

# Documenting musical surrogate languages

LAURA MCPHERSON, Linguistics, Dartmouth College, US, [laura.emcpherson@gmail.com](mailto:laura.emcpherson@gmail.com)

## ABSTRACT

Musical surrogate languages are systems of communication that transpose linguistic content onto a musical instrument. These are found all over the world on a wide range of instruments, yet they remain severely underdocumented in linguistics. This paper is a call to action to include musical surrogate languages in language documentation projects and linguistic theory, collaboratively building the foundation of a new subfield of linguistics: surrogate linguistics. The paper provides a basic typology of musical surrogate systems, offers examples of what we can learn about language by studying musical surrogates, and provides a brief how-to guide for documenting these systems, including ethical and methodological issues around recording, transcribing, and archiving them.

**Keywords:** music; verbal arts; oral literature; speech surrogacy

## 1. INTRODUCTION

Language documentation promotes the recording of diverse genres of speech (Woodbury 2003), including genres that cross boundaries into other scholarly disciplines (Holton 2018). An often overlooked genre is musical surrogate speech: the use of musical instruments to transmit linguistic messages. A related

*Language Documentation and Description* is a peer-reviewed open access journal published by Aperio. © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.



practice, whistled speech, is also a speech surrogate, though it has received more scholarly attention (e.g., Meyer 2015; Riailand 2005), and I will only address it as it relates to musical surrogate speech in this paper. The reasons these practices are overlooked are likely varied. Perhaps they do not seem “linguistic” enough if there are no words articulated by human vocal tracts or hands. Perhaps researchers are not musically inclined and feel out of their depth. Or researchers may simply be unaware of the phenomenon in general and never thought to ask if these form a part of the community’s linguistic practices.

This article is a clarion call to the language documentation community to seek out and make records of these traditions around the world. As I will detail below, musical surrogate languages offer unique insights into linguistic structure and the human language faculty, and they often fill unique communicative niches, shaped by cultural, spiritual, and physical considerations. The recognition and conceptualization of linguistic structure that is required to encode it on a musical instrument makes these traditions a form of ethnosience: linguistic analysis from an emic perspective. They also tend to be highly endangered practices (McPherson 2018; Seifart et al. 2018; Sicoli 2016; Struthers-Young 2022), even if the spoken language itself is stable (Eze & McPherson 2025). This makes it even more important to create records of them now. The creation of these documentary materials has enormous potential to revolutionize how we think about language and music as interrelated forms of human expression and how people develop speech technologies (and relatedly, emic theories of language) in response to a wide range of physical, cultural, and spiritual challenges. Moreover, the structure, acquisition, usage, and comprehension of surrogate languages touch on nearly every aspect of linguistic theory in unique and potentially transformative ways. Thus, I suggest that such research on musical surrogate languages and whistled speech should be considered a new subfield of linguistics: *surrogate linguistics*.

From what we know, musical surrogate languages are found all around the world on a vast array of instruments, especially but not exclusively among speakers of tonal languages (McPherson et al. 2025). It is likely that many more traditions are entirely undocumented, particularly in West/Central Africa and Papua New Guinea. Indeed, in 2025, I launched a survey of Nigerian surrogate languages in collaboration with Dr. Samuel Akinbo (University of Toronto), and virtually every language we have asked about either has active musical surrogates or is remembered to have had such systems in the past. The vast majority of the systems we are learning about are completely undocumented and figure nowhere in the published literature. My hope in writing this paper is that by approaching this as a scholarly community we will be able to catalogue and record many more musical surrogate languages around the globe before it is too late.

The remainder of the paper is structured as follows. In Section 2, I define and exemplify musical surrogate languages, including indisputable cases as well as cases at the margins. Section 3 addresses the question of why linguists should be interested in musical surrogate languages, and why we should be including them in our documentary corpora. Section 4 sketches out a brief “how-to” guide, with discussion of recording, elicitation, transcription, ethics, and archiving. Section 5 concludes.

## 2. WHAT ARE MUSICAL SURROGATE LANGUAGES?

Musical surrogate languages are systems in which linguistic content is encoded in the sounds of a musical instrument in order to transmit or communicate a message. The term “linguistic content” is chosen

explicitly to encompass the sounds of language, morphological or syntactic structure, and of course meaning/semantics; the definition is deliberately broad in that a system need not encode all of these levels of linguistic structure to qualify as a musical surrogate. This diverges from Nketia (1971), who treats only those systems that emulate speech sounds (see Section 2.1 below) as true speech surrogates. As James (2021) points out, it is often a blurry line at best between a truly phonemic system and one based on arbitrary musical signs, and many systems combine elements of the two. Hence, I will cast a broad net, which is arguably even more important when documenting systems whose structural underpinnings have yet to be explored and analyzed.

The definition I offer here aligns with one of two categories that Niles (2010) proposes for speech surrogate systems more broadly: what he calls “instrumental systems” (Niles 2010: 2). The other category he proposes are “somatic systems” that use only the body itself, which would include whistled speech but also singing and chanting. One could debate whether singing and chanting should be considered surrogates or simply a different mode of language use, with divergences in phonological, syntactic, and lexical structure from everyday speech. Indeed, going further, this debate could be extended to whistled speech too, especially for non-tonal languages where the oral articulators are in largely the same position that they would be for speech but the airstream mechanism differs (see e.g. Busnel & Classe 1976; Cowan 1972). However, as Meyer (2015) notes, whistling is “not easily identified as a speech act by untrained and unaware speakers” (3) and “requires more training than... other speech registers” (2). It thus fits more squarely with a definition of speech surrogates as being the transfer of linguistic form to what is traditionally not a linguistic modality (i.e., not [only] the vocal tract and/or hands).

Meyer & Manfredi (2025) use the term “non-voiced auxiliary speech” to combine whistled speech with only those instrumental traditions that emulate phonemic aspects of speech, thus excluding instrumental surrogates grounded in arbitrary relationships between sound and meaning (i.e. the “lexical ideograph” systems discussed in Section 2.2 below, which are common in Oceania, the part of world focused on by Niles 2010). Sebeok & Umiker-Sebeok (1976) likewise combine whistled speech and instrumental speech in their seminal volume *Speech surrogates: Drum and whistle systems*, but they do not exclude lexical ideograph systems. It is clear that whistling and musical surrogate systems have many elements in common, and indeed, some whistled traditions can be carried out either somatically (with the hands and mouth) or with the aid of an external tool (e.g., whistled Spanish of the Canary Islands vs. in Andalusia, where a small wood or clay whistle known as a *pito* is used, Meyer 2015). Their definition thus draws the line in a different place than Niles (2010), excluding some somatic systems like chanting and singing while including whistled speech.

Slicing the domain yet another way, one could propose a class of “musical speech registers” that would include instrumental traditions as well as singing, which have in common their perception as music, to the exclusion of whistled speech, which is generally not intended to be musical. A final distinction could be between systems that reproduce a linguistic contrast in a different way from how it is regularly articulated vs. those that do not. In the domain of whistled languages, this would distinguish non-tonal whistling from tonal whistling, since in the latter, fundamental frequency ( $f_0$ ) contrasts normally produced by the larynx are instead transferred to tongue position (what Meyer & Manfredi 2025 call “articulatory transfer”). A similar distinction could be made for musical instruments like the jaw harp

that use the oral cavity as a resonator. These instruments are likewise used for speech surrogacy in both non-tonal and tonal languages, but the mechanism for reproducing speech sounds differs. For instance, in a non-tonal language like Altai (Levin & Süzükei 2006), the jaw harp acts simply as the external sound source for otherwise regularly articulated speech (i.e., the musician mouths the words into the jaw harp), whereas in Hmong, the backbone of the jaw harp’s speech surrogacy is the language’s tone system reproduced not by the larynx but through oral harmonics, with a secondary consonantal contrast reproduced in the plucking pattern of the instrument. In other words, the articulation of Hmong surrogate speech on the jaw harp differs radically from regular spoken language, which is not the case for Altai.

In any case, it is clear that there exist many interrelated phenomena of speech modes that fall outside the bounds of regular day-to-day speech.<sup>1</sup> In what follows I will focus only on musical surrogate languages that use an instrument to create new pathways for expressing linguistic form and meaning, and especially those in which speech is not merely mediated with an instrument as an artificial larynx or vocoder. (Systems at the margins of this definition are discussed in Section 2.3.)

A final definitional note concerns my use of the term “speech” in the discussions that follow. I use this to refer to all genres and registers of spoken language, as opposed to signed language. This includes poetry, conversational speech, announcements, chants, etc. The relationship between speech genre and musical surrogate languages are discussed in Section 3.5.

While the best known musical surrogate languages may be “talking drums”, this phrase can refer to numerous kinds of percussion instruments in unrelated languages and areas of the world. These include tension drums such as Yorùbá *dùndún* or *gáangan* (Akinbo 2019; Durojaye et al. 2021), sets of barrel drums such as Akan *atumpan* (Bagemihl 1988; McPherson & Obiri-Yeboah 2024; Nketia 1963, 1971), or slit-log drums such as the Bora *manguaré* (Seifart et al. 2018) or the Bauro *tamtam* (Snyders 1976). Instruments used for speech surrogacy go far beyond (drummed) percussion. They include flutes such as Igbo *oja* (Carter-Ényì et al. 2021; Eze & McPherson 2025), clarinets such as Gavião *totoráp* (Meyer & Moore 2021), fiddles like Hmong *xim xaus* (Poss 2012), xylophones such as Toussian *balafon* (Struthers-Young 2022), jaw harps such as Huli *híriyùla* (Pugh-Kitingan 1977), mouth bows such as Huli *gáwa* (Pugh-Kitingan 177), and horns such as Asante *mmentia* (Kaminski 2008; McPherson & Obiri-Yeboah 2024), among others. For the past several years, a team of students and I have been compiling a database of speech surrogates around the world, the *Online Database of Speech Surrogates* or ODSS (McPherson et al. 2025). We currently catalogue over 449 surrogate traditions (including whistled speech) based on 168 languages spoken in 48 countries, and this number continues to grow as we conduct further surveys and fieldwork.<sup>2</sup> See **Figure 1** for a screenshot of the current map view showing the distribution of speech surrogate traditions in the database.

---

1 And this does not even consider play languages, ludlings, or secret languages, which may likewise pattern with speech surrogates in different ways (see, e.g., Bagemihl 1988 on “alternate phonologies and morphologies” in speech surrogates and language games).

2 The ODSS database is a work in progress and should not be considered exhaustive of all published or non-published work on speech surrogates worldwide. In fact, it is intended to be a collaborative database, so other researchers are invited and encouraged to submit their own entries.



**Figure 1:** Distribution of speech surrogates in the *Online Database of Speech Surrogates* (McPherson et al. 2025). Blue markers represent whistled speech, red markers represent instrumental surrogates, and yellow markers indicate languages with multiple modalities and/or instruments.

Stern’s (1957) classic typology of speech surrogates divided them into three categories based on how they encode speech. “Abridging” systems are phonemic and iconic: the sounds of speech are emulated on the instrument; these systems are also referred to in the literature as “emulated speech” (Seifart et al. 2018). “Encoding” systems are phonemic but not iconic: the sounds of speech form the basis of the musical surrogate, but they do not need to physically resemble them. Perhaps unsurprisingly, there are very few cases in the literature offered as “encoding” rather than “abridgement”, since what would be the purpose of basing a system on sounds but in a totally arbitrary manner?<sup>3</sup> Catlin (1982) refers to Hmong *qeej* surrogacy as encoding, given the polyphonic nature of the music, but as I will detail in Section 2.1, there is a spectrum of how closely a phonemic system emulates the sounds of speech. Thus, I group together both “abridging” and “encoding” systems under the heading of “phonemic” speech surrogates. Stern’s third category, “lexical ideogram” systems, stand in contrast to these: such systems pay no heed to the sounds of language but rather represent concepts directly through arbitrary patterns. I will discuss these systems in Section 2.2. Finally, Section 2.3 discusses systems that are technically included under the definition of musical surrogate languages developed here, but which have features that stretch this definition in different ways.

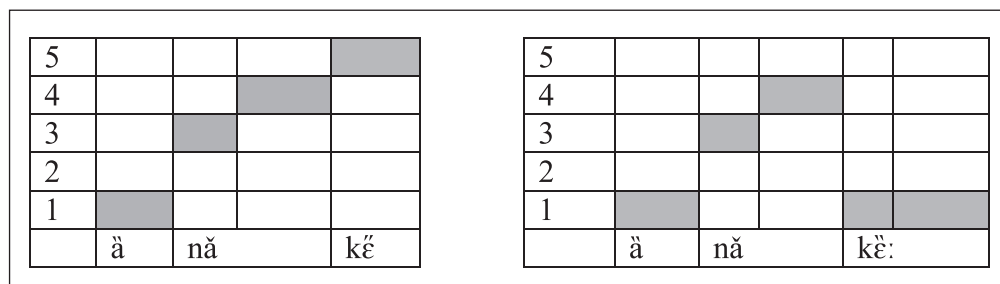
<sup>3</sup> Stern offers Morse code as an “encoding” type of code, though this is not a true speech surrogate, especially since Morse code essentially is a surrogate of a surrogate (written letters); according to Sebeok & Umiker-Sebeok (1976), it is a “second order” system, unlike musical surrogate languages, which are based on speech directly.

## 2.1 Phonemic systems

The most uncontroversial musical surrogate languages are those that emulate, to a varying degree, the sounds of the spoken language through the notes, rhythms, and timbres of a musical instrument.<sup>4</sup> Whether the sounds in question are segments or tones, I refer to these systems as “phonemic”. In every known case, musical surrogate languages reduce the number of contrasts in the spoken language to a smaller set, with the most typical contrasts encoded being prosodic ones: tone and speech rhythm. I illustrate this principle with a brief description of the Sambla balafon surrogate system; for deeper analysis and discussion, see McPherson (2018).

The Sambla balafon is a resonator xylophone from southwestern Burkina Faso, used to encode the community’s language, Seenku (McPherson 2020). Seenku is a Mande language with robust segmental inventory and a complex tone system, but balanced by a largely mono- and sesquisyllabic vocabulary (Matisoff 1990). In other words, Seenku words are short, but they are packed full of phonemic contrasts. However, the balafon eschews all vowel and consonant contrasts, encoding only tone and a binary syllable structure contrast: simplex (CV) and complex (CVV, CV:, CəCV, CəCVV, CəCV:), with CVN syllables varying in their treatment on the balafon between simplex and complex. Simplex syllables (with a level tone) receive one note strike, corresponding to the tone, while complex syllables (and any contour tone, including those on simplex syllables) receive a flam stroke. (In percussive terminology, this is two note strikes in close succession; in melodic traditions, this might be considered a grace note.) Example (1) contrasts the single strike used for a short CV syllable *kě* (1a) with the flam used for a CV: syllable *kè:* (1b).<sup>5</sup> A transcription of these examples is shown in **Figure 2**, with a the strike corresponding to *kě* in the lefthand figure (corresponding to 1a) and the flam strikes corresponding to *kè:* in the righthand figure (corresponding to 1b). A difference in tone encoding is also audible here, which will be discussed below.

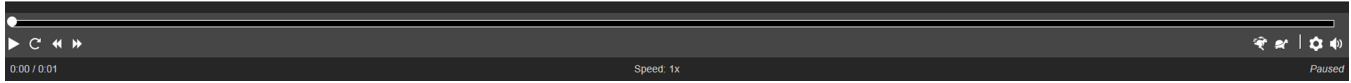
- (1) a.    à     nǎ     kě  
          3SG PROSP go  
          ‘s/he will go’
- b.    à     nǎ     kè:  
          3SG PROSP dry.up  
          ‘it will dry up’



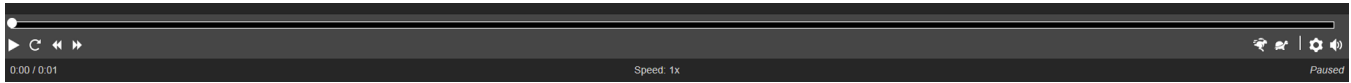
**Figure 2:** Transcription of examples in (1). Notes 1–5 indicate scale degrees of the balafon’s pentatonic scale. In the tabular transcription, the narrow width of the cell for the first note of a flam indicates its shorter duration.

4 I know of no cases of a musical surrogate for a signed language. This is not surprising, given the auditory modality.

5 The abbreviations used in this paper are as follows: 3 third person; ART article; COMPL completive; DEF definite; PL plural; PROSP prospective; REL relative; SG singular.



AUDIO CLIP: Mamadou Diabate plays ‘she will go’ on the Sambla balafon. For the full audio, visit the online journal at <https://iddjournal.org/>.



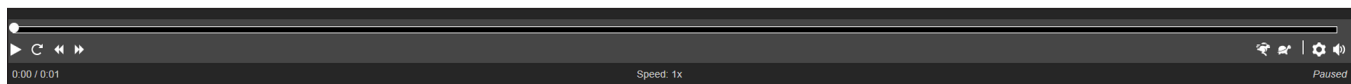
AUDIO CLIP: Mamadou Diabate plays ‘it will dry’ on the Sambla balafon. For the full audio, visit the online journal at <https://iddjournal.org/>.

The neutralization created by this binary syllable shape encoding principle can be seen in example (2), where the same musical surrogate encoding could represent many different words, depending upon the context. Two possibilities are given here, with their balafon transcription in **Figure 3**.

- (2) a. səgà    bâ  
          sheep male  
          ‘ram’
- b. bɛ̀:    kərê  
          pig    man  
          ‘boar’

5				
4				
3				
2				
1				
	səgà bɛ̀: ...		bâ kərê ...	

**Figure 3:** Ambiguity of complex syllable shape and lack of segmental encoding on the Sambla balafon.



AUDIO CLIP: Mamadou Diabate plays ‘boar/ram’ on the Sambla balafon. For the full audio, visit the online journal at <https://iddjournal.org/>.

The flam on the same note could indicate either a sesquisyllabic word like səgà ‘sheep’ or a long vowel like bɛ̀: ‘pig’. The flam on different notes indicates a contour tone, but this contour could be on either a simplex syllable like bâ ‘male’ or a complex syllable like kərê ‘man’ (or on a long vowel, not shown here).

Turning to tone encoding, Seenku's tone system boasts four phonemic levels (Extra-low, Low, High, and Super-high), which can combine to create numerous contour tones. These tones are represented by notes in the balafon's pentatonic scale, with the tone-note mapping dependent upon musical mode (McPherson 2018). The verb in (1a) is a S(uper-high)-toned syllable and thus played on the highest note of the scale, while the verb in (1b) is an X(tra-low)-toned syllable and thus played on the lowest note of the scale; this X tone can also be seen on the 3sg pronoun *ã* and in the head nouns in (2). Contour tones are played by playing each component tone as a flam. The auxiliary verb *nã* in (1) is a LS rising tone, and the note corresponding to each tone is struck in close succession, while both *bâ* and *kəɾɛ* in (2) show a HX falling tone. The final X is audibly played on the same note as the initial X tone of *səgã* and *bɛ̃*.

Through this combination of speech rhythm (dictated by syllable shape) and tone encoding, the Sambla balafon player is able to say any Seenku utterance on the instrument. **Video 1** illustrates fluent Sambla balafon speech with subtitles.



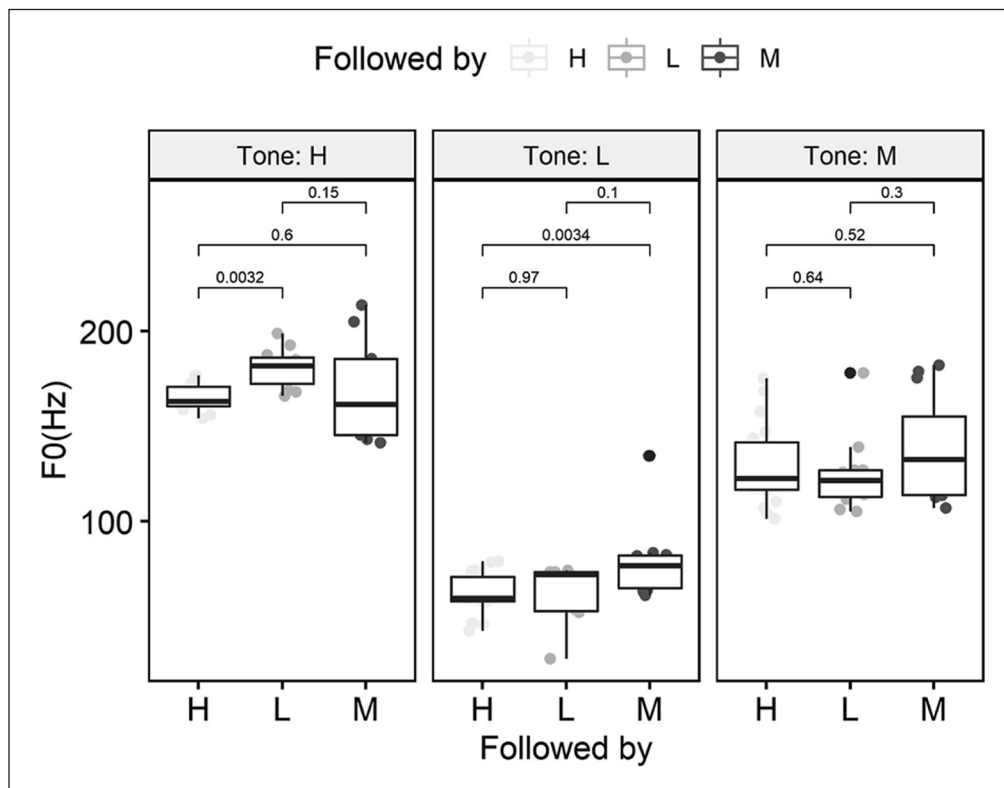
**Video 1:** Sadama Diabate and Sabwee Diabate demonstrate speech surrogacy on the Sambla balafon. This video was recorded by the author in 2017 in the village of Toronosso, Burkina Faso. For the full video, visit the online journal at: <https://www.lddjournal.org/>.

While the Sambla balafon neutralizes all segmental contrasts, this is not always true for musical surrogate languages. Two Yorùbá drumming traditions, the *dùndún* tension drum (Akinbo 2019; Durojaye et al. 2021) and the *bàtá* drum ensemble (González & Oludare 2021; Villepastour 2010), are also based primarily upon tone and speech rhythm. But in addition, they encode a binary vowel height contrast known emically as “soft” vs. “intense”, which corresponds to the linguistic distinction between high and non-high. This emic distinction appears to be grounded in intrinsic amplitude, as high vowels have a lower intrinsic amplitude than non-high vowels (Lehiste and Peterson 1959).<sup>6</sup> On the *dùndún*, this binary

<sup>6</sup> Intrinsic properties of speech sounds, like intrinsic  $f_0$ , intrinsic amplitude, etc. refer to characteristics that are thought to follow largely from their articulatory properties. For instance, low vowels are intrinsically louder (have higher intrinsic amplitude) than high vowels because the jaw is more open.

vowel distinction is encoded as a difference in amplitude, with a softer stroke for soft/high vowels and a louder stroke for intense/non-high vowels. Among mid vowels, Villepastour (2010) notes that there is some variation in whether they are treated as soft or intense. On the *bàtá* drum ensemble, this difference is encoded more categorically, with intense vowels involving the addition of a sharp strike on a double-sided barrel drum's second membrane with a leather strap. This likewise iconically increases the amplitude of the strike combination.

These two Yorùbá traditions demonstrate two further ways in which a musical instrument can encode tone. As we saw above, the Sambla balafon has a fixed musical scale, and tones map onto these fixed notes. The *dùndún*, on the other hand, is a tension drum, meaning that it can produce gradient pitch distinctions depending upon the tension applied by the player's arm to the cords that attach the membranes to the body of the drum. Akinbo (2021) demonstrates that this continuous pitch mechanism is able to encode phonetic detail of Yorùbá tone, including raising of H tone before L and lowering of L before H, in addition to closely emulating contour tones created by tonal carryover in H-L and L-H tone sequences. These results can be seen in **Figure 4**.

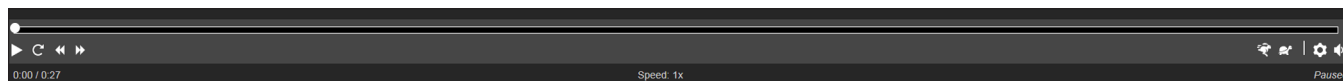


**Figure 4:** Phonetic realization of Yorùbá's three tone levels on the *dùndún* in different tonal contexts.

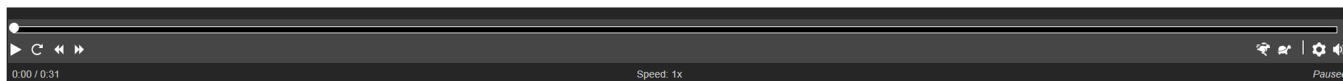
In contrast, the *bàtá* drum ensemble is not a pitched instrument in a strict sense. Rather, the drums produce a range of timbres depending upon how they are struck. The three tones of Yorùbá are mapped to different timbral combinations, though it should be noted that the timbre used for L tone (an open resonating strike) has a lower spectral center, i.e., a lower overall frequency, than that used for H tone (a sharp, high frequency slap mute).

Neither the *dùndún* nor the *bàtá* typically encode any consonants, with the exception of the phoneme /r/, which can at times be played with a double strike on both instruments (González & Oludare 2021). The different sonic nature of these two musical surrogate systems encoding the same Yorùbá material can be heard in the audio corresponding to example (3), included with permission from Amanda Villepastour (text from Villepastour 2010: 152).

- |     |                       |                                   |
|-----|-----------------------|-----------------------------------|
| (3) | Èṣù Láaróyè           | ‘Èṣù Láaróyè (praise name)’       |
|     | Èṣù tí bélékún sunkún | ‘Èṣù who weeps with the bereaved’ |
|     | Èlèkún ní sunkún      | ‘While the bereaved is weeping’   |
|     | Láaróyè (ní)yò        | ‘Láaróyè is rejoicing’            |



AUDIO CLIP: An excerpt of praise poetry (oriki) played on the Yorùbá *bàtá*, corresponding to the text in example (3). For the full audio, visit the online journal at: <https://iddjournal.org/>.



AUDIO CLIP: An excerpt of praise poetry (oriki) played on the Yorùbá *dùndún*, corresponding to the text in example (3). For the full audio, visit the online journal at <https://iddjournal.org/>.

Neither the *dùndún* nor the *bàtá* typically encode any consonants, with the exception of the phoneme /r/, which can at times be played with a double strike on both instruments (González & Oludare 2021). Indeed, most phonemic systems I am aware of encode little (e.g., Yorùbá drums) or no (e.g., Sambla balafon) segmental information. An exception is the Hmong *ncas*, a brass jaw harp traditionally used for courtship (Catlin 1982). In an ongoing study with my collaborators Neng Thao and Ford Springer, the *ncas* was found to encode information on vowel quality and consonants in addition to tone, though the number of contrastive dimensions for both is greatly reduced. This ability is thanks to the unique nature of jaw harps, which use the oral cavity as a resonator to amplify and attenuate the rich harmonics produced by plucking the instrument. Because it uses the oral cavity, the filter is the same as that used for speech. However, as mentioned in the introduction to Section 2, there is a distinction between the use of jaw harps for tonal and non-tonal languages. For non-tonal languages, the jaw harp is literally an external larynx, a vocoder, and speech is articulated normally in the vocal tract. For tonal languages like Hmong, however, the tone contrast remains primary, and these tones are produced not by laryngeal  $f_0$  but by selectively amplifying harmonics of the  $f_0$  produced by the vibration of the instrument by changing the configuration of the oral cavity.<sup>7</sup> While this might seem to preclude the articulation of vowels, we have found some minor front-back distinctions across some but not all tone categories, suggesting that the articulation of at least some harmonics can accommodate front-back movement of the tongue. Consonants are not articulated with normal speech gestures, but

7 Specifically, the seven contrastive tones of Hmong are neutralized to four musical notes, corresponding to H6–H9.

rather are reduced to a binary contrast between those that receive a pluck on the instrument and those that do not. Interestingly, this distinction does not neatly track the sonorant vs. obstruent distinction noted in Rialland's (2005) work on Hmong whistling, with voiceless sibilants receiving a pluck but all other voiceless fricatives and all voiced fricatives receiving no pluck. An example of Hmong *ncas* surrogacy can be seen in **Video 2**.



**Video 2:** Xiong Thor plays a love song on the Hmong *ncas* (jaw harp). For the full video, visit the online journal at: <https://iddjournal.org/>.

All of the examples thus far have been drawn from tonal languages, with good reason, seeing as phonemic musical surrogate languages are far more common among tonal languages than non-tonal ones (James 2021). An exception would be the use of jaw harps among non-tonal languages such as Ifugao (Blench & Campos 2010), Altai (Levin & Süzükei 2006), or Khmu (Proschan 1994; see also Nikolsky 2020). (I return to these cases in Section 2.3.) We saw that the segmental contrast most likely to be encoded in musical surrogate speech for tonal languages was vowel height, and this appears to hold true for non-tonal systems as well. For example, Snyders (1969/1976) describes a slit-log drumming system for Bauro, a language of the Solomon Islands. Like better described tonal slit-log drumming systems of Central Africa (Carrington 1949) or the Amazon (Seifart et al. 2018), the Bauro slit log drum produces two pitches, one on either lip of the slit log. Rather than mapping these high and low pitches to high and low tone, however, the Bauro system instead maps the low to pitch to the low vowel /a/ and the high pitch to all non-low vowels (/i, e, o, u, ê/), as shown in the following example, reproduced from Snyders (1969/1976: 1297).<sup>8</sup> L and H indicate low- and high-toned drum strokes, placed above their corresponding vowel to make room for a gloss:

---

8 It is not clear from Snyders (1969/1976) what vowel quality *ê* refers to.

- (4) L L L H L L H H LH L L L L  
 Ha-ha-ga-ngi-si-a na go-go i-a-ni ta-nga-go na  
 make.resonate ART drum this arrive? ART  
 ‘I am beating the drum [for you]...’<sup>9</sup>

The underscore indicates positions where a H drum stroke might be expected for the vowel /i/, but where none is found. It appears that there is variation in how vowel-vowel sequences such as *ia* or *ai* are treated in Bauro drummed speech.

Whether vowel height is a privileged contrast in other non-tonal systems remains to be seen with future documentation.

## 2.2 Lexical ideogram systems

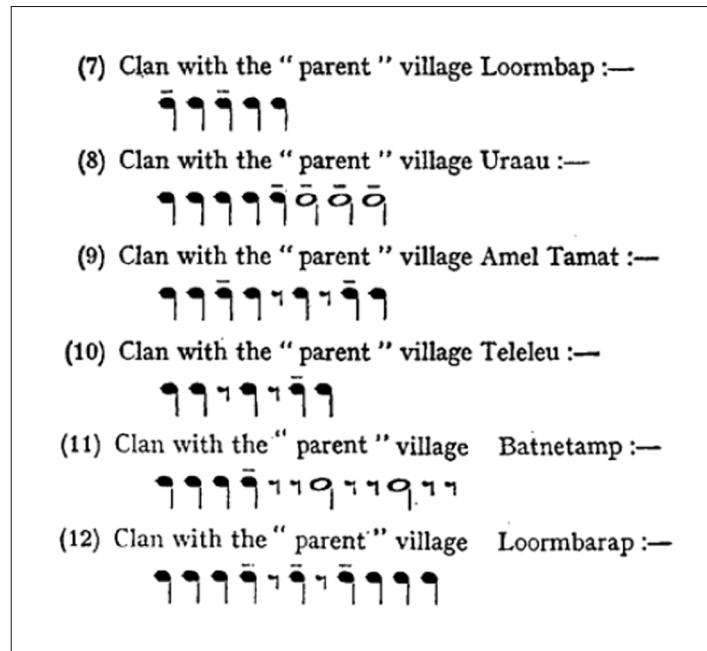
If we define musical surrogate languages as systems of communication that encode linguistic content in the sounds of a musical instrument, then it is not strictly necessary that these messages *sound like* speech. Lexical ideogram systems encode words, concepts, and sometimes entire messages through arbitrary musical sequences, but their linguistic content is still understood by listeners. These systems are sometimes referred to as “signal drums” (Burrige 1959; Nketia 1958), and indeed, lexical ideogram systems are to the best of our knowledge most common on drums, whether membranophones or struck idiophones like slit-log drums.

The Bauro drum discussed in Section 2.1 is an interesting case of a phonemic slit-log drumming system in Melanesia. Most documented systems, like slit-log *garamut* drums in Papua New Guinea, are lexical ideogram systems. Detailed description of these systems and documentation of their signals remain scant, especially as it relates to linguistic structure, likely because these systems have been historically treated as only tenuously linguistic in nature, but see Niles (2010) for an excellent overview of existing work. This is a shame, as we may be missing opportunities to analyze, for instance, how concepts are strung together and the extent to which this mirrors or diverges from the syntactic structure of the spoken language.

I will offer here brief illustrations of two slit-log drumming systems from Melanesia that have been described in some detail, though audio-video documentation is lacking, as they are both older studies. First, Deacon (1934/1976) describes slit-log drumming in Malekula, Vanuatu. Deacon notes that beating these slit-log drums (which he terms “gongs”), creates “a very efficient ‘gong language’ by means of which they are able to communicate with their fellows in distant villages, for every important object or happening of every day life has its motif, and even time can, to a limited extent, be expressed in the same way” (Deacon 1976: 1207). The examples in **Figure 5** show a selection of “motifs” corresponding to different clan names. I have explicitly chosen this excerpt to include both the village of Loorbap and Loorbapap, which show us that there is no discernible relationship between the rhythm played and the spoken names of villages.

---

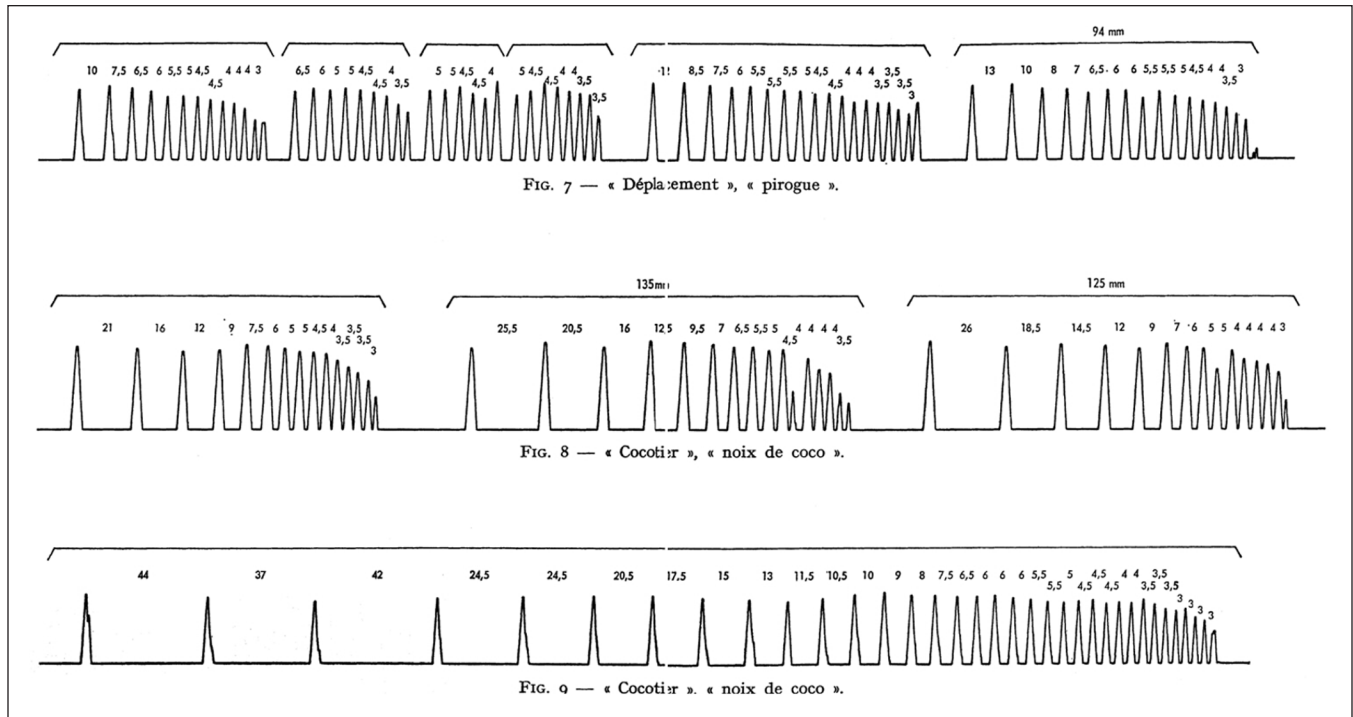
<sup>9</sup> Snyders (1969/1976) does not explicitly gloss the example, instead providing a glossary of terms and a free translation. This example is drawn from Line 1, which in the free translation flows into Line 2, reading “(Je) fais résonner le tambour pour vous, l’homme qui est au village de Getemaringa” (‘I am beating the drum for you, the man who is in the village of Getemaringa’). It is possible that the Bauro line needs to be grammatically connected to the following one.



**Figure 5:** A selection of drummed Malekula clan names (Deacon 1976: 1200).

More complex messages can be sent by combining different signals. For instance, Deacon describes a rhythm known by the name *iuswus ngileo*, whose meaning is roughly “where is...” or “is it at your place?” Interestingly, even the spoken name of this rhythm is arbitrary (or at the very least, so archaic as to be uninterpretable by speakers at the time). This rhythm is then followed by a rhythm that refers to what has been lost, such *naai tamap*, which refers to a certain grade of pig.

From the description in Deacon (1934/1976), only one side of these slit-log drums is beaten, thus only one pitch is produced, unlike the Bauro drumming, where each side of the slit produces a different pitch. The same situation is found for the Kwoma slit-log drum known locally as *mee* (Whiting & Reed 1938; Zemp & Kaufmann 1976), or in Papua New Guinea’s lingua franca, Tok Pisin, *garamut*. Like the Malekula system, arbitrary drum sequences differing only in their rhythm/spacing of strikes stand for concepts such as clan names, commands to come or go, nouns, or whole announcements like “a man is sick” or “a man has died”. Interestingly, Zemp & Kaufmann note that it is not the absolute number of strikes that characterizes a sign (this number can vary) but rather the temporal relation between the strokes and the overall duration of the message. Thus, the signals for “journey/canoe” (top line in **Figure 6**) and “coconut” (bottom two lines in **Figure 6**) both consist of a series of strokes that increase in velocity, but the strokes begin closer together for “journey/canoe” and the overall length of the message is shorter than for “coconut”, which begins with more widely spaced strokes and lasts considerably longer (though still of variable duration, as can be seen by comparing the middle and bottom lines).



**Figure 6:** Transcriptions of drum stroke spacing for “journey/canoe” and “coconut” in Kwoma, taken from Zemp & Kaufmann 1969/1976: 56–57).

As noted in Section 2, Nketia (1971) excludes such “signal drumming” from musical surrogate languages, given the lack of a concrete relationship between the sounds of speech and the musical signal. Seifart et al.’s (2018) designation of Bora slit-log drumming as “emulated speech” would likewise exclude these systems, as the speech signal is not clearly emulated. Niles (2010), on the other hand, clearly includes lexical ideogram systems in his definition of musical surrogate languages. I am very much in agreement with this inclusion and would argue (as does James 2021) that we would be remiss to exclude these systems from our documentation of musical surrogate languages, since in fact many otherwise phonemic systems include cases of lexical ideograms, and even disambiguation strategies that are ostensibly based on speech sounds can result in an arbitrary musical sequence that differs from regular spoken language.

A simple example of a lexical ideogram in an otherwise phonemic system can be found on the Sambla balafon. Though most messages transparently encode the speech sounds of Seenku, a common message is an alternating minor third (the exact number of strokes can vary, as in Kwoma) that means something like “thank you” or “correct”, when someone has understood and appropriately responded to a message (McPherson 2018). This lexical ideogram stands out from regular surrogate speech in making use of the note of the scale corresponding to roughly a minor third in Western musical tradition, which is otherwise avoided in regular Sambla surrogate speech. Similar kinds of lexical ideograms are often found as opening and closing elements in a drummed message, for instance in Lokele (Carrington 1949/1976). In short, to gain a holistic understanding of a phonemic system like the Sambla balafon, we cannot exclude lexical ideogram elements, and to better understand how lexical ideogram elements work in a musical communication system, we likewise need to document and study them in cases where they form the basis of the system. Description and documentation of these lexical ideogram systems are sorely lacking, and I urge anyone working in an area like Papua New Guinea that has been known to have such systems

to ask about them and document whatever knowledge might remain as these brilliant long distance communication systems fall into disuse.

A more ambiguous case of arbitrariness in an otherwise phonemic system is the use of a disambiguation strategy known as “enphrasing”, attested in a number of different surrogate traditions. Reducing speech to only tone and rhythm produces a large number of homophonous words, as exemplified in (2) above for the Sambla balafon. To resolve the ambiguity, in some traditions, words are placed in conventional longer phrases that are tonally distinct from one another. For instance, in Lokele (Carrington 1949), *likɔndo* ‘plantain’ and *lomata* ‘manioc’ are both trisyllabic L-toned words that would be played the same way on the drums. To make matters worse, they are both starch crops, and so simple linguistic context would be unlikely to differentiate them. To solve the problem, *likɔndo* is always enphrased as *likɔndo libotúmbela* ‘plantain to be propped up’ and *lomata* as *lomata otíkala kóndo* ‘manioc remaining in fallow land’. The tonal structure of these two longer phrases is distinct, and so the correct word can be identified. Although these enphrasings are played phonemically, by encoding the tones associated with the words in the phrase, the phrases themselves become conventional and arbitrary. A listener would have to know, for instance, that the tonal sequence L-L-L-H-L-H-L-L maps to ‘plantain’, via ‘plantain to be propped up’. Thus we cannot use arbitrariness alone as a yardstick for whether a system should be seen as phonemic (a “true” speech surrogate) or as lexical ideogrammatic (“signal drumming”). This kind of enphrasing is also seen in Bora drumming in the Amazon (Seifart et al. 2018), along with conventionalized morphological markings to distinguish nouns from verbs; more details are provided in Section 3.3.

### 2.3 Systems at the margins

As we have seen, scholars differ in how they define speech surrogates or musical surrogate languages, sometimes excluding lexical ideogram systems, sometimes excluding vocal music, or in the case of my definition, excluding whistled speech. But there is a core set of traditions that everyone agrees are musical surrogate systems: those that use a musical instrument to reproduce at least some phonemic aspects of more-or-less regular speech (for more on genre, see Section 3.5). This core includes long-distance drumming traditions like Bora (Seifart et al. 2018), the Yorùbá *dùndún* (Akinbo 2019; Durojaye et al. 2021; González and Oludare 2021), as well as the Sambla balafon (McPherson 2019). But as we step away from this center, there are many musical surrogate languages whose messages are more poetic in nature, adding yet another musical or aesthetic layer to the encoding of speech, such as the many Hmong speech surrogates (Catlin 1982; Poss 2012), Gavião flutes, mouth bows, and clarinets (Meyer & Moore 2021), and the Igbo *òjà* (Eze & McPherson 2025).<sup>10</sup> At the periphery, we find two sets of practices that push the boundaries of what can be considered speech surrogacy altogether: instrumental renditions of vocal music and traditions where speech is fully articulated with the vocal tract, with the exception of the larynx, which in turn is replaced by an instrument. My working definition of “systems in which linguistic content is encoded in the sounds of an instrument in order to communicate a message” would include both of these kinds of systems at the margins, but as I will show, they operate in somewhat different ways and may tell us less about how speech is mapped to music and how people conceptualize their linguistic categories than the core surrogates do.

---

<sup>10</sup> I would argue that it is no accident that these more melodic instruments tend towards styles that are closer to vocal music, though the Sambla balafon shows that we can also find melodic instruments that encode ordinary speech.

Many speech surrogacy traditions like the Sambla balafon have a direct “speech mode” (what was detailed above) as well as a “singing mode”, where the instrument is conceived of as singing the lyrics of songs. In the case of the Sambla balafon, these are songs that can also be sung vocally, and the instrument reproduces the sung melodies; in other words, the resulting notes are not created from a strict linguistic mapping, as is found in the speech mode (McPherson 2018), but rather after being passed through a filter of vocal composition. This first step of relating language to music is more a question of (tonal) text-setting rather than a process of speech surrogacy per se (e.g., Ladd & Kirby 2020; McPherson & Ryan 2018; Schellenberg 2012). Instrumental renderings of vocal music can similarly be found on the *pion*, a side-blown flute. The practice is recognized by the Sambla community as “singing” the lyrics of songs, and in that sense, it is a form of instrumental communication of linguistic content where the melodies are linguistically informed through relatively strict tone-melody mapping (McPherson & James 2021). However, if these traditions are considered speech surrogates, which seems to be their conceptualization within the community, then should instrumental versions of Western songs be considered speech surrogates as well? Or perhaps closer to the *pion*, what about Carnatic music of South India, where an instrument like a violin often mimics vocal music (Viraraghavan & Sankaran 2023), with repercussions for bowing patterns, among other stylistic choices (Charumathi Raghuraman, personal communication, 2025-09-25)? Neither of these cases is ever discussed as speech surrogacy, yet there are clear parallels. As Niles (2010: 5) also attests, it is an extremely difficult task to draw clean lines through what is in reality a cloud of interrelated musico-linguistic systems, and depending upon how we define musical surrogate languages, these marginal systems may in fact be found in most musical cultures.

The other set of systems at the margins involves the full production of speech, through either whispering or mouthing words, but with an external rather than internal sound source. One case with more commonalities with other musical surrogate languages is the *inanga chuchotée* (Fales 2017). The *inanga* is a zither used by Kirundi speakers to accompany songs, but the singer whispers rather than sings the words (thus the French *chuchoter* ‘to whisper’). The vocal melodies, which, like the Sambla *pion*, are tightly related to the tones of Kirundi, are transferred to the strings of the zither. Fales (2017) describes how the right combination of performance elements (e.g. a quiet setting, the singer’s mouth close to the strings) produces an auditory illusion where the  $f_0$  of the strings is decoupled from the instrument and perceived as being part of the human voice. This tradition is at face value simple vocal music accompanied by an instrument, but the fact that the perceived pitch/tones of the words are produced not by the larynx but by an external instrument brings it under the umbrella of musical surrogate languages. However, through the combination of the whispered words and the musical notes, it does not result in the reduction of phonemic contrasts typically found in speech surrogacy.

Further towards the margins are non-tonal jaw harp traditions, where the jaw harp acts simply as a vocoder or sound source for the otherwise normal articulation of words, though often poetic ones (Levin & Süzükei 2006; Proschan 1994). Apart from basic voicing, these cases involve no transfer of a linguistic feature to a non-linguistic modality, such as playing different strings to follow the  $f_0$  of a vocal melody. Thus, I would argue that they are cognitively quite different from other musical surrogate languages, even if they overlap in terms of acoustic outputs or communicative niches (e.g. jaw harps are often used for courtship in both tonal and non-tonal contexts). These traditions fall under the definition of musical surrogate languages taken in this paper, but I think it is noteworthy that it is almost exclusively systems like these that are found for non-tonal languages.

I do not mean to diminish any of these traditions that are further from the “core” of musical surrogate languages; all these forms of oral culture are still very much worth documenting (see Section 3.4 for further discussion of cultural value), but one should be aware of the different kinds of traditions that we find and what implications that has for the sorts of analyses and conclusions we can draw from them. Just as linguists engaged in language documentation ought to have some knowledge of linguistic typology to understand and contextualize their work, so too is it helpful to have a basic understanding of the typology of musical surrogate languages and other related traditions.

### 3. WHY DOCUMENT MUSICAL SURROGATE LANGUAGES?

In this section, I provide a few reasons why linguists should be interested in documenting musical surrogate languages. There is indeed an imperative to document these systems since many, if not most, are rapidly disappearing from communities around the world. Musical surrogate languages that were used for long distance communication are falling into disuse with the advent of cell phones and other forms of communication technology. Many more that were tied to traditional religious practices are being lost due to conversion (though in other cases, they can evolve into two practices, like the Ewondo slit log drums used to call the faithful to church, Neeley 1996). Still others are threatened by changes to traditional lifestyle, globalized music tastes, or even environmental changes that threaten natural materials used in instrument construction. But why should linguists (as opposed to anthropologists or musicologists) feel compelled to document them? Perhaps most obviously, because musical surrogate languages are systems of communication based upon spoken languages, it takes a linguist (or a speaker of the language) to have the necessary knowledge of the linguistic structures to understand the relationship between the spoken and musical forms.

Furthermore, since there are linguists working all around the world in different language communities already, we have the access and community ties to ask about these traditions. But we also have a lot to learn about spoken language, language communities, and local emic understandings of language through work on musical surrogate languages, as I will detail below.

#### 3.1 Musical surrogate languages as ethnoscience

Musical surrogate languages are concrete manifestations of Indigenous theories of language. They show us which contrasts are the most salient to speakers, and how they are perceived and categorized. We can even learn how they are conceptualized as we engage speakers in discussions about these traditions. For instance, we saw in Section 2.1 that Yorùbá musicians have an emic understanding of vowel contrasts that is grounded in intrinsic amplitude and discussed in terms of “soft” and “intense”. This terminology relates to the speech sounds themselves. But in other cases, the instrument itself may shape the understanding of the speech sounds, as in the case of the Sambla balafon, where lower pitches are referred to as “high” because the gourds used as resonators are larger and thus keep the wooden bars of the balafon further from the ground, whereas higher pitches (called “low”) use smaller gourds which are closer to the ground, as can be seen in **Video 1**. For more discussion of cross-linguistic differences in pitch metaphors, see Sapir (1975) and Shayan et al. (2011).

Linguistics has long been interested in other areas of ethnoscience, such as ethnobiology (e.g., Shepard 1997; Si 2011) or color terminology (Kay & McDaniel 1978). It is thus surprising that linguists have not taken much interest in musical surrogate languages, seeing as these represent local scientific theories

of language itself, which should be crucial in guiding and shaping our understanding of the languages which we study and document.

Indigenous theories of language are an important source of evidence for metalinguistic awareness at various levels of linguistic structure. For instance, as we will see in Section 3.2, the Sambla balafon surrogate language distinguishes between lexical and postlexical tone, exceptionlessly encoding the former while eschewing the latter. These strata have been proposed (or perhaps presupposed) in phonological theory, but it is important to find evidence for their psychological reality. The fact that balafonists do not encode contour tone simplification, a postlexical process, constitutes one such piece of evidence.

Surrogate languages also provide good evidence for the psychological reality of natural classes. Consider the case of vowel height: as noted, Yorùbá drummers (on both the *dùndún* and the *bàtá*) distinguish between high and non-high vowels (“soft” vs. “intense”), which aligns with the use of the phonological feature [high] in Western linguistic terms. Bauro drummers, on the other hand, distinguish between low and non-low vowels, which aligns with the phonological feature [low]. It would be interesting to investigate whether the choice between encoding [high] or [low] in the surrogate system (given that both languages also have mid vowels) is correlated with any other phonological processes in the language that might show that one or the other vocalic feature is more active or salient to listeners.

Though consonants typically play a far smaller role in surrogate speech than tones or vowels, the cases where they are encoded can also reveal information about natural classes, or the lack thereof. The Hmong *ncas* (jaw harp) makes a binary distinction between notes where the jaw harp is plucked and notes without a pluck. With some minor variation, plucked notes are all plosives and voiceless sibilants. In Western phonological feature theory, plosives form a natural class, but there is no natural class that includes plosives and voiceless sibilants but excludes other voiceless fricatives and voiced fricatives. While subconscious, the Hmong emic categorization of speech sounds into continuous and interrupted ones deviates from Western feature theory, but it is very likely grounded in a perceived acoustic trait (a distinct break in the voiced sound stream, perhaps?). Whether this can tell us anything about the behavior of spoken Hmong phonology, I am uncertain, but the deviation from natural classes is at the very least an interesting point raised by musical surrogate speech. Not all theories of language classify speech sounds in the same way, and treating musical surrogate languages as the ethnosciences of language allows us to consider linguistic structure in a new light.

### 3.2 Tonal analysis

As we saw in Section 2.1, the majority of musical surrogate languages are based on tonal languages, and tone is generally the primary basis of the surrogate system. Musical surrogates can be of enormous benefit in analyzing and understanding tonal systems, aided by musicians’ own conceptualizations of their categories and salient acoustic traits (typically  $f_0$ , but perhaps also amplitude or duration, e.g., Rouget 1964). Linguists may be daunted by tone or feel that they cannot reliably hear tonal contrasts (McPherson 2019a). But these contrasts can be amplified when transposed to a musical instrument, especially one with categorical contrasts (unlike the continuous  $f_0$  of, say, a tension drum).

Even as a tonologist, I originally struggled with Seenku’s complex tone system. Based on other languages in the area, I expected a three-tone system, and indeed, most lexical contrasts do employ only three levels (as well as contour tones). However, this analysis left some alternations (and some lexical contrasts)

poorly explained, as I struggled to understand why a putative Mid tone would sometimes surface as lower or higher than an adjacent tone. As I describe in detail in McPherson (2018), the balafon surrogate language provided evidence that Seenku employs four tonal contrasts rather than three, and that one of the alternations I failed to fully grasp is due to a contour tone simplification process, whereby a Low-Superhigh rise simplifies to Low in some contexts (which is higher than the lowest tone, Extra-low) and to downstepped Superhigh in others (lower than an adjacent Superhigh but higher than High). It was possible to hear this underlying form because Sambla balafonists distinguish between lexical/grammatical tone and postlexical/surface tone in the surrogate language: the former is always faithfully encoded, and the latter is almost never encoded.

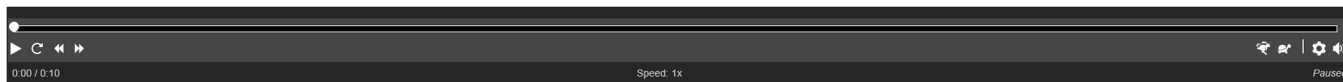
An instrument with continuous  $f_0$  like the Yorùbá *dùndún* can encode not just the phonemic contrasts but also phonetic details of tone, as I described in Section 2.1 above. This suggests that the encoding of tone—and what the linguist can learn about it—arises from an interplay between the language’s tonal system and the instrument’s capabilities. Fixed pitches and a complex tone system with four levels are likely to lead to phonemic encoding, as on the Sambla balafon; continuous pitch is more likely to lead to surface emulation of  $f_0$ , as on the Yorùbá *dùndún*; a larger scale of fixed pitches and fewer tone contrasts allows the possibility of encoding downstep and downdrift, which we find on the Igbo *òjà* (Carter-Ényì et al. 2021; Eze & McPherson 2025). This last case is demonstrated in (5), taken from Eze & McPherson (2025):

- (5) Ókú 'gbárá àchàlà àgbááná  
 fire burnt.REL bamboo burnt.COMP  
 ‘The fire that burnt the bamboo has burnt it completely’

Igbo is fundamentally a two-tone language, though it also shows extensive (and contrastive) downstep of H tone as well as downdrift, in which the second H in a HLH sequence is lower than the first (Connell 2011). In this example, we see a tonal sequence H-!H-L-H, and this stepwise lowering is mimicked in the notes of the *òjà*, shown in **Figure 7**.



**Figure 7:** *òjà* rendition of example (5).



AUDIO CLIP: Gerald Eze plays a warning on the Igbo *oja*, corresponding to the text in example (5) and the musical transcription in **Figure 7**. For the full audio, visit the online journal at <https://iddjournal.org/>.

If, on the other hand, the language has more tone contrasts than available pitches, then the ways in which contrasts are maintained or neutralized can provide insight into their underlying representations and/or speaker’s awareness of them. This is the case on the Akan *atumpan*, a set of two barrel drums that encode

the H and L tones of Akan (Bagemihl 1988; McPherson & Obiri-Yeboah 2024; Nketia 1971). Like Igbo, Akan makes use of a downstepped H, but there is no third distinct pitch category on the drums to encode this surface tone. Nketia (1971) states that it is simply encoded as H, but McPherson & Obiri-Yeboah's (2024) analysis of a corpus of drummed phrases revealed that downstepped H is sometimes played as H, sometimes “unpacked” into an underlying LH sequence, and sometimes played as L, especially in phrase-final position, presumably due to the cumulative effect of phrase-final lowering and downstep.

The documentation and analysis of more tonal surrogate languages will allow us to further probe the relationship between tonal structure and instrumental constraints and look for any universal tendencies in how tone systems are mapped to the sounds of musical instruments.

### 3.3 Morphosyntactic structures

While it may appear that musical surrogate languages are solely the domain of phoneticians and phonologists, this is absolutely not the case. We can likewise find phenomena of interest to (morpho) syntacticians and those interested in morphosyntactic typology.

First, bridging the gap between morphosyntax and phonology, musical surrogate languages based on tonal languages can be expected to encode grammatical tone. West Africa is a natural testing ground for this hypothesis, as there exists a high density of both grammatical tone as well as musical surrogate traditions. Some examples of studies confirming the presence of grammatical tone in speech surrogacy include McPherson (2018) for the Sambla balafon, Akinbo (2021) for the Yorùbá *gáangan*, Struthers-Young (2022) for the Northern Toussian balafon, and McPherson & Obiri-Yeboah (2024) in Akan speech surrogacy. Unfortunately, many morphosyntactic descriptions and analyses continue to omit tone marking (McPherson 2019a), making it more of a challenge to identify and confirm the presence of grammatical tone in musical surrogate language data. On the other hand, musical surrogate languages can offer an important tool for hearing and understanding grammatical tone contrasts, since they often render tone contrasts as discrete and more easily distinguished, as detailed in Section 3.1.

In most cases of a phonemic surrogate language, we expect the morphosyntax of the surrogate message to mirror that of the spoken language, and indeed, that is typically the case (e.g., Yorùbá, Igbo, Seenku, etc.). But we can also find cases where they differ in interesting ways, especially in enphrasing. First, Bora slit log drumming distinguishes between nouns and verbs in the surrogate message by means of conventionalized morphology. Nouns are marked with the suffix *-úβù*, which in the spoken language means ‘deceased’, though it does not carry this meaning in surrogate speech; verbs are marked with the suffix *-ʔh̄k'á*, indicating repeated action, in the spoken language, but again, it is devoid of these semantics in surrogate speech. Though tonally identical, they are rhythmically distinct, as the timing of drum strikes on the Bora slit log drums carefully tracks vowel-to-vowel intervals (Seifart et al. 2018). This shows how existing morphological structure can be co-opted for its rhythmic properties in surrogate speech.

Enphrasing also shows cases where morphological structure that is unattested in regular speech is used in the surrogate language. We find examples of co-hyponym compounds (compounds that combine two specific terms to refer to an overarching category) used to support enphrasing across multiple Bantu slit log drumming systems (Carrington 1949, described in James 2021), such as Lokele *mbóli sango* ‘news’ made up of two words for ‘news’ or Mbòle *tofulú áncɔli* ‘little bird’ made up of two kinds of little birds. What makes these cases particularly interesting is that Bantu languages do not generally have productive

co-hyponym compounding, which shows that speakers are able to draw upon linguistic structures that are attested cross-linguistically, though not in their own spoken languages.

Another such case is found in Senegalese *sabar* drumming. Winter (2014) describes *sabar* drumming as likely a lexical ideogram system—there are but tenuous connections between spoken Wolof and the words and phrases played on the *sabar* drums (Ros 2021). In principle, a lexical ideogram system could follow the same morphosyntactic patterns as the spoken language, simply with arbitrary signs, but this need not be the case. An interesting divergence between the morphosyntax of spoken Wolof and the *sabar* drum system can be found in the plural. Wolof employs noun classes, and plural number is marked on the noun class determiner, e.g., singular definite *bi* vs. plural definite *yi*. On the *sabar* drums, however, pluralization is marked through reduplication (or retriPLICATION) of the last drum stroke, as shown in the examples in (6) from Winter (2014: 658), where the syllables in the righthand column reflect onomatopoeia for different drum strokes:

- |     |    |                 |                  |                     |
|-----|----|-----------------|------------------|---------------------|
| (6) | a. | <i>weex bi</i>  | white DEF.SG     | rwe gin             |
|     |    | <i>weex yi</i>  | white DEF.PL     | rwe gin gin         |
|     | b. | <i>rafet bi</i> | beautiful DEF.SG | gin tac tac tac     |
|     |    | <i>rafet yi</i> | beautiful DEF.PL | gin tac tac tac tac |
|     | c. | <i>tuuti bi</i> | little DEF.SG    | pax ce gin          |
|     |    | <i>tuuti yi</i> | little DEF.PL    | pax ce gin gin gin  |

Like the Bantu co-hyponym case, *sabar* reduplication is interesting in that it draws upon a morphological process for plural marking—reduplication—that is typologically common but not attested in the musicians’ spoken language. One could use such an observation to make a bold claim that this provides evidence for universal grammar or other tendencies towards common linguistic structures.

Lexical ideogram systems remain virtually undescribed from a linguistic perspective. Documenting these systems could provide a rich source of data for probing how linguistic messages are structured when they diverge phonemically from the spoken language.

### 3.4 Oral literature and cultural importance

Oral literature is a common feature of documentary corpora. Folk tales, prayers, and songs are often among the first genres collected, and yet musical surrogate languages do not make this list, despite similarly being a form of oral literature. Like other genres, musical surrogate messages and songs are intimately intertwined with cultural institutions and societal structure, and so their documentation can also capture important aspects of the culture that may otherwise be missed.

Eze & McPherson (2025) describe the role of the Igbo *òjà* in the *Mm̀̀nẁ̀* masquerade cult. The aerophone is used to send messages of warning to the masked spirit, such as that shown in (5) and **Figure 7**, as well as to shower praises upon titled men and women in the audience. These messages are poetic and rich in proverbs, and as such, they are only fully understood by intellectuals in Igbo society. They also draw upon other aspects of Igbo culture and metaphysics, such as numerology. For example, a common praise is *ńwó'ké téghété* ‘a man worth nine’, where the number nine is of spiritual significance, as a multiple of the already important and balanced number 3 (Umeh 1997).

Documenting the Sambla balafon as part of a larger project on Seenku (McPherson 2018, 2020) has taught me an enormous amount about Sambla culture that I would not otherwise have learned. For example, I learned that every family belongs to a different clan or totem that each has its own unique song (itself rich with proverbial lyrics). Whenever balafonists visit the house of that family, they must play that song. These songs could also be requested by family members at weddings or festivals where balafonists are playing. Similar songs exist for the chiefs of each Sambla village. Without documenting the surrogate language, I would likely have never learned about this aspect of Sambla societal structure.

The rich cultural content of musical surrogate languages, as well as its integration with other aspects of the culture (dance, religion, social relations, etc.), is one of the reasons why communities may take greater interest in the documentation of these systems as opposed to everyday speech. (Although it may be counterintuitive to linguists, music, it turns out, generates more interest than verb conjugation.) Indeed, at least in Africa, many Indigenous scholars have taken great interest in musical surrogate languages (particularly “talking drums”) and promoted their value as a source of knowledge about history, literature, and the arts. For instance, renowned Ghanaian musicologist J.H. Kwabena Nketia dedicated several articles and monographs to exploring the linguistic, musical, and literary contributions of Akan and other African drummed speech traditions (Nketia 1958, 1963, 1971). In Côte d’Ivoire, Prof. Georges Niangoran-Bouah proposed the field of *drummologie* as an anti-colonial endeavor, recognizing the technological and scholarly achievement that these drum systems represent (Arnaut & Blommaert 2009; Niangoran-Bouah 1980). Building upon this, in Burkina Faso, the writer Titinga Pacéré developed the notion of *bendrologie* for the study of drum literature in the Mooré language. In each of these cases, the tight connection between musical surrogate languages and their literary elements is recognized to be inseparable.

Thus, not only does the inclusion of musical surrogate languages in language documentation offer novel opportunities for understanding linguistic structure, it also opens up a world of oral literature, local history, and emic understandings of language science. The inclusion and promotion of this kind of Indigenous scholarship in language documentation represents a step towards anti-colonial linguistics, taking practitioners of musical surrogate languages, with their theories of cross-modal mapping, as linguists in their own right, rather than simply language consultants who provide data to Western science.<sup>11</sup>

On the flip side, the close connection between musical surrogate languages and Indigenous religious traditions can be one of the reasons why these practices are endangered or have already fallen out of use. Due to changing religious affiliations (or external pressures from religious groups), former practitioners may be hesitant to publicly play them. For example, in Igbo territories, Christian groups have invited priests and pastors to burn the giant slit log *ikoro* drums, which are sacred drums used for speech surrogacy (Eze & McPherson 2025). The destruction of these instruments and the musical-linguistic practices that depend on them represents the destruction of Indigenous science and ways of thinking. In other communities, musical surrogate traditions have adapted to new religious practices, as in the case of Ewondo drumming used to call the faithful to church (Neeley 1996). The researcher must remain sensitive to the cultural and spiritual implications of musical surrogate languages and the repercussions these have for recording and disseminating materials, as also mentioned in Section 4.1 and Section 4.6

---

11 I would like to thank LDD editors Mark Sicoli and Lise Dobrin for raising this important point.

below. If it is safe for all involved, linguists and language documentarians have an opportunity to act in solidarity with communities to preserve and promote the physical instruments and cultural practices that form the heart of Indigenous language science.

### 3.5 Genres and communicative ecology

A final reason to document musical surrogate languages is that they help us gain a fuller understanding of a society's communicative ecology. Spoken language is just one piece of the puzzle, though arguably the dominant one. If we wish to understand how people communicate with one another (and with the physical and spiritual world around them) at different times and in different contexts, then seeking whether a community has instrumental communicative practices is of utmost importance. To understand their role in the communicative ecology, we must ask not simply how do they work, but also when, where, and why are they used? What speech genres do they represent? What impact does their use have on comprehension or discourse strategies? This can likewise help us understand why musical surrogate languages may be structured the way that they are.

I present here a case study from Hmong culture, which has surrogate traditions on a number of different instruments, each with its own communicative niche. The *ncas* jaw harp, discussed in Section 3.2, is used primarily for courtship (Catlin 1982; Mareschal 1976). Why would a culture use speech surrogacy for courtship? A first reason is that music, if not speech surrogacy, is routinely used in human cultures for courting, and musical surrogate languages are a form of music. A further reason, cited in Catlin and echoed by some of our Hmong consultants, is that using an instrument allows a form of dissociation between the player as an individual and their message, such that if the courtship is rejected, it can be viewed as a rejection of the song rather than the self. But a convenient side effect of using an instrumental surrogate is that it disguises the voice of the lover from others who might be listening in.

So why use a jaw harp instead of a drum, flute, or xylophone?<sup>12</sup> The sound of the jaw harp is quiet and intimate, and so it can only be properly heard in close proximity to the intended listener (rather than declaring messages of love across the hills). It is also small and portable, so can easily be tucked into a pocket and brought on (sometimes clandestine) courtship missions. I note also an interesting connection between speech surrogacy and a society's material culture: jaw harps are only used for courtship in communities whose houses have bamboo walls, which allows the sound to easily pass through (Neng Thao, personal communication, 2025). Our Hmong consultants describe that a young man would go to the house of the young woman he was courting and start playing on the other side of the wall from where she slept until she heard him. At this point, she could either ignore it (being so quiet, the instrument gives her plausible deniability) or she can begin to play in response. Lovers get to know the playing style of their partners, and so even if the *ncas* disguises the person's voice to others listening in, the pair will be able to recognize each other.

As the exchange above suggests, the *ncas* can be used by both men and women, a relative rarity for speech surrogacy traditions in my own experience (most tend to be male-dominated). The genre of language that is encoded on the *ncas* is not regular everyday speech, but rather poetic verses around the themes of love and longing. This heightens the emotional impact of the messages and also helps the

---

12 Indeed, the use of jaw harps in courtship in particular has been attested across a vast number of unrelated cultures, perhaps due to its quiet intimate nature, or its small size, or even its physical construction being "symbolic of procreative imagery" (Morgan 2008: 28).

listener make better guesses about what is being said, since it restricts the hypothesis space to phrases that fall within those semantic and aesthetic bounds.

We can compare the use of the *ncas* to the use of the *qeej*, a free-reed mouth organ with six bamboo pipes (Catlin 1982; Falk 2004). The *qeej* is a considerably louder instrument that also produces polyphonic music by combining a drone tone with other musical notes. Catlin (1982) argues that this obscures direct tone-note correspondences to the extent that the *qeej* should be considered an encoding rather than abridging phonemic system in Stern's (1957) terminology. According to Falk (2004), this more obscured relationship between language and music may be beneficial, since the instrument is used to encode a lengthy funeral repertoire, with messages that give instructions to the spirits of the deceased. In this sense, the messages are meant not so much for human listeners but rather for those in the spirit realm, and these "dangerous messages" could lead a living person's spirit astray. This funeral repertoire is fixed and learned by rote, unlike the more creative poetic language used on the *ncas*. However, it should be noted that the *qeej* is also a popular instrument outside of funeral contexts, such as at New Years festivities. Thus, its messages need not be dangerous ones, but this louder more melodious instrument is considerably better suited to both contexts than the quiet (and more speech-like) *ncas*. Traditionally, only men play the *qeej*.

Meyer & Manfredi (2025) likewise note that "talking and singing with a same instrument can be robustly distinguished in performance thanks to rhythmic and melodic cues", citing cases including the Bora drums (Seifart et al. 2018), Gavião instruments (Meyer & Moore 2021), and Yorùbá drumming (Durojaye et al. 2021), among others. This same distinction was noted for the Sambla balafon in Section 2.3, where speech genre had ramifications for tone encoding. We must thus remember that musical surrogate speech, even on a single instrument in a single culture, may be not monolithic. It may not represent a single genre, but may itself be multifaceted. Different settings or styles may involve different registers, rules of encoding, or types of language.

Apart from distinctions in speech genre (poetry, song, speech, chants, proclamations, etc.), musical surrogate languages can also differ in terms of their discourse structure. Many long-distance drumming traditions are meant to communicate one-way: They send out a message, but do not expect to receive a (real time) response. The message often follows a strict template to render it more comprehensible, as detailed for the Bora slit log drum (Seifart et al. 2018) and many of the Papuan *garamut* signaling systems (Niles 2010). More local musical surrogate languages can either be performance-oriented, as in the Hmong *qeej*, with no expected verbal interaction from bystanders, or conversational, as in the case of the Sambla balafon. In a performance, the balafonist can stop the music and converse with a member of the audience, the musician using notes of the balafon and the audience member speaking normally in Seenku (McPherson 2018). This conversational usage also means that the kind of speech encoded on the balafon in these instances is less poetic and song-like.<sup>13</sup> In the Yorùbá *dùndún* drumming tradition, it is common for the drum to directly repeat and mimic the speech of a "hype man" or other participant, rather than being used conversationally (Akinbo 2021). The use of speech surrogates can also place limits on the syntactic complexity of messages (Sicoli 2016), likely to mitigate the difficulties in comprehension that arise from the reduced set of contrasts.

---

13 Relatedly, whistled speech is often used for long-distance conversation, encoding more ordinary conversational speech and thus requiring sometimes interesting turn-taking strategies that differ from regular speech (Sicoli 2016).

These examples show that, in the documentation of musical surrogate languages, it is important not just to record musical speech (and ideally analyze its connection to the spoken language), but also to document where, when, and why it is used, how the messages are structured (both in terms of genre as well as in their discourse practices), and how the system fits into a broader communicative ecology.

## 4. A BRIEF “HOW TO” GUIDE

For the linguistic researcher who is willing and interested to document musical surrogate languages but doesn't know where to start (or doesn't feel fully equipped to do so), this section offers some brief suggestions. Obviously, this methodology can be adapted depending upon the needs, interests, and material capabilities of the researcher.

### 4.1 Working with musicians: Practical and ethical considerations

Knowledge of musical surrogate languages can be highly specialized, mastered by only a (sometimes small) subset of the population. In some cases, performance of musical surrogate speech may be their livelihood, and so the collection of these materials may mean something very different from the kind of data collection routinely done in language documentation. It can take years, or indeed a lifetime, of practice to understand the depths of the tradition, and these true masters can be highly sought after by the community; thus, it may be very difficult to work with these individuals, as their time is at a premium.

All of these factors come with financial and ethical considerations. Typically in language documentation, linguists compensate consultants for their time and participation, either through modest financial contributions, returning the time and effort on community projects, or gifts to individuals or the community at large. For expert musicians, however, who are sharing their craft (and their intellectual material—more on this below), higher payments are merited (and often necessary). The challenge for the researcher is to determine what is a reasonable payment based on local norms, the musician's skill level and seniority, and the research budget. Here, there are no easy answers.<sup>14</sup> I have found in my own work that the best course of action is to find musician collaborators who see the value in the documentary work and so are themselves stakeholders in its success. They too must be paid appropriately for their expertise, but when they understand the scientific (rather than media) motivations for the work and the researcher's budget, then these payments tend to be reasonable. My general approach has been to pay musicians between two and four times what I would pay a regular speaker for the equivalent time of work. This standard acknowledges the artistry, expertise, and cultural value of what the musician is sharing. The exact amounts, as with any research compensation, depend upon the local economy, the relationship with the community, and the project resources.

Sometimes very well-known or senior musicians may simply be out of budget. I have had the experience of a musician (not based in the United States!) asking for over \$2000 to film a concert, which was infeasible. In cases like this, my recommendation would be to identify skilled musicians (with the help of local community members) who are a bit younger or perhaps more scientifically inclined, even if they are not the recognized as being the top expert in the tradition. They still have valuable skills and insights,

---

<sup>14</sup> This is of course also true in language documentation more broadly. For resources on the ethics of language documentation, see Austin (2010), Dobrin (2008), Dwyer (2006), Good (2018), Sarvasy (2025), and Woods (2023), among many others.

and when the limits of their knowledge are reached, it may be possible to negotiate a short session with a more senior expert (at a potentially higher cost) to fill any gaps. This also offers the opportunity to document how musical surrogate languages are learned and transmitted from one generation to the next.

Another consideration in recording musical surrogate languages is whether any of the information is secret, sacred, or only meant to be understood by a particular (e.g., initiated) sector of society. In such cases, it may not be appropriate to research the traditions at all, or there may need to be special provisions made for archiving or disseminating results. Close consultation with the community is needed to determine appropriate boundaries in these (and indeed all!) cases.

Finally, the researcher should remain sensitive to the potential implications of recording the musical tradition at all. For instance, in her work on recording Garlu folk songs in Nepal, Blumenthal (2007) found that the availability of recordings rendered the caste of musicians who traditionally played them expendable. There can thus be a delicate balance between documenting and preserving threatened oral literature and preserving the cultural ecosystem that has kept it alive.

For issues of intellectual property that arise when recording (and archiving) music, see Section 4.6.

## 4.2 Recording considerations

Linguistic training in recording techniques is tailored to the human voice. Musical instruments present a unique challenge, especially if the recording also includes a spoken interview. Many instruments, especially percussive ones, will be much louder than the human voice and so will result in extensive clipping if a microphone is set to capture speaking. If available, two (or more) microphones should be set up with different input levels, one best suited for the instrument and one for the human voice. The two recordings can either be mixed together later or both can be archived, so as to have access to both recorded streams.

Some instruments are used naturalistically in larger ensembles, but this presents a challenging recording environment. If the researcher wants to capture the whole musical performance, which allows for documentation of important aspects of participation structure and discourse rather than isolated messages, then ideally multiple microphones would be used for different instruments, with musicians spaced as far from one another as feasible to reduce cross-talk, microphones picking up input from other instruments. **Figure 8** shows one such set up from fieldwork I carried out in Benin, where I also had the benefit of working with a professional sound engineer. (Would that we would always be so lucky in language documentation!) Here, having the musicians so spaced out was an unnatural set up for musical performance, but provided the best possible recording conditions while allowing musicians to play simultaneously. To avoid this situation, if one particular instrument is the target of research, then the microphone could be placed closest to this instrument so that its sounds would stand out above the rest.

Ultimately, it would be ideal to collect different kinds of recordings to answer different kinds of questions: To focus solely on the instrument as a means of surrogacy, i.e., to probe how the linguistic message is translated into musical performance, it is best to record it and its messages played solo, separate from a larger ensemble. This will be especially true for structured elicitation, discussed in Section 4.4. However, to capture participation structure both between musicians as well as between musicians and those around them (audience members, dancers, spectators, etc.), then wide angle video, little to no interference in

terms of natural positioning of musicians, and separate microphones on the primary instruments of communication would provide the best data to showcase how the speech surrogate is naturally used.



**Figure 8:** Distancing of musicians in a recording set-up in Savalou, Benin.

### 4.3 Naturalistic phrases and naturalistic performance

In analyzing a musical surrogate language, we can ask three questions: 1. What does the instrument generally say, i.e. what kinds of messages does it naturally send in the culture? 2. How productive is the system? Could it, in principle, say anything? 3. How is the surrogate actually performed, and how does it fit into the community's broader communicative ecology?

There are (at least) two ways to answer these questions, each with their pros and cons. The first is to conduct formal interviews with musicians asking about their usual repertoire of words and phrases; the second is to record natural instances of musical surrogate languages being used. The former makes it easier to address the recording considerations discussed in the last subsection, rendering the phrases easier to hear and analyze from a structural perspective, but this set up makes it very difficult to answer the third question—how is speech surrogacy actually performed? As noted above, if feasible, both kinds of recordings should be made.

Let us first assume that we are staging a formal interview with musicians. Answering the first question involves recording naturalistic phrases that are part of the musical surrogate's typical repertoire. This tends to be much easier for musicians than structured elicitation to probe productivity, discussed in Section 4.4. To record these phrases, I would set up an interview with the musician and their instrument and begin by asking about the contexts in which they play their instrument and what kinds of messages it can send. This offers the player the opportunity to demonstrate phrases and provides some answers to the third question of communicative niche. For structural analysis, it is of utmost importance that the phrase played by the instrument also be spoken out loud to determine the ways in which speech maps to music. Ideally, this would be done by the musician, since it is their understanding (and perhaps pronunciation) of a phrase that is being mapped to the instrument, but in cases where this isn't possible, another speaker well-versed in the musical tradition could speak them instead. In Sambla culture, for instance, there are taboos against the musician speaking the words (McPherson 2018). **Video 3** demonstrates this kind of back-and-forth for the Igbo *ufie* (slit-log drum). One should aim to have the musical phrases and spoken phrases repeated more than once to safeguard against errors or to capture variation.



**Video 3:** Matthias Chinwendu Ugwoke, Michael Odome, and Lawrence Oke of the group Ozor Amauwenu demonstrate speech surrogacy on the Igbo *ufie* (slit log drum). This video was recorded in Nigeria in 2021 by Aaron Carter-Ényì and the Africana Digital Ethnography Project (ADEPT).

By conducting an interview with the musicians about cultural contexts and modes of use, the researcher has the opportunity to both learn about the uses of the instrument as well as to probe different kinds of phrases and messages. For instance, messages are likely to differ in a funeral setting vs. a marriage setting, and by talking about both contexts, there is the opportunity to demonstrate more phrases. For the language documentarian, conducting interviews of this sort in the target language also produces valuable narratives to enrich the documentary corpus (McPherson 2018). It is also likely to present new (musical) vocabulary to add to a lexicon.

When asking for naturalistic phrases, one should also probe whether the instrument has a distinct speech mode for directly encoding linguistic form vs. a musical or singing mode, which may be associated with songs and their lyrics but not so directly, as discussed in Section 2.3. Though the latter may fall at the margins of what could be considered a musical surrogate language, it is still a related and valuable cultural tradition worthy of documentation.

Of course, it would also be desirable to record speech surrogacy in its natural environment, say, in a wedding setting. This would allow us to better understand how musical surrogate languages perform their communicative function and how (or whether) they interact with other linguistic modes (regular speech, song, etc.). However, this raises the recording issues pointed out in Section 4.2 and does not offer the added benefits of a narrative about cultural practices, direct repetition of phrases, and so on. Both forms of recordings offer benefits and drawbacks.

#### 4.4 Structured elicitation

Capturing naturalistic phrases shows what kinds of speech the musical surrogate language is used to naturally produce. However, there are at least two reasons to also carry out some structured elicitation of novel or non-canonical phrases on the instrument. The first is that naturalistic phrases may be rote, proverbial, and potentially rife with archaic forms that differ from how the message would actually be uttered in the spoken language. Oftentimes musicians may know the verbal (spoken) forms of these musical messages, but it is also conceivable that they do not—they simply know that this is the message that is played to convey this particular information. Thus, it can be difficult to determine how or whether aspects of spoken language are actually encoded on the instrument. The second reason is that focusing only on naturalistic phrases cannot tell us whether or not the musical surrogate language is productive: do musicians have actual rules for translating speech to the instrument in their head, or is it all rote? This itself may be a spectrum, from fully rote systems (most likely the case for lexical ideogram systems with arbitrary rhythms) to fully productive ones (like the Sambla balafon or the Yorùbá *dùndún*, where conventions of practice in the surrogate system regularly produce novel utterances); many systems may fall somewhere in between, with mostly rote or conventional phrases but encoding that is transparent enough for musicians to be able to produce novel phrases as well.

As a first pass to test productivity, I will often ask musicians to say an everyday statement like “bring me a beer” or “I want to go lie down” that is not already part of the instrument’s repertoire. If they are able to do this, a few other phrases can be offered, of varying length and morphosyntactic complexity, as well as containing different phonological contrasts (tone, vowel length, etc.).

If there is time and desire to more systematically analyze the structure of the musical surrogate language, then a more structured elicitation is possible. For instance, if it appears that the musical surrogate language encodes tone (as they often do), then an elicitation list containing words with different tones (and combinations of tones) can be put together. If one is interested in morphological structure in the surrogate language, then the elicitation list might include singulars and plurals or verbs in different tenses.

It should be noted, though, that doing elicitation with a surrogate language presents some unique challenges when compared to traditional spoken elicitation. First, some instruments may require quite a bit of physical effort to play, and so care should be taken to not tire the musician with an exceedingly long elicitation session. Further, musicians may find it highly unnatural to play a single word in isolation

when the instrument is typically used for longer phrases. Carrier sentences may thus be important in such cases. Longer phrases should also be included for elicitations targeting the tone system, since a word played in isolation cannot tell you whether the mapping from tone to notes is consistent or whether it shifts over the course of a musical phrase, whether due to linguistic reasons like downdrift/declination or musical considerations.

I offer the following workflow as a suggestion, taking into account these different kinds of recordings and what they teach us about musical surrogate languages. First, if a community is still actively practicing speech surrogacy, then begin by recording it in its natural context, which provides crucial information about participation and discourse structure, as well as naturalistic data for structural analysis. Second, further probe these naturalistic phrases with targeted recording, even going so far as to rewatch the original recording with either the same or another musician, asking them to reproduce the phrases that they hear. Third, based on this naturalistic repertoire, begin to form hypotheses about the grammar of the musical surrogate language and how it encodes linguistic structure, which can then be tested by structured elicitation of novel phrases. For instance, if vowel length and speech rhythm appear to play a role, then the elicitation list can include words with different vowel lengths or metrical configurations.

#### 4.5 Transcription

Linguists are not usually trained in musical notation, nor should we assume that Western musical transcription is appropriate for all musical cultures around the world. How, then, should we transcribe musical surrogate speech for analysis and archiving?

I do not pretend to have all the answers, nor do I believe that there is a single best way to transcribe musical surrogate languages. How they are transcribed depends on the instrument and on the aims and capabilities of the researchers. If a faithful musical transcription is desired, then something akin to Western musical staff notation might be the most apt, as it can capture various aspects of musical performance, including notes, rhythms, and dynamics. The issue, however (beyond the general issues with imposing Western musical notation on non-Western music), is that this staff notation does not combine well with standard linguistic annotations, such as those produced in ELAN. Sleeper and Basurto (2022) offer a methodology for including Western musical notation for vocal music in ELAN through the use of a plain-text musical annotation system known as ABC notation (Walshaw 2011), which could certainly be extended to musical surrogate languages, if the researcher has some familiarity with musical transcription.

What I have often done, however, is to create a shorthand for the musical system that can capture the instrument's contrastive categories (i.e. notes or timbres) with some coarse-grained information about rhythm. Consider the simple example in **Figure 9**, taken from McPherson and Obiri-Yeboah (2024). This is a sample transcription of Akan *atumpan* barrel drums, which have just two contrastive categories: a strike on the high-pitched drum, abbreviated H, and a strike on the low-pitched drum, abbreviated L.<sup>15</sup> Distinct beats are separated by a hyphen, while two beats in rapid succession (a flam) contain no

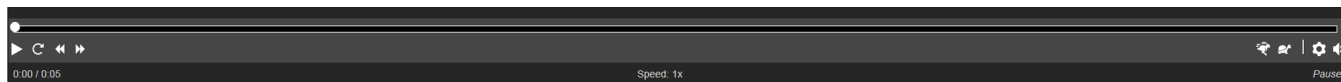
---

<sup>15</sup> As noted in Section 3.1, not all communities conceptualize pitch using the same metaphors. My use of High and Low here is representative of English terminology, with higher f0 being “High” and lower f0 being “Low”. For the Sambla balafon, lower pitches are conversely referred to as “high” because their large resonating calabashes put them physically higher off the ground. For more discussion, see Shayan et al. (2011).

hyphen. Spaces indicate word boundaries (and separate beats) to allow for easy tokenizing of the tier into annotations corresponding with individual words.

Surrogate_Transcription-txt-twi [186]	Pàpá Kòfí kàsá kyèré nè bà?	Pàpá Kòfí kàsá kyèré nè bà?
Surrogate_Music-txt-twi [185]	L-HH L-H L-H L-H L HL	L-HH L-H L-H L-H L HL
Surrogate_Music-words-txt-twi [795]	L-HH L-H L-H L-H L HL	L-HH L-H L-H L-H L HL
Surrogate_Words-txt-twi [802]	Pàpá Kòfí kàsá kyèré nè bà?	Pàpá Kòfí kàsá kyèré nè bà?
Surrogate_Tones-txt-en [800]	L-H L-H L-H L-H L HL	L-H L-H L-H L-H L HL
Surrogate_Translation-gls-eng [185]	Father Kofi talks to his child?	Father Kofi talks to his child?

**Figure 9:** ELAN transcription of Akan *atumpan*; L and H refer to both drums and linguistic tones.



AUDIO CLIP: Osei Korankye demonstrates speech surrogacy on the Akan *atumpan*, corresponding to **Figure 9**. For the full audio, visit the online journal at <https://iddjournal.org/>.

For scaled instruments like the Sambla balafon or the Igbo *òjà*, the scale degrees can be numbered, with a note in the document laying out the frequencies or closest approximate Western notes for each note of the scale. This allows the transcription to be abstract and represent the capabilities of the instrument, rather than imposing a Western interpretation, which may suggest the existence of intermediate notes that do not, in fact, exist on the instrument and obscure generalizations about the mapping of linguistic structure to musical form. For instance, a H to !H tone transition may be mapped onto two adjacent notes of the scale regardless of whether the pitch interval between the two notes is roughly one or two semitones (as is the case for the Igbo *òjà*).

Timbrally complex drums pose a challenge, with different outputs depending upon how the drum membrane is struck. In our current work, my colleagues and I are developing a shorthand for these kinds of strokes, such as D for a strike with the fingers/digits, T for a strike with the thumb, P for a strike with the palm, etc. “Dm” could refer to a muted finger strike, “DE” could refer to a finger strike on the edge of the drum, etc.

Certain traditions might include an emic method for representing sounds of the musical instruments, whether visual or oral (e.g. the use of vocables). These could also be taken into consideration by the research team in deciding how best to represent the recordings.

So long as the transcription conventions are clearly laid out and explained, then precisely how they are done can be tailored to the needs of the instrument and those involved in the research.

#### 4.6 Archiving: Metadata and intellectual property

As with any archiving, metadata is crucial in the documentation of musical surrogate languages. The metadata should clearly state in the topics or keywords “speech surrogacy” or “musical surrogate

language”, so that someone interested in speech surrogates will be able to find these recordings on a search for them. I suspect that such recordings already exist in language archives, but without some sort of clear indication of surrogacy, they are unfindable without wading through all instances of “drums” or “music” to find ones that have communicative functions.

Another consideration specific to musical surrogate languages is that these are forms of musical expression, and recordings may also include original music. This should be explicitly acknowledged in the metadata as being the intellectual property of the musicians and subject to all applicable local and international copyright laws. Explicit care must be taken in securing informed consent from musicians to determine what level of access they want to grant to their recordings. This may differ greatly between communities and between different kinds of musical surrogate languages, depending upon their level of secrecy, their ties to spiritual beliefs, and their use as public entertainment, among other things. If there is a tight connection between the surrogate tradition and spiritual or sacred matters, extra care must be taken in securing the recordings, as deemed appropriate by the community.

## 5. CONCLUSIONS

This paper lays the groundwork for a new subfield of linguistics—*surrogate linguistics*—by providing a brief introduction to the world of musical surrogate languages, their varied structures, and what they can teach us about linguistic structure and communicative practices. It is my hope that this will inspire documentary linguists working in communities around the world to take a moment to ask about these traditions and make, at the very least, some preliminary recordings.

Time is of the essence in this work, even more than it is for spoken languages, given that musical surrogate speech is often specialized knowledge and tied to aspects of culture and lifestyle that are undergoing rapid changes: long-distance communication technology, religious changes, even changes to building materials that affect the transmission of sound. There are many communities where it is already too late; there is knowledge that such a system existed, but no one today knows how to play it.

By leveraging our existing language documentation projects and community connections, we have the opportunity to fill in significant gaps in our knowledge of musical surrogate languages around the world and better understand how these unique amalgams of language and music have evolved, are passed down, and are understood by listeners. In doing so, we are likely to gain an even deeper understanding of the spoken language and its community of users. To quote Laycock (1977:144, quoted in Niles 2010:39): “It is only when a linguist knows at least some of the ‘special languages’ that he can be said to be fully ‘initiated’ into the linguistics of the New Guinea area.” This sentiment rings true regardless of the geographic location of the communities with which we work.

## ACKNOWLEDGMENTS

This work has been over ten years in the making, and as such, has been shaped, supported, and inspired by more people than I could ever list by name. I am especially grateful for my collaborators, Samuel Akinbo, Andre Batchelder-Schwab, Aaron Carter-Ényì, Mamadou Diabaté, Gerald Eze, Idiongouwem Obong, Michael Obiri-Yeboah, Anthony Struthers-Young, Neng Thao, and Yoad Winter, for driving this work forward. I owe a great debt to many Dartmouth students who have worked on this project, but especially Lucas James for his immense typological work in the early stages of this project, Emma

Rafkin, Alina Chadwick, and Josué Godeme for their management of the website and database, Ford Springer, for his collaboration on the Hmong *ncas*, and Ryan Beveridge, who helped bring the database up to speed for the publication of this article. Thank you to Ted Levin for helping with musicological perspectives and to Jim Stanford for inspiring the term “surrogate linguistics”, as well as to sound engineers Burt Gregory and Kamal Humphrey for their technical expertise and enthusiastic collaboration over the years. This research was supported by the National Science Foundation (BCS-1943896). All figures or recordings taken from other published sources have been used with explicit permission of the author and/or publisher. Final thanks go to two anonymous reviewers and to the editors Mark Sicoli and Lise Dobrin, whose comments and suggestions have greatly enriched and strengthened this manuscript. All errors are my own.

## COMPETING INTERESTS

The author has no competing interests to declare.

## AUTHOR AFFILIATIONS

Laura McPherson  [orcid.org/0000-0002-9027-0375](https://orcid.org/0000-0002-9027-0375)  
Department of Linguistics, Dartmouth College, US

## REFERENCES

- Akinbo, Samuel. 2019. Representation of Yorùbá tones by a talking drum: An acoustic analysis. *Linguistique et Langues Africaines* 1: 11–23. <https://doi.org/10.4000/lla.347>
- Akinbo, Samuel. 2021. The language of gángan, a Yorùbá talking drum. *Frontiers in Communication* 6: 650382. <https://doi.org/10.3389/fcomm.2021.650382>
- Arnaut, Karel & Jan Blommaert. 2009. Chthonic science: Georges Niangoran-Bouah and the anthropology of belonging in Côte d’Ivoire. *American Ethnologist* 36(3): 574–590.
- Austin, Peter K., 2010. Communities, ethics, and rights in language documentation. *Language Documentation and Description* 7: 34–54. <https://doi.org/10.25894/idd226>
- Bagemihl, Bruce. 1988. *Alternate phonologies and morphologies*. Vancouver: University of British Columbia PhD dissertation. <http://hdl.handle.net/2429/28617>
- Blench, Roger & Fredeliza Campos. 2010. Recording oral literature in a literate society: A case study from the Northern Philippines. *Language Documentation and Description* 8: 1–16.
- Blumenthal, Katherine. 2007. Hierarchical transformation and song circulation: Documenting the Garlu folksongs of Lo Monthang, Nepal. Charlottesville, VA: University of Virginia master’s thesis.
- Burridge, Kenelm O. L. 1959/1976. The slit-gong in Tangu, New Guinea. In Thomas A. Sebeok & Jean Umiker-Sebeok (eds.), *Speech surrogates: Drum and whistle systems*, 1236–1247. Berlin/Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110804423-031>
- Busnel, René-Guy & André Classe. 1976. *Whistled languages*. Berlin: Springer-Verlag.
- Carrington, John. 1949/1976. *The talking drums of Africa*. London: Carey-Kingstate Press. Partially reprinted in Sebeok & Umiker-Sebeok (1976).
- Carter-Ényì, Aaron, Nnaemeka C. Amadi, Quintina Carter-Ényì, Charles Chukwudozie, Jude Nwankwo & Ebruphiyor Omodoro. 2021. Igbo speech surrogacy: Preliminary findings based on the oja flute. *Frontiers in Psychology* 12: 653068. <https://doi.org/10.3389/fpsyg.2021.653068>

- Catlin, Amy. 1982. Speech surrogate systems of the Hmong: From singing voices to talking reeds. In Bruce T. Downing & Douglas P. Olney (eds.), *The Hmong in the West: Observations and reports*, 170–197. Minneapolis, MN: University of Minnesota Center for Urban and Regional Affairs.
- Connell, Bruce. 2011. Downstep. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 824–847. Malden, MA: WileyBlackwell. <https://doi.org/10.1002/9781444335262.wbctp0035>
- Cowan, George M. 1972. Segmental features of whistled Tepehua. *Proceedings of the Seventh International Congress of Phonetic Sciences*, Montreal, 695–698.
- Deacon, A. B. 1934/1976. *Malekula*. London: Routledge and Sons, Ltd. Partially reprinted in Sebeok & Umiker-Sebeok (1976).
- Dobrin, Lise M. 2008. From linguistic elicitation to eliciting the linguist: Lessons in community empowerment from Melanesia. *Language* 84(2): 300–324. <https://doi.org/10.1353/lan.0.0009>
- Durojaye, Cecilia, Kristina L. Knowles, K. Jakob Patten, Mordecai J. Garcia & Michael K. McBeath. 2021. When music speaks: An acoustic study of the speech surrogacy of the Nigerian dũndún talking drum. *Frontiers in Communication* 6: 652690. <https://doi.org/10.3389/fcomm.2021.652690>
- Dwyer, Arienne M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 31–66. Berlin: Mouton de Gruyter.
- Eze, Gerald & Laura McPherson. 2025. From masquerades to modern crises: Endangered messages in the Igbo oja surrogate language. In Alireza Korangy & Evegenia Guzhova (eds.), *Multilingualism, identity, and language endangerment in Africa*, 575–602. Singapore: Springer. [https://doi.org/10.1007/978-981-96-4729-3\\_21](https://doi.org/10.1007/978-981-96-4729-3_21)
- Fales, Cornelia. 2017. Issues of timbre: The *Inanga Chuchotée*. In Ruth M. Stone (ed.), *The Garland encyclopedia of world music: Africa*, 164–207. New York: Routledge.
- Falk, Catherine. 2004. Hmong instructions to the dead: What the mouth organ *qeej* says (Part one). *Asian Folklore Studies* 63(1): 1–29. <https://www.jstor.org/stable/30030310>
- González, Mariano & Olupemi Oludare. 2021. The speech surrogacy systems of the Yoruba dũndún and bàtá drums: On the interface between organology and phonology. *Frontiers in Communication* 6: 652542. <https://doi.org/10.3389/fcomm.2021.652542>
- Good, Jeff. 2018. Ethics in language documentation and revitalization. In Kenneth Rehg & Lyle Campbell (eds.), *Oxford handbook of endangered languages*, 419–440. Oxford: Oxford University Press.
- Holton, Gary. 2018. Interdisciplinary language documentation. In Kenneth Rehg & Lyle Campbell (eds.), *Oxford handbook of endangered languages*, 739–760. Oxford: Oxford University Press.
- James, Lucas. 2021. Systems of communication: Aspects of culture and structure in speech surrogates. *Frontiers in Communication* 6: 653268. <https://doi.org/10.3389/fcomm.2021.653268>
- Kaminski, Joseph. 2008. Surrogate speech of the Asante ivory trumpeters of Ghana. *Yearbook for Traditional Music* 40:117–135.
- Kay, Paul & Chad K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. *Language* 54(3): 610–646. <https://doi.org/10.2307/412789>
- Ladd, D. Robert & James Kirby. 2020. Tone-melody matching in tone language singing. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford Handbook of Language Prosody*, 676–688. Oxford: Oxford University Press.

- Laycock, Donald C. 1977. Special Languages in Parts of the New Guinea Area. In Stephen Wurm (ed.), *New Guinea area languages and language study, Vol. 3: Language, culture, society and the modern world*, 133–149. Pacific Linguistics series C 40. Canberra: Australian National University.
- Lehiste, Ilse & Gordon Peterson. 1959. Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America* 31(4): 428–435.
- Levin, Theodore & Valentina Süzükei. 2006. *Where rivers and mountains sing*. Bloomington, Indiana: Indiana University Press.
- Mareschal, Eric. 1976. *La musique des Hmong* [Hmong music]. Paris: Musée Guimet.
- Matisoff, James A. 1990. Bulging monosyllables: Areal tendencies in Southeast Asian diachrony. In David J. Costa (ed.), *Proceedings of the 16<sup>th</sup> annual meeting of the Berkeley Linguistics Society*, 543–559. Berkeley: Berkeley Linguistics Society.
- McPherson, Laura. 2018. The talking balafon of the Sambla: Grammatical principles and documentary implications. *Anthropological Linguistics* 60(3): 255–294.
- McPherson, Laura. 2019a. Tone: the present state and future potential. *Language: Commentaries* 95(1): e188–e192.
- McPherson, Laura. 2019. Musical adaptation as phonological evidence: Case studies from textsetting, rhyme, and musical surrogates. *Language and Linguistics Compass* 13(12): e12359.
- McPherson, Laura. 2020. *A Grammar of Seenku*. Berlin: De Gruyter Mouton.
- McPherson, Laura & Lucas James. 2021. Artistic adaptation of Seenku tone: Musical surrogates vs. vocal music. In Akinbiyi Akinlabi, Lee Bickmore, Michael Cahill, Michael Diercks, Laura J. Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected papers from the 50<sup>th</sup> Annual Conference on African Linguistics*, 203–223. Berlin: Language Science Press.
- McPherson, Laura, Lucas James, Alina Chadwick, Josué Godeme & Emma Rafkin. 2025. *Online Database of Speech Surrogates*. <http://www.speechsurrogates.org>
- McPherson, Laura & Michael Obiri-Yeboah. 2024. Akan tone encoding across musical modality. *Studies in African Linguistics* 52(2): 160–187.
- McPherson, Laura & Kevin Ryan. 2018. Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94(1): 119–156.
- Meyer, Julien. 2015. *Whistled languages: A worldwide inquiry on human whistled speech*. Berlin: Springer-Verlag.
- Meyer, Julien & Victor Manfredi. 2025. Auxiliary speech by vocal tract modulation and musical surrogacy. In Daniela Sammler (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780192894700.013.0021>
- Meyer, Julien & Denny Moore. 2021. A flute, musical bows and bamboo clarinets that “speak” in the Amazon rainforest: Speech and music in the Gavião language of Rondônia. *Frontiers in Psychology* 12: 674289. <https://doi.org/10.3389/fpsyg.2021.674289>
- Morgan, Deirdre. 2008. *Organs and bodies: The jew’s harp and the anthropology of musical instruments*. Vancouver: University of British Columbia master’s thesis.
- Neeley, Paul. 1996. Drummed transactions: Calling the church in Cameroon. *Anthropological Linguistics* 38(4): 683–717.
- Niangoran-Bouah, Georges. 1980. *Introduction à la drummologie* [Introduction to drummology]. Abidjan: Université Nationale de Côte d’Ivoire Institut d’ethnosociologie.

- Nikolsky, Aleksey. 2020. Talking jew's harp and its relation to vowel harmony as a paradigm of formative influence of music on language. In Nobuo Masakata (ed.), *The Origins of Language Revisited: Differentiation from music and the emergence of neurodiversity and autism*, 217–322. Singapore: Springer.
- Niles, Don. 2010. Editor's introduction to Zemp and Kaufmann. *Kulele Occasional Papers in Pacific Music and Dance* 4: 150.
- Nketia, Kwabena J. H. 1958. The poetry of drums. In Henry Swazee (ed.), *Voices of Ghana: Literary contributions to the Ghana Broadcasting System, 1955–57*, 17–23. Accra: Ministry of Information.
- Nketia, Kwabena J. H. 1963. *Drumming in Akan communities of Ghana*. Edinburgh: Published for the University of Ghana by Thomas Nelson.
- Nketia, Kwabena J.H. 1971. Surrogate languages of Africa. In Thomas Sebeok (ed.), *Current Trends in Linguistics vol. 7: Linguistics in Sub-Saharan Africa*, 699732. The Hague: Mouton. <https://doi.org/10.1515/9783111562520-028>
- Poss, Nicholas F. 2012. *Hmong music and language cognition: An interdisciplinary investigation*. Columbus, OH: The Ohio State University PhD dissertation.
- Proschan, Frank. 1994. Khmu Play Languages. *Mon-Khmer Studies* 23: 43–65.
- Pugh-Kitingan, Jacqueline. 1977. Huli language and instrumental performance. *Ethnomusicology* 21(2): 205–232. <https://doi.org/10.2307/850944>
- Rialland, Annie. 2005. Phonological and phonetic aspects of whistled languages. *Phonology* 22(2): 237–271. <https://doi.org/10.1017/S0952675705000552>
- Ros, Sofiya. 2021. Rhythm-speech correlations in a corpus of Senegalese drum language. *Frontiers in Communication* 6: 643683. <https://doi.org/10.3389/fcomm.2021.643683>
- Rouget, Gilbert. 1964. Tons de la langue en gùn (Dahomey) et tons du tambour [Speech tones in Gùn (Dahomey) and drum tones]. *Revue de Musicologie* 50(128): 3–29.
- Sapir, J. David. 1975. Big and thin: Two Diola-Fogny meta-linguistic terms. *Language in Society* 4(1): 1–15. <https://doi.org/10.1017/s0047404500004462>
- Sarvasy, Hannah S. 2025. Ethical budgets in (psycho-)linguistic fieldwork. *Linguistics* 63(2): 487–510.
- Schellenberg, Murray. 2012. Does language determine music in tone languages? *Ethnomusicology* 56: 266–278.
- Sebeok, Thomas A. & Jean Umiker-Sebeok. 1976. *Speech surrogates: Drum and whistle systems*. Berlin: De Gruyter Mouton.
- Seifart, Frank, Julien Meyer, Sven Grawunder & Laure Dentel. 2018. Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science* 5(4): 170354. <https://doi.org/10.1098/rsos.170354>
- Shayan, Shakila, Ozge Ozturk & Mark Sicoli. 2011. The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *Senses & Society* 6(1): 96–105. <https://doi.org/10.2752/174589311X12893982233911>
- Shepard, Glenn Jr. 1997. Noun classification and ethnozoological classification in Machiguenga, an Arawakan language of the Peruvian Amazon. *The Journal of Amazonian Languages* 1: 29–37.
- Si, Aung. 2011. Biology in language documentation. *Language Documentation & Conservation* 5: 169–186.
- Sicoli, Mark. 2016. Repair organization in Chinantec whistled speech. *Language* 92(2): 411–432.
- Sleeper, Morgan & Griselda Reyes Basurto. 2022. Musicolinguistic documentation: Tone & tune in Tlahuapa Tù'un Sàví songs. *Language Documentation & Conservation* 16: 168–208.

- Snyders, J. 1969/1976. Le langage par tambours à San Cristoval, British Solomon Islands. *Journal de la Société des Océanistes* 24: 133–138. Reprinted in Sebeok & Umiker-Sebeok (1976).
- Stern, Theodore. 1957. Drum and whistle “languages”: An analysis of speech surrogates. *American Anthropologist* 59(3): 487–506.
- Struthers-Young, Anthony. 2022. A preliminary account of the Northern Toussian balafon surrogate language. *Frontiers in Communication* 6: 652740. <https://doi.org/10.3389/fcomm.2021.652740>
- Umeh, John Anenechukwu. 1997. *After God is Dibia* (Vol 1). London: Karnak House.
- Villepastour, Amanda. 2010. *Ancient text messages of the Yorùbá bàtá drum: Cracking the code*. Farnham: Ashgate Publishing, Ltd.
- Viraraghavan, Venkata Subramanian & Sridharan Sankaran. 2023. Salient features of melody, rhythm, and lyrics in Carnatic music. In Preeti Rao, Hema A. Murthy & S.R.M. Prasanna (eds.), *Indian Art Music: A Computational Perspective*, 75–118. Sriranga Digital Software Technologies Pvt. Ltd.
- Walshaw, Christopher. 2011. *The abc music standard 2.1*. (<http://abcnotation.com/wiki/abc:standard:v2.1>)
- Whiting, J. W. M. & S. W. Reed. 1938. Kwoma culture: Report on field work in the mandated territory of New Guinea. *Oceania* 9: 170–216.
- Winter, Yoad. 2014. On the grammar of a Senegalese drum language. *Language* 90(3): 644–668. <https://doi.org/10.1353/lan.2014.0061>
- Woodbury, Anthony. 2003. Defining documentary linguistics. *Language Documentation and Description* 1: 35–51.
- Woods, Lesley. 2023. *Something’s gotta change: Redefining collaborative linguistic research*. Canberra: ANU Press.
- Zemp, Hugo & Christian Kaufmann. 1969. Pour une transcription automatique des “langages tambourines” mélanésiens: Un exemple kwoma, Nouvelle-Guinée [Toward an automatic transcription of Melanesian “drum languages”: A Kwoma example, New Guinea]. *L’Homme* 9(2): 38–88.

Submitted: 15 July 2025

Accepted: 06 November 2025

Published: 06 May 2026