Musical adaptation as phonological evidence: Case studies from textsetting, rhyme, and musical surrogates

Laura McPherson (Dartmouth College)

Abstract

Music and language are both inherently human, rule-governed systems. Language-based music, including vocal music and musical surrogate languages, offers the opportunity to look at how musical and linguistic grammars interact, and how lingusitic structure is adapted to another modality. This article focuses on how this musical adaptation of language can provide valuable evidence for phonological structure and theory, drawing on case studies from textsetting (tonal and non-tonal), rhyme, and musical surrogate languages.

1 Introduction

Recognition of the connection between language and music is by no means new. Ancient philosophers like Plato claimed that the power of music lay in its resemblence to the sounds of speech (Neubauer 1986, cited in Patel 2008), and 18th century enlightenment thinkers like Diderot and Rousseau speculated about the role of music in the emergence of human language (Thomas 1995). The rise of generative linguistics in the 20th century saw those ideas spread to the study of music, with foundational works like Lerdahl and Jackendoff's (1983) *A Generative Theory of Tonal Music*. More recently, a thriving research program on the language-music connection has been working to disentangle similarities and differences in cognitive and neural processing (Patel 2012), ranging from pitch perception (e.g. Schön et al. 2004, Bidelman et al. 2010), to prosodic processing (e.g. Patel et al. 1998, Hausen et al. 2013), to syntactic parallels (e.g. Fiveash and Pammer 2014, Asano and Boeckx 2015).

The broader literature on structural, cultural, and cognitive parallels between language and music is vast and far beyond the scope of a single article. Here, I focus on connections between music and phonology, and more narrowly on how phonologists can use data from music and

musical adaptation to advance phonological theory.[1] The comparison of music and phonology in particular is a natural one. Both music and phonology involve the organization of sound into structured units, following a set of either implicit or explicit rules. In what I call **language-based music** (musical genres that involve the adaptation of linguistic form to musical settings), we have the opportunity to see how these systems of organization (let's call them grammars) accommodate one another.

In many ways, musical adaptation of language is similar to metrical verse; in both systems, linguistic structure must be properly chosen and aligned with an externally imposed system of organization. In music, this is melody, phrasing, and rhythm, while in metrical verse, it is the alignment of linguistic material to a metrical grid of strong and weak positions. Both metrical verse and music are sources of so-called "external evidence" for phonological grammar (Churma 1979); external evidence can be crucial in determining the productivity of phonological patterns and generalizations, especially when existing words or linguistic constructions displaying them ("internal evidence") may be limited.

Generative metrics—the formal study of the rules mapping linguistic structure to metrical grids—has long enjoyed a symbiotic relationship with phonological theory; for an overview of the topic, see Blumenfeld (2016). However, despite many commonalities, musical adaptation of language has remained closer to the fringes of phonological theory. In this paper, I hope to demonstrate that language-based music is a largely untapped source of evidence for phonological structure and the phonetics-phonology interface. I consider three main areas of research into the phonology-music connection. First, in §2, I address textsetting studies, both tonal and non-tonal, showing how they have provided evidence for phonological constituents, syllable weight, phonetic duration, and postlexical tone processes. Next, I turn to studies of rhyme in §3. While rhyme in principle is tied to the text itself and not its musical adaptation, I show how broadening the scope of studies from traditional poetics to more musical genres like rap has resulted in novel findings on the role of perceptual similarity and the phonetics-phonology interface. Finally, in §4, I look at musical surrogate languages, cases where musical form has

[1] For a comprehensive comparison of syntactic structure in language and music, see Katz and Pesetsky 2011 and references therein.

become a medium of linguistic communication. Though less widely studied than textsetting or rhyme, results of these studies provide evidence for psychologically real distinctions between lexical and postlexical phonology as well as evidence for V-to-V intervals rather than syllables in rhythmic organization. I lay out conclusions and directions for future research in §5.

Throughout the paper, I will not be focusing on the grammatical systems of these language-based musical genres—that is, the subconscious rules that musicians follow to map linguistic systems to musical form. These systems make an interesting field of study in their own right, with universal tendencies emerging as the body of literature grows (see citations in each section below). My main focus in this paper is how the study of language-based music fits into linguistic and specifically phonological theory.

2 Textsetting

Studies of textsetting look at the way a linguistic text is set or aligned to a tune or a musical melody. In musical traditions with explicit meter (with meter here referring to the regular alternation of strong and weak beats), textsetting is the musical equivalent of generative metrics and uses much of the same methodology;[2] see, for instance, Halle and Lerdahl (1993) on matching musical and linguistic grids, and Halle (2011) and Hayes and Schuh (to appear) as examples of work explicitly referring to the two under a single heading. However, with the introduction of a new modality (music), there are additional parameters modulating how a line is set. For instance, music displays isochronous meter, where a cycle of strong and weak beats of varying levels could in principle roll continuously for as long as the musician wants. The level of the "beat" (what we tap along to) is often referred to as the tactus (Lerdahl and Jackendoff 1983), but subdivisions and higher groupings can play a role musically and in textsetting. When texts are set to this rhythmic grid, syllables can take up varying numbers of beats, resulting in notes of varying durations; this duration parameter arguably plays no direct role in traditional metrics,

[2]The two traditions differ in that in traditional metrics, the rhythm is created by the syllables of the text itself rather than from an independent musical beat, where syllables able to be stretched or compressed into its bounds. Classic metrics evaluates the degree to which the actual attested lines of poetry match the abstract pattern of light and heavy (or stressed and unstressed) syllables demanded by the meter.

where a metrical template simply stipulates the distribution of stressed and unstressed or heavy and light syllables with no explicit reference to duration. Musical syntax also introduces hierarchical structure at the melodic level such that certain notes are more prominent than others, and this too can influence the ways in which a text is set to music.

As pointed out by Proto (2015), textsetting studies in tonal languages tend to focus on the question of how linguistic tone is set to musical melody, while textsetting in other languages tends to focus more on questions of general prominence and phrasing, including a greater emphasis on rhythm, though see the discussion of Wee (2007) below in §2.2 for the role of prominence in Mandarin folk songs. I will address these latter studies first under the heading of **non-tonal textsetting** then return to studies of **tonal textsetting**.

2.1 Non-tonal textsetting

The rules governing textsetting have been studied for a great many non-tonal languages, including English (e.g. Palmer and Kelly 1992, Halle and Lerdahl 1993, Hayes and Kaun 1996, Hayes 2009), French (e.g. Dell 1989, Dell and Halle 2009), Spanish (e.g. Janda and Morgan 1988, Rodríguez-Vázquez 2010), German (e.g. Proto 2013), Italian (e.g. Proto and Dell 2013), Japanese[3] (e.g. Hayes and Swiger 2008, Starr and Shih 2017), Tashlhiyt Berber (e.g. Dell and Elmedlaoui 2008, Dell 2011), Warlpiri (e.g. Turpin and Laughren 2013), among others. As even this selection shows, there is a vast literature on this form of musical adaptation, including comparative studies that seek to identify universals in the rules that govern how text is set to music. For a good overview of the issues, I refer the reader to Proto (2015).

Here, I will provide two examples of how textsetting studies can shed light on questions of phonological theory.

Hayes and Kaun (1996) investigated textsetting in English folk songs, including both their original forms (as transcribed by Sharp 1916) and in extemporaneously chanted form by ten native English-speaking consultants. Many of their results regarding the mapping between

[3]One could argue that Japanese is a tone language, given the presence of contrastive pitch accent, but I will group it with non-tonal languages here as the studies that have focused on it look not at pitch but rather at rhythm and the alignment of prosodic constituents.

prosodic structure and the metrical grid mirror findings from generative metrics, including stricter mappings within words than across them and stricter mapping at the ends of constituents (with strictness increasing at higher levels of structure). As I outline here, their study also highlights how music (here, sung verse) can provide a source of data for phonological theory.

First, the textsetting data provide evidence for the layers of prosodic structure (word, clitic group, phonological phrase, intonational phrase) proposed based on non-musical data; in the mid-90s, at the time their study was published, the clitic group was a relatively recent (and sometimes controversial) proposal outside of metrics (e.g. Nespor and Vogel 1982, Zec 1988, Hayes 1989, Inkelas 1989)—Hayes and Kaun's textsetting results bolster the arguments for its existence, as they find a boost in strictness of mapping at boundaries correlated with clitic groups, i.e. intermediate between the word and the phonological phrase. For an overview of the role of prosodic structure in textsetting, see also Palmer and Kelly (1992). Further, their results show that textsetting is sensitive not just to phonological configurations but also phonetic details of speech in those configurations. In particular, studies have shown that syllables are phonetically lengthened at the right edge of constituents, with the amount of lengthening sensitive to the level of the hierarchy of the boundary (higher constituents result in more lengthening); see Ladd and Campbell (1991), Wightman et al. (1992), among others. Hayes and Kaun show that this phonetic lengthening plays a determining role in textsetting, with syllables found at the right edges of progressively higher constituents more and more likely to be textset across multiple beats. The effect of phonetic duration is likewise demonstrated for syllable weight; heavy syllables like *town* are more likely to be textset on longer notes than light syllables, like the first syllable of *city*. What makes the textsetting findings particularly interesting is that it shows that speakers are at least subconsciously aware of gradient durational differences, which can exert an influence on how language is mapped to musical form. A similar sensitivity to phonetic duration in textsetting has been demonstrated for Hausa *rajaz* (Hayes and Schuh in press).

Textsetting studies have also been used to provide evidence for phonological structure below the word level, namely the syllable and the mora. The mora has played a key role in generative metrics and textsetting, since it is often assumed that the distinction between heavy and light

syllables arises from the presence of one vs. two moras associated with the syllable (see e.g. Minkova and Stockwell 1994, Ryan 2011, Hayes and Schuh in press). Whether all languages have the mora as a unit, regardless of evidence for phonological weight, remains an open question, but it has long been assumed that the syllable level is universal (Selkirk 1982, Kubozono 2003, Hyman 2008, etc.). This is true even in a language like Japanese, where the mora has been shown to play a more active role in phonological organization (Vance 1987, Otake et al. 1993, Inaba 1998, etc.). Nevertheless, in a recent paper, Labrune (2012) made the provocative claim that Japanese shows no evidence for the syllable, thus drawing into question its universality as a prosodic constituent.

One of Labrune's arguments against the syllable came from textsetting, where she claimed that "the mora is the metric unit of Japanese verse in poetry and singing" (116). This is despite the fact that earlier work on textsetting in Japanese children's songs showed that both the mora and the syllable play a role (Vance 1987). Following this earlier work, Starr and Shih (2017) looked at textsetting in a corpus of songs from animated features including both native Japanese anime theme songs and translated songs from Disney movies to determine which prosodic constituents (syllables or moras) play a role in setting text to music. Both native and translated songs were included in the corpus to vary the likelihood of finding moraic vs. syllabic settings: Native Japanese singing is predicted to be more moraic, since moraic textsetting (e.g. singing *suki* 'like' across two notes) is more sonorous (avoiding consonant clusters and realizing each vocalic mora on a note) than a syllabic setting like *ski* on a single note, though this more closely approximates the spoken realization [sɯ̥ki], where the high vowel devoices and becomes nearly inaudible between the two voiceless consonants. The translated songs, on the other hand, face the challenge of fitting the same meanings into roughly the number of notes and syllables used for English, which has a much higher syllable-based information density and hence shorter words. Thus, it may be predicted that translated songs will rely on syllabic textsetting in an effort to pack more meaning into the available space. In order for this to be possible, though, Japanese must have the unit "syllable" as part of its prosodic hierarachy.

