No cover image available

The Oxford Handbook of Language and Music

(In Progress)

Daniela Sammler (ed.)

https://doi.org/10.1093/oxfordhb/9780192894700.001.0001

Published: 20 March 2025 - **Online ISBN:** 9780191915581 **Print ISBN:** 9780192894700

Search in this book

CHAPTER

Auxiliary speech by vocal tract modulation and musical surrogacy

Julien Meyer, Victor Manfredi

https://doi.org/10.1093/oxfordhb/9780192894700.013.0021

Published: 19 November 2025

Abstract

In situations of telecommunication and proximal artistic performance, human groups have engineered diverse, ingenious formats of non-voiced auxiliary speech: (i) modulating the vocal tract to enhance selected acoustic features of a sound source alternative to the vocal cords; or (ii) adapting musical instruments to simulate aspects of the spoken phonetic signal. Both types of 'surrogate language' (Nketia 1971) employ techniques of 'speech abridgment' (Stern 1957), and coin novel 'acoustic icons' (Sebeok and Umiker-Sebeok 1976). Auxiliary speech phenomena attest non-literate traditions of text transmission and implicit mental capacities of natural language processing. This chapter provides a comparative view of these traditional special speech types worldwide. It shows that their various expressions depend on (a) the acoustic nature of the modified speech form, (b) the grammar of the base language, (c) the expertise of the performers, and (d) but also whether they imitate spoken dialogues, declamative verbal arts, or musical sung forms of speech. Case studies explore more details in the second part of the chapter.

Keywords: surrogate speech, acoustic iconicity, auxiliary speech, instrumental speech, whistled language, drummed language, talking drum, musical surrogacy, whistled speech, music-language relation

Subject: Phonetics and Phonology, Psycholinguistics, Linguistics

Series: Oxford Handbooks

Collection: Oxford Handbooks Online

1 Introduction

Human groups of all inhabited continents of the planet have developed diverse traditional unvoiced auxiliary speech modalities, whether by modulating the vocal tract or by substituting a musical instrument. These special speech forms represent a way to adapt human spoken and sung speech modes for long-distance communication and public announcing, or for verbal art. They still express a local spoken language, but in a modified acoustic form. Such auxiliary speech modes occur with languages of diverse phonetic and morphological types. They exploit the adaptability and flexibility of human languages by transforming some selected linguistic features of speech into sounds that do not use the voice as a source[1] but that have notable advantages for propagation in natural environments (telecommunication) or for the production of esthetical/musical effects (verbal art). These acoustic signals are as diverse as whistles (e.g. 'whistled speech'), drumbeats (e.g. 'talking drums'), or musical notes and timbres produced with various types of traditional musical instruments. They can be observed not only in 'endangered' settings, where rural lifeways still resist the pressures of encroaching industrial economy, but also in urban mass movements of ethnic cultural revitalization.

A small sample shows that the phenomenon is quasi-universal:

> *West and central Africa*: A large panel of musical instruments + whistled forms (Carrington 1949; Kawada 2014; Labouret 1923; Locke and Agbeli 1981; McPherson 2018; Nketia 1963; Nketia 1971; Pepper 1950 1954; Sebeok and Umiker-Sebeok 1976; Winter 2014; Zemp and Soro 2004).
>
> East and southern Africa: Some types of musical instruments + whistled forms (Dargie 2007 2011; Fournel 2002; Wedekind 1983).
>
> *Europe and North Africa*: Whistled forms (Busnel 1967 1970; Busnel et al. 1962; Cañadas and Jiménez 1996; Charalambakis 1994; Ridouane et al. 2018)
>
> *Central and North America*: Whistled forms (Busnel 1974; Cowan 1948 1972; Meyer 2015; Ritzenthaler and Peterson 1954; Sicoli 2016; Voorhis 1971).
>
> *South American Amazon*: Several types of musical instruments + whistled forms (Beaudet 1997; Ermel 2004; Everett 1983; Franchetto and Montagnani 2011; Hill 1993; Izikowitz 1935; Mindlin et al. 2001; Moore and Meyer 2014; Thiesen 1969; van der Voort 2016; Whiffen 1915).
>
> *Southeast Asia*: A large panel of musical instruments + whistled forms (Busnel et al. 1989; Catlin 1982; Meyer 2007; Poss 2005 2012; Stern 1957).
>
> *Oceania*: Several types of musical instruments + whistled forms (Niles 2010; Persoon and Schefold 2017; Schefold 1973; Snyders and Haudricourt 1968).

