the word (including compounds) as well as any affixes or clitics. In §5.2, it was reported that no boundary, the clitic boundary, and the compound boundary were not significantly different from each other in terms of strictness, while all three were significantly

⁹ $AIC_C = 2p - 2L + \frac{2p(p+1)}{n-p-1}$, where p is the number of constraints, n is the number of data points, and L is the log-likelihood of the data given the model (Burnham and Anderson 2004). Because n is 2,183 here, the correction term is negligible.

stricter than the full prosodic word boundary. However, we leave open the possibility that other languages (or a larger corpus of Tommo So) might reveal a more articulated hierarchy of boundary sensitivity.

- (12) *CONTRARY_{CG}: Same as *CONTRARY in (11), except evaluated only if the transition is contained within the same clitic group.

Note that *CONTRARY and *CONTRARY_{CG} comprise a stringency hierarchy (cf. Prince 1999, de Lacy 2004): When the latter is violated, the former is also violated, but not vice versa. This means that under any ranking or weighting, mapping is at least as strict within clitic groups as it is between them. If this turns out not to be a universal, the definitions would have to be adjusted, but available evidence suggests that it may well be a universal (e.g. it is found also in English metrics, as noted in §5.2).

Second, mapping is stricter for lexical tone than for grammatical tone. In the latter case, a grammatical tonal overlay overwrites the underlying, lexical tone.¹⁰ As discussed in §5.2, grammatical tone may appear to be less strict because singers continue to show some sensitivity to the underlying tone. For example, if a bigram has a flat surface tone due to a grammatical overlay, it is more likely to be set to rising music if the underlying tone is rising. Therefore, we define *CONTRARY_{Lex} as being identical to *CONTRARY except that mapping is evaluated with respect to the underlying tone. Outside of the context of grammatical tone, *CONTRARY and *CONTRARY_{Lex} incur the same violations, because the underlying and surface tones are the same. See §7.1–7.2 below for further discussion and (15) for a tableau example.

- (13) *CONTRARY_{Lex}: Same as *CONTRARY in (11), except evaluated with respect to the underlying rather than surface tones.

While §5.2 also established that rote material is stricter than improvised material, this tendency is not implemented as a constraint in the present model, which is intrinsically synchronic, that is, a production and/or acceptability grammar. As mentioned in §5.2, the strictness of rote material may be due to evolutionary selection of lines across multiple sittings, singers, and possibly generations. Since the present model is not diachronic, such community-level selection effects are beyond its scope.

Finally, mapping becomes stricter towards the ends of lines (see Figure 8 in §5.2). *CONTRARY might be modified in various ways to implement this tendency. First, for every locus of violation, the penalty could be multiplied by the absolute position in the line (say, in syllables; e.g. five violations for a one-step contrary mapping starting with the fifth syllable). Second, and similarly, the multiplier could be the proportion of the way through the line (e.g. 0.5 for a single violation that occurs half of the way through the line). These two

¹⁰Tommo So does not permit us to test whether a level intermediate between underlying tone and grammatical tone is relevant for mapping. There are only two levels to consider, namely, the grammatical tone, if any, and the lexical tone, which is what would surface if no grammatical overlay applied. In other languages, the derivational picture could conceivably be more complex. See, for example, the cases of Seenku and Kalevala Finnish in §7.2.

approaches are both scalar multiplier approaches (cf. distance effects in harmony, e.g. Zymet 2015, McPherson and Hayes 2016); a nonscalar alternative will be discussed in the next paragraph. The absolute and proportional multipliers make different predictions. For example, the absolute version predicts that longer lines will tend to have stricter endings than shorter lines. As Figure 13 reveals, this is not the case: Long and short lines exhibit strictness increases of roughly the same magnitude (if anything, shorter lines are stricter). Therefore, a proportional multiplier is more appropriate. But this would be an unorthodox constraint, since its violations are fractional (cf. Flemming 2001). Furthermore, even putting aside the issue of fractional violations, counting gradience is sometimes argued to be unnecessary, at least in Optimality Theory (McCarthy 2003).¹¹

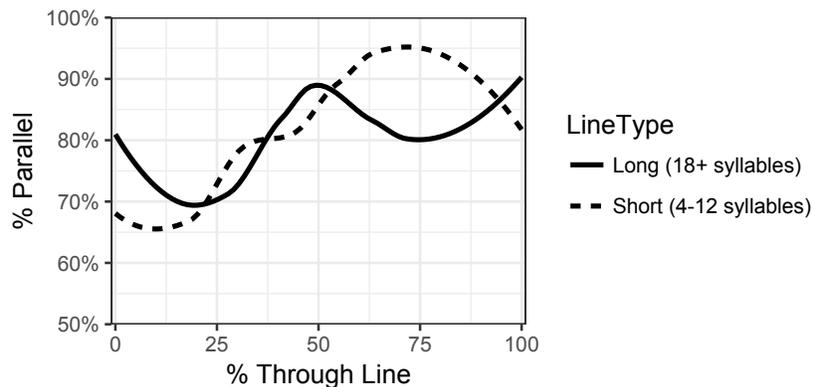


Figure 13: Increasing strictness in long vs. short lines (LOESS smoothers).

An alternative approach to increasing strictness that does not involve a distance multiplier is to index *CONTRARY to the right branches of prosodic structures. For example, imagine that a (long) line has the partial prosodic structure in Figure 14, where subscript S is ‘strong’ and W is ‘weak’. Lines can be divided into hemistichs (half-lines), hemistichs into cola (phrases), and so forth. Assuming that right branches are strong (as is usually the case in metrics for supraword structure; Hayes 1988, Golston and Riad 2000, Hayes et al. 2012, Blumenfeld 2015), *CONTRARY can be indexed to strong branches of prosodic constituents to implement greater stringency towards endings. This indexation could be general, as with *CONTRARY_S: ‘Penalize a contrary mapping for every strong node dominating it’. In the final colon, for example, this constraint would incur violations from the strong colon, the strong hemistich, and the strong line, whereas positions intermediate in the line would receive only a subset of these violations. Alternatively, *CONTRARY could be indexed to each prosodic level separately, e.g. *CONTRARY_{Hemistich_S}. This would allow different domains to be weighted differently. Nevertheless, because our corpus is not annotated for higher-level prosodic structure, we cannot implement this version of the constraint, and put aside positional modulation in the maxent model.

¹¹A fractional scale could be converted to a whole number scale using multiplication and rounding (e.g. the nearest percent). But the conversion factor then becomes an issue: Why not deciles, or thousandths, etc.? Moreover, conversion to integers would not obviate the gradient nature of the constraint.

