The Role of Music in Documenting Phonological Grammar: Two Case Studies from West Africa

Laura McPherson
Dartmouth College

1 Introduction

While the connection between language and music has long been recognized by scholars and philosophers, the last two decades has seen a surge in interest in the topic. Much of this research has been in the realm of cognitive science rather than formal linguistics, with studies revealing heavy overlaps in sound structure (Lerdahl and Jackendoff 1983, Lehral 2001, Patel and Daniele 2002, Iversen et al. 2008, i.a.), syntax (Maess et al. 2001, Patel et al. 1998a, i.a.), and processing (Besson and Schön 2001, Zatorre et al. 2002, Schön et al. 2004, Patel et al. 1998b, Kölsch et al. 2004, i.a.).

The field of phonology has much to gain from looking at musical data, and yet music remains peripheral to the study of phonological grammar. This is despite the fact that other forms of artistic adaptation of language, such as poetic verse, have long informed phonological theory (Jakobson 1960, Kiparsky 1973 et seq., Halle and Keyser 1969, 1971, Hayes 1988, Hayes and Kaun 1996, Hayes and Moore-Cantwell 2011, Ryan 2014, 2017, i.a.), especially with regards to stress systems and syllable weight.

In this paper, I consider what musical practices can tell us about phonological structure. I focus in particular on what I call language-based music, i.e. sung music or musical surrogate languages, though studies have also suggested an influence of spoken language phonology even on purely instrumental music (e.g. Patel 2008). I draw on two case studies from my fieldwork in West Africa: tone-tune association in Tommo So folk songs (vocal music) and surrogate speech on the Sambla balafon (ostensibly instrumental music). I demonstrate how these musical traditions provide evidence for phonological organization in spoken language and allow us to probe the interfaces between morphology, phonology, and phonetics. In addition to advancing phonological theory, I show how studies of language-based music can be a key tool in language documentation.

This paper is structured as follows: In §2, I briefly summarize some previous studies on music and phonological structure. I then turn to the case studies, first of Tommo So tonal textsetting in §3 and then of the Sambla balafon surrogate language in §4. §5 discusses whether differences between the two systems is a result of their modality and presents the results of a preliminary study of Seenku tonal textsetting. Finally, I conclude with suggestions for future directions in §6.

2 Previous studies on music and phonology

Previous work on the relationship between phonology and music has tended to focus on four main topics: textsetting, whether tonal (Herzog 1934, Leben 1983, Wong and Diehl 2002, Schellenberg 2012, Ladd and Kirby to appear, etc.) or non-tonal (Halle and Lerdahl 1993, Shih 2008, Hayes 2009, Calder 2013, Starr and Shih 2017, etc.); grouping and phrasing (Lerdahl and Jackendoff 1993, Katz and Pesetsky 2011, Katz (m.s.), etc.); rhyme (Zwicky 1976, Holtman 1996, Hanso 2003, Kawahara 2007, Katz 2015, etc.); or lastly rhythm (Patel and Daniele 2003, Seifart et al. 2018, etc.). To illustrate how data from music can advance phonological...
Their musical study investigated the role of the mora vs. the syllable in Japanese textsetting. For decades, the mora has played an important role in the analysis of Japanese phonology, with seminal works like Vance (1987), Otake et al. (1993), Inaba (1998), among others, but the assumption remained that syllables are a universal part of human language organization and are thus still attested in Japanese. In a recent paper, however, Labrune (2012) makes the provocative claim that Japanese provides no evidence for syllables, thus calling into question their universality.

Given the central role of the syllable in studies of textsetting in other languages, Starr and Shih (2017) looked to Japanese music to determine whether the syllable played any role or if textsetting could be accounted for with the mora alone. They used a corpus of songs from animated shows and movies, including anime songs written in Japanese and Disney songs translated from English. Figure 1 summarizes the results for different mora configurations in both sets of songs.