Figure 1 Composite photo of whistler/flute/aerophone/chordophone/log 'drum'/'talking drum', different techniques used to transform speech in auxiliary speech forms. Upper row, left: whistling in El Hierro Island whistled Spanish (Canary Islands, see also Figures 3 and 6). Upper row, right: flute used among the Gavião of Rondônia (Brazil) to emulate song lyrics (see also Figures 2 and 10). Middle row, left: mouth arch played by Gavião women of Rondônia (Brazilian Amazon) to utter courtship songs. Middle row, right: ensemble of three bamboo 'clarinets' used to sing in Gavião language (see also Figure 9). Lower row, left: double hollow log drums played by the Bora/Miraña people of West Amazon to send spoken messages to distant villages or villagers (Peru/Colombia) (see also Figure 7). Lower row, right: Yorùbá *dùn-dún* 'talking drum' used to recite proverbs and pass messages to a large audience in Nigeria.

Auxiliary speech can emulate spoken/sung fundamental frequency and/or acoustic resonance, as well as speech-specific rhythmic patterns to encode phonetic/phonological properties, morphosyntax, and conversational turn-taking (Busnel and Classe 1976; Dentel and Meyer 2020; James 2021; Meyer and Moore 2021; Nketia 1971; Ridouane et al. 2018; Seifart et al. 2018; Sicoli 2016; Winter 2014). One notable common characteristic of auxiliary signals is that they are not as polyvalent and malleable as the spoken voice modulated by the vocal tract. The nature of the linguistic features emulated in this way and the degree of precision with which they are rendered by each auxiliary practice depend on various criteria. These include the type of signal used to modify/emulate speech, their production mode, the grammar of the base language, and the expertise of the performers, but also whether they imitate spoken or sung forms of speech, or whether the auxiliary speech type serves for dialogues, public announcements, or verbal art performances.

As a result, some speech surrogates diverge from the surface realization of a base utterance, and instead encode more abstract features of the phonology of the language (as described for example in Ewondo Drumming (Neeley 1999) or Sambla Balafon (McPherson 2018)), whereas others rather emulate some aspects of the surface form (such as found in oral whistled speech traditions (Meyer 2021) or in Gavião various musical instruments;

see Section 3.2.3). The following sections of this chapter propose a general overview of the different acoustic and perceptual strategies guiding the speech modifications at play in the vocal-to-auxiliary transformations characterizing these special speech forms (Section 2); and detailed case studies from languages of Africa, Europe, and Latin America (Section 3).

2 Modalities, strategies, and goals of auxiliary speech

In principle, any sound-producing technique alternative to vocal cords can be used for the purpose of the auxiliary speech forms reviewed here (see examples in Figure 1). Articulation techniques bifurcate as follows:

> *Vocal tract modulation*. A primary acoustic source substituting the vocal cords resonates in the vocal tract of the speaker/player who modulates with gestures approaching articulation of spoken words and sentences. The source might be whistling compressed airstreams through the lips, teeth, and hands or various instrumental acoustic generators placed between the lips or inside the mouth: a leaf, a 'jew's harp' lamellophone, the strings of a hunting bow or a tube zither, or even insect wings (Arom 1970; Bonhomme 2014; Busnel 1974; Busnel et al. 1989; Cowan 1972; Meyer 2007; Meyer 2015; Meyer and Moore 2021; Moore and Meyer 2014; Nikolsky 2020; Niles 2010; Yegnan-Touré 2008).
>
> *Full musical surrogacy.* A non-somatic sound source is co-opted both as acoustic generator and modulator. In an enabling context, virtually any musical instrument can 'talk' this way—not just a two-head hourglass membranophone with variable pitch (so-called 'talking drum') but also a *bala* xylophone, log idiophone, bell or gong, horn or wooden whistle, flute, mouth organ or bamboo clarinet (Akínbọ̀ 2019 2021; Catlin 1982; Eboué 1935; Hill 1993; McPherson 2018; Meyer and Moore 2021; N'Guessan 2020; Niangoran-Bouah 1981; Pacere 1987; Pepper 1950 1954; Ritzenthaler and Peterson 1954; Zemp 2005; Zemp and Soro 2004), or a chordophone like the Ethiopian *kar* guitar or the Akan *serepewa* lute (Nketia 1994; Wedekind 1983).