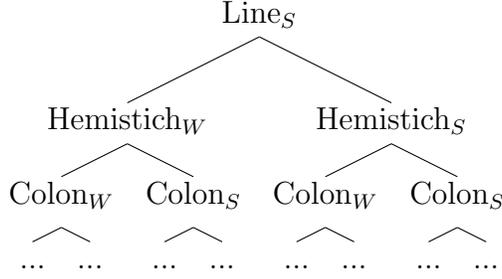


Figure 14

Beyond *CONTRARY, which penalizes opposing mappings, we posit *NONPARALLEL, which penalizes any mapping that is not parallel. Recall from §5 that a parallel mapping is defined as a rising melody on rising tone, falling melody on falling tone, or flat melody on flat tone. *NONPARALLEL therefore penalizes any contrary or oblique mapping, as summarized in Table 8. We take the penalty to be scalar in terms of the number of steps, as with *CONTRARY above, except when the violation involves a flat tune, in which case the minimum violation of one is assigned. These constraints once again comprise a stringency hierarchy: A structure violating *CONTRARY always also violates *NONPARALLEL, but not vice versa. This stringency guarantees that under any ranking/weighting, contrary mappings are at least as penalized as oblique mappings, putatively a universal of tone-tune mapping. Languages are, however, permitted to differ in the degree to which oblique mappings are avoided. Tommo So, for its part, is highly tolerant of oblique mappings, such that they are treated similarly to parallel mappings (§5.1). Certain other languages, such as Vietnamese (Kirby and Ladd 2016) and Cantonese (Ho 2006), show a greater avoidance of oblique mappings. Finally, given that *CONTRARY comes in variants (*CONTRARY_{CG}, etc.), we assume that these same indices are available for *NONPARALLEL.

Mapping Type	Tone	Tune	*CONTRARY	*NONPARALLEL
Contrary:	Rising	Falling	*	*
	Falling	Rising	*	*
Oblique:	Flat	Rising		*
	Flat	Falling		*
	Rising	Flat		*
	Falling	Flat		*
Parallel:	Rising	Rising		
	Flat	Flat		
	Falling	Falling		

Table 8: Two mapping constraints and the tone-tune mappings that they penalize.

The weights of the six mapping constraints and the two musical constraints posited above are given in Table 9. Weights were estimated by maxent learning software by Wilson and George (2008) with the default settings of $\mu = 0$ and $\sigma^2 = 100,000$. In general, contrary mappings are strongly penalized, while oblique mappings are only weakly penalized (especially in the indexed conditions, where the weights for *NONPARALLEL are virtually zero).

*STEP	0.869
*UP	0.467
*CONTRARY	0.864
*CONTRARY _{CG}	0.282
*CONTRARY _{Lex}	0.163
*NONPARALLEL	0.349
*NONPARALLEL _{CG}	0
*NONPARALLEL _{Lex}	0.004

Table 9: Constraints and their weights.

A sample tableau is provided in (14). Constraints are now arranged by weight and inactive/nonsignificant constraints are omitted.¹² In this case, the input (upper left cell) is a rising tone sequence <L, H> contained within a word and whose lexical tone is also <L, H>. The output candidates are tonal steps. The harmony of each candidate (\mathcal{H}) is the weighted sum of its violations (i.e. $\sum_j w_j c_j$, where w_j is the weight of constraint j and c_j is the violations that the candidate incurs on j). ‘Gen’d p ’ in the fourth column is the generated (or predicted) probability, which, as mentioned above, is $e^{\mathcal{H}_0} / \sum_i e^{\mathcal{H}_i}$, where \mathcal{H}_i ranges over candidates. These predictions can be compared to the observed proportions (‘Obs’d p ’), which reflects the empirical counts (‘Obs’d N ’). A close match between ‘Obs’d p ’ and ‘Gen’d p ’ suggests that the grammar is performing well, though because the training data are a finite sample, a perfect match is generally not achievable or even desirable (see below on overfitting). The smaller a sample is, the more susceptible it is to sampling error. Moreover, the weights were optimized based on the whole corpus, not just the subset of the data in this tableau.

L, H <small>lex=LH, intraword</small>	Obs’d N	Obs’d p	Gen’d p	\mathcal{H}	*STEP	*CONTRARY	*UP	*NONPARALLEL	*CONTRARY _{CG}	*CONTRARY _{Lex}
					.869	.864	.467	.349	.282	.163
a. -2	0	.000	.005	-5.054	-2	-2	0	-2	-2	-2
b. -1	35	.099	.069	-2.527	-1	-1	0	-1	-1	-1
c. 0	176	.496	.606	-0.349	0	0	0	-1	0	0
d. 1	112	.315	.226	-1.336	-1	0	-1	0	0	0
e. 2	32	.090	.095	-2.205	-2	0	-1	0	0	0

Tableau (15) illustrates a case of grammatical tone conflicting with lexical tone in the crossword condition. Once again, while the fit is hardly perfect, the observed sample is small. Candidate (c) has no violations in (15), but is still apportioned $p < 1$, since harmonic bounding does not exist in maxent HG (Jesney 2007).

¹²Although *NONPARALLEL_{Lex} was assigned a nonzero weight of 0.004, its inclusion did not significantly improve AIC_C (as described above), and it is therefore omitted.

L, L <small>lex=LH, crossword</small>	Obs'd	Obs'd	Gen'd	\mathcal{H}	*STEP	*CONTRARY	*UP	*NONPARALLEL	*CONTRARYCG	*CONTRARY _{Lex}
	N	p	p		.869	.864	.467	.349	.282	.163
a. -2	0	.000	.054	-2.436	-2	0	0	-2	0	0
b. -1	1	.026	.182	-1.218	-1	0	0	-1	0	0
c. 0	28	.737	.616	-0.000	0	0	0	0	0	0
d. 1	9	.237	.114	-1.685	-1	0	-1	-1	0	0
e. 2	0	.000	.034	-2.903	-2	0	-1	-2	0	0

Though we have shown only two tableaux here, the full set of tableaux is available at the article website at <URL>. As for assessing the overall performance of this model, its log-likelihood is -2,521 and its AIC_C is 5,064, though these metrics are not particularly useful outside of the context of model comparison. The issue of R^2 or explained variation is a complex one for multinomial models (even for the simpler case of binomial logistic regression, more than a dozen methods have been proposed; Mittlbock and Schemper 1996, Menard 2000). A simple correlation (whether parametric or nonparametric) between observed N and generated p across all tableaux is inappropriate because, for one thing, the data are not distributed evenly among tableaux; constraint weights were optimized to fit the data, and thus more sensitive to populous tableaux than to sparse ones.¹³ A raw correlation would weight all rows (i.e. N, p pairs) the same, and therefore overpenalize discrepancies from sparse tableaux. Weighted Spearman's ρ , for its part, where each row is weighted by the total N for its tableau, is 0.614 (cf. unweighted $\rho = 0.498$). Another problem for assessing maxent models trained on empirical data is that sampling error can be considerable, especially for sparse tableaux; therefore, overfitting is a danger. Indeed, the optimization formula employed here invokes a Gaussian prior whose purpose is to militate against overfitting (Johnson et al. 1999, Goldwater and Johnson 2003). Without this prior, the N, p fit would improve, but the prior is nevertheless still included in most maxent analyses.