These predictions are borne out: translated songs show greater proportions of syllabic textset-

ting, but interestingly, even native Japanese songs can sometimes have upwards of 25% syllabic textsetting, depending on the kind of bimoraic sequence being considered (e.g. long vowels vs. Vi sequences); see Figure 1, reproduced with permission from Starr and Shih (2017).

[Figure 1 here]: Proportion of syllabic and moraic settings in Disney and animated songs; reproduced from Starr and Shih (2017:17)

The authors take these results as evidence that contra Labrune (2012), Japanese phonology must have the syllable level as part of the prosodic hierarchy, and as such, the proposed universality of the syllable remains (for the time being) unchallenged. See Kawahara (2016) for a review of other evidence for the syllable in Japanese.

Most textsetting studies are based on relatively well studied languages, like English, French, Italian, or even Hausa (see references to each above), allowing them to presuppose the prosodic structure of the language and hence focus on the mapping between that structure and the metrical grid. But as the Starr and Shih (2017) study shows, we can use patterns of textsetting to work backwards and adduce those very stuctures. If a larger body of work emerges showing strong trends or even universals in how prosodic structure is mapped to musical structure, then we could confidently take textsetting data as evidence for prosodic structure in the phonology of a wide range of languages, including underdescribed languages.

2.2 Tonal textsetting

While studies of textsetting in non-tonal languages focus on questions of phrasing and prominence, correlated with rhythm and duration, studies of tonal languages tend to look instead at how linguistic tone interacts with or is set to musical melody. Hence, these studies often leave aside rhythm in favor of looking at the notes of the music. The general question remains the same: How is linguistic structure, in this case tone, matched with musical structure, and what do these mappings tell us about either the language, the music, or the grammar of textsetting?

Studies in tonal textsetting date back nearly a century. In a foundational paper, Herzog (1934) investigated the relationship between speech tone and musical melody in Jabo (Kru) and Navajo (Athabaskan) music. His was one of the first studies to show that tone is neither

"slavishly" adhered to by the singer nor is it irrelevant in textsetting, but also that the degree of correspondence varies by song genre. Studies of tonal textsetting in a variety of languages spanning the continents have followed, with an equally wide range of results, from languages like Cantonese with 92% parallel[4] mappings between tone and tune (Wong and Diehl 2002) to Shona with only 53% parallel mappings (Schellenberg 2009). For literature reviews of tonal textsetting, see Schellenberg (2012a), McPherson and Ryan (2018), and Ladd and Kirby (forthcoming).

Investigation of tonal textsetting can provide insights into the tonal structure of the language. First, textsetting studies reveal that in many languages, it is a surface or postlexical level of tone, rather than a deeper underlying or lexical level, that determines the musical match. For instance, in an early study of Hausa tonal textsetting, Richards (1972) argued that Hausa singers pay little attention to the relationship between tone and tune, with only 49% of tonal sequences matched by musical trajectory. However, Leben (1983) reassessed the data and showed that if we consider not the underlying tone, as Richards did, but rather the surface realization of that tone in speech, then the match is close to 100%. There are two crucial facts about Hausa intonation that he argues account for the supposed "non-matches": the first is downdrift, where the second H in a HLH sequence is pronounced substantially lower (often at the same level as the preceding L). The second is that L tones are quite variable in their pronunciation in Hausa, falling somewhere in an acceptable range rather than being clearly tied to a target. Leben proposes that singers are sensitive to this phonetic range and assign a L tone to an extreme within it when singing. If this is the case, we may be able to use data from Hausa textsetting to examine whether there is a pattern to where in the range a L tone is assigned and make predictions about the realization of L tone in spoken Hausa (to be corroborated by further study). Evidence for similar sensitivity to surface rather than underlying tone can be found in a preliminary study of Seenku (Mande, Burkina Faso) vocal music (McPherson and James 2019), which I will briefly return to in §4 below.

In a study of tonal textsetting in Tommo So (Dogon) folk songs, McPherson and Ryan (2018)

[4]In the tonal textsetting literature, "parallel" refers to the configuration in which the pitch change between two tones and their corresponding notes goes in the same direction, e.g. a rising tonal transition from L to H sung on a rising melody or a falling tonal transition sung on a falling melody.

show two other tonal phenomena brought to light by textsetting. Tommo So is a two tone language, contrasting H and L tone (along with surface underspecified syllables, McPherson 2011). Though the authors looked at the mappings of all tonal bigrams (i.e. L-H, H-H, H-L, L-L), statistical analysis revealed that singers appear to be sensitive instead to a binary distinction between rising tonal sequences (L-H) versus non-rising (everything else); the former can be sung on a rising or level melody, and the latter can be sung on either level or falling melodies. Looking back at spoken language data, this effect could be explained either by the largely binary pattern of lexical tone melodies (/LH/ vs. /H/, McPherson 2013) or by the phonetic realization of tone, wherein both level and falling tone sequences tend to fall in f0 and only rising sequences are allowed to rise; both rising sequences and level sequences can, depending upon the environment, be pronounced with level f0. Further study could tease apart what exactly Tommo So speakers are sensitive to when singing. For instance, systematic study of how the tones of lyrics are pronounced when speaking could show it is the phonetic realization of tone that determines tonal textsetting. Conversely, psycholinguistic experiments on perceived similarity between rising, level, and falling tonal melodies could reveal that the binary categorical distinction between rising and non-rising tone sequences is psychologically real and part of the phonological grammar.

The second phenomenon found by McPherson and Ryan (2018) is that singers show stricter adherence to lexical tone than grammatical tone. Grammatical tone in Tommo So takes the form of replacive tonal melodies that overwrite lexical tone in different syntactic configurations (Heath and McPherson 2013, McPherson 2014, McPherson and Heath 2016). For instance, the LH-toned noun *bàbé* can surface with an all {H} melody after a pronominal possessor (*mí* ^{H}*bábé* 'my uncle') but an all {L} melody before an adjective (*bàbè*^{L} *kómmó* 'skinny uncle'). Deeper analysis of grammatical tone sequences in the sung corpus shows that many of the mismatches can be explained by taking into consideration the underlying lexical tone of the words in question. For instance, a grammatically (surface) L-toned word with underlying LH tone is more likely to be sung on a rising melody than a grammatically (surface) L-toned word with underlying HL tone. This result raises interesting questions about the mental representation of gram-

matical tone, particularly the replacive grammatical tone that characterizes Dogon languages. Are both lexical and grammatical melodies activated when speaking a word with a replacive overlay? Is there a surface phonetic effect (incomplete neutralization) of this dual activation in speech? And if so, are singers sensitive to surface phonetics or to the activation that underlies it? This example highlights how results of musical study can provide direction for deeper study of phonetic and phonological structure in spoken language. For further discussion, see McPherson (2019b).

In some languages, tonal textsetting provides evidence of prominence effects, both tonally and metrically. Drawing on a study of textsetting in ten folk songs, Wee (2007) argues for the notion of "tonal headedness" in Mandarin Chinese. He shows that the four tones of Mandarin can be understood in terms of low and high ([l] and [h]) elements, such that Tone 1 is [h], Tone 2 is [lh], Tone 3 is [l], and Tone 4 is [hl]. He takes the rightmost pitch element in each case as the head of the tone, as evidenced by the fact that it is this latter pitch that is held when the syllable is lengthened. This divides the tones into two classes, +H (1 and 2) and -H (3 and 4). Wee shows that textsetting in Mandarin folk songs is sensitive not to every tonal and melodic transition but rather to transitions to and from the head tones on metrically prominent syllables (head syllables) in the music, unless the head syllable is reduplicated, in which case it has greater melodic freedom; 97% of the corpus adheres to these rules. Thus, even in tonal textsetting, metrical prominence may play a role.

In all of these studies, it should be noted that the degree of tone-tune association depends not just on the language but also on the genre. To take Mandarin as an example, other authors have found no effect of lexical tone on melody in popular music (Ho 2006, Schellenberg 2012b), while lexical tone essentially defines the melody in genres like chants or Beijing opera (Chao-Pian 2000), though even in this latter case, musical considerations may override linguistic ones (Stock 1999). Thus, while there is a clear connection between a language's phonology and its music, this connection may be mediated by cultural considerations.

Many open questions remain regarding tonal textsetting: First, how should contour tones be treated in calculating tone-tune association? Do both initial and final components play a

role? Results thus far have been varied, with Wong and Diehl (2002) arguing that Cantonese textsetting is sensitive to the endpoint of contours, while Schellenberg and Gick (2018) show that singers actually represent rising contour tones with microtonal variation; for Mandarin, on the other hand, Wee (2007) points to the role of tonal heads rather than endpoints per se. More studies across a more typologically diverse set of languages will be required to identify any universal trends in the textsetting of contour tones. Second, is the sensitivity to surface rather than underlying melodies universal? In other words, can we generalize Leben's (1983) findings on the role of intonation in Hausa? Third, is there a correlation between tonal and musical interval size (i.e. are larger tonal intervals mapped to larger musical intervals)? Preliminary results from Seenku suggest this is the case (McPherson and James 2019), and Tommo So tonal textsetting likewise showed that contrary mappings were more heavily penalized on larger musical intervals (McPherson and Ryan 2018). Finally, does musical syntax and its resulting prominence asymmetries between notes play any role in tonal textsetting? That is, do singers show any special treatment of the tonic vs. the leading tone, etc. that would influence the way in which the text is set to music? This last question also suggests that prominence, phrasing, and rhythm ought to be considered together with melody in the textsetting of tonal languages, where, with some exceptions (e.g. Wee 2007), such questions are typically set aside.

3 Rhyme

The topic of rhyme falls partially outside of the scope of music and phonology, since it is in large part a question of the text itself and not of music per se. I will briefly summarize two studies looking specifically at rhyme in musical genres rather than written poetry, specifically studies of rap lyrics. These examples further highlight the role that data from music can play in phonological analysis.

We can divide rhyme into two classes, perfect rhyme and imperfect or half rhyme. For a perfect rhyme, the material in the two "rhyme domains" (typically the rime of a stressed syllable and any unstressed syllables following it, Holtman 1996) must be an exact match, e.g. *toaster* and *poster* . Half rhymes show some mismatch between the corresponding rhyme domains. For

instance, Katz (2015) gives the example of *right* vs. *pipe* (from Nas, *The World is Yours*), which share the same nucleus but whose coda consonants differ in place (but not manner or voicing).

Rhyme is a particularly good source of external evidence for phonetic perceptibility, i.e. that perceptual similarity can shape phonological patterns (see Steriade 2009 [2001] *et seq.*). Steriade (2003), for instance, shows that half rhyme in Romanian poetry systematically reflects common cross-linguistic phonological processes like final devoicing, even though those patterns are not an active part of Romanian phonology. She takes this as evidence that phonetic perceptibility must be part of a speaker's linguistic competence.

In the more musical domain, this result is echoed in studies of rap lyrics. Looking at Japanese rap, Kawahara (2007) shows how phonetic similarity (calculated in part by looking at shared features) positively correlates with the observed over expected values for the correspondence between consonants in rhyming environments. In other words, the more phonological features a pair of consonants like [s] and [h] share (looking at surface allophones rather than underlying forms), the more likely they are to be found standing in rhyming correspondence. He goes on to show that **all** features contribute to phonetic similarity and hence rhymability; in other words, it is not the case that place, manner or voicing features are privileged in determining similarity. However, similarity can also be acoustic; for instance, rhyming correspondence is more likely between dorsals and pharyngeals, despite the fact that they are not featurally similar. Finally, the half rhyme data from rap corroborate a finding from cooccurrence restrictions in Yamato Japanese vocabulary: while homorganic consonants are generally underrepresented in adjacent onset positions, coronal sonorants and obstruents are *not* underrepresented, suggesting that they are treated as different classes of sounds in Japanese grammar. This would predict that coronal sonorants and obstruents make a worse rhyming pair than other homorganic consonant pairs, a prediction that is upheld by the data. In the conclusion, Kawahara highlights an important point: Japanese folk tradition for rhyme states that consonants do not play any role, hence the attested distribution of half rhyme is not the result of explicitly learned poetic rules. Rather, rap artists' sensitivity to principles of phonetic similarity tells us something about their underlying linguistic system and their subconscious knowledge of it.

Katz (2015) undertakes a similar study looking at African-American English rap lyrics. His general finding echoes that of Kawahara (2007), namely that sounds that are phonetically similar are more likely to stand in rhyming correspondence. However, he argues that what is behind the patterns of half rhyme is **perceptual** rather than featural similarity, with the interesting result that pairs that are more likely to be treated as rhyming are the same pairs that are cross-linguistically likely to display neutralization. For instance, languages are most likely to neutralize voicing distinctions in obstruents before other obstruents, followed by word-final position, and they are the least likely to lack a voicing distinction between sonorants. In the rap corpus, this typology is mirrored: words with a voicing mismatch before obstruents were the most likely to form half rhyme pairs, followed by words with a voicing mismatch word-finally, and only then by words with a voicing mismatch between sonorants. In short, like Steriade (2003), this study shows that musical adaptation of language (here, rhyme in rap) provides evidence for perceptual universals, even in a language where their effects have not been phonologized (English).

For other studies of rhyme in musical genres (as opposed to non-musical poetic form), see Zwicky (1976), Holtman (1996), and Horn (2010).

4 Surrogate languages

The last form of musical adaptation I will consider is more musical in form if less musical in function: musical surrogate languages. In a musical surrogate language (Stern 1957, Sebeok and Umiker-Sebeok 1976, etc.), linguistic form is mapped to a musical modality, with the resulting system used for communication. Thus, its substance may be purely musical but its purpose is largely communication rather than artistic.

Speech surrogates are found all around the world on a wide variety of instruments. The best known cases involve drums (e.g. Beier 1954 on the Yorùbá of Nigeria, Locke and Agbeli 1981 on the Ewe of Ghana and Togo, Winter 2014 on the Sabar of Senegal, Thiesen 1969, Seifert et al. 2018 on the Amazonian Bora), giving rise to the moniker “talking drums”. But surrogate systems are also attested on flutes (e.g. Hurley 1968 on the Mexican Kickapoo, Moore and Meyer 2014 on the Amazonian Gavião, etc.), jaw harps (e.g. Khmu, Proschan 1994; Hmong,

Falk 2008, etc.), horns (e.g. Asante, Kaminski 2008, Sambla, McPherson in press), xylophones (e.g. Senoufo, Zemp and Soro 2014, Sambla, McPherson in press) and fiddles (Bebey 1975 quoted in Arhine 2009). There is also a sizable literature on whistled surrogate languages, which I will not focus on here; for an overview, see Rialland (2003, 2005).

Stern (1957) divides speech surrogate systems into two types: "abridging" systems that encode phonemic aspects of the language (stress, rhythm, pitch, etc.) and "lexical ideogram" systems that encode meanings directly without any explicit reference to the way the language sounds. For the purposes of this article, I focus on the former, which stand to shed light on the nature of phonological representations.

Until recently, surrogate speech has remained more of an area of study for ethnomusicologists than for linguists, but this is a missed opportunity (see also McPherson 2019b, McPherson in press). Not only can linguists offer valuable insights into what phonological aspects are being encoded, but the surrogate data themselves can be a source of evidence for phonological representations and theory, analagous to data from textsetting. I will highlight two case studies here.

The first case study looks at surrogate speech of the Sambla balafon, a kind of resonator xylophone (McPherson in press). The system is based on the spoken language, a Mande language known as Seenku, whose complex tone systems includes four phonemic levels and multiple contour tones (McPherson 2019c). These tones are mapped to the notes of the Sambla pentatonic scale (consisting of roughly notes 1, b3, 3, 5 and 6 of a Western major scale), with the exact relationship between tone and notes dependent on the mode of the song. In addition to encoding the language's tone system, balafon surrogate speech also encodes rhythmic aspects of the language, including the distinction between short and long vowels as well as the distinction between monosyllables and "sesquisyllables" (words like *məni̋* 'woman' or *kərɛ̂* 'man', which begin with a short half syllable, Matisoff 1990, Pittayaporn 2015). Characteristically of many musical surrogate systems (at least tone-based systems), segmental identity (consonants and vowels) is not encoded. For details of the grammar of the speech surrogate, see McPherson (in press).

Most of the textsetting and rhyme studies discussed above highlighted speakers' sensitivity to phonetic details of the language—gradient lengthening, perceptual similarity, intonational effects in tone. The balafon surrogate language provides evidence for deeper phonological structure, namely a level like the underlying form or lexical representations. In surrogate speech, both lexical tone melodies and grammatical tone (including ostensibly meaningless but morphosyntactically constrained sandhi alternations, McPherson 2019c) are exceptionlessly encoded in the notes of the balafon. In contrast, postlexical tonal alternations, including downstep and tonal absorption, are never encoded despite being almost obligatory in spoken language. Thus, if the studies summarized above would suggest that artistic adaptation applies to a surface level of language, the Sambla balafon surrogate demonstrates that this is not universal and that deeper levels of phonological structure are both psychologically real and accessible to speakers; see McPherson (2019b).

Interestingly, this may be a matter of musical modality rather than a language- or culture-wide parameter. Preliminary work by McPherson and James (2019) compares the Sambla balafon to Sambla vocal music, showing that in this modality, singers do encode a surface level of tone, including the output of postlexical tone rules like downstep and tonal absorption. These results are consistent with tonal textsetting principles seen in other languages, which suggests that the same linguistic system can give rise to different levels of musical encoding depending upon modality. In short, this study points to the benefit of analyzing multiple facets of musical adaptation within a single language, which can reveal a highly nuanced structure, with speakers able to tap into different layers depending upon the demands of the task.

Other musical surrogate systems show more evidence for phonetic detail. Seifart et al. (2018) study a drum surrogate system used by the Bora people of the northwestern Amazon. It is similar to the Sambla balafon system in that it encodes tone (in the case of Bora, only two tones, Low and High) and rhythm. Broadly speaking, the number of drum beats equals the number of syllables in a word, but the interbeat durations differ based on syllable shape. In a quantitative study of interbeat duration, the authors show that Bora drummers are closely matching the length of V-to-V intervals in the spoken language (V, VC, VVC, VCC) rather than

syllable types (V, CV, CVV, CVC, VV, VC). They argue that this adds to the growing body of evidence that rhythmic structure in language, including stress systems, is based around V-to-V intervals rather than syllables, given the perceptual and acoustic prominence of the vowel as compared to its surrounding consonants (Steriade 2012, Hirsch 2014, Ryan 2014, Lunden 2017). From a rhythmic perspective, the Bora drumming system confirms results from textsetting, that is that speakers (and musicians) are sensitive to fine-grained duration distinctions in the spoken language, though in the case of the Bora V-to-V intervals, these represent potentially contrastive categories whereas in the case of textsetting, the lengthening effects were non-contrastive and non-categorical, tied instead to phrasing. From a tonal perspective, Bora drumming corroborates the findings from the Sambla balafon system, with intonational effects like downdrift not represented, though unlike the balafon with its wide range of possible notes, the Bora slit log drums set strict limits on how tones can be produced.

Compared to textsetting, detailed linguistic study of musical speech surrogates remains in its infancy. These systems are likely to provide fertile ground for future discoveries on what musicians are sensitive to when encoding linguistic form.

5 Conclusions

With language and music both universal aspects of human expression, it should come as little surprise that there is a tight connection between the two. In this paper, I have shown how musical adaptation of linguistic form can provide valuable data for phonological theory and the phonetics-phonology interface. Studies of textsetting, for instance, give evidence of phonological constituents, from phonological phrases down to the level of the syllable or the mora. In cases where the prosodic hierarchy for a particular language is called into question, textsetting can help settle the score, as Starr and Shih (2017) showed for Japanese. Many studies also reveal a high level of sensitivity to phonetic detail of rhythm and duration, whether due to segmental make up (as in the V-to-V intervals in Bora drumming, Seifart et al. 2018) or phrasing effects (as in English folk songs, Hayes and Kaun 1996). Speakers' behavior with respect to tone appears to depend in part on the genre of musical adaptation: tonal textsetting appears to apply to

a relatively surface level of tone, taking into account intonational effects and postlexical rules (as in Hausa, Leben 1983), while speech surrogate systems encode a deeper phonemic level of tone (as in Seenku, McPherson in press).

Much work remains to be done on the relationship between music and phonology. More case studies will allow us to develop a better typology of cross-linguistic musical adaptation and identify potential universals. These universals can then in turn be used to "work backwards" from musical adaptation to identify constituents, structures, and phonological rules in the languages themselves, especially understudied languages or languages for which only a limited range of (mostly musical) data are available.

References

Arhine, Adwoa. 2009. Speech surrogates of Africa: a study of Fante Mmensuon. *Legon Journal of the Humanities* 20:105–122.

Asano, Rie, and Cedric Boeckx. 2015. Syntax in language and music: what is the right level of comparison? *Frontiers in Psychology* 6:942.

Beier, Ulli. 1954. The talking drums fo the Yoruba. *African Music* 1:29–31.

Bidelman, Gavin M., Jackson T. Gandour, and Ananthanarayan Krishnan. 2010. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience* 23:425–434.

Blumenfeld, Lev. 2016. Generative metrics: an overview. *Language and Linguistics Compass* 10:413–430.

Chao-Pian, Rulan. 2000. Tone and tone: applying musical elements to Chinese words. *Journal of Chinese Linguistics* 28:181–200.

Dell, Francois. 1989. Concordances rythmiques entre la musique et les paroles dans le chant. l'accent et l'e muet dans la chanson francaise. In *Le souci des apparences*, ed. M. Dominicy, 121–136. Brussels: Editions de l'Université de Bruxelles.

Dell, Francois. 2011. Singing in Tashlhiyt Berber, a language that allows vowel-less syllables. In *Handbook of the syllable*, ed. Charles E. Cairns and Eric Raimy, 173–193. Leiden: Brill.

Dell, Francois, and Mohamed Elmedlaoui. 2008. *Poetic meter and musical form in Tashlhiyt Berber songs*. Cologne: Köppe.

Dell, Francois, and John Halle. 2009. Comparing musical textsetting in French and in English songs. In *Towards a typology of poetic forms: form language to metrics and beyond*, ed. Jean-Louis Aroui and Andy Arleo, 63–78. Amsterdam: John Benjamins.

Falk, Catherine. 2008. "if you have good knowledge, close it well tight": Concealed and framed meaning in the funeral music of the Hmong qeej. *British Journal of Ethnomusicology* 12:1–33.

Fiveash, Anna, and Kristen Pammer. 2014. Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music* 42:190–209.

Halle, John. 2011. Poetic meter and musical form in Tashlhiyt Berber songs (review). *Language* 87:181–189.

Halle, John, and Fred Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55:3–23.

Hausen, Maija, Ritva Torppa, Viljami R. Salmela, Martti Vainio, and Teppo Särkämö. 2013. Music and speech prosody: a common rhythm. *Frontiers in Psychology* 4:566.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In *Rhythm and meter*, ed. Paul Kiparsky and Gilbert Youmans, 201–260. San Diego: Academic Press.

Hayes, Bruce. 2009. Textsetting as constraint conflict. In *Towards a typology of poetic forms: form language to metrics and beyond*, ed. Jean-Louis Aroui and Andy Arleo, 43–62. Benjamins.

Hayes, Bruce, and Abigail Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13:243–304.

Hayes, Bruce, and Russell Schuh. To appear. Metrical structure and sung rhythm of the Hausa Rajaz. *Phonological Analysis* .

Hayes, Bruce, and Tami Swiger. 2008. Two Japanese children's songs. Ms. UCLA.

Heath, Jeffrey, and Laura McPherson. 2013. Tonosyntax and reference restriction in Dogon NPs. *Language* 89:265–296.

Herzog, George. 1934. Speech-melody and primitive music. *Musical Quarterly* 20:452–466.

Hirsch, Aron. 2014. What is the domain for weight computation: the syllable or the interval? In *Proceedings of the 2013 Annual Meeting on Phonology*, DOI: http://dx.doi.org/10.3765/amp.v1i1.21.

Holtman, Astrid. 1996. A generative theory of rhyme. Doctoral Dissertation, Utrecht Institute of Linguistics, Utrecht.

Horn, Elizabeth. 2010. Poetic organization and poetic license in the lyrics of Hank Williams, Sr. and Snoop Dogg. Doctoral Dissertation, University of Texas at Austin.

Hurley, William M. 1968. The Kickapoo whistle system: a speech surrogate. *Plains Anthropologist* 13:242–247.

Hyman, Larry. 2008. Universals in phonology. *The Linguistic Review* 25:83–137.

Inaba, Seiichiro. 1998. Moras, syllables and feet in Japanese. *Language, information and computation* 106–117.

Inkelas, Sharon. 1989. Prosodic constituency in the lexicon. Doctoral Dissertation, Stanford University.

Janda, Richard, and Terrell Morgan. 1988. El acentó dislocadó – pues cantadó – castellanó: on explaining stress-shift in song-texts from Spanish (and certain other Romance languages). In *Advances in Romance linguistics*, ed. David Birdsong and Jean-Pierre Montreuil, 151–170. Dordrecht: Foris.

Kaminski, Joseph. 2008. Surrogate speech of the Asante ivory trumpeters of Ghana. *Yearbook for Traditional Music* 40:117–135.

Katz, Jonah. 2015. Hip-hop rhymes reiterate phonological typology. *Lingua* 160:54–74.

Katz, Jonah, and David Pesetsky. 2011. The identity thesis for language and music. Ms., MIT.

Kawahara, Shigeto. 2007. Half-rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16:113–144.

Kawahara, Shigeto. 2016. Japanese has syllables: a reply to Labrune (2012). *Phonology* 33:169–194.

Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In *The syllable in Optimality Theory*, ed. Caroline Féry, 99–122. Cambridge University Press.

Labrune, Laurence. 2012. Questioning the universality of the syllable: evidence from Japanese. *Phonology* 29:113–152.

Ladd, D. Robert, and W.N. Campbell. 1991. Theories of prosodic structure: evidence from syllable duration. In *Actes du XIIème congrès international des sciences phonétiques*, volume 2, 290–293. Aix-en-Provence: Services des Publications, Université de Provence.

Ladd, D. Robert, and James Kirby. In press. Tone-melody matching in tone language singing. In *The Oxford handbook of prosody*, ed. Carlos Gussenhoven and Aoju Chen. Oxford University Press.

Leben, William R. 1983. The correspondence between linguistic tone and musical melody. In *Proceedings of the Ninth Annual Meeting of the Berkeley Linguistics Society*, 148–157.

Lerdahl, Fred, and Ray Jackendoff. 1983. *A generative theory of tonal music*. Cambridge: MIT Press.

Locke, David, and Godwin Agbeli. 1981. Drum language in Adzogbo. *The Black Perspective in Music* 9:25–50.

Lunden, Anya. 2017. Syllable weight and duration: a rhyme/intervals comparison. In *Proceedings of the Linguistic Society of America*, DOI: http://dx.doi.org/10.3765/plsa.v2i0.4084.

Matisoff, James A. 1990. Bulging monosyllables: Areal tendencies in Southeast Asian diachrony. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, 543–559.

McPherson, Laura. 2011. Tonal underspecification and interpolation in Tommo So. Master's thesis, UCLA.

McPherson, Laura. 2013. *A Grammar of Tommo So*. MGL 62. Berlin: De Gruyter Mouton.

McPherson, Laura. 2014. Replacive grammatical tone in the Dogon languages. Doctoral Dissertation, UCLA, Los Angeles, CA.

McPherson, Laura. 2018. The talking balafon of the Sambla: grammatical principles and documentary implications. *Anthropological Linguistics* 60:255–294.

McPherson, Laura. 2019a. Illustration of Seenku. *Journal of the International Phonetic Association* .

McPherson, Laura. 2019b. The role of music in documenting phonological grammar: two case studies from West Africa. In *Proceedings of the 2018 Annual Meeting on Phonology*.

McPherson, Laura. 2019c. Seenku argument-head tone sandhi: allomorph selection in a cyclic grammar. *Glossa* 4:22.

McPherson, Laura, and Jeffrey Heath. 2016. Phrasal grammatical tone in the Dogon languages. *Natural Language and Linguistic Theory* 34:593–639.

McPherson, Laura, and Lucas James. 2019. Artistic adaptation of Seenku tone: musical surrogates vs. vocal music. Handout, 50th Annual Conference on African Linguistics, UBC.

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

Minkova, Donka, and Robert Stockwell. 1994. Syllable weight, prosody and meter in Old English. *Diachronica* 11:35–64.

Moore, Denny, and Julien Meyer. 2014. The study of tone and related phenomena in an Amazonian tone language: Gavião of Rondônia. *Language Documentation and Conservation* 8:613–636.

Nespor, Marina, and Irene Vogel. 1982. Prosodic domains of external sandhi rules. In *The Structure of Phonological Representations*, ed. Harry van der Hulst and Neil Smith, 224–255. Dordrecht: Foris.

Neubauer, John. 1986. *The emancipation of music from language: Departure from mimesis in eighteenth-century aesthetics*. New Haven: Yale University Press.

Otake, Takashi, Giyoo Hatano, Anne Cutler, and Jacques Mehler. 1993. Mora or syllable? speech segmentation in Japanese. *Journal of Memory and Language* 32:258–278.

Palmer, Caroline, and Michael H. Kelly. 1992. Linguistic prosody and musical meter in song. *Journal of Memory and Language* 31:525–542.

Patel, Aniruddh. 2008. *Music, language, and the brain*. Oxford University Press.

Patel, Aniruddh. 2012. Language, music, and the brain: a resource-sharing framework. In *Language and music as cognitive systems*, ed. Patrick Rebuschat, Martin Rohrmeier, John A. Hawkins, and Ian Cross, 204–223. Oxford University Press.

Pittayaporn, Pittayawat. 2015. Typologizing sesquisyllabicity. In *Languages of mainland Southeast Asia: The state of the art*, ed. Nick Enfield and Bernard Comrie, 500–528. Berlin/Boston: De Gruyter Mouton.

Proschan, Frank. 1994. Khmu play languages. *Mon-Khmer Studies* 23:43–65.

Proto, Teresa. 2013. Singing in German: text-setting rules and language rhythm. In *Atti del IX convegno AISV*, ed. Vincenzo Galatà, 9–10. Roma: Bulzoni.

Proto, Teresa. 2015. Prosody, melody and rhythm in vocal music: The problem of textsetting in a linguistic perspective. *Linguistics in the Netherlands* 2015:116–129.

Proto, Teresa, and Francois Dell. 2013. The structure of metrical patterns in tunes and in literary verse: Evidence from discrepancies between musical and linguistic rhythm in Italian songs. *Probus* 25:105–138.

Rialland, Annie. 2003. A new perspective on Silbo Gomero. In *Proceedings of the 15th Congress of Phonetic Sciences*, ed. Maria-Josep Solé, Daniel Recasens, and Joaquin Romero, 2131–2134. Barcelona: Causal Productions.

Rialland, Annie. 2005. Phonological and phonetic aspects of whistled languages. *Phonology* 22:237–271.

Richards, Paul. 1972. A quantitative analysis of the relationship between language tone and melody in a Hausa song. *African Language Studies* 13:137–161.

Rodríguez-Vázquez, Rosalía. 2010. Text-setting constraints: a comparative perspective. *Australian Journal of Linguistics* 30:19–34.

Ryan, Kevin. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology*

28:413–454.

Ryan, Kevin. 2014. Onsets contribute to syllable weight: statistical evidence from stress and meter. *Language* 90:309–341.

Schellenberg, Murray. 2009. Singing in a tone language: Shona. In *Selected Proceedings of the 39th Annual Conference on African Linguistics*, ed. Akinloye Ojo and Lioba Moshi, 137–144. Somerville, MA: Cascadilla Proceedings Project.

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

Schellenberg, Murray. 2012b. Tone realization in sung Mandarin. In *Tonal Aspects of Languages: Third International Symposium*.

Schellenberg, Murray, and Bryan Gick. 2018. Microtonal variation in sung Cantonese. *Phonetica* 1–24.

Schön, Daniele, and Cyrille Magnea nd Mireille Besson. 2004. The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41:341–349.

Seifart, Frank, Julien Meyer, Sven Grawunder, and Laure Dentel. 2018. Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society of Open Science* 5:170354.

Selkirk, Elizabeth. 1982. The syllable. In *The structure of phonological representations*, volume 2, 337–383. Dordrecht: Foris.

Sharp, Cecil J., ed. 1916. *One hundred English folksongs*. Oliver Ditson.

Starr, Rebecca, and Stephanie Shih. 2017. The syllable as a prosodic unit in Japanese lexical strata: evidence from text-setting. *Glossa* 2:1–34.

Steriade, Donca. 2003. Knowledge of perceptual similarity and its phonological uses: evidence from half-rhymes. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 363–366. Barcelona: Causal Productions.

Steriade, Donca. 2009 [2001]. The phonology of perceptibility effect: The P-map and its consequences for constraint organization. In *The nature of the word: Studies in honor of Paul Kiparsky*, ed. Kristin Hanson and Sharon Inkelas, 151–180. MIT Press.

Steriade, Donca. 2012. Intervals vs. syllables as units of linguistic rhythm. Handout, EALING, Paris.

Stern, Theodore. 1957. Drum and whistle "languages": an analysis of speech surrogates. *American Anthropologist* 59:487–506.

Stock, Jonathan P. 1999. A reassessment of the relationship between text, speech, and melody, and aria structure in Beijing Opera. *Journal of Musicological Research* 18:183–206.

Thiesen, Wesley. 1969. The Bora signal drums. *Lore* 19:101–103.

Thomas, Downing A. 1995. *Music and the origins of language: Theories from the French enlightenment*. Cambridge: Cambridge University Press.

Turpin, Myfany, and Mary Laughren. 2013. Text and meter in a Lander Warlpiri. In *Selected papers from the 44th conference of the Australian Linguistic Society*, ed. Lauren Gawne and Jill Vaughan. University of Melbourne Digital Repository.

Vance, Timothy. 1987. *An introduction to Japanese phonology*. Albany, NY: SUNY Press.

Wee, Lian Hee. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics* 35:128–143.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91:1707–1717.

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.

Zec, Draga. 1988. Sonority constraints on prosodic structure. Doctoral Dissertation, Stanford University.

Zemp, Hugo, and Sikaman Soro. 2010. Talking balafons. *African Music* 8:7–24.

Zwicky, Arnold. 1976. Well, this rock and roll has got to stop. *Chicago Linguistics Society* 12:676–697.