Inherent phonetic redundancies of spoken modal speech permit some features of natural language to be elided and others enhanced or simulated depending on the goal: either to boost signal strength for spontaneous communication in constrained acoustic environments or to heighten the aesthetics of performance by retrieving fixed (proverbial, lyrical) texts from collective memory. Distant dialogue between hunters, shepherds, or hill agricultors is most often achieved by whistling, while drums are the most widely used instruments of one-way public broadcast.

The importance of perceived *pitch* (the perceptual attribute of F0, 'fundamental frequency') in all these functions is unsurprising, given its central role in whistles and in most musical instruments, as well as its key role in the evolution of vocal learning by many species including our own (Fitch 2010; Ohala 1984); but another possible target of emulation is vocal *timbre*—the distinct *Gestalten* of vowel and consonant quality (and perceptual attribute of the whole frequency spectrum). Consequently, both segmental and suprasegmental aspects of spoken speech are generally transposed, but in different proportions, depending on various factors. Our current knowledge of systems emulating speech with musical instruments—and thus systems that are not based on orally modulated somatic whistles—suggests that they overwhelmingly develop in 'tonal' languages (Bagemihl 1988; Sebeok and Umiker-Sebeok 1976; Seifart et al. 2018), although exceptions exist. Indeed, a few drummed practices concern 'non-tonal' languages (e.g. Schefold 1973; Snyders and Haudricourt 1968; Winter 2014; Wojtylak 2019), whereas orally modulated jew's harp practices are rather evenly distributed between 'tonal' and 'non-tonal' languages (Nikolsky 2020) (just like what happens for speech transposition in whistles, Meyer 2015).

Pitch and timbre are controlled rather independently through their acoustic correlates in the human vocal tract for spoken speech production, and they are separable in speech perception (Fant 1960; Risset 1968; Stevens

1998). Few musical instruments manipulate pitch and timbre autonomously, and most of them use only one of these perceptual frequency levels as a target for speech emulation. For example, whistles and flutes are based exclusively on pitches to encode speech information (see Figures 2 and 3), whereas several types of drums, such as Bendré skin drums, are based on unpitched sounds (Kawada 2014, and see Figure 4).

The choice of pitch vs timbre as primary target in the vocal-to-auxiliary transformation is thus mediated by: (1) acoustic limitations of the auxiliary production mode (as just mentioned); but also (2) the role of intonation in the language (for instance, 'tonal' languages clearly deploy pitch for lexical meaning more than 'non-tonal' ones); and (3) the type of speech that is emulated (sung vs spoken) because song lyrics emphasize particularly pitch through the musical melodies. Finally, this choice yields four possible configurations:

> *Vocal timbre ⇒ auxiliary timbre.* For instance, the vowels and consonants of a 'non-tonal' language are signalled with a jew's harp while preserving the same oral articulatory gestures that shape the resonance spectra of primary speech (Nikolsky 2020). Vocal timbre can also be imitated externally to the body with hand and/or stick attacks on a drum like the Wolof single-head *sabar* (Ros 2021), the Duala slit-log 'drum' (Schneider 1976, see Figure 5), and the Yorùbá lead bàtá drum (Section 3.2.2).
>
> *Vocal pitch ⇒ auxiliary timbre.* Spoken pitch is emulated by varying the timbre of the auxiliary technique. Here, the spectral distribution of the auxiliary timbre mimics the changes of spoken pitch. For example, this occurs when spoken F0 contours in 'tonal' languages are imitated with a jew's harp by modifying the oral mouth shape (Nikolsky 2020; Poss 2012), or else evoked external to the body by contrasting drum attacks as in Gùn-gbè and Mooré traditions (Rouget 1964, Section 3.2.1; see also Dentel and Meyer 2020; Kawada 2014, and an example in Figure 4).
>
> *Vocal timbre ⇒ auxiliary pitch.* For example, when spoken vowel qualities, controlled by oral mouth shape, are transposed in whistled forms of 'non-tonal' languages (Meyer 2015; Meyer 2021, and see explained examples in Figures 3 and 6).
>
> *Vocal pitch ⇒ auxiliary pitch*. This is the most widespread strategy that consists in a direct prosodic correspondence. It is found when pitches produced by musical instruments emulate speech prosody encoded in spoken pitch (e.g. Hill 1993; Meyer and Moore 2021; Nketia 1971; Seifart et al. 2018; van der Voort 2016; and see Figure 2, lower row), but also when F0 contours of a 'tonal' language are imitated by whistling (Cowan 1948; Figure 2, middle row; Moore and Meyer 2014).