Maxent HG is closely related to logistic regression, as in §5.2. For starters, both use weighted criteria to define a probability distribution over outcomes. The goal of the previous section was to establish empirically which factors modulate the strictness of mapping, and in which directions, whereas this section seeks to establish a generative production (or acceptability) grammar by which tone can be mapped onto tune (or vice versa; see §7.4). The maxent grammar is multinomial (in this case, operationalized as five candidates, though other cases might involve many more), whereas logistic regression in §5.2 was binomial. Furthermore, at least as a matter of convention, constraints in (maxent) HG are usually regulated in several ways that predictors in a regression are not. Three such discrepancies are mentioned here, though this list is not exhaustive. First, constraint weights are normally not allowed to be negative in HG, whereas predictors in regression are permitted to have both positive and negative coefficients. Second, constraint violations all have the same sign in HG, being penalties rather than rewards (but cf. Kimper 2011 concerning the latter). Third, constraint violations are normally assumed to be discrete, unlike predictor values (but cf. Flemming

¹³For example, if the observed counts in our first tableau are multiplied by 100, the learned weights are different, even though the observed proportions remained unchanged. Hayes and Wilson (2008) use the correlation r , but the caveat does not apply, since their data are not organized into different tableaux for different inputs, but model only surface phonotactics.

2001 on real-valued violations). More generally, the violation functions of constraints are supposed to be reined in by Universal Grammar, for which numerous restrictions have been put forth. The question of directionality and the nature of the input will be addressed further in the next section (§7.4).

7 Discussion

In this final section, we discuss some of our main findings and their implications for the study of tone-tune association and metrics. We also speculate on open questions that remain.

7.1 The varying rate of tone-tune association

One of the main results of our study was to show that a single aggregate rate of tone-tune association calculated across a whole corpus, such as 90% nonopposing, obscures a great deal of complexity. Numerous factors, both linguistic and nonlinguistic, affect this rate, and these effects provide telling glimpses into a speaker’s subconscious knowledge of his or her language and its textsetting. We comment on a few of the factors here.

First, in §5.2, we showed that the rate of tone-tune association is greater in rote lyrics than it is in improvised lyrics. As we suggested in §5.3 above, this may not be surprising if there is a certain amount of evolutionary selection of melodies and lyrics, with the best combinations more likely to be passed down from one singer or generation to the next. This raises the question of the role of diachrony vs. synchrony in tonal textsetting. If improvised lyrics showed no significant correlation between tone and melody, we could conclude that tonal textsetting does not belong in the synchronic grammar. However, singers do in fact match tone and melody at significantly greater than chance levels even when improvising (in fact, one singer matched equally strictly in rote and improvised material; see Figure 10). It is part of a singer’s linguistic competence, albeit with minor interspeaker differences. Even rote lyrics may be regulated by the synchronic grammar (aside from the trivial sense in which they must have been composed anew at some point): Singers produce slight variations and improvisations in the basic melody of rote portions of song, but these variations still respect the principles of tone-tune association. See also §7.4 below regarding productivity.

The rate of tone-tune association was also found to be higher within words than across them (§5.2). Specifically, bigrams contained within a word or a clitic group showed significantly stricter mapping than bigrams straddling a prosodic word boundary. Intraword strictness conceivably has at least two functional sources. First, it may follow from experience. Take, for example, a disyllabic LH-toned word such as *gìné* ‘house’. Every time that a speaker encounters this word, whether in speaking or listening, LH is reinforced (modulo grammatical tone). The transition across the word boundary, by contrast, varies according to what follows, meaning that speakers may have fewer expectations for that tonal transition. Second, depending on phrasing, a pitch reset and other boundary phenomena might occur across prosodic words, interfering with the relative pitch levels of the adjacent tones. For example, an L#L sequence across a reset might be slightly rising phonetically. More

generally, given that tone-tune mapping is not required across line boundaries, line-internal boundaries above a certain level may represent an intermediately strict situation (see further §7.2). Our analyses above distinguished between # (a prosodic word boundary or higher) vs. all lower level boundaries (clitic, compound, or none), but did not distinguish among the higher-level boundaries that instantiate #. For example, a boundary between two intonation groups would be coded as #, just as would the boundary between two prosodic words within a phonological phrase. Therefore, it remains possible that strictness is relaxed more for higher-level # boundaries. This remains a question for future research.

The last factor that we will address here is the difference between lexical and grammatical tone, which has not been previously tested to our knowledge. Bigrams in which both tones are grammatical were found to show significantly lower rates of tone-tune association than bigrams in which one or both tones are lexical. We offer our thoughts on some possible explanations for this asymmetry. First, recall that grammatical tone in Tommo So almost exclusively involves replacive overlays in particular morphosyntactic contexts. It may be the case, especially for nouns, that words with a grammatical overlay are less frequent than words with their lexical tone, tying into the functional explanation for juncture strength discussed above. This explanation would likely not hold for verbs, where in fact forms with lexical tone are rare, but the exact grammatical overlay applied depends on TAM, so multiple overlays will be attested for any given verb (e.g. {H}, {L}, {HL}, {LH}), whereas nouns will typically exhibit only one ({L}). Another possibility is that singers at least sometimes ignore grammatical overlays in tonal textsetting, matching instead lexical tone. The latter was supported by the finding in §5.2 (Table 6) that lexical tone significantly affects melody even when surface, grammatical tone is controlled. *CONTRARY_{Lex} in §6 formalized this latter motivation.

There may be still other factors affecting the rate of tone-tune association for which we did not code, such as phonological phrasing or whether a high tone is found in a downdrift environment. We simply hope to have shown here that the notion of ‘a rate of tone-tune association’ is more complex than previously recognized.

7.2 Tonal textsetting and metrics

A consistent finding of our analysis is that tonal textsetting obeys many of the same principles and constraints as does metrics, pointing towards a more general theory of the artistic adaptation of language. While metrics has been mentioned in passing in the tone-tune literature, no explicit parallels have been drawn, despite the fact that both involve the artistic adaptation of suprasegmental material (tone vs. stress or weight).

For starters, consider the core tone-tune mapping constraint family proposed here, *CONTRARY. Hayes et al. (2012:701), following Jespersen (1933, et seq.), posit a comparable constraint family for the English iambic pentameter, namely, *RISE FROM S, which penalizes rising stress on falling meter (i.e. an SW sequence, where S is a strong metrical position and W a weak one), and *FALL FROM W, which penalizes falling stress on rising meter. For example, *the tree* placed in a falling sequence in the meter violates the former constraint. Thus, the poetic meter is analogous to the music here, and stress is analogous to tone. Both

types of mapping evaluate an artistic dimension of height/prominence (musical or metrical) against the suprasegmental height/prominence inherent in linguistic material (tone, stress, or weight). Moreover, Hayes et al. (*ibid.*) propose a constraint family akin to our *NONPARALLEL, which they term *NO FALL FROM S and *NO RISE FROM W. These constraints are in a stringency relationship with the previous set, just as our *NONPARALLEL contains *CONTRARY. For example, *the tree* in SW violates both *RISE FROM S and *NO FALL FROM S, but *to a* in SW, being flat, violates only the latter. Put differently, *NO FALL FROM S is violated by both contrary and oblique mappings. Of course, not all analyses of the iambic pentameter employ these particular constraints (cf. e.g. the STRESS MAXIMUM PRINCIPLE of Halle and Keyser 1966), but the general outlook for parallelism remains the same.

