frontiers in Communication

ORIGINAL RESEARCH
published: 17 May 2021
doi: 10.3389/fcomm.2021.652635

Categoricity, Variation, and Gradience in Sambla Balafon Segmental Encoding

Laura McPherson *

Department of Linguistics, Dartmouth College, Hanover, NH, United States

This paper analyzes the musical surrogate encoding of Seenku (Mande, Burkina Faso) syllable structure on the balafon, a resonator xylophone used by the Sambla ethnicity. The elements of syllable structure that are encoded include vowel length, sesquisyllabicity, diphthongs, and nasal codas. Certain elements, like vowel length and sesquisyllabicity, involve categorical encoding through conscious rules of surrogate speech, while others, like diphthongs and nasal codas, vary between being treated as simple or complex. Beyond these categorical encodings, subtler aspects of rhythmic structure find their way into the speech surrogate through durational differences; these include duration differences from phonemic distinctions like vowel length in addition to subphonemic differences due to phrasal position. I argue that these subconscious durational differences arise from a "phonetic filter", which mediates between the musician's inner voice and their non-verbal behavior. Specifically, syllables encoded on the balafon may be timed according to the perceptual center (p-center) of natural spoken rhythm, pointing to a degree of phonetic detail in a musician's inner speech.

Keywords: phonology, phonetics, speech surrogate, rhythm, Burkina Faso, Mande

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco, Sevilla University, Spain

Reviewed by:
Lei He, University of Zurich, Switzerland
Gertraud Fenk-Oczlon, University of Klagenfurt, Austria

***Correspondence:**
Laura McPherson
Laura.e.mcpherson@dartmouth.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *12 January 2021*
Accepted: *26 April 2021*
Published: *17 May 2021*

Citation:
McPherson L (2021) Categoricity, Variation, and Gradience in Sambla Balafon Segmental Encoding. Front. Commun. 6:652635. doi: 10.3389/fcomm.2021.652635

INTRODUCTION

For linguists studying musical surrogate languages, the main question of interest is which linguistic features get encoded into musical form and how? Many older studies of surrogate languages, written largely by anthropologists or ethnomusicologists, contain general statements on the connection between spoken and musical language (e.g. "drums encode tone and speech rhythm") with no further explanation. This leaves us to wonder, which tones are encoded, lexical or grammatical? Which aspects of rhythm—only the phonemic ones, such as vowel length or syllable shape, or also more subtle rhythmic effects?

Through careful comparison of spoken and surrogated language, we can unveil these patterns of encoding. However, I suspect that in many cases we can draw a distinction between conscious elements of encoding—let's call these the grammatical rules of the speech surrogate—and subconscious patterns of encoding. Determining which linguistic elements are consciously encoded and which are subconscious stands to shed light on a speaker's metalinguistic knowledge and the division between phonology and phonetics.

This paper explores these questions by looking at segmental encoding strategies on the Sambla balafon, a West African resonator xylophone used among the Seenku-speaking population of southwestern Burkina Faso. Previous work (McPherson 2018) on the surrogate system focused largely on the encoding of tone, showing a distinction between lexical/morphological and postlexical tone, with the former encoded and the latter ignored by musicians. In this paper, I take a closer look

at segmental encoding, specifically syllable structure (as individual consonant and vowel phonemes are not distinguished in the speech surrogate). I show that musicians' conscious rules of surrogate encoding marks a categorical distinction between simplex and complex syllables, with complex defined as CV: (long vowels) and C(ə)CV (sesquisyllables; Matisoff 1990; Pittayaporn and Enfield, 2015). Two other syllable shapes, namely CVN (nasal codas) and CVV (diphthongs) are variably treated as complex. In addition to these conscious encoding strategies, syllable structure and speech rhythm more broadly also influence the Sambla balafon surrogate in subtler subconscious durational differences that closely mirror spoken language.

This range of encoding behavior—categorical, variable, gradient—can shed light on the phonetic and phonological structure of the spoken language, and what speakers know about that structure. First, the variable encoding of diphthongs is shown to be the result of conscious disambiguation on the part of the musician, demonstrating a recognition of the distinction between monophthongs and diphthongs when such a contrast is needed. The variation with nasal codas, on the other hand, is argued to arise from their weak phonological representation in the language itself (McPherson 2020a; McPherson 2020b; McPherson 2020c), with more "complex" encodings correlated with environments in which they surface as true coda consonants, and more "simplex" encodings correlated with environments in which they are subsumed into the following onset or the preceding vowel. Finally, the presence of gradience in the system suggests that surrogate encoding passes through a subconscious "phonetic filter", in which surrogate speech is mediated by the musician's inner voice. Specifically, syllable timing on the balafon appears to be tied to the perceptual center (p-center) of spoken syllables (Morton et al., 1976; Ryan 2014), despite the fact that musicians are not speaking as they play. This in turn points to a relatively high degree of phonetic detail in inner speech (Corley et al., 2011; Oppenheim 2013; Martin et al., 2018, *inter alia*), demonstrating yet another way in which musical surrogate languages can serve as a crucial source of data for understanding the human language faculty.

The paper is structured as follows. In *Background on Sambla Music and the Seenku Language*, I provide background information on Sambla music and the spoken language, Seenku. *Encoding Strategies* turns to the different encoding strategies, covering categorical (*Categoricity*), variable (*Variability*), and gradient (*Gradience*) encoding in turn. *Discussion* discusses the results in light of the phonetic filter and p-centers, and *Conclusion* concludes.

BACKGROUND ON SAMBLA MUSIC AND THE SEENKU LANGUAGE

The Sambla people are a Mande ethnicity with a population of less than 20,000 living in southwestern Burkina Faso. The name "Sambla" (also spelled Sembla) is an exonym, but is used by the people to refer to themselves when speaking French or Jula. I will be referring to the spoken language by its endonym, Seenku (IPA [sɛ̃́ː-kû]), though it is likewise referred to as Sambla or Sembla in the literature.

In this section, I will briefly lay out the pertinent background information on Sambla balafon music and Seenku phonology, which will set the stage for an analysis of segmental encoding.

Sambla Balafon

Arguably the most important instrument in the Sambla musical tradition is the balafon. The term "balafon" is applied to a range of resonator xylophones found throughout West Africa, where tuned gourds are hung beneath each wooden key to amplify the sound. The balafon is played at all major cultural events, including weddings, funerals, communal work parties, and religious festivals.

Balafon playing is a hereditary profession, passed from father to son in just a small number of families belonging to the griot caste. Unlike some balafon traditions, such as the Senufo balafon orchestras (Zemp and Soro 2010), Sambla balafon music involves just a single instrument at any event. However, three people play at the same time: a simple middle part sets the tempo, while a more complicated bass part creates a layered polyrhythmic backdrop to the "soloist" on the treble. It is this soloist who practices speech surrogacy, through both melodic lines that stand for proverb-rich lyrics of songs (the "sung mode" of surrogacy) and improvised surrogate utterances that communicate directly with the audience or other musicians (the "speech mode" of surrogacy). This paper focuses solely on speech mode, which, as the name suggests, displays a much tighter connection to the spoken language.

The Sambla balafon is tuned to a pentatonic scale, though the exact notes/frequencies vary instrument to instrument; as musicians put it, just like people, each instrument has its own voice. The intervals between the notes, however, remain fixed. The names of the notes, along with their closest corresponding Western scale degrees, are shown in **Table 1**.

As this table shows, the Sambla scale is rather unusual, with the presence of both the minor and major third. However, the minor third, the "fetish balafon key", is only rarely used. It is reserved for spiritual uses and typically does not figure in the speech surrogate, with the exception of one lexical ideogram expression corresponding to "yes" or "that's correct", in which the player alternates between ♭3 and 5. It should be noted that the spatial relations of "under" and "above" in the key names refer to physical location, since low notes have larger gourds beneath them and are hence "higher" from the ground than high notes; in other words, the keys of the balafon slope downwards as one moves from the bass to the treble.

For more information on the Sambla balafon, including its history, construction, and social uses, see Strand (2009) and McPherson (2018).

Seenku Phonology

Seenku is a member of the Samogo group of Western Mande languages. It is phonologically rich, with complex consonantal, vocalic, and tonal inventories. Balancing this complexity is a largely monosyllabic vocabulary, though syllable structure itself

TABLE 1 | Notes and scale degrees of the Sambla balafon.

Western scale degree	Seenku note name	Abbreviation	Gloss
1	*bâa̰-ɲǎ / sərǎ-kŭa*	B / S	"balafon mother" / (no translation)
♭3	*ĵĭo-ba̰a̰-dɛ̌n*	J	"fetish balafon key"
3	*bâa̰-ɲǎ-gŭ-nɔ̰̀n*	Bg	"the one under the balafon mother"
5	*tərɔ́n-tərɔ́n*	T	(no translation)
6	*sərà-kùa-kɔ̰̀n-nɔ̰̀n*	Sk	"the one above the *sərǎ-kŭa*"

TABLE 2 | Seenku consonant inventory.

	Bilabial	Alveolar	Palatal	Velar	Labiovelar
Plosive	p b	t d	c ɟ <j>	k g	kp gb
Nasal	m	n	ɲ	ŋ	ŋm
Affricate		ts dz			
Fricative	f	s			
Approximant		l	j <y>		w

can be complex. For more detail on Seenku's phonological system, see McPherson (2020a), McPherson (2020b).

Segmental Inventories

The consonant inventory is shown in **Table 2**, where any orthographic deviations from IPA are shown in angled brackets.

As this table shows, Seenku contrasts five places of articulation, including labiovelars. Unusually for Mande, it displays voiced and voiceless alveolar affricates. There is no contrast between [l] and [ɾ], with the realization depending upon the context.

The oral vowel inventory contains either eight or nine vowel qualities, depending upon the speaker; this inventory results from an [ATR] contrast among high and mid vowels, though for some speakers, I have seen no evidence of high [-ATR] /ɪ/. Thus, the vowel inventory can be summarized as /i (ɪ) e ɛ a ɔ o ʊ u/. These oral vowels can be either phonemically short or long.

Seenku also has phonemic nasal vowels, though the inventory is more constrained, with just a five-way distinction: /ĩ ɛ̃ ã ɔ̃ ũ/; these nasal vowels may also be short or long.

In addition to these monophthongs, Seenku displays an impressive array of diphthongs, including (but not limited to) /ia iɛ ie ua uo uɔ ɔɛ oe/, etc. A striking fact of Seenku phonology is that diphthongs can also be either short or long, due to the non-moraic nature of the diphthong-initial vocalic element (McPherson, 2020a; McPherson, 2020b).

Tonal Inventory

Seenku has four contrastive tone levels, which I call Superhigh (S), High (H), Low (L), and Extralow (X), marked with double acute (ă), acute (á), grave (à) and double grave (ǎ) diacritics, respectively. The four tone levels are distinguished solely by f0, with no significant role played by vowel length or phonation. A minimal set for level tones is shown in (1):

(1) a. sa̋ 'tree sp.'
b. sá 'reciprocal'
c. sa̖ 'first son (birth order name)'
d. sa̗ 'water jar'

These four level tones can combine to create a vast array of two- and three-tone contours. The most common lexical contours are HX (â), LS (ă), and HS (áă); contours created by grammatical tone and clitic elision include SX (ä), XH (ăá), SH (ă̋á), and HL (áà). Three-tone contours include XHX (e.g. *gɔ̏ɔ̂n* "sorrel") and LSX (e.g. *nàä* "come (perfect)").

Syllable and Word Structure

Most Seenku vocabulary is monosyllabic, but there is a wide range of possible syllable shapes, including sesquisyllabic words (Matisoff 1990; Pittayaporn and Enfield, 2015), i.e. a short minor syllable (Cə) followed by a full syllable; it may also be possible to analyze these as syllables with a complex onset, broken up by an epenthetic or even excrescent vowel (McPherson 2020a). Seenku syllable structure is schematized in (2):

(2) (C)(əC)V(V)(ː)(N)

The only obligatory element is the syllable nucleus, V (though in the case of the 1sg *ń*, it can also be a nasal). Only a small number of pronouns (e.g. 3sg *a̰*, 2pl *í*, etc.) are vowel-initial (indeed, consist of only V). All other vocabulary must have an onset consonant (e.g. *kâ* "griot", *mi̋* "1pl" etc.). Syllables can be either mono- or sesquisyllabic, with the addition of əC after the initial onset consonant (e.g. *mənɔ̋* "woman", *səgà* "sheep", etc.). The syllable nucleus can be either a monophthong, as seen so far, or a diphthong (e.g. *kŭɛ* "others", *səgùa* "stack (v.)", etc.). Both monophthongs and diphthongs can be either short or long, with long vowels indicated here by doubling the vowel (e.g. *kâa* "fight", *kùaa* "farm (intr.)", etc.). Finally, the only permissible coda is a nasal whose place of articulation is non-contrastive. In fact, in most cases, it is realized as either late nasalization of the preceding vowel (contrastive with phonemically nasal vowels) or as nasalization of a following sonorant. Only in rare instances is it realized fully as a nasal stop; for more details, see McPherson (2020c). The repercussions of this weak phonological element for the balafon surrogate language will be discussed below.

Any of these syllable elements may co-occur, with lexical frequency decreasing as complexity increases (McPherson 2020a). In other words, simple CV syllables like *sɔ̏* 'horse' are vastly more frequent than complex syllables like CəCVVːN, of which we find just a single instance *təgŭaan* 'tree sp. (*Carapa procera*)'.

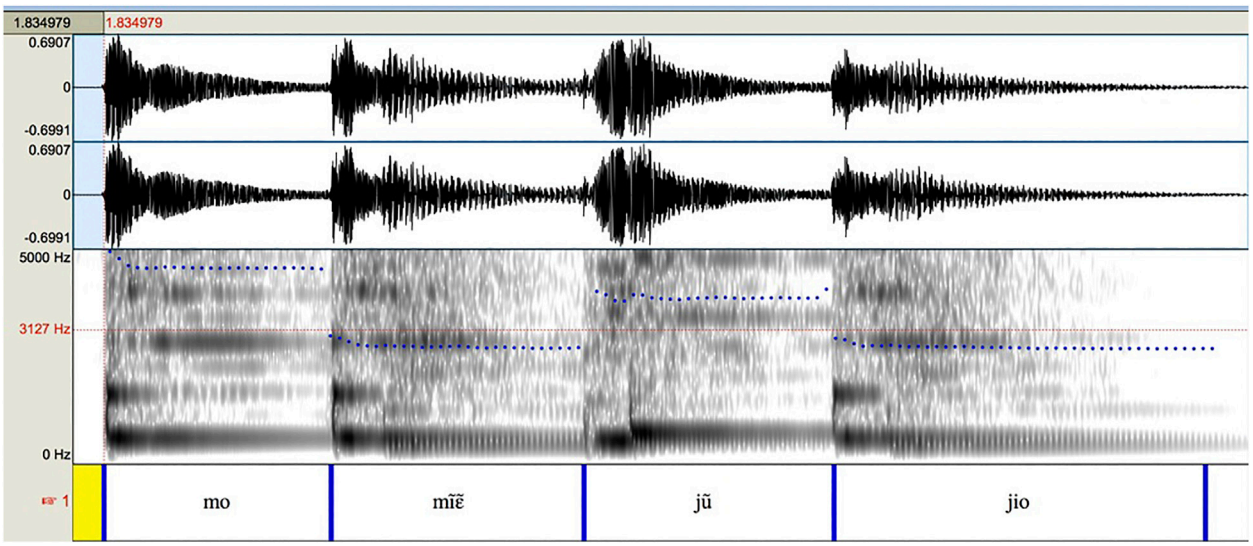

FIGURE 1 | Praat TextGrid of balafon phrase *mó mi̯ɛ̃ jũ̯ ji̯o* "I saw a long snake".

ENCODING STRATEGIES

With this musical and phonological background in place, we can now turn to their confluence in the balafon surrogate language.

The findings in this paper are drawn from a corpus of data collected by the author since 2014. The majority of the data is elicited, though the corpus also includes some naturally occurring phrases from field recordings. All phrases are in the balafon speech mode—the surrogate mode most closely emulating speech—rather than musical or sung mode; for more details, see McPherson (2018) and McPherson and James (forthcoming). In total, the corpus contains 1259 syllables, only slightly higher than the number of words given the largely monosyllabic nature of Seenku vocabulary. Data were recorded with four musicians, all from the same Diabate family of griots from Toronsso, Burkina Faso, though the majority of the data are from a single musician, Mamadou Diabate.

In order to determine the encoding strategies, each syllable in the corpus was coded for a number of binary linguistic features related to syllable structure and tone: Sesquisyllabicity, Diphthong, Long vowel, Coda nasal, Contour tone. In addition, the syllables were coded for the binary musical feature Flam, that is, whether the syllable was played with a flam (two strikes) or a single strike.

The total duration in milliseconds of each syllable played on the balafon was also included. Measurements were made by creating TextGrids in Praat (Boersma and Weenink 2017), with boundaries placed at the attack of the first strike of each word (or the only strike for single-strike syllables). The interbeat duration between the two strikes of a flam is not measured but rather the two are considered as a single unit for the purposes of measurement, since the two strikes together encode aspects of syllable structure as a whole. **Figure 1** shows one such example, where both *mi̯ɛ̃* "snake" and *jũ̯* "long" are played with a flam (visible in the spectrogram) which are treated as single interbeat durations. The final word of the phrase, here *ji̯o* "saw", is marked in the TextGrid but its measurement is not included in the database since it is not followed by another beat and as such, no explicit endpoint exists for measuring its duration. For audio recordings of this and all other examples in the paper, see the **Supplementary Materials**.

Position in the line was also coded by numbering each syllable in a line and dividing that number by the total number of syllables in the line. Some syllables were excluded from the final analysis since their original line membership had been lost and hence their position could not be calculated, and all line-final syllables were removed since their duration could not be measured. This left 1053 syllables for the durational analysis in *Gradience*.

This paper focuses on the encoding of segmental aspects of Seenku rather than tone. Briefly, tone is encoded in the notes of the balafon, with the four notes of the scale excluding the "fetish balafon key" (♭3) corresponding to the four phonemic tone levels; the highest tone, S, is generally pegged to the center of the musical mode and the remaining three tone levels—H, L, X—are played subsequently on each lower note. For more details of tonal encoding, see McPherson (2018).

Consonant and vowel qualities are not encoded in the surrogate language, i.e. there is no way to encode the difference between /a/ and /i/ or between /m/ and /t/. As such, segmental encoding consists of the encoding of the different aspects of syllable structure described in *Syllable and Word Structure*, namely vowel length, diphthongs, sesquisyllabicity, and nasal codas. When we look at the data corpus, we find three patterns of encoding: categorical encoding, variable encoding, and gradient encoding. In the subsections below, I will describe each pattern in turn and the elements of syllable structure for which they are employed.

Categoricity

The conscious rules of balafon segmental encoding involve a single formal distinction: The word can be played with a single

strike, or with a flam (two strikes in quick succession, with the first significantly shorter than the second, like a grace note). This choice is made categorically for long vowels and for sesquisyllabic words, both of which are exceptionlessly played with a flam (resulting in potential ambiguity). For instance, **Figure 2** shows a short phrase that could encode either *bɛ̏ɛ kərɛ̂* 'male pig/boar' or *səgà bâ* 'male sheep/ram' (among other things). Both words are played with a flam; the first involves two strikes on the same note, corresponding to the level-toned long vowel or sesquisyllable, while in the second, the strikes are on two different notes to encode the HX contour tone. In the case of *bâ,* this brings up the important point that if a word has a contour tone, it will always be encoded as a flam, regardless of syllable structure.

It would be considered ungrammatical on the balafon to play the first word of this phrase with just a single strike, and the corpus contains no instances in which a sesquisyllable or long vowel is played with a single strike.

Note that this is a binary opposition: Long vowels result in two strikes, and so do sesquisyllables, but a syllable type with both complexities like CəCV: still falls into the two-strike category. Likewise, the presence of any one complexity (like sesquisyllabicity) will put the syllable into the two-strike category. That is, even though CəCV has a short vowel, it is encoded with two strikes due to its sesquisyllabic nature.

Variability

While long vowels and sesquisyllabicity are exceptionlessly encoded with a flam, other elements of syllable structure are subject to variation in their encoding. Specifically, diphthongs and nasal codas are variably encoded either with one strike or two; I present the data below and suggest reasons why these elements may behave differently from those we saw in the last section.

Diphthongs

As stated in *Syllable and Word Structure*, the syllable nucleus in Seenku may be occupied by either a monophthong or a diphthong. All diphthongs are either rising sonority (higher to lower vowels) or back to front at the same vowel height. The first vocalic (**V**V) element is non-moraic, acting like a glide, leaving the second (V**V**) to carry contrastive vowel length. In addition,

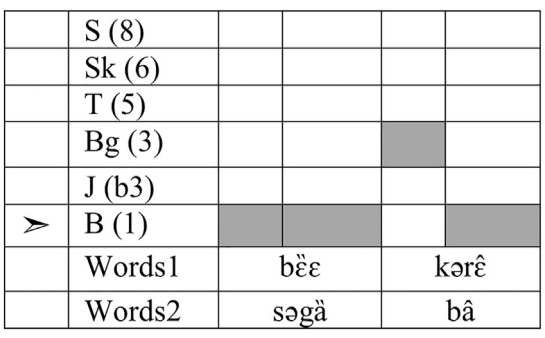

FIGURE 2 | Balafon encoding of "male pig" or "male sheep". Arrow by B(âa-ɲǎ) indicates the center of the mode.

diphthongs may be either oral or nasal, though vowel nasality plays no role in the balafon surrogate language.