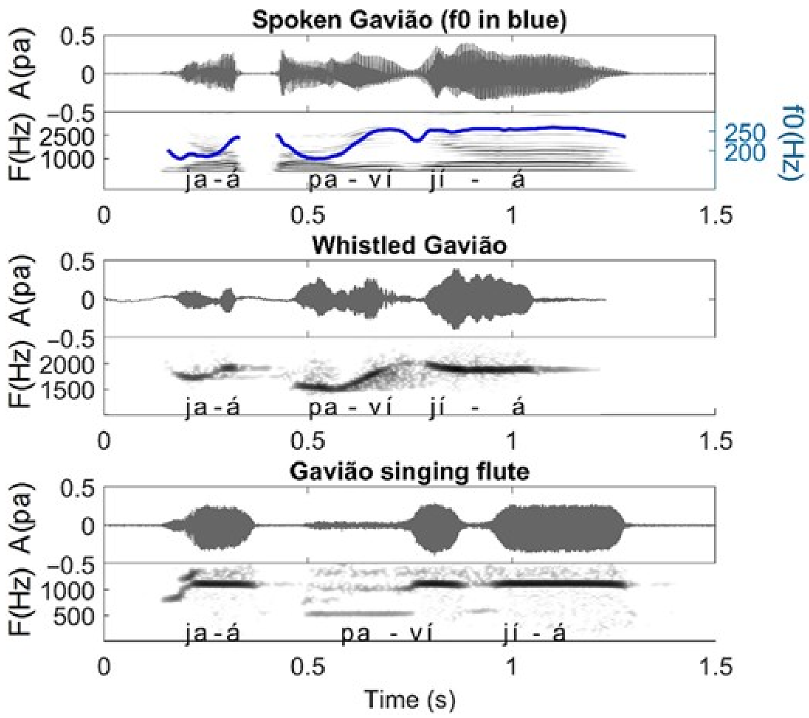

Figure 2 Waveforms and spectrograms of the sentence *jaá pavíjíá* (meaning 'Let's go to bathe' in the Gavião language of Rondônia, Brazil) uttered in spoken speech (upper row, Audio 1 <insert play symbol>), in whistled speech with lip protrusion (middle row, Audio 2 <insert play symbol>), and with the flute as in traditional flute songs (lower row, Audio 3 <insert play symbol>) (listen also to sound extracts in https://soundcloud.com/user-28976943/sets/sounds-in-meyer-manfredi-2023). The bold line on the spectrogram of spoken speech represents the extracted f0, highlighting the surface tonal line of speech. This figure illustrates two cases of pitch-to-pitch emulation (from the f0 of spoken speech either to mouth modulated whistles or to flute whistled pitches). The singing flute form in Gavião language uses an alternative flattened encoding of spoken pitches, which preserves relations between consecutive tones, as described in Meyer and Moore (2021).