Beyond basic mapping, the strictness-modulating factors that we discussed in §5.2 are largely paralleled in metrics. First, mapping is stricter within tighter phrasal constituents in both tone-tune setting and metrics (on the latter, see Magnuson and Ryder 1970, Kiparsky 1975, Hayes et al. 2012), with possible rationales in the work just cited and in §7.1. Second, mapping is (if anything) stricter towards the ends of constituents than it is at their beginnings. This appears to be a universal asymmetry in metrics when strictness differentials are observed (e.g. Hayes 1983:373), but has not been previously documented for tone-tune alignment to our knowledge. The rationale for this asymmetry may relate to prosodic headedness being right-oriented, at least for higher-level domains in the relevant languages, as discussed in §6, though it remains an open question. Next, we demonstrated that rote material is stricter than improvised material, presumably due to community-level, diachronic selection effects. While we are not aware of an exact parallel from metrical improvisation, we note that meters are often observed to vary in strictness according to the genre in which they are deployed: For example, the Ancient Greek iambic trimeter is notably stricter when used in tragedy than in comedy (Halporn et al. 1980:14). We also investigated interspeaker differences; in metrics, authors are known to vary in their strictness and licenses (see e.g. Kiparsky 1977 and Hayes et al. 2012 on Milton vs. Shakespeare).

One final factor that we demonstrated to affect strictness is that of lexical vs. grammatical tone, where the former is stricter. We are not aware of this factor being investigated previously in metrics or textsetting. It is logically possible that there is an analog in metrics, for example, if stress or weight determined by late phrasal phonology (e.g. the Rhythm Rule in English; Selkirk 1981, Hayes 1984) were not mapped as strictly as their lexical counterparts, though this is an open question. We argued that the lexical-grammatical discrepancy in Tommo So may arise from the fact that speakers remain sensitive to the underlying tone even when it is replaced on the surface. Metrical mapping and rhyme have been claimed to sometimes be sensitive to presurface representations (Kiparsky 1968, 1972, Malone 1982). For example, Kiparsky (1968) claims that Kalevala Finnish metrification is sensitive to an intermediate phonological representation, prior to the application of at least five late phonological rules. Intermediate (or at least pre-postlexical) representations may also be invoked by other forms of artistic mapping. For instance, in a xylophone surrogate language for the four-tone Mande language Seenku, it is presurface tone that is mapped to notes on the

xylophone, before postlexical phonological processes such as tonal absorption have applied (McPherson 2016). That said, a meter based purely on underlying forms appears to be impossible; in general, meters are surface-oriented. The extent to which presurface material is available to artistic mapping therefore remains an issue for future research, though we have brought to light one empirical area in which it can be probed.

7.3 Bigram vs. sequence modeling

While we analyze only bigrams (i.e. transitions) in this article, as with most previous research on tone-tune association (§2), the problem is ultimately one of sequence modeling, that is, selecting and aligning a time series of notes with a time series of tones. That said, the constraints developed here can be applied to longer sequences without modification. To do so, the input and candidates merely have to be specified as ordered sets of transitions, such that correspondence is coindexed (cf. McCarthy and Prince 1995). A related question concerns whether tone-tune mapping needs to make reference to sequences larger than the bigram. We are not aware of this being necessary. For example, whether the tonal bigram 0H patterns as rising or not depends on the preceding tone, as demonstrated in §5.1; however, this case is not actually trigram-driven, assuming that 0H is realized locally as rising due to interpolation. Line-level effects also clearly obtain, such as increasing strictness (Figure 8) and increasing flatness (Figure 9). But these can be analyzed locally by invoking structure, as shown above for the former. Moreover, to the extent that long-distance dependencies occur (e.g. in the form of the musical line, or in the distribution of linguistic tone), they seem generally to arise from other components, such as the musical grammar, rather than from tone-tune mapping per se. Nevertheless, it remains possible that tone-tune correspondence is sensitive to sequences larger than bigrams in respects that we do not treat in this article (e.g., to give a hypothetical example, if a contrary mapping were especially strongly avoided immediately following another contrary mapping).

7.4 Directionality

The model of grammar developed in §6 takes the input to be linguistic material (the tonal bigram) and the output to be a musical transition; in short, tone \Rightarrow tune. As a production model, this grammar takes language to be prior, which is then mapped onto music. The music itself would be constrained not only by the musical constraints *STEP and *UP, as illustrated in §6, but also by a more complex musical production grammar characterizing Tommo So music, not treated in this article. But one could also imagine the grammar defined above operating in reverse: Given a melodic transition size, what is the probability distribution over tone sequences?

Additionally, the grammar could be envisaged as bidirectional, omitting the input altogether and populating the candidates with every possible tone \Leftrightarrow tune mapping. For example, combining the tableaux in (14)–(15) into a bidirectional model yields tableau (16), in which each candidate is a tone-tune mapping. Subscripts indicate lexical tone and # is a prosodic word boundary. This model assumes nothing about the priority of tone or tune,

but rather defines the distribution of mappings between the two. As a production model, it would involve the speaker selecting a tone pattern and melody simultaneously, in a mutually constrained fashion. Constraint weights were reoptimized for (16) based not only on the shown candidates but the full set of possible mappings.¹⁴ The critical point here, however, is not the particular weights, but the fact that the same mapping constraints are compatible with any directionality.

		Obs'd <i>N</i>	Obs'd <i>p</i>	Gen'd <i>p</i>	\mathcal{H}	*STEP .987	*CONTRARY .764	*UP .282	*CONTRARY _{Lex} .280	*NONPARALLEL .132
a.	LH _{LH} ↔ -2	0	.000	.004	-4.326	-2	-2	0	-2	-2
b.	LH _{LH} ↔ -1	35	.089	.036	-2.163	-1	-1	0	-1	-1
c.	LH _{LH} ↔ 0	176	.448	.278	-0.132	0	0	0	0	-1
(16) d.	LH _{LH} ↔ 1	112	.285	.089	-1.269	-1	0	-1	0	0
e.	LH _{LH} ↔ 2	32	.081	.033	-2.256	-2	0	-1	0	0
f.	L#L _{LH} ↔ -2	0	.000	.034	-2.238	-2	0	0	0	-2
g.	L#L _{LH} ↔ -1	1	.003	.104	-1.119	-1	0	0	0	-1
h.	L#L _{LH} ↔ 0	28	.071	.317	-0.000	0	0	0	0	0
i.	L#L _{LH} ↔ 1	9	.023	.078	-1.401	-1	0	-1	0	-1
j.	L#L _{LH} ↔ 2	0	.000	.026	-2.520	-2	0	-1	0	-2

A larger conceptual question arising from this work, then, is what exactly is happening online when singing? Do singers come to the performance with a melody in mind and fit words to that melody, or do they come with the words and fit the melody to them? The constraints posited here are compatible with any directionality, but a bidirectional model, as just sketched, appears to be justified. On the one hand, there are a core set of melodies, passed down from generation to generation, and a core set of words that match these melodies. On the other hand, during a performance, singers may improvise, drawing names and inspiration from the audience or recent events, and melodies can be tweaked and elaborated upon as the singer sees fit. As mentioned above, even rote lyrics whose melodies are in principle learned (thus, any tone-tune association would be mostly diachronic) can be sung on slightly differing tunes. Consider the first line of ‘An elephant gave birth’:

- (17) gwé nàl-è (èm)-kémíjé sàm-è gwé nàl-è
 elephant give.birth-PFV milk-colostrum be.plentiful-PFV elephant give.birth-PFV
 ‘An elephant gave birth, there was so much colostrum, an elephant gave birth.’

¹⁴*CONTRARY_{CG} received zero weight in the bidirectional model here, and is therefore omitted from the tableau. This unexpected result might be due to the fact that the bidirectional model is missing purely linguistic constraints, such as phonotactics and tonology. Such constraints were moot in §6, where language was the input; as a general principle of constraint-based grammars, the input is unconstrained. However, in the bidirectional model, language is now part of candidates, which means not only that purely linguistic constraints are viable, but that their omission could distort the weights of tone-tune constraints, since a major factor affecting the frequencies of candidates is ignored.

The data corpus contains as many as eight different melodic variations of the line (with slight lyrical variations as well, for instance variation between *èm kémíjé* and *kémíjé*). Differences in melody can be subtle, such as the sequences of notes 5-3-2-1, 4-3-2-1, or 4-4-2-1 for the final words *gwé nàl-è* ‘an elephant gave birth’ (with melisma on the last syllable), all of which equally satisfy tone-tune correspondence constraints. They can also be more dramatically different, even violating constraints, such as the difference between melodies 6-6-4 and 3-3-4 for the initial repetition of *gwé nàl-è* in the line, where the latter exceptionally maps a low-low sequence to a rising melody. In these cases, the singer presumably deviates from the set melody for aesthetic purposes, but in most cases deviations are minor and still obey mapping principles. For musical notation of four of these variations, see the musical score in the appendix.

This suggests to us that performance is a negotiation between desired lyrics, desired melodies, and a synchronic set of principles mapping the two. Lyrics are improvised in real time, hence words must be mapped to melodies; melodies are likewise subtly changed in each repetition, showing that melodies cannot be fully fixed and must be fit to the lyrics. Ultimately, a bidirectional model may be the most logical choice to capture these results, though questions remain about the best way to implement a model whose candidate set consists of maps between two domains. This article proposes mapping constraints as well as some basic constraints on musical form (i.e. *STEP and *UP). However, because a bidirectional grammar adds linguistic form to candidates, an adequate bidirectional grammar would also require constraints on linguistic form, which we do not treat here. The mapping constraints proposed here can still be implemented in a bidirectional setting, but without constraints on linguistic form, such a model is incomplete, and weights might be misleading. Thus, in order to keep the focus of this article on mapping constraints, we do not pursue a bidirectional grammar more fully than the sketch in this section.

7.5 Nonnative singers

We raise one last issue as a topic of future work. There are roughly half a million ethnically Dogon people living in Mali, speaking approximately twenty languages in the Dogon family. Despite this linguistic diversity, reportedly nearly all Dogon sing in Tommo So, regardless of their native language (Hochstetler et al. 2004, personal experience of the first author). Since most of these singers do not speak Tommo So, they would be expected to lack intuitions on many tonal forms. Analysis of these songs would shine light on the question of synchronic versus diachronic textsetting. Assuming that segmental pronunciations do not differ significantly, it would also offer the opportunity to probe native Tommo So speakers’ aesthetic judgments of music based on tone-tune association. Can a native speaker identify a nonnative singer based on differences in rate of tone-tune association? Are these singers judged as more or less talented? Unfortunately, the current situation in Mali does not allow us to return to test this question, but we offer it in hopes that someone may one day pick up the charge.

8 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 melody, 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.

APPENDIX.

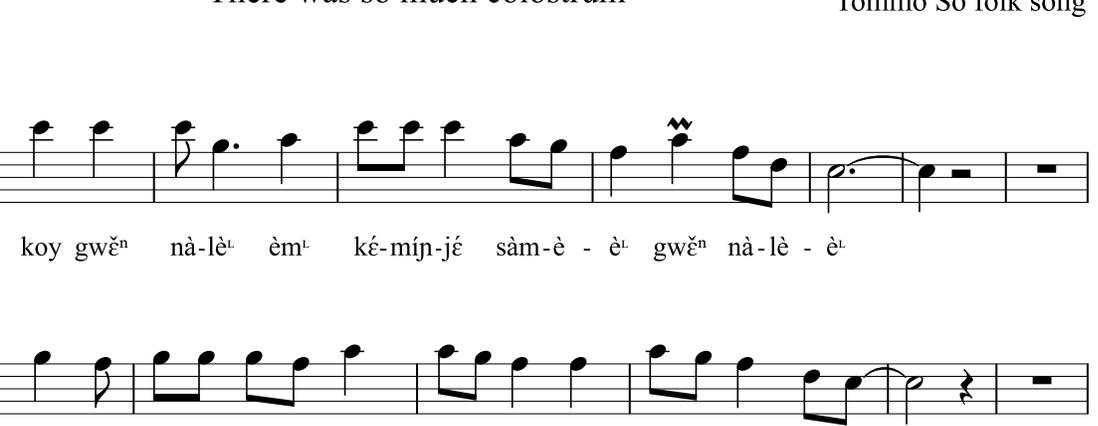
An elephant gave birth

There was so much colostrum

Tommo So folk song

Bass 

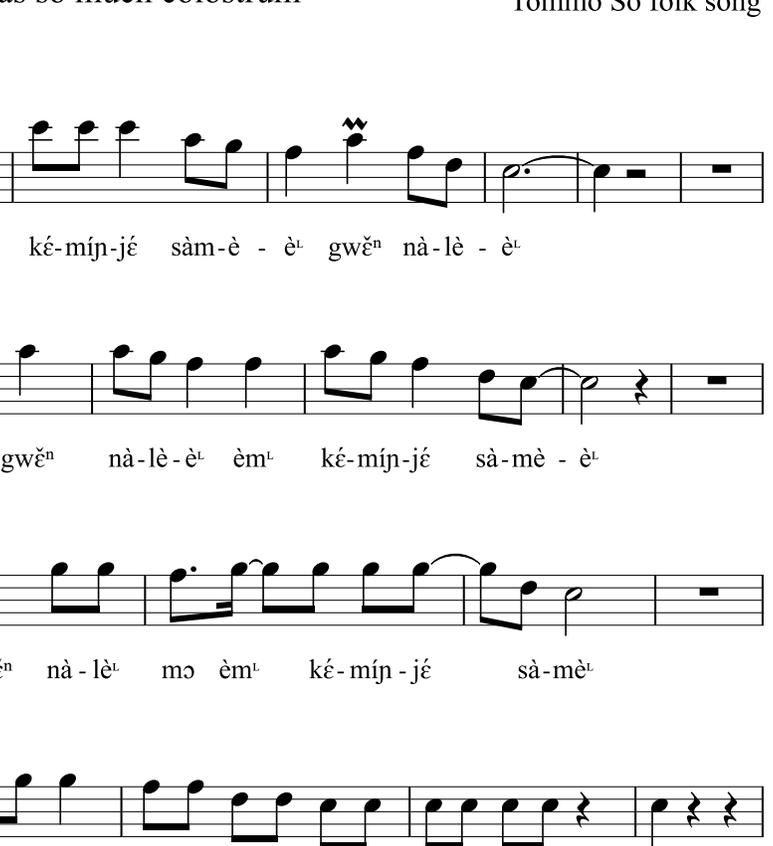
ha koy gwěⁿ nà-lè^t èm^t ké-míj-jé sàm-è - è^t gwěⁿ nà-lè - è^t