While moraic textsetting is more common in Japanese, particularly in the native Japanese corpus, the syllable also accounts for relatively large proportions of data. As the authors argue, these settings would not be an option if the syllable were not part of Japanese phonology. This result is supported by non-musical evidence as well, for instance in the work of Kawahara (2016).

Thus, as this study shows, phonologists can turn to musical adaptation of language as a source of evidence about phonological structure. In this paper, I ask what we can learn about the phonological structure of understudied languages through the investigation of traditional musical forms.

3 Case study 1: Tommo So tonal textsetting

The first case study looks at vocal music in Tommo So, a Dogon language spoken by about 60,000 people in central Mali (McPherson 2013). In joint work with Kevin Ryan (McPherson and Ryan 2018), we studied a corpus of women’s folk songs to determine the relationship between linguistic tone and musical melody (tonal textsetting). For background, Tommo So has two contrastive tonal primitives, H and L, which can combine to form the contour tones HL and LH on heavy syllables. Minimal pairs contrasting H and L are shown in (1):

(1) a. /H/ dàmmá /LH/ dàmmá
   village    hoe

b. /H/ ìsé /LH/ ìsé
   empty     dog

In addition, syllables may be surface-underspecified for tone, receiving pitch via interpolation from surround-
McPherson

The role of music in documenting phonological grammar

Table 1: Relationship between tonal bigrams and musical bigrams in Tommo So vocal music
The role of music in documenting phonological grammar

Rising Tone (LH)  Flat Tone (LL or HH)  Falling Tone (HL)

Figure 2: Results of tonal textsetting by tone sequence and interval size

<table>
<thead>
<tr>
<th>Surface tone sequence</th>
<th>Underlying tone sequence</th>
<th>% rising music</th>
<th>Out of N total bigrams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrising</td>
<td>Nonrising</td>
<td>17%</td>
<td>116</td>
</tr>
<tr>
<td>Nonrising</td>
<td>Rising</td>
<td>36%</td>
<td>42</td>
</tr>
<tr>
<td>Rising</td>
<td>Nonrising</td>
<td>82%</td>
<td>11</td>
</tr>
<tr>
<td>Rising</td>
<td>Rising</td>
<td>100%</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2: Influence of surface and underlying tone in tonal textsetting

Even split between level and rising melodies, while non-rising tonal transitions strongly avoid rising melodies and are predominantly mapped to level and falling melodies. The avoidance of contrary mappings becomes stronger with larger musical intervals.

We looked beyond the basic rate of tone-tune association in Tommo So to determine the extent to which other factors modulate this mapping. Our results show that tonal textsetting, at least in Tommo So, follows many of the same principles as non-musical verse: mapping is stricter near the ends of lines (cf. Chen 1979, Hayes 1983), stricter within words than across across words (cf. Magnuson and Ryder 1970, Kiparsky 1975, Hayes et al. 2012), and stricter in rote lyrics rather than improvised lyrics. A novel finding, specific to tonal textsetting, is that lexical tone is more strictly adhered to than grammatical tone. This appears to be due to a latent effect of lexical tone, as demonstrated in the following table:

This table shows that though surface tone is the largest determinant of tonal textsetting, the underlying tone (where it differs from a surface grammatical tone melody) still exerts an influence. For instance, a word with an underlying LH tone like ginè ‘house’ is more likely to be sung on a rising melody, even when the lyrical context places it in a construction where it would surface with a {L} grammatical overlay, such as the possessive (e.g. Sànà ‘ginè, where superscript L indicates grammatical tone). I will return to this latent effect of grammatical tone below.

We modeled Tommo So tonal textsetting in Maximum Entropy Harmonic Grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008, i.a.). The grammar contains both a general textsetting constraint, *Contrary (assign a penalty to contrary mappings), as well as versions of *Contrary tagged specifically for grammatical tone, sequences within the clitic group, etc., which form a stringency hierarchy with the more general constraint; see McPherson and Ryan (2018) for further discussion and exemplification.
3.3 What can tonal textsetting teach us about Tommo So phonology? While the textsetting results are interesting in their own right and are suggestive of universal tendencies in artistic adaption—that is, tendencies that go beyond the bounds of metrical verse or tonal textsetting specifically—the question remains: What insights do the textsetting results provide into Tommo So phonology? I highlight two possibilities here.

First, textsetting data revealed a binary distinction between rising and non-rising tone sequences, despite the fact that the tone inventory predicts three categories: rising, level, and falling. Inclusion of a constraint targeting this binary definition of textsetting improved model fit (the AIC improved by 6). Looking back to the spoken language, it turns out that rising vs. non-rising may be an organizing principle of Tommo So tonal phonology more broadly. First, native vocabulary shows primarily two lexical tone melodies: /LH/ (rising) and /H/ (non-rising); falling melodies /HL/ make up only 6% of nouns, most of which are loanwords from Fulfulde. Further, the rising vs. non-rising distinction may play a role in grammatical tone overlays, where only non-rising melodies (/H/, /L/, /HL/) are attested.

While this may suggest a structural/phonological explanation for the binary distinction between rising and non-rising textsetting, an alternative explanation may lie in the phonetic realization of Tommo So tone. Both HH and HL sequences are phonetically falling (approximately .6 semitones for HH and 3.5 semitones for HL; McPherson 2011). The phonologically rising sequence LH, however, can be phonetically rising or in some environments even level, due to near total downdrift in the language (i.e. the last H in HLH is pronounced at the same level as the preceding L).1 In other words, rising pitch is salient to Tommo So speakers, found only in a single phonological configuration, whereas level and falling are unremarkable. This may lead singers to favor rising melodies in just the single context of rising tone, whereas level and falling melodies are generally less constrained.

Thus, the binary division in Tommo So textsetting could arise from two distinct facts about the language’s structure: It could either reflect singers’ inherent knowledge of phonetic realization or it could reflect a deeper categorical divide between rising tonal melodies and non-rising ones, supported by both lexical and grammatical tone distribution.

The second interesting insight from textsetting is the latent effect of underlying lexical tone in contexts where grammatical overlays have applied. In McPherson and Ryan (2018), we model this effect by having both the surface tone and the underlying tone as part of the input. The grammar contains both a general textsetting constraint *Contrary as well as a version specific to lexical tone *Contrary_lex, which form a stringency hierarchy. In cases where surface tone and lexical tone differ, the lexical tone can still exert an influence in how the tones are set to music.

The textsetting results led me to ask whether we can find evidence of simultaneous activation of underlying lexical tone and surface grammatical tone in regular speech as well. Recall that grammatical tone in Tommo So consists primarily of replacive overlays that completely overwrite a word or stem’s lexical tone. This means that lexical tone contrasts are, in principle, neutralized in grammatical tone environments, as shown in (3) with a demonstrative, which triggers a {L} grammatical overlay on the c-commanded noun (McPherson 2014, McPherson and Heath 2016):

(3)  
  a.  dàmmà
     ‘village’
  b.  dàmmà
     ‘hoe’
  c.  dàmmà nɔ́
     ‘this village/hoe’

Unfortunately, I do not have extensive recorded data of grammatical overlays to systematically test whether this neutralization is complete or incomplete. Nevertheless, investigation of available recordings suggests that it may be incomplete. As an illustration, consider the f0 trace in Figure 3 for the phrase jàndùlù pilù ‘white donkey’. The noun ‘donkey’ is underlyingly LH-toned (jàndú/, with a toneless epenthetic [u] at the end). If the {L} overlay were totally applied, then the pitch should remain flat and low throughout the noun, only

---

1 Unfortunately, we were unable to test whether downdrift plays a role in tonal textsetting. The process appears to be phonological phrase-bounded, but how phonological phrases are determined remains an open question.
McPherson

The role of music in documenting phonological grammar

Figure 3: f0 trace of 'white donkey'; lines show level of initial L tone and final H tone.

Figure 4: f0 trace of incomplete application of relative clause grammatical tone; lines show level of initial L and first H tone.

raising on pilu. Instead, Figure 3 shows that the syllable dûl—the underlying H-toned syllable—is pronounced at a pitch intermediate between the L tone of jân and the H tone of pilu.

In this case, the intermediate pitch could be viewed as interpolation between the preceding L and the following H, rather than as an effect of underlying lexical tone. In other cases, though, the interpolation argument is unavailable, and yet we still see what looks like the effect of underlying tone. This is shown in Figure 4, an f0 trace of the following relative clause:

(4) yàà-nàᶫ sàgàràᶫ jàndúlu mî bênd-ɛ̀=ge
    woman-hum.sg young donkey 1sg hit-rel.pfv=def
    ‘the young woman whose donkey that I hit’ (cf. yàà-nàᶫ ságárá ‘young woman’)

The figure once again shows a line from the level of the initial L tone and another line marking the level of the H tone of jàndúlu. We can see that both yàà-nàxlabel and the initial L-tone of jàndúlu are pronounced at the same level, but interestingly, the grammatically lowered tone of sàgàràxlabel (lexically ságárá) is pronounced at a higher level than the surrounding L tones.

I speculate that when a speaker is going to utter a sentence with a grammatical overlay, both the lexical and grammatical tone are activated; the grammatical tone is the intended target, but the underlying lexical tone may “bleed through”, producing the effects seen here. Of course, these data are preliminary, and more systematic study would be required to test whether it holds of the language more broadly.

Returning to tonal textsetting, we are once again left with the question of whether the latent effect of lexical tone in textsetting reflects subtle details of phonetic realization or whether they arise from the same
Table 3: Seenku’s four tone levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Symbol</th>
<th>Syllable</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Super-high</td>
<td>(S)</td>
<td>sî</td>
<td>“Terminalia (sp.)”</td>
</tr>
<tr>
<td>High</td>
<td>(H)</td>
<td>sî</td>
<td>“reciprocal”</td>
</tr>
<tr>
<td>Low</td>
<td>(L)</td>
<td>sî</td>
<td>“second son (proper name)”</td>
</tr>
<tr>
<td>Extra-low</td>
<td>(X)</td>
<td>sȗ</td>
<td>“water jar”</td>
</tr>
</tbody>
</table>

facts of mental activation.

3.4 Interim summary In this section, I have shown that musical melody is constrained by linguistic tone in Tommo So folk music. While a phonological analysis of the language is a prerequisite to carrying out a study of tonal textsetting, observations from textsetting also raise new questions for linguistic analysis. In the case of Tommo So, behavior of tone and musical melody may have revealed a deeper structural division between rising and non-rising tone, active in the inventory of tonal melodies, and it has also raised the possibility of incomplete neutralization of grammatical tone overlays. Both of these questions would be fertile areas of further research that deepen our understanding of the language’s phonological structure and the connections between phonetics and phonology.

In short, musical adaptation provides a window into the speaker’s implicit knowledge of their phonology that phonologists can capitalize on to direct questions of phonological theory.

4 Case study 2: Sambla talking balafon

The second case study broadens the definition of language-based music by focusing on a musical surrogate language—the encoding of linguistic messages in (instrumental) music. Specifically, I consider the Sambla balafon surrogate tradition from Burkina Faso. “Sambla” is an exonym for the language and ethnicity; the people themselves call the language [sɛ̰ ́ɛ̰ -kû] (language of the Séекû), which I will transliterate as Seenku here, but refer to the culture and people with the term Sambla.

Seenku is a Western Mande language of the Samogo group spoken in southwestern Burkina Faso by approximately 15,000 people. It has two primary dialects, Northern and Southern; the latter has been the focus of my fieldwork since 2013 and is the subject of a forthcoming reference grammar (McPherson forthcoming). The language has a rich tone system, with four contrastive level tones, illustrated in (5):

These level tones can combine to create multiple contour tones both lexically and grammatically. These include the falling tones SX, SH, HX, and HL, the rising tones HS, LS, and XH, as well as complex contour tones like XHX and LSX.

Most Seenku vocabulary is either mono- or sesquisyllabic—a full syllable preceded by a “half” or “minor” syllable (Matisoff 1990, Pittayaporn 2015). Both monophthongs and diphthongs can be either short or long, and syllables are all open except for a coda nasal whose place is non-contrastive. Examples of typical Seenku vocabulary are illustrated in (5):

(5) a. Monosyllabic:
   - tɛ̀ ‘who’
   - gɔ̋ɛɛ ‘woods’
   - dȁân ‘basket holder’

b. Sesquisyllabic:
   - tagè ‘chicken’
   - dagɔ̋ɛɛ ‘place(s)’
   - ɲagàn ‘guinea fowl’

4.1 Background: Surrogate languages and the Sambla balafon Surrogate languages are defined as systems of communication in which linguistic form is mapped to non-linguistic modality. These modalities include whistling, drumming, melodic instruments, and even writing. Stern (1957) defines two types: abridgement systems, which encode phonemic aspects, and lexical ideogram systems, which symbolize concepts.

Musical surrogate languages are found worldwide on a broad range of instruments, including drums (e.g.
McPherson

The role of music in documenting phonological grammar

<table>
<thead>
<tr>
<th>Scale degree</th>
<th>1</th>
<th>b3</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>22</td>
<td>38</td>
<td>158</td>
</tr>
<tr>
<td>H</td>
<td>4</td>
<td>0</td>
<td>72</td>
<td>85</td>
<td>154</td>
<td>28</td>
</tr>
<tr>
<td>L</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>73</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>X</td>
<td>101</td>
<td>2</td>
<td>93</td>
<td>57</td>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 4: Relationship between tones and notes of the scale

Traditional Sambla balafons have 23 keys (Strand 2009), spanning four and a half octaves, and are played by three players simultaneously. Surrogate speech is largely confined to the treble, played by the most senior musician. For further discussion of the balafon and cultural contexts of surrogacy, see McPherson (in press).

4.2 Phonological encoding in the surrogate language

The data corpus used in this study consists of 135 phrases on the balafon transcribed for notes and inter-strike duration; the data include both elicited and naturally occurring phrases, yielding a corpus of approximately 823 words. Each of these words was coded for a number of factors, including the tone(s), the note(s) on the balafon, total duration of all strikes comprising the word, whether or not the word contains a contour tone, a long vowel, a diphthong, or a coda nasal, and finally its word shape (mono- or sesquisyllabic).

Analysis of the data reveals the following principles of encoding:

(7) Principles of encoding:
1. Tone (both lexical and grammatical) is always encoded.
2. Vowel length is always encoded.
3. Sesquisyllabicity is always encoded.
4. Segmental identity is never encoded.
5. Coda nasals are sometimes encoded.
6. Postlexical tone is sometimes encoded.

Beginning with tonal encoding, we find that the four lexical tones of the language roughly correspond to four notes of the scale. There is not a one-to-one relationship, however; exact encoding depends on the mode of the song, with S typically tied to the tonal center of the mode. In most cases, this is the octave of the tonic, shown in the following chart with 8, but other modes are also attested (e.g. centered on note 6).

With the mode centered on tones 1 and 8 of the scale (octaves) as the most general, we can see a clear pattern emerging where most S tones are played on note 8, most H on note 6, most L on note 5, and X tones split between 3 and the lower octave tonic. Interestingly, we can see in this chart that b3 is almost universally avoided. This note, known in Seenku as jîo-bȁ̰a̰-dȅn ‘fetish balafon key’, is said to be reserved for more spiritual uses and is seldom used in surrogate speech (Strand 2009).

To illustrate how tone encoding works, consider Figures 5-6. These illustrate how the phrases in (8-a-b) are played on the balafon, with notes of the scale on the y-axis, and words of the phrase running along the x-axis. Gray cells indicate a note strike. The central note of the mode is indicated with an arrow on the left, showing that these examples are both played in the regular “bûg-ɲû” (‘balafon mother’) mode centered on note 1 of the scale.

Yoruba, Beier 1954), flutes (e.g. Gavião, Moore and Meyer 2014), trumpets (e.g. Asante, Kaminski 2008), jaw harps (e.g. Khmu, Proschan 1994), and most importantly here, balafons (e.g. Senufo, Zemp and Soro 2010).

The balafon is a kind of resonator xylophone with many versions and variations throughout West Africa. Under each key hangs a gourd tuned to resonate at the same frequency, thus amplifying the note. Like Tommo So music, Sambla music employs a pentatonic scale, but with different intervals shown in (6):

(6) Sambla pentatonic scale, in the closest equivalent Western scale degrees
1 b3 3 5 6
Figure 5: ‘Don’t hit me!’ (NMD: July 13, 2015)

Figure 6: ‘Don’t hit us!’ (NMD: July 13, 2015)

(8) a. á mó bū ŋé
   2SG 1SG.EMPH hit.IRREAL NEG
   ‘Don’t hit me!’

   b. á mī bū ŋé
   2SG 1PL hit.IRREAL NEG
   ‘Don’t hit us!’

As Figures 5 and 6 show, the melody played on the balafon follows the pattern laid out by the tones of the language; H is one step below S, and X is two steps below H, just as we would expect from this four tone system.

Contour tones are encoded by playing each component tone of the contour, with the first tone played as a grace note (i.e. with a significantly shorter duration than the second tone). Figure 7 represents the balafon rendition of the following phrase:

(9) jō-marī nā mó bō tagọn-tagön
   water-drink.ANTIP.NOM PROSP 1SG.EMPH kill.IRREAL RED-COMpletely
   ‘I am dying of thirst.’ (Lit. ‘Thirst will kill me completely.’)

There are two contour tones in this example, first a HL contour tone on jō (indicated with the dangling L tone, since the circumflex is used to represent the more common HX contour), and second a LS rising tone on the prospective auxiliary nā. Both of these contour tones are encoded on the balafon by playing a sequence of two notes, the first of which is shorter. The duration of 73 milliseconds for the L of LS-toned nā is more
representative of grace note length in the corpus.

This example also illustrates the encoding of sesquisyllabic words: they are also played with two note strikes, but on the same note here, since the words *mar̆̆* and *tagón-tagón* have level tone. Interestingly, vowel length is encoded in precisely the same way, i.e. with two note strikes. This gives rise to ambiguity: If two note strikes are on a single note, then it could either represent a sesquisyllabic word or a long vowel; if they are on distinct notes, then it represents either a short, long, or sesquisyllabic contour tone. For further discussion and illustrations, see McPherson (in press).

4.3 What can the balafon surrogate language teach us about Seenku phonology? I now turn to the same question posed above for Tommo So: What can studying the balafon surrogate language tell us about the phonological structure of Seenku? First, and most simply, the balafon surrogate language provided key evidence that Seenku has four rather than three contrastive tones, as I originally believed. In my original analysis of the language, there were three contrastive levels, L, M, and H, but there were two featurally-distinct but phonetically identical M tones, one found lexically and the other derived through plural formation (McPherson 2017). At times, I would hear what sounded like upstep or downstep between two adjacent “M tones”, but I couldn’t discern a pattern. When I began eliciting phrases on the balafon, it was immediately evident that the “derived” and “lexical” M tones were consistently played differently—the derived M always lower than its supposedly lexical equivalent. Around the same time, I discovered words with this lower “derived” tone lexically (such as the numeral ‘five’ *nɔ̀*, always lower than ‘one’ *sóen* but higher than *sȗɛ* ‘three’). Taken together, this led to a reanalysis of Seenku as a four-toned language, with the former M category split into L and H (and the extreme tones reclassified as X and S).

Beyond the big questions of phonemic inventory, the balafon surrogate language also provides evidence for the structure of the morphophonological component. Once again, the clearest case comes from the behavior of different kinds of tone. While in Tommo So, lexical tone and grammatical tone are treated differently for the purposes of textsetting, in Seenku, both are exceptionlessly encoded in the balafon surrogate language. Even borderline cases of grammatical tone like argument-head tone sandhi (McPherson 2019) are encoded, as can be seen in Figures 5 and 6, where ‘hit’ becomes X-toned *bȁ̰* after the H-toned 1sg but S-toned *ba̰̋* after the S-toned 1pl.

On the other hand, postlexical tone processes are only rarely encoded. The clearest example of a postlexical tone process in Seenku is contour tone simplification, whereby rising LS is simplified to either L or IS (downstepped S), depending on the context. Figure 8 shows the balafon rendition of (10), which illustrates clearly the encoding of grammatical tone and the lack of encoding of rising tone simplification:

(10) mó nā bī sū

1SG.EMPHT PROSP GOAT.PL BUY.IRREAL

‘I will buy goats.’

There are two cases of grammatical tone in this example: First, singular *bī* ‘goat’ is raised to S in the plural, which I argue to be a case of featural affixation (McPherson 2017). Second, this S tone then triggers *sū* to raise to S through argument-head tone sandhi (McPherson 2019). As Figure 8 shows, the output of both processes is represented on the balafon. However, in the spoken language, *nā* in this context would simplify to L-toned *nà* (via tonal absorption, Hyman and Schuh 1974). This process does not apply in the balafon surrogate language; both L and S are played as they normally would be.

These results suggest a psychologically real distinction between lexical/grammatical tone on the one hand...
Table 5: Tone distribution by scale degree in the vocal song Kôkô tê sô

<table>
<thead>
<tr>
<th>Scale degree</th>
<th>1</th>
<th>b3</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>1</td>
<td>0</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>H</td>
<td>7</td>
<td>0</td>
<td>22</td>
<td>32</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>L</td>
<td>9</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>X</td>
<td>10</td>
<td>0</td>
<td>25</td>
<td>19</td>
<td>13</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 6: Relationship between tonal bigrams and musical bigrams in Seenku vocal music

5 A matter of modality?

Results of Tommo So textsetting shows that singers may be sensitive to phonetic details of the language and thus are encoding a very surface level when singing. In contrast, the Sambal balafon language shows that musicians encode phonology at the lexical or morphophonological level, but crucially before postlexical tone. The two cases differ not only by language (and tonal complexity), but also by modality, with vocal music in the case of Tommo So and musical surrogate speech in the case of Seenku. In vocal music, segmental information is available in addition to tone, whereas in a musical surrogate language like the one described here, listeners can rely only on tonal and rhythmic information to decipher the message. Could this difference in modality then be related to the level at which tone is encoded?

To answer the question, I conducted a pilot study on tonal textsetting in Seenku vocal music. This preliminary study looks at just one song, Kôkô tê sô ‘Koko’s Song’. First, I arranged the tones in the lyrics according to notes of the scale, as I did for the balafon surrogate language. The results show no clear patterns: however, if we instead look at tonal transitions, as we did for Tommo So tonal textsetting, the system becomes clear:

We see strict directional tone-tune association. As in Tommo So, contrary mappings (in gray cells) are strongly avoided, but unlike Tommo So, oblique mappings are also heavily dispreferred. In other words, Seenku vocal music follows very strict principles of tonal textsetting, with melodies generally following the prosodic contours laid out by the linguistic tone.

Thus far, Seenku tonal textsetting looks more like Tommo So tonal textsetting than it does the encoding of tone in the surrogate language. Closer inspection reveals that the level of encoding is likewise more akin to the Tommo So case study than the balafon language. Specifically, we see a reemergence of postlexical tone processes in sung music, including contour tone simplification as well as downdrift. Figure 9 shows the second line from the song transcribed musically; dashed lines show progressive downdrift, with the ceiling for S lowering after intervening L and X tones. Additionally, the LS-toned auxiliary sị shows what appears to be contour tone simplification—it is sung on a single note at the same level as the preceding S tone, rather than showing any rising melodic trajectory.

Figure 9: Downdrift (represented with dotted lines) and contour tone simplification in Seenku vocal music
These results, although preliminary, suggest that tonal textsetting in sung music takes place at a higher level, nearer to the surface, than surrogate speech, rather than the differences being related to Tommo So vs. Seenku. This could be due to the fact that segmental information is available in sung music, allowing the artist to stray from underlying lexical or grammatical forms that may help the listener retrieve the message. But if this were the case, we might expect more artistic leeway in general, i.e. that singers could deviate from even the prosodic contours of the language since the consonants and vowels are also sung. Instead, it may be the case that by using the same apparatus as speaking (i.e. the vocal tract rather than a musical instrument), singers employ the same phonological grammar including phonetic implementation that they would when speaking.

More cross-linguistic data are required to test these hypotheses. This is especially true for surrogate languages, where much of the literature is approached from an ethnomusicological angle rather than a linguistic one, meaning that subtler aspects of the phonology may not be addressed. But more work is still needed for tonal textsetting as well; though it has a relatively large literature, many studies simply report statistics for rates of association without investigating the effects of phonological and grammatical factors that can help pinpoint the level at which the singers are musically encoding phonological structure. By asking more detailed questions about how linguistic structure is musically adapted, we can in turn use the results to ask more detailed questions of the phonological structure itself.

6 Conclusion

In this paper, I have shown how musical adaptation of language can provide a unique source of evidence for speakers’ implicit knowledge of phonological structure (e.g. the mora and the syllable in Japanese, Starr and Shih 2017, or differences between lexical and postlexical tone in the balafon surrogate language), allophonic variation and postlexical processes (Seenku tonal textsetting), and even surface phonetic details (Tommo So tonal textsetting; see also Katz 2015 on the phonetic details of imperfect rhyme in hip-hop). Musicians’ sensitivities to these different levels of representation can provide evidence for boundaries between components of grammar (e.g. morphonology vs. phonology vs. phonetics).

In addition to providing a unique source of data to advance phonological theory, the study of musical adaptation is also a valuable tool in language documentation and description. By failing to encode postlexical phonology, musical surrogate languages may provide a glimpse at underlying forms that can aid in phonological analysis, while their discrete notes can amplify hard-to-hear tonal contrasts. Furthermore, language-based music is a linguistic practice and thus falls well within the bounds of language documentation, which seeks to document the linguistic and communicative practices of a speech community (Himmelmann 1998). Its inclusion in language documentation serves at least two other purposes: 1. Engaging the community, who may be more interested in music than pure language, and 2. documenting traditional music, which in many cases may be even more endangered than the language itself (Marett and Barwick 2003).

Future work will look deeper into Seenku tonal textsetting, expanding the data corpus and investigating the relationship between tonal and melodic interval size. This is an important task both for what it can teach us about Seenku phonological structure and for the fact that Sambla vocal music is a disappearing tradition. Hypotheses arising from musical study about the phonetics-phonology interface or different levels of phonological grammar will also need to be corroborated with non-musical evidence in both Tommo So and Seenku. Finally, we will need a larger cross-linguistic and cross-modal study of musical adaptation. This will allow us to answer questions like, do surrogate languages typically encode underlying structure? Is textsetting of vocal music usually sensitive to surface realization? Where does an “oral” surrogate like whistled speech (e.g. Riallant 2005) fall?

As phonologists, we are uniquely suited to study musical adaptation. In exchange, it stands to open up new avenues and insights into phonological structure.

References


Hanson, Kristen (2003). Formal variation in Robert Pinsky’s The Inferno of Dante. Language and Literature 12, 309–337.