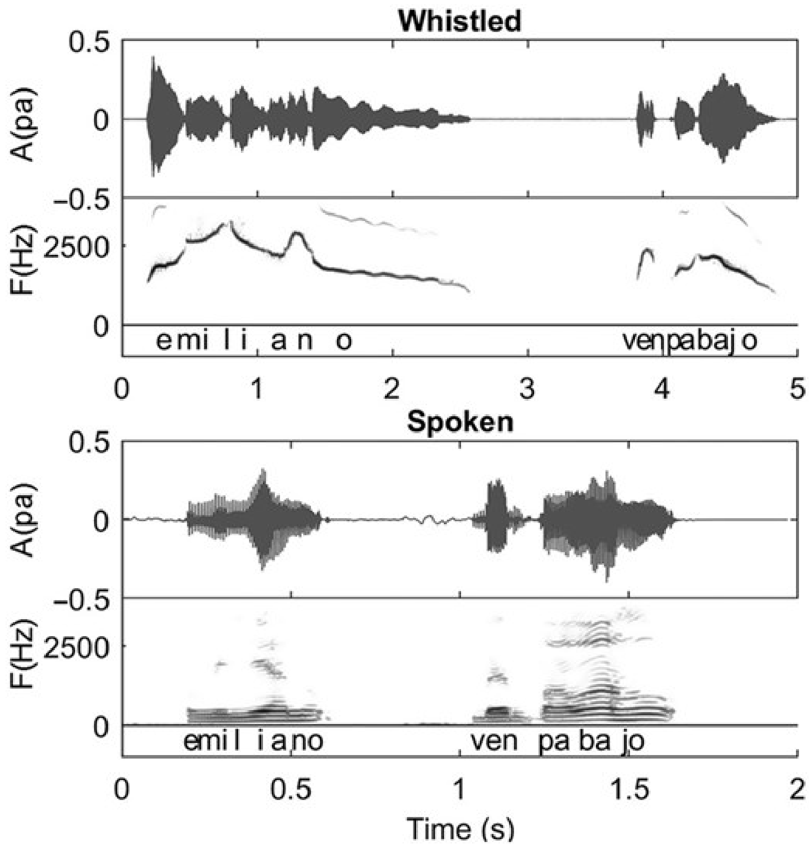

Figure 3 Waveform and spectrogram of the whistled (upper row) and spoken (lower row) forms of the Spanish sentence *Emiliano ven pa' bajo* [emiliano ben pa βaxo] meaning 'Emiliano come down, where *pa' bajo* is the fast speech fusion of *para bajo*, i.e. an allegro version. The x-axis of the whistled and spoken forms are not at the same scale because the respective whistling tempos are relatively slower. Whistling as produced here by a traditional whistler (Emiliano Fernandez Armas, from El Hierro Canary Island) relies on over-articulations and lengthening—mostly vowels—to ensure long-distance transmission of the signal. While being a frequency simplification of the spoken acoustic signal, the whistled form mirrors different phonetic aspects of modal speech: vowels are the most stable parts of the signal, and different qualities are whistled at different frequency ranges (typically /i/ higher than /e/ higher than /o/). They are modulated by coarticulation with consonants, and their whistled frequency may sometimes be influenced by stress (see e.g. the frequency of /a/ in /ba/ higher than the /a/ in /pa/). Listen to sound extracts in Audio 4 <insert play symbol> or https://soundcloud.com/user-28976943/sets/sounds-in-meyer-manfredi-2023.

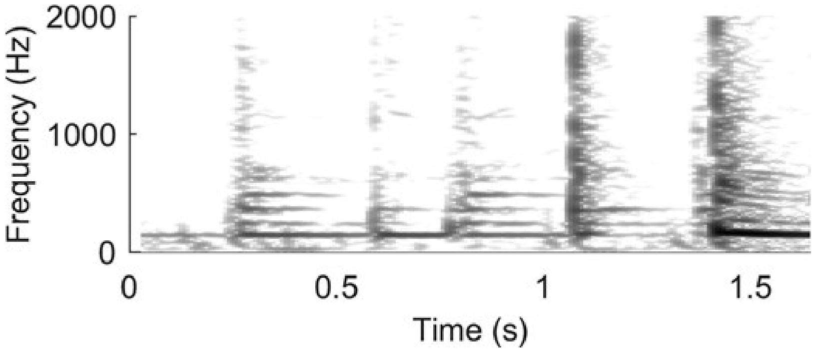

Figure 4 Spectrogram of a sentence drummed on the single-membrane Bendré drum of the Mossi people of Burkina Faso (the sentence is *wẽnd tẽntumd dolle*, see Kawada 1985, and Kawada and Jun-Ichi 1996, for text transcription with corresponding tone categories (H-L-H-H-H) in the Mooré language). The 5 strikes correspond to the 5 syllable nuclei (the first strike is around 0.25 sec). The ending strike is special and does not code the H spoken tone; it is instead played with an emphasis on low partials to mark the end of the sentence, as often in this tradition (see Kawada and Jun-Ichi 1996). The 1st and the 3rd strikes are of the same type and encode/emulate H tones. For these two, the beat on the drum activates simultaneously metal rings of which we see the partials. Drummed beats are inharmonic and some are without any pitch, such as the 'Slap' of the penultinian strike, which also encodes/emulates a H tone. The 2nd strike makes an emphasis on low partials and encodes/emulates a Low tone.

Source: Adapted from Dentel and Meyer (2020).