B. ⁸ 

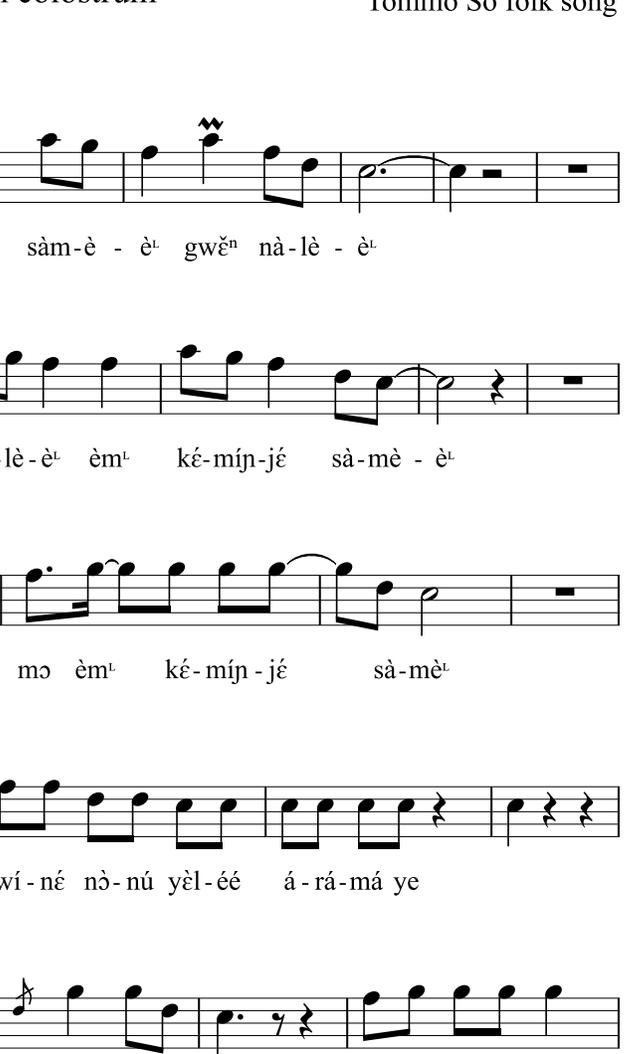
nàà^t yàà - ndó gi - né ú-wò= ne gwěⁿ nà-lè-è^t èm^t ké-míj-jé sà-mè - è^t

B. ¹⁴ 

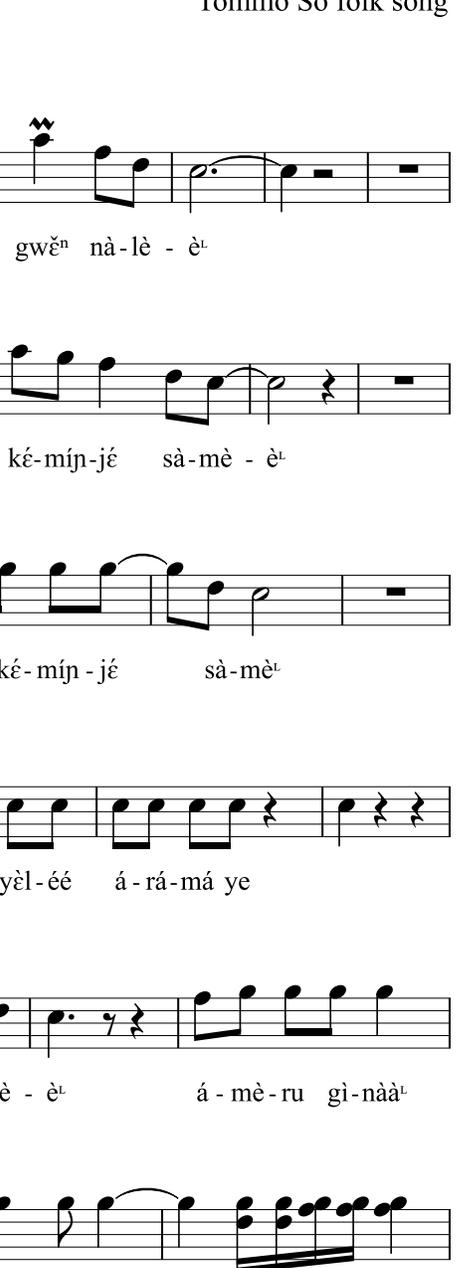
á - mí - ru ge gi - nè^t ge ne gwěⁿ nà - lè^t mò èm^t ké-míj - jé sà-mè^t

B. ¹⁹ 

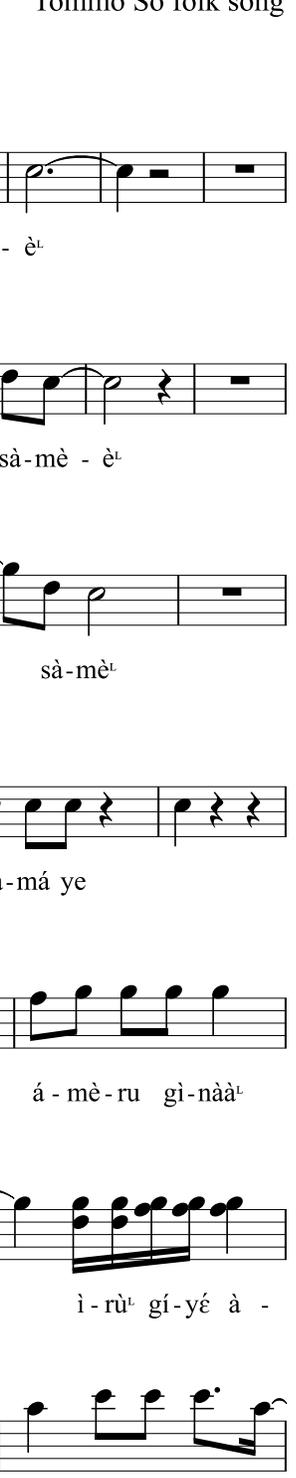
nòm-báá yèl-éé ì - rù^t gí-yé ò - dè^t á - wí - né nò - nú yèl-éé á - rá - má ye

B. ²⁴ 

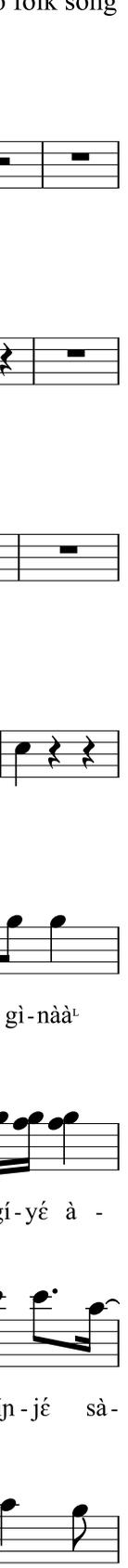
gwěⁿ nà-lè^t ké-míj-jé sà-mè - è^t gwěⁿ nà-lè - è^t á - mè - ru gi-nàà^t

B. ²⁹ 

gwěⁿ nà - lè^t èm^t ké-míj-jé sà - mè^t m-báá yèl - é ì - rù^t gí-yé à -

B. ³⁴ 

wè^t yo m-báá yè - léé á - rá - má ye gwěⁿ nà - lè - è^t èm^t ké-míj - jé sà -

B. ³⁹ 

mè^t gwěⁿ nà - lè - è^t tò - gò^t bí - lu dàm - má é - wò gwěⁿ nà -

Don't Forget Us

Tommo So folk song



nàà^L - gú ye Àm - báá Sàà - lú jám le dèw^L yo yò - gó é - m - méjì



nàà^L - gú ye bàà^L gí - né jám le dèw yo yò - gó é - m - méjì nàà^L - gú ye ú



báá^H gí - nè^L jám le dèw yo yò - gó é - m - méjì nàà^L - gú ye



nàà^L - gú ú bàà^L gí - né jám le dèw yo yò - gó é - m - méjì nàà^L - gú ye



nàà^L - gú ye Yàà - ndó yò - gó é - m - méjì nàà^L - gú ye nàà^L - gú bàà^L gí - né



jam le dèw... nàà^L - gú ye sáá Àll - à yò - gó ém - méjì nàà^L - gú ye



yò - gó ém - méjì bàà^L gí - né jám le dèw yo yò - gó ém - méjì nàà^L -



gú ye nàà^L - gú ú báá^H gí - nè^L jám le dèw^L yo yò -

The Crane

Who hears her?

Tommo So folk song

Bass

kùw^{ná}n á éǵ - wɔ á éǵ - wɔ kù - w^{ná}n á éǵ - wɔ

Bass

kù - w^{ná}n kù - w^{ná}n á éǵ - wɔ kù -

5

B.

á éǵ - wɔ kù - w^{ná}n júú le pí - yè^{HL} - dè á éǵ - wɔ -

B.

w^{ná}n kù - w^{ná}n á éǵ - wɔ kù - w^{ná}n kù -

9

B.

ɔ kù - w^{ná}n náy le pí - yè^{HL} - dè á éǵ - wɔ - ɔ

B.

w^{ná}n á éǵ - wɔ kù - w^{ná}n kù - w^{ná}n á éǵ - wɔ

13

B.

mò - lù - gí - yé míḡ bò - déé - lè à - báá mbe sòm^L ye á éǵ - wɔ - ɔ

B.

kù - w^{ná}n kù - w^{ná}n á éǵ - wɔ

References

- Agresti, Alan. 2002. *Categorical data analysis*. New York: John Wiley & Sons, 2nd edition.
- Baart, Joan L. G. 2004. Tone and song in Kalam Kohistani (Pakistani). *On speech and language: Studies for Sieb G. Nooteboom*, ed. by Hugo Quené and Vincent van Heuven, 5–15. Utrecht: Netherlands Graduate School of Linguistics.
- Bakker, Egbert J. 1997. *Poetry in speech: Orality and Homeric discourse*. Cornell University Press.
- Bakker, Egbert J. 2005. *Pointing at the past: From formula to performance in Homeric poetics*. Harvard University Press.
- Blench, Roger. 2005. A survey of Dogon languages in Mali: an overview. *OGMIOS: Newsletter of the Foundation for Endangered Languages* 3.14–15.
- Blumenfeld, Lev. 2015. Meter as faithfulness. *Natural Language and Linguistic Theory* 33.79–125.
- Bright, William. 1957. Singing in Lushai. *Indian Linguistics* 17.24–28.
- Burling, Robbins. 1966. The metrics of children’s verse: A cross-linguistic study. *American Anthropologist* 68.1418–1441.
- Burnham, Kenneth P., and David R. Anderson. 2004. Multimodal inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research* 33.261–304.
- Chao, Yuen Ren. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. *For Roman Jakobson. essays on the occasion of his sixtieth birthday*, ed. by Morris Halle, Horace Lunt, and Hugh McLean, 52–59. The Hague: Mouton.
- Chen, Matthew Y. 1979. Metrical structure: Evidence from Chinese poetry. *Linguistic Inquiry* 10.371–420.
- Cutler, Anne, and Hsuan-Chih Chen. 1997. Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics* 59.165–179.
- Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.7–44.
- Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spender, Anders Eriksson, and Osten Dahl, 111–120.
- Golston, Chris, and Tomas Riad. 2000. The phonology of Classical Greek meter. *Linguistics* 38.99–167.
- Halle, John, and Fred Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55.3–23.
- Halle, Morris, and Samuel Jay Keyser. 1966. Chaucer and the study of prosody. *College English* 28.187–219.
- Halle, Morris, and Samuel Jay Keyser. 1971. *English stress: Its form, its growth, and its role in verse*. New York: Harper and Row.
- Halporn, James W., Martin Ostwald, and Thomas G. Rosenmeyer. 1980. *The meters of*

- Greek and Latin poetry: Revised edition*. Norman, OK: University of Oklahoma Press.
- Hanson, Kristin, and Paul Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72.287–335.
- Hayes, Bruce. 1983. A grid-based theory of English meter. *Linguistic Inquiry* 14.357–393.
- Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15.33–74.
- Hayes, Bruce. 1988. Metrics and phonological theory. *Linguistics: The Cambridge survey, vol. ii, Linguistic theory: Extensions and implications*, ed. by Frederick J. Newmeyer, volume 2, chapter 12, 220–249. Cambridge University Press.
- Hayes, Bruce. 1989. The prosodic hierarchy in meter. *Rhythm and meter: Phonetics and phonology, volume I*, ed. by Paul Kiparsky and Gilbert Youmans. San Diego, CA: Academic Press.
- Hayes, Bruce, and Abigail Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13.243–303.
- Hayes, Bruce, and Margaret MacEachern. 1998. Quatrain form in English folk verse. *Language* 74.473–507.
- Hayes, Bruce, and Claire Moore-Cantwell. 2011. Gerard Manley Hopkins’s Sprung Rhythm: Corpus study and stochastic grammar. *Phonology* 28.235–282.
- Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.
- Hayes, Bruce, Colin Wilson, and Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88.691–731.
- Herzog, George. 1934. Speech-melody and primitive music. *Musical Quarterly* 20.452–466.
- Ho, W. S. V. 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)*, ed. by M. Baroni, A. R. Addessi, R. Caterina, and M. Costa, 1414–1422. Alma Mater Studiorum University of Bologna.
- Hochstetler, J. Lee, J.A. Durieux, and E.I.K Durieux-Boon. 2004. *A sociolinguistic survey of the Dogon language area*. SIL International.
- von Hornbostel, E.M. 1928. African negro music. *Africa* 1.30–62.
- Huron, David Brian. 1996. The melodic arch in Western folksongs. *Computing in Musicology* 10.3–23.
- Huron, David Brian. 2006. *Sweet anticipation: Music and the psychology of expectation*. MIT Press.
- Jesney, Karen. 2007. The locus of variation in weighted constraint grammars. Poster presented at the Workshop on Variation, Gradience, and Frequency in Phonology, Stanford, CA.
- Jespersen, Otto. 1933. Notes on meter. *Linguistica: Selected papers in English, French, and German*, 249–274. Copenhagen: Levin and Munksgaard.
- Johnson, Mark, Stuart Geman, Stephen Canon, Zhiyi Chi, and Stefan Riezler. 1999. Estimators for stochastic ‘unification-based’ grammars. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 535–541. Association for Computational

Linguistics.

- Johnson, Randolph B., David Huron, and Lauren Collister. 2013. Music and lyrics interactions and their influence on recognition of sung words: An investigation of word frequency, rhyme, metric stress, vocal timbre, melisma, and repetition priming. *Empirical Musicology Review* 9.2–20.
- Jones, A.M. 1959. *Studies in African music*. London: Oxford University Press.
- Karlsson, Anastasia, Hakan Lundström, and Jan-Olof Svantesson. 2014. Preservation of lexical tones in singing in a tone language. *Interspeech 2014*, 1357–1360.
- Kimper, Wendell A. 2011. *Competing triggers: Transparency and opacity in vowel harmony*. Doctoral Dissertation, University of Massachusetts-Amherst.
- Kiparsky, Paul. 1968. Metrics and morphophonemics in the Kalevala. *Studies presented to Roman Jakobson by his students*, ed. by Charles Gribble. Cambridge, MA: Slavica.
- Kiparsky, Paul. 1972. Metrics and morphophonemics in the Rigveda. *Contributions to generative phonology*, ed. by Michael K. Brame, 171–200. Austin: University of Texas Press.
- Kiparsky, Paul. 1975. Stress, syntax, and meter. *Language* 51.576–616.
- Kiparsky, Paul. 1977. The rhythmic structure of English verse. *Linguistic Inquiry* 8.189–248.
- Kirby, James, and D. Robert Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. MS., University of Edinburgh.
- de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21.145–199.
- Ladd, D. Robert. 2014. Singing in tone languages: An introduction to the kinds of things we might expect to find. *Jahrbuch des Phonogrammarchivs der Österreichischen Akademie der Wissenschaften*, ed. by Gerda Lechleitner and Christian Liebl, volume 4, 13–28. Göttingen: Cuvillier Verlag.
- Leben, William R. 1983. The correspondence between linguistic tone and musical melody. *Proceedings of the Ninth Annual Meeting of the Berkeley Linguistics Society*, 148–157.
- List, George. 1961. Speech melody and song melody in Central Thailand. *Ethnomusicology* 5.63–86.
- Magnuson, Karl, and Frank G. Ryder. 1970. The study of English prosody: An alternative proposal. *College English* 31.789–820.
- Malone, Joseph L. 1982. Generative phonology and Turkish rhyme. *Linguistic Inquiry* 13.550–553.
- McCarthy, John. 2003. OT constraints are categorical. *Phonology* 20.75–138.
- McCarthy, John, and Alan Prince. 1995. Faithfulness and reduplicative identity. *Papers in Optimality Theory*, ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, volume 18 of *University of Massachusetts Occasional Papers in Linguistics*, 249–384. University of Massachusetts, Amherst: Graduate Linguistics Student Association.
- McPherson, Laura. 2011. Tonal underspecification and interpolation in Tommo So. Master's thesis, UCLA.
- McPherson, Laura. 2013. *A grammar of Tommo So*. Berlin: De Gruyter Mouton.
- McPherson, Laura. 2014. *Replacive grammatical tone in the Dogon languages*. Doctoral

- Dissertation, University of California, Los Angeles.
- McPherson, Laura. 2016. The talking xylophone of the Sambla: Seenku phonology in a speech surrogate system. Handout, Annual Meeting of the Linguistic Society of America.
- McPherson, Laura, and Bruce Hayes. 2016. Relating application frequency to morphological distance: the case of Tommo So vowel harmony. *Phonology* 33.125–167.
- McPherson, Laura, and Bruce Hayes. 2016. Relating application frequency to morphological structure: The case of Tommo So vowel harmony. *Phonology* 33.125–167.
- McPherson, Laura, and Jeffrey Heath. 2016. Phrasal grammatical tone in the Dogon languages. *Natural Language and Linguistic Theory* 34.593–639.
- Menard, Scott. 2000. Coefficients of determination for multiple logistic regression analysis. *The American Statistician* 54.17–24.
- Mittlbock, Martina, and Michael Schemper. 1996. Explained variation in logistic regression. *Statistics in Medicine* 15.1987–1997.
- Morey, Stephen. 2010. The realisation of tones in traditional Tai Phake songs. *North east indian linguistics*, ed. by Stephen Morey and Mark Prost, 59–74. Foundation Books.
- Prince, Alan. 1999. Paninian relations. Handout, University of Marburg. Available July 2010 at <http://ling.rutgers.edu/people/faculty/prince.html>.
- Richards, Paul. 1972. A quantitative analysis of the relationship between language tone and melody in a Hausa song. *African Language Studies* 13.137–161.
- Rousseeuw, Peter J., and L. Kaufman. 1990. *Finding groups in data*. Wiley Online Library.
- Ryan, Kevin M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28.413–454.
- Rycroft, David. 1959. African music in Johannesburg: African and non-African features. *Journal of the International Folk Music Council* 11.25–30.
- Rycroft, David. 1979. The relationships between speech-tone and melody in Southern African music. *South African music encyclopedia*, ed. by J.P. Malan, volume 2, 301–314. Cape Town: OUP for Human Sciences Research Council.
- Saurman, Mary E. 1999. The agreement of Thai speech tones and melodic pitches. *Notes on Anthropology* 3.15–24.
- Schellenberg, Murray. 2009. Singing in a tone language: Shona. *Selected Proceedings of the 39th Annual Conference on African Linguistics*, ed. by Akinloye Ojo and Lioba Moshi, 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*. Doctoral Dissertation, University of British Columbia.
- Schirmer, Annett, Siu-Lam Tang, Trevor B. Penney, Thomas C. Gunter, and Hsuan-Chih Chen. 2005. Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience* 17.1–12.
- Schneider, Marius. 1961. Tone and tune in West African music. *Ethnomusicology* 5.204–215.
- Selkirk, Elisabeth. 1981. On prosodic structure and its relation to syntactic structure. *Nordic*

- phonology II*, ed. by Thorstein Fretheim, 111–140. Trondheim: Tapir.
- Sollis, Michael. 2010. Tune-tone relationships in sung Duna Pikono. *Australian Journal of Linguistics* 30.67–80.
- Starke, Amy. 1930. The relation between the intonation of song and of speech among the Amaxosa. Master's thesis, University of Cape Town.
- Wee, Lian Hee. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics* 35.128–143.
- Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–982.
- Wilson, Colin, and Benjamin George. 2008. Maxent grammar tool. Software package. University of California, Los Angeles.
- Wong, Patrick M., and Randy L. Diehl. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30.202–209.
- Zymet, Jesse. 2015. Distance-based decay in long-distance phonological processes. *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, ed. by Ulrike et al. Steindl, 72–81. Cascadilla Proceedings Project.