In total, the corpus contains just 66 instances of diphthongs. In order to isolate the effect of the diphthong on surrogate encoding, we must restrict our analysis to only level-toned syllables of the shape CVV. Contour tones, long vowels, or sesquisyllables will all result in categorical flam encoding independently of the diphthong. For the time being, I will also exclude diphthongs carrying a nasal coda, which as we will see below also can trigger a flam. This leaves 41 CVV syllables in the corpus. Of these, 9 are played with a flam and 32 with a single strike (22% flam). However, while we have 41 individual tokens, these represent just 18 distinct words, with many words showing multiple repetitions. Out of these 18 unique words, only 3 are responsible for the flams. These three words are shown in (3):

(3) kùa 'farm/cultivate (tr.)'
sòe 'horses'
sùɛ 'three'

None showed free variation; in other words, they were uniformly encoded with flams. Thus, it is possible that what we find is not free variation in diphthong encoding but rather lexical variation, with certain words encoded with a single strike and others with a flam. However, in this data set, multiple repetitions were given by the musician back-to-back, and so there may be a priming effect, such that once a decision is made to play a word with a flam, the same encoding will be used in subsequent repetitions. These phrases would need to be elicited on another day, and also ideally with different musicians to test whether there is any inter-player variability (which I reiterate is unattested for either long vowels or sesquisyllables).

Why should diphthongs be subject to variable encoding when long vowels and sesquisyllables are not? I hypothesize that there is competition between two contrasts that musicians may try to maintain: the contrast between a monophthong and a diphthong, but perhaps more importantly, the contrast between short and long diphthongs. From a vowel length standpoint, CVV is a simplex syllable (while CVV: is complex); but at the same time, CVV is still more complex than CV. Overall, musicians appear to err on the side of treating CVV as simplex. But as the examples in (3) show, this is not always the case.

We can ask whether there is anything special about the words in (3) that may lead them to be played with a flam rather than a single strike. For *kùa* 'farm (tr.)' (underlyingly *kùa,* but with tone raising in its phrasal context), it isn't clear. It is possible that there is interference from the intransitive *kùaa,* which would be played with a flam, though it seems unlikely to me that musicians would confuse this grammatical vowel length contrast. For *sòe* 'horses' and *sùɛ* 'three', though, a functional explanation may be available.

First, *sòe* 'horses' is the plural form of *sɔ̀* 'horse'. In isolation, the two are distinguished both by the diphthong and by tone raising in the plural (McPherson 2017). In the phrasal context elicited on the balafon, however, a process of tone sandhi raises the X-tone of *sɔ̀* 'horse' to the same L-tone as the plural. These two phrases are shown in (4):

(4) a. mó nă sò sa̰n
1SG PROSP horse buy.IRR
'I will buy a horse'
b. mó nă sòe sa̰n
1SG PROSP horse.PL buy.IRR
'I will buy horses'

When I elicited the phrase 'I will buy horses', the consultant first offered the singular phrase in (4a), in which *sò* 'horse' was represented with a single strike, as expected due to its level-toned CV nature. In playing the plural form immediately afterward, the musician chose to encode *sòe* as a flam, likely to disambiguate the two phrases. It is an open question whether a musician would think to disambiguate in this way if the singular and plural were not immediately juxtaposed

A similar explanation may be at play for *sùɛ* 'three' (lexically *sŭɛ*, but once again raised to L tone in the phrasal context). Played on a single strike, *sùɛ* would be confusable with *nɔ̀* 'five' (lexically L-toned), a relatively common numeral played on the balafon thanks to the monetary system, and as such, the musician may have chosen to explicitly encode the diphthong with a flam. Nevertheless, this flam encoding would then make *sùɛ* 'three' confusable with other single digit numerals like *nàa* 'four' (also lexically X-toned *nâa*). My impression is that *nâa* is less common on the balafon and hence would be less salient as a potential confound to a musician. More data are required to test this functional hypothesis, such as the numeral 'three' in a context without tone raising so that *sŭɛ* and *nɔ̀* would be tonally distinct.

Musicians are keenly aware of ambiguity in the signal, which they have reported to me in other instances. Curiously, sometimes they do not offer disambiguation by flam. For instance, one consultant told me that saying *bĭ̃ fŭɛ* 'twenty goats' on the balafon would be unlikely, because it would be confused for *bĭ̃ fĭ̃* 'two goats'; *fŭɛ* 'twenty' was not played with a flam. Then again, it may be unlikely to speak of twenty goats in general (who would offer to slaughter so many?), so less thought was put into how to make the signal clear, whereas 'three' vs. 'five' is a more realistic difference to need to make.

Nasal Codas

The other variable aspect of syllable structure is the nasal coda, which will be transcribed with <n>, though its realization and place of articulation are variable and non-contrastive. In the spoken language, this is a weak phonological element whose realization depends heavily on context (segmental, phrasal) but which is also subject to free variation. In phrase-final position, it is variably realized as late nasalization of the preceding vowel, with or without weak lingual articulation (palatal for front vowels, velar or uvular for back vowels, but not achieving closure), or it goes unrealized. If followed by a sonorant /l/ or /w/, it will be realized on the sonorant, yielding either nasal stops [n] and [m] or nasalized sonorants [l̃] and [w̃]. Before nasals, it goes unrealized. Only before obstruents is it variably realized as a nasal stop. For more details, see McPherson (2020a, 2020b 2020c).

Given the variable nature in the spoken language, it is unsurprising that it should also receive variable treatment in the balafon surrogate language. The corpus contains 109 instances of nasal codas, but once again we must focus only on those cases where another syllable element cannot be responsible for its balafon encoding. Of the 109 nasal codas, 70 are found on level-toned CVN syllables, without the influence of long vowels, contour tones, sesquisyllables, or diphthongs. 7 out of 70, or 10%, are played with flams. Another word, *wĭ̃ɛn* 'money', varies in its spoken pronunciation between *wĭ̃ɛn* and *wĭ̃ɛ(n)*; it is a very common word on the balafon, since asking for money is one of the surrogate language's main functions, and it is uniformly played with a flam. Since it cannot be determined whether this is the result of the nasal coda, vowel length, or both, I exclude it from consideration here.

Unlike in the case of diphthongs, we do see some free variation in the corpus for nasal codas. For instance, the word *dán* '10,000' appears four times in the same recording, three of which are played with flams and one of which is not. Similarly, the word 'sauce', which appears as both *nàn* and *nán* depending on the phrasal context, appears eight times, twice with a flam and otherwise as a single strike. These two flam realizations are found in the phrase in (5), but here too we see free variation, with one repetition of the phrase showing just a single strike for *nàn*:

(5) gɔ̀ɔ́ǹ nàn kɔ̀rɔ̀ sa̰ mó tɛ̃
sorrel sauce desire be 1SG GEN
'I want sorrel sauce'

In contrast, the phrasal variants in (6a) and (6b) never see the word *nán* played with a flam:

(6) a. mó nă à nán lɛ̂ kʊ́ nɔ̃
1SG PROSP DEF sauce DEM D.DEF eat.IRR
'I will eat that (aforementioned) sauce'
b. mó nă à nán lɛ̂ à nɔ̃
1SG PROSP DEF sauce DEM DEF eat.IRR
'I will eat that sauce'

While there are just six tokens of these phrases (three each) and thus we should take any patterns with some degree of caution, the fact that these cases are never played with a flam while those in (5) are may find a linguistic explanation: In (5), the nasal coda on *nàn* appears before an obstruent and is pronounced as a velar nasal [nàŋ kɔ̀rɔ̀], whereas in (6), the nasal appears before a sonorant and is realized entirely on this following segment [ná nɛ̂]. In other words, the surface realization in (6) renders *nán* a simple CV.

It is by no means the case that every coda nasal before an obstruent will be realized with a flam—in fact, the majority are not; out of nine repetitions of the phrase *à sa̰ sã́n kã́* 's/he is in God's hands', none were produced with a flam. But at the same time, all seven instances of nasal coda flams are found before an obstruent, suggesting that surface realization does have a role to play.

Thinking about why nasal codas should be variably rather than categorically encoded, the first explanation relates to its weak representation. Given its variable pronunciation even in the spoken language, it may be that nasal codas are not fully activated phonological representations, in the sense of

Gradient Symbolic Representation (Rosen 2016; Smolensky and Goldrick 2016; Zimmermann 2019). As such, when musicians come to represent speech on the balafon, they may not be activated enough to be consciously registered as contributing to syllabic complexity. Another possible explanation is that in the case of nasal codas, we have a mismatch between the timing of syllable complexity (after the nucleus) and the timing of surrogate syllable complexity (at the beginning of the word, with a flam). This stands in contrast to diphthongs and sesquisyllables, both of which find their "weaker" or "faster" element at the left edge. A larger corpus will reveal whether the higher rates of flam encoding in contexts where the nasal coda is realized as a true coda are statistically significant.

We could also ask whether there are any musical or phrasal effects on variation, such as the length of a line or the presence of other flams in the vicinity of the target syllable. For instance, we might predict that in longer phrases, musicians may counter this complexity by using fewer flams (cf. Fenk-Oczlon et al., 2009). As far as the current data show, there appears to be no effect of phrase length; both flam and single strike realizations are found in phrases of varying lengths and in both the first and second half of the phrase. However, as with the linguistic explanations above, the current dataset is too small to draw any firm conclusions. With more data, we could also test whether the presence of other flams in the immediate vicinity either encourages or inhibits a flam encoding. I leave these questions to future research.

Before leaving the topic of variable encoding, I want to briefly point out that it may be the case that words that contain two variably encoded elements—that is, a diphthong followed by a nasal coda—are more likely to be encoded as a flam. The dataset only contains one such word, *sôen* 'one', but in 8 out of 9 tokens, it is encoded with two strikes. If we think about variable flam encoding as a probability, then it makes sense that two factors that result in such an encoding would increase its odds. Since we are dealing with just a single type, though, itself a high frequency word in the natural surrogate language, we should not draw any strong conclusions about the interaction of these two probabilities. Like the word *wɛ̃n* 'money' discussed above, musicians may have a lexically specific representation of this word as a flam regardless of the productive rules of surrogate encoding. Further cases of level-toned CVVN words must be sought to test the rest of flam encoding with these two variable elements of syllable structure.

Gradience

Thus far, all encoding we have discussed has been what could be considered conscious rules of the balafon surrogate language, a reduction of contrastive elements of syllable structure to a binary choice between a single strike and a flam on the balafon. Musicians even report rampant neutralization due to these rules of encoding, such as the example shown in **Figure 2**.

Nevertheless, while the notes of the instrument are fixed, and the number of strikes (single vs. flam) is likewise a categorical feature, duration is gradient. Specifically, the timing between strikes of the balafon (interbeat durations, Seifart et al., 2018) is not isochronous and varies depending upon the word that is encoded. These gradient distinctions appear to be subconscious, since at the level of conscious encoding musicians report neutralizations where duration measurements suggest a distinction.

As noted at the beginning of **Section 4**, each syllable in the corpus was coded for a variety of binary factors, in addition to its phrasal position and its duration in milliseconds. To investigate the influences on gradient duration, a linear mixed effects regression analysis was carried out in R using the lme4 package (Bates et al., 2015), with duration in milliseconds as the dependent variable. Independent variables included the position of the word in a phrase, whether the nucleus of the syllable was a diphthong or a monophthong, whether the nucleus vowel was phonemically short or long, whether the syllable was sesquisyllabic or not, whether the syllable had a coda or not, whether the syllable's tone was a contour or level, and finally whether the syllable was played on the balafon with a flam or a single strike. The interaction between phonemic vowel length and the presence of a coda was also included in the model. Random intercepts for player and word were included. The package lmerTest (Kuznetsova et al., 2017) was used to test the statistical significance of the results.

First, there is a significant interaction between phonemic vowel length and codas ($\beta = -52.3 \pm 21.4$, $t(930) = -2.4$, $p = 0.02$). When the syllable has no coda, the difference in duration between long and short vowels is 77 ms (Long: 302±54 ms, Short: 225±69 ms). However, when the syllable does have a coda, the difference is much smaller, at only 12 ms (Long: 284±38 ms, Short: 272±77 ms). This result captures the fact that segments tend to be compressed and shorter in more complex (and especially closed) syllables (e.g. Fowler 1983; Maddieson 1985; Clements and Hertz 1996). As main effects, both Vowel length and Coda were significant ($\beta = 25.0 \pm 7.9$, $t(935) = 3.1$, $p < 0.01$ for Vowel length and $\beta = 39.5 \pm 6.9$, $t(940) = 5.7$, $p < 0.001$ for Coda).

The model found a main effect of position ($\beta = 100.2 \pm 11.3$, $t(430) = 8.8$, $p < 0.001$); syllables closer to the end of the phrase have longer durations. There is also a main effect of diphthong ($\beta = 17.9 \pm 7.8$, $t(937) = 2.3$, $p = 0.02$); syllables with a diphthong nucleus are significantly longer than those with a monophthong. Sesquisyllabicity was also significant ($\beta = 17.1 \pm 8.6$, $t(939) = 2.0$, $p = 0.05$); sesquisyllabic syllables had significantly longer durations than simple C(V)V syllables. Finally, syllables played with a flam were significantly longer than those played with a single strike ($\beta = 52.2 \pm 8.3$, $t(939) = 6.3$, $p < 0.001$). The factor of contour tone was not significant ($p = 0.38$).

Figure 3 plots duration of syllables in the surrogate language by position. This gradient effect differs from the others (vowel length, sesquisyllabicity, etc.) in that it does not arise from a phonemic contrast in the underlying words being encoded. Instead, we may attribute this phrase-final lengthening to one (or both) of two sources: 1) the phonetic realization of spoken phrases, which would be subject to cross-linguistic patterns of phrase-final lengthening (Klatt 1975; Scott 1982, *i.a.*), or 2) the same effect of phrase-final lengthening attested in musical performance and perception (Todd 1985; Palmer 1989, *i.a.*). In other words, gradient durational differences in the balafon surrogate language are controlled by more than just contrastive categories; we may also see the effects of low-level phonetic differences, lying beneath the level of a musician's

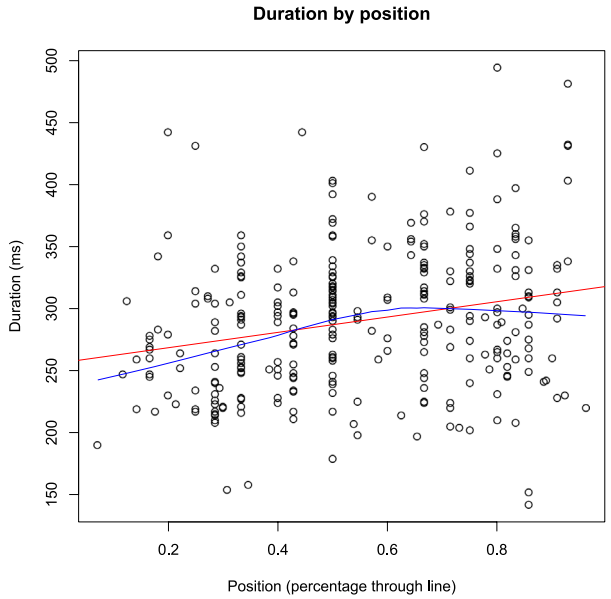

FIGURE 3 | Duration plotted by phrasal position (syllable #/total number of syllables in the line).

consciousness. I return to this point in the next section and offer some thoughts on how to conceptualize these different influences on musical surrogate speech.

DISCUSSION

In this paper, I have shown that encoding of syllable structure in the Sambla balafon surrogate language is more complex than first meets the eye. In terms of conscious encoding, i.e. the explicit grammatical rules of the surrogate language, there is just a binary contrast of a single strike vs. a flam. The flam can be employed categorically—for contour tones (not discussed here), vowel length, or sesquisyllabicity—or it can be employed variably, with diphthongs and coda nasals vacillating between a simple and complex syllabic treatment. Beneath these conscious encodings, we also find a range of subconscious, gradient durational effects, ranging from gradient encoding of phonemic contrasts like vowel length to phonetic or musical phrasing effects. Musicians appear to be unaware of these durational contrasts, at least consciously. Future work could investigate whether they play any role in comprehension, e.g. in a forced choice task asking musicians to choose between two interpretations of a supposedly ambiguous balafon phrase.

How can we account for these different levels of surrogate speech encoding? I suggest that the encoding of surrogate speech passes through a "phonetic filter", formalized in the schema shown in **Figure 4**.[1]

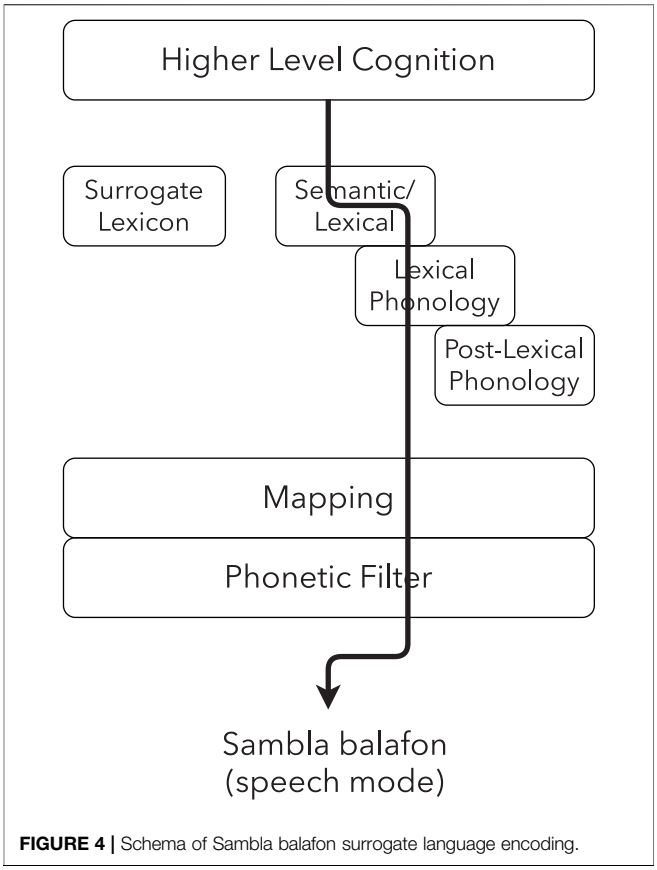

FIGURE 4 | Schema of Sambla balafon surrogate language encoding.

The idea of what a musician wants to play first comes from "higher level cognition", and its linguistic content is selected from the "semantic/lexical" component.[2] These words are then passed through the lexical phonology, where they receive their phonological form. In the Sambla balafon surrogate language, it is the underlying lexical and morphological level of tone that is encoded, not surface tone resulting from post-lexical tone rules (McPherson 2018), and so the encoding does not pass through post-lexical phonology. However, given the possible influence of the surface realization of nasal codas (for instance), the model may require more nuance, separating tonal from segmental phonology; in other surrogate traditions where the post-lexical output is uniformly encoded, the arrow would pass through that component. The output of the phonological component(s) is then passed through the "mapping" component, containing the conscious rules of surrogate encoding. It is here that the notes of the balafon and the choice between a single strike and a flam are selected by the musician. Finally, on the way to being played, the surrogate encoding passes through the phonetic filter where subconscious phonetic elements of speech can influence the surrogate realization, depending upon instrumental constraints. On a variably pitched instrument, this could be

[1]Schema credit: Lucas James.

[2]In a lexical ideogram (Stern 1957) surrogate system, with arbitrary musical signs standing in for words or phrases, there would be a separate step of going to the "surrogate lexicon".

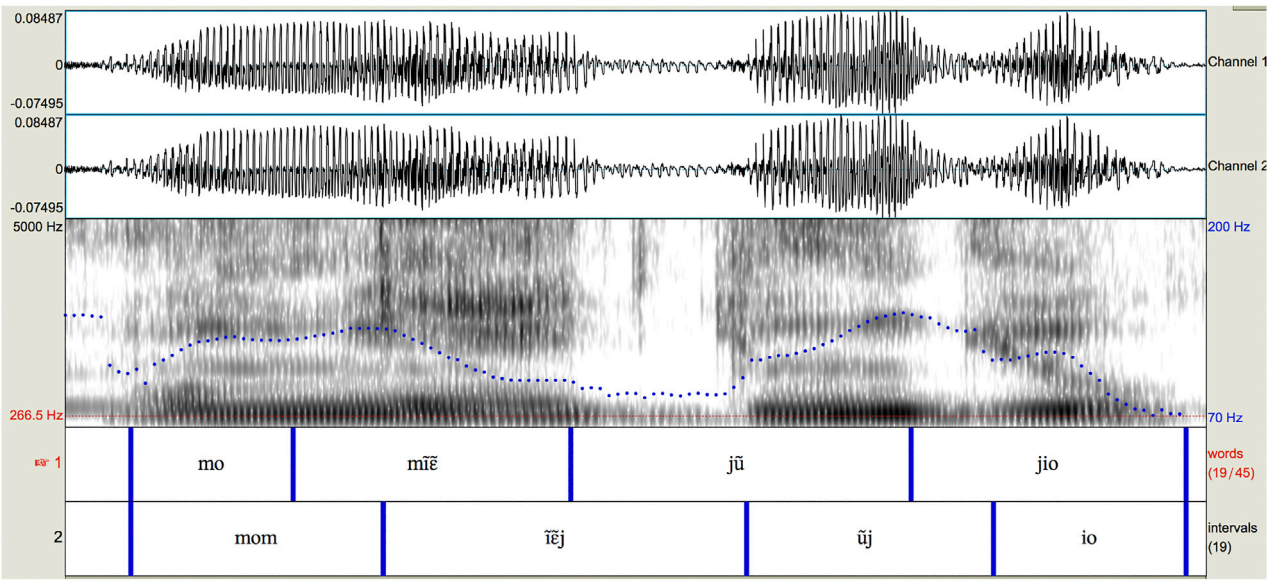

FIGURE 5 | Syllable **(top tier)** and interval **(bottom tier)** timing of spoken phrase *mó mi̯ɛ̃ ju̯ ji̯o* "I saw a long snake".

subtle phonetic aspects of tonal realization (see Akinbo 2019 for one such possibility on the Yorùbá *dùndún*). In the case of the balafon, the influence can be seen in micro-timings.

I propose that these phonetic effects exert influence over surrogate speech as musicians think about the words as they play, which controls the timing of their strokes. This suggests that the timing of surrogate speech should largely mirror the timing of spoken language, a finding reported for the Bora *manguaré* slit log drumming tradition (Seifart et al., 2018). In the case of Bora, the authors determined that beat timing tracked vowel-to-vowel intervals rather than syllables.

To test this explanation behind the phonetic filter for the Sambla balafon language, I ran a pilot study comparing the relative timing of words in natural speech and surrogate speech, spoken and played by the same individual. For natural speech, measurements were taken for both syllables and vowel-to-vowel intervals, as shown in **Figure 5**, the spoken phrase corresponding to the balafon phrase in **Figure 1**.

Final syllables were excluded due to the inability to measure an interbeat duration for their balafon correspondents. To time-normalize the balafon and spoken phrases, the total duration of all beats/syllables, excluding the final, was calculated, then each individual duration measure was divided by this total to indicate its individual percentage of the overall phrase. 125 words/syllables were compared, each with three corresponding measures: 1. Interbeat balafon duration; 2. Syllable duration; 3. Vowel-to-vowel interval duration. The deviation between the interbeat duration and each of the syllable and interval durations was calculated to determine which measurement provided a closer match to the balafon timing. The average absolute deviation for syllables was over twice as high as that for intervals, 0.055 (5.5% off) vs. 0.0268 (2.68% off). Visual inspection of plotted phrases,

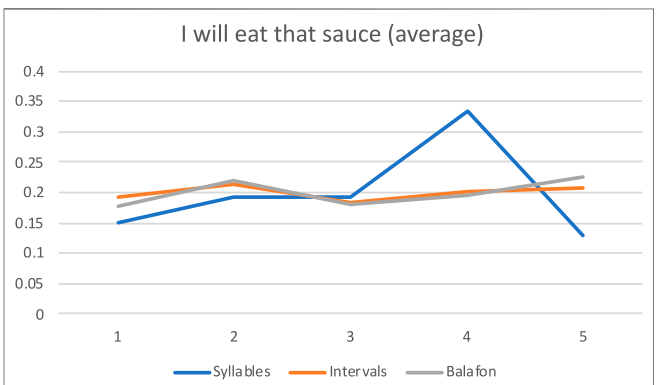

FIGURE 6 | Relative timing of first five words of *mó nă kú nän bɛ́ nɔ̃* "I will eat that sauce", averaged across three spoken and played iterations.

such as that shown in **Figure 6**, corroborates these results, with the interval data much more closely tracking the relative balafon interbeat durations.

In all likelihood, vowel-to-vowel intervals are simply an approximation of the perceptual center, or p-center, of the syllable, which is aligned closely with the left edge of the rime but which may be pulled leftwards into the onset by consonant clusters or otherwise long onsets (Morton et al., 1976; Ryan 2014). We see a suggestion of this effect in the pilot data: Sesquisyllables and the consonant /f/ are both longer onset events in Seenku, and all of their instances in the data show an interval duration shorter than the balafon duration; if it were the p-center that controlled the timing, the left edge of the interval would be anticipated, rendering the whole

measurement longer and bringing it closer to that of the balafon. Seifart et al. (2018) make a similar nod to the p-center but do not explore the hypothesis further.

What is interesting about the phonetic filter on the Sambla balafon is that it comes solely from mental speech—the musician is not speaking while playing, and yet precise relative timings of words in spoken language appear to control the timing of instrumental rhythm. This suggests a fair amount of phonetic detail in either the lexicon, the inner voice, or both. Even though the oral articulators are not being used, their influence is felt on this musical modality.

Most studies of the p-center focus on speaking syllables aloud to an isochronous beat, in other words, a beat controlling the timing of spoken syllables. The study of musical surrogate languages like the Sambla balafon offer the opportunity to flip the script and study the effect of natural non-isochronous syllable timing on a beat. In future work, it would be interesting to compare interbeat duration on the balafon when the musician is speaking aloud vs. when they are simply thinking the words in their head. This could help us tease apart exactly how much phonetic detail is present in the inner voice.

CONCLUSION

At first glance, the Sambla balafon surrogate language is like most African surrogate languages, based largely on tone. The rich syllable structure of the spoken language, Seenku, is ostensibly collapsed to a binary distinction between "simple" and "complex". But a deeper look reveals that musicians are sensitive to a host of phonological and phonetic factors that inform their playing. Long vowels and sesquisyllables are both invariably treated as complex syllables, while diphthongs and nasal codas vary in their encoding. The variation seen with diphthongs seems to reflect to the tension between the complex nucleus and the pressure to maintain a short vs. long contrast even within diphthongs. Variation for nasal codas, on the other hand, may relate to their weak phonological representation and variation in surface form in the spoken language. At a subconscious level, the mental timing of speech influences the timing of words on the balafon, suggesting gradient distinctions between words that musicians report to be neutralized.

These details of encoding demonstrate the various ways in which musical surrogate languages act as a unique window onto phonological and phonetic structure of the spoken language. Differing treatment of phonemic contrasts can provide a source of evidence for differing phonological representation or for competition in the phonological component. The presence of phonetic detail in the surrogate signal raises questions about how it gets there: Could this provide evidence for an exemplar model of phonology, with mental storage of detailed tokens of speech (Pierrehumbert 2002; Johnson 2006)? Does thinking of the phrases activate the motor neurons for articulation (Tian and Poeppel 2010; Pickering and Garrod 2013), without reaching the threshold of producing the gestures but still replicating the timing patterns of external speech?

Further detailed studies of musical surrogate languages will allow us to better understand the deep connections between the human language faculty and these unique modalities of communication.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Center for the Protection of Human Subjects, Dartmouth College. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This material is based on work supported by the National Science Foundation-National Endowment for the Humanities Documenting Endangered Languages Program under National Science Foundation Award No. BCS-1664335.

ACKNOWLEDGMENTS

I would first like thank Mamadou Diabate and the other members of the Diabate family for sharing their tradition and expertise with me. I would like to thank Kevin Ryan and Lucas James for helpful conversations in the preparation and revision of this work. Finally, I am deeply grateful to Rolando Coto Solano for his help with statistical analysis. Any remaining errors are my own.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.652635/full#supplementary-material

REFERENCES

Akinbo, S. (2019). Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis. *Linguistique et Langues Africaines* 1, 11.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *J. Stat. Soft.* 67, 1. doi:10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (2017). Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.28, Available at: http://www.praat.org/ (Accessed January 2, 2017).

Clements, G. N., and Hertz, S. R. (1996). An Integrated Model of Phonetic Representation in Grammar. *Working Pap. Cornell Phonetics Lab.* 11, 34–116.

Corley, M., Brocklehurst, P. H., and Moat, H. S. (2011). Error Biases in Inner and Overt Speech: Evidence from Tongue Twisters. *J. Exp. Psychol. Learn. Mem. Cogn.* 37 (1), 162–175. doi:10.1037/a0021321

Fenk-Oczlon, G., Fenk, A., and August, F. (2009). Some Parallels between Language and Music from a Cognitive and Evolutionary Perspective. *Musicae Scientiae*. 13 (2 Suppl. l), 201–226. doi:10.1177/1029864909013002101

Fowler, C. A. (1983). Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet. *J. Exp. Psychol. Gen.* 112 (3), 386–412. doi:10.1037/0096-3445.112.3.386

Johnson, K. (2006). Resonance in an Exemplar-Based Lexicon: The Emergence of Social Identity and Phonology. *J. Phonetics* 34 (4), 485–499. doi:10.1016/j.wocn.2005.08.004

Klatt, D. H. (1975). Vowel Lengthening Is Syntactically Determined in a Connected Discourse. *J. Phonetics* 3 (3), 129–140. doi:10.1016/s0095-4470(19)31360-9

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* 82, 1–26. doi:10.18637/jss.v082.i13

Maddieson, I. (1985). Phonetic Cues to Syllabification. in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Editor Fromkin (New York, NY: Academic Press), 203–221. doi:10.18637/jss.v082.i13

Martin, S., Iturrate, I., Millán, J. d. R., Knight, R. T., and Pasley, B. N. (2018). Decoding Inner Speech Using Electrocorticography: Progress and Challenges toward a Speech Prosthesis. *Front. Neurosci.* 12. 422. doi:10.3389/fnins.2018.00422

Matisoff, J. (1990). Bulging Monosyllables: Areal Tendencies in Southeast Asian Diachrony. *Proc. 16th Annu. Meet. Berkeley Linguistics Soc.*, 543–559. doi:10.3765/bls.v16i0.1680

McPherson, L. (2017). Multiple Feature Affixation in Seenku Plural Formation. *Morphology* 27 (2), 217–252.

McPherson, L. (2018). The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. *Anthropological Linguistics* 60 (3), 255–294.

McPherson, L. (2020a). A Grammar of Seenku. *MGL 83*. Berlin/Boston: De Gruyter Mouton

McPherson, L. (2020b). Illustration of Seenku. *J. Int. Phonetic Assoc.* 50 (2), 220–239.

McPherson, L. (2020c). On Latent Nasals in Samogo. *Mandenkan* 63, 68–88.

McPherson, L., and James, L. (2020). "Artistic Adaptation of Seenku Tone: Musical Surrogates vs. Vocal Music,"in. Selected Proceedings of the 50th Annual Conference on African Linguistics. Editor Akinlabi (Berlin, Germany: Language Science Press).

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual Centers (P-Centers). *Psychol. Rev.* 83, 405–408.

Oppenheim, G. M. (2013). Inner Speech as a Forward Model. *Behav. Brain Sci.* 36 (4), 369–370. doi:10.1017/S0140525X12002798

Palmer, C. (1989). Mapping Musical Thought to Musical Performance. *J. Exp. Psychol. Hum. perception Perform.* 15 (2), 331. doi:10.1037//0096-1523.15.2.331

Pickering, M. J., and Garrod., S. (2013). An Integrated Theory of Language Production and Comprehension. *Behav. Brain Sci.* 36, 329–392. doi:10.1017/S0140525X12001495

Pierrehumbert, J. (2002). *Word-specific Phonetics*. Laboratory Phonology 7

Pittayaporn, P. (2015). "Typologizing Sesquisyllabicity," in *Languages of Mainland Southeast Asia: The State of the Art*. Editors Enfield and Comrie (Berlin/Boston: De Gruyter Mouton), 500–528.

Rosen, E. (2016). "Predicting the Unpredictable: Capturing the Apparent Semi-regularity of Rendaku Voicing in Japanese through Harmonic Grammar," in. Proceedings of BLS. Editors Clem (Berkeley, CA: Berkeley Linguistic Society), 42, 235–249.

Ryan, K. (2014). Onsets Contribute to Syllable Weight: Statistical Evidence from Stress and Meter. *Language* 90 (2), 309–341.

Scott, D. (1982). Duration as a Cue to the Perception of a Phrase Boundary. *J. ASA* 71, 996. doi:10.1121/1.387581

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing Language to Rhythm: Amazonian Bora Drummed Language Exploits Speech Rhythm for Long-Distance Communication. *R. Soc. Open Sci.* 5, 170–354. doi:10.1098/rsos.170354

Smolensky, P., and Goldrick, M. (2016). *Gradient Symbolic Representations in Grammar: The Case of French Liaison, Ms.* Johns Hopkins Univeristy and Northwestern University. Available at: http://roa.rutgers.edu/content/article/files/1552_smolensky_1.pdf (Accessed September 15, 2019).

Stern, T. (1957). Drum and Whistle "Languages": an Analysis of Speech Surrogates. *Am. Anthropologist* 59, 487–506.

Strand, J. (2009). The Sambla Xylophone: Tradition And Identity in Burkina Faso. PhD Dissertation. Middletown (CN): Wesleyan University.

Tian, X., and Poeppel, D. (2010). Mental Imagery of Speech and Movement Implicates the Dynamics of Internal Forward Models. *Front. Psychol.* 1, 166. doi:10.3389/fpsyg.2010.00166

Todd, N. (1985). A Model of Expressive Timing in Tonal Music. *Music Perception*. 3 (1), 33–57.

Zemp, H., and Soro, S. (2010). Talking Balafons. *Afr. Music.* 8 (4), 7–24.

Zimmermann, E. (2019). Gradient Symbolic Representations and the Typology of Ghost Segments. Proceedings of the Annual Meeting on Phonology, Vol. 7.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.