The choice between spoken timbre and pitch as targets emulated in auxiliary modality is not necessarily exclusive. For example, all experts of the Duala log 'drum' (idiophone) contrast low/high pitch by striking thick/thin wood lips of the drum respectively, but some players also encode vowel quality in parallel: 'My most accurate informant drummed in such a way that the vowels ***e*** and ***i*** were struck at the outside of the lip (of the drum), ***a*** at the edge and ***u*** and ***o*** in the outer part' (Schneider 1976: 672; Figure 5 below). Note that in this case, the spoken closed vowels /i/ and /e/ share a somehow similar—rather acute—organization of their harmonics as amplified in the front oral part of the mouth (emulated by a high note of the beat played by striking at the very end of the drum-slit). This contrasts with /o/ and /u/, which emphasize lower partials in the voice (played further away from the slit-edge). Interestingly, this kind of emulation of vowel qualities follows similar frequency similarities as when vowels are whistled in non-tonal languages (see Figures 3 and 6).

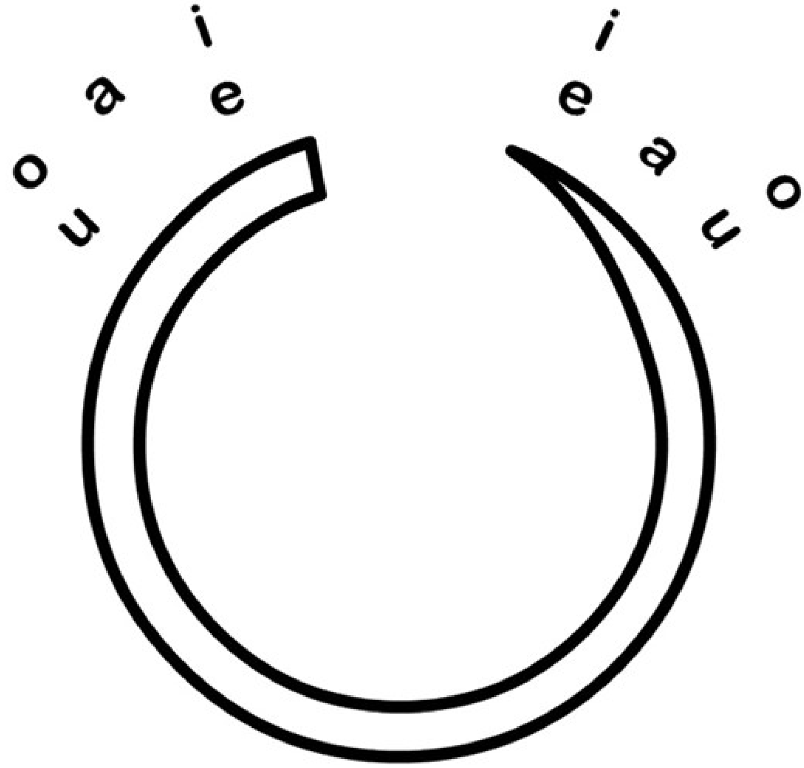

Figure 5 Schematized Duala log 'drum' producing low pitch (thick edge on left) and high pitch (thin edge on right). Above each lip are marked locations where a skilled drummer distinguishes three vowel timbres (oral 'heights').

Source: adapted from Schneider (1976: 676).

The acoustic constraints and limitations of auxiliary speech can be compensated semiotically by boosting information redundancy in several ways (cf. Jakobson 1960; Shannon 1948). A musical instrument can call a fixed text from memory; a short word *songe* 'moon' can be replaced with a longer, less ambiguous *songe li tange la manga* 'the moon looks down on the earth' (Carrington 1949); and information can be framed in a sequential template of *opening* > *addressee name* > *message type* > *message content* > *closing*. Memorized fixed texts are common in auxiliary speech forms based on verbal arts relying on song lyrics, proverbs, or poems (e.g. Kawada 1985 2014; Locke and Agbeli 1981). Framing sentences and enphrasing words generally compensate for the very reduced acoustic distinctiveness of log drums, instruments commonly chosen for sending public long-distance messages—to call someone, ask for something, organize an event (as was well described for Kele language (Carrington 1949), Bulu (Alexandre 1967), Banda Linda (Cloarec-Heiss 1999), and Bora (Seifart et al. 2018)). Whistling speech generally enables more acoustic distinctiveness than log drums, and whistled forms of words are rarely enphrased. Still, a whistler can substitute a synonym with a stronger syllabic pattern, e.g. Spanish *abrigo* 'coat' > *chaqueta* 'jacket' (Diaz 2017). Whistled sentences also tend to be simple, with single-proposition utterances and few conversational repairs to help regulate this constraint (Diaz 2017; Sicoli 2016). A whistled message can open with a 'phatic' call *aaaa* or *oye* in Canary Island Spanish, *oooo* in Pyrenean Bearnese, or a rising tone in Moroccan Tamazight (Busnel and Classe 1976; Diaz 2017; Meyer and Moore 2021; Rialland 2005; Trujillo 1978).

There is a marked difference between imitating modal spoken speech vs sung speech, even when the two tasks are accomplished by the same instrument in the same language (e.g. Seifart et al. 2018). Talking and singing with the same instrument can be robustly distinguished in performance thanks to rhythmic and melodic cues like gross repetition, canonical 'tonal' patterns, and vibrato that are more present in singing (Meyer 2015; Meyer and Moore 2021; Nikolsky 2020), but also thanks to fine acoustic microstructures (Dúrójáyé et al. 2021). Whistlers of many cultures can actively switch between 'speaking' and 'singing' modes, as can players of Asian jew's harps, Amazonian log idiophones, and West African drums and *bala* xylophones (McPherson 2019; Meyer 2015; Meyer and Moore 2021; Nikolsky 2020; Nketia 1963; Seifart et al. 2018; Thiesen 1969).

Perhaps it should not be surprising that musical performance is routinely employed for auxiliary speech in traditions where, conversely, the rhythms and melodies of musical performance directly emulate the cadences of spoken text. As the texts are often ancient and may refer to traditional myths and belief, a subset of this music is based on archaic forms of the local language (e.g. Kawada 1985; Marsh 2013; Meyer and Moore 2021; and see an example in Figure 9). This adds difficulty but also interest to the analysis, promoting approaches taking historical linguistics into consideration.

The influence of texts is not uniform in such practices: some musical styles related to auxiliary speech are fully based on texts (Dentel and Meyer 2020; Kawada 1985 2014), whereas others intersperse syllables without meaning or musical series of notes with no corresponding vocables (Dúrójáyé et al. 2021; Meyer and Moore 2021). In other cases, only a few words are emulated while the rest is music without underlying speech (for example, names in some Kuikuro flutes of the Xingu park in the Amazon; see Franchetto and Montagnani 2011). Another example of text/no-text shifts is illustrated by the Yorùbá and Ìgbo masked dancers simulating dead ancestors who can consciously shift from melodically modulated music without underlying speech to musical surrogate speech (Adédéjì 1972; Babáyẹmí 1980; Ùgońnà 1984). Note that many languages with speech surrogate practices, such as Yorùbá and Ìgbo, lack a direct translation for English 'music'—a 'discursive [...] historical and cultural' label that has no generally agreed intentional definition (Cross 2012)—and default instead to metonyms like 'song' (Yorùbá *orin*, Ìgbo *ábụ̀*), 'dance' (Yorùbá *ijó*), 'drum' (Yorùbá *ìlù*, Ìgbo *ńkwa*) or 'play' (Ìgbo *égwu*), cf. Nigerian Educational Research Development Council (1991: 164). These adjacent concepts illuminate the relationships between music and other domains of thought and behaviour by referring to circumstances of use, including language.

3 Case studies

Some auxiliary speech examples from the literature further illustrate local adaptations of the aforementioned techniques, strategies, and goals. In order to cover the main contexts of use, we present two telecommunication practices followed by several verbal art musical surrogacy speech modes. Note that tone-marking in this part of the manuscript is detailed in a footnote.[2]

3.1 Telecommunication speech forms

In isolated rural environments, text transmission by whistles and musical instruments extends the distance of intelligibility. The limit of shouting—a few hundred metres—is extendable by whistling to a kilometre or more in a favourable environment (Busnel and Classe 1976; Meyer 2015 2020) and by drumming to several kilometres (Seifart et al. 2018). The greater the distance to reach, the more the acoustic encoding is reduced to concentrate energy in powerful sounds that contrast with natural background noise. As mentioned above, this reduction in signal complexity entails greater redundancy of the message texts through procedures of framing and enphrasing. Conversely, such methods can also be used to conceal information from outsiders, and even close dialects may become less mutually intelligible when transposed into the reduced acoustic contrasts of whistles or drumbeats (Meyer 2021; Seifart et al. 2018).

3.1.1 Whistled speech in Spanish (Canary Islands) and Gavião of Rondônia (Brazil)

Present mostly in isolated forests and mountains which create a strong demand for long-distance communication in rural settings because dense vegetation and rough topography often lead to physical isolation and constrain spoken communication, whistled speech imposes high performance pressure on the whistlers to transmit the meaning of the text. This is the reason why whistled sentences highlight some salient features of consonants, vowels, and/or tones of spoken languages through which the listeners can cognitively reconstruct the message. To maintain a whistling airstream while pronouncing the words, the sound source is limited to the front of the vocal tract, where a strong air pressure is easier to apply. This results in a drastically reduced frequency band in comparison to modal spoken speech. Articulatory constraints are thus more restrictive than in whispering or shouting. This limitation imposes a choice between pitch and timbre as a primary target in the vocal-to-whistle transformation.

For a 'tonal' language, whistlers transpose the vocal pitch (corresponding to the aforementioned category ***vocal pitch* ⇒ *auxiliary pitch***; one example is the Gavião language in Figure 2 above), otherwise the target is vocal timbre (aforementioned category ***vocal timbre => auxiliary pitch***; one example is Spanish, see Figures 3 and 6). The fact that such whistled sounds proceed only from the front/oral part of the mouth has notable consequences in terms of surrogacy.

While transposing vocal timbre into whistles, speakers–whistlers of 'non-tonal' languages maintain similar oral articulatory gestures for whistled vowels and consonants (even if more limited to sustain the whistle). The difference is mainly at the acoustic level, from many spectral frequencies defining the vocal spoken timbre to one simple modulated frequency of the whistles. Whistles map only key salient acoustic/cognitive cues of the original spoken signal, and particularly frequency shapes of different speech formants of modal speech. Examples on how the whistles map these formants can be seen in Figures 3 and 6 (details as a function of the vowel type are further explained in Meyer 2015; Ridouane et al. 2018; Tran Ngoc et al. 2022).

By contrast, people transposing vocal spoken pitch into whistles in tonal languages operate both a sound source change and an articulatory transfer from the larynx to the oral cavities. While the acoustic transposition is simple (pitch-to-pitch) (Figure 2), the articulatory difference is high because the dynamics of spoken sounds produced originally in the larynx are now transposed to the upper oral cavities.

Moreover, in all languages surveyed, secondary cues are also targeted, and whistled speech frequency variations may also be partly sensitive to stress when it is encoded by intonation (as shown in Figure 6; see also Diaz 2017). Finally, in any whistling transposition strategy, phonetic reduction imposes a context of the conversation, but whistlers are still able to exchange complex (not stereotyped) sentences in real dialogues (Moles 1970).

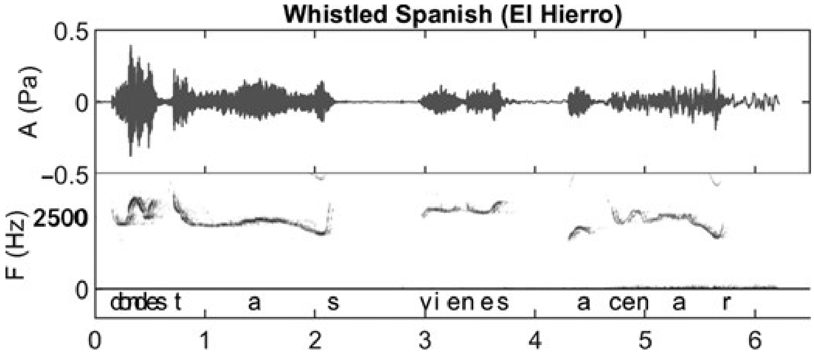

Figure 6. Spanish sentences *¿ donde estas? ¿vienes a cenar?*, pronounced [dondestas bienes a senar], with the fast speech fusion of the two first words based on the same vowel. This whistled extract, meaning 'Where are you? Are you coming for dinner?', was recorded by a traditional whistler (Emiliano Fernandez Armas) of El Hierro Canary Island who was targeting long-distance communication. The El Hierro variant of Canarian Spanish is characterized by the pronunciation of [s] before the end of a speech group (differing from most other variants in other islands, e.g. Alvar 1975), which is whistled by a quick frequency rise at the end of the word, as in /estas/. Note that, by contrast, the final /s/ of /vienes/ is whistled differently with a flattened end, which is the whistled surface manifestation of a slight aspiration (see other examples typical of El Hierro in Diaz 2017). Listen to sound extracts in Audio 5<insert play symbol> or https://soundcloud.com/user-28976943/sets/sounds-in-meyer-manfredi-2023.

3.1.2 Bora drums (Amazonia)

Drumming exploits short beats, and drummed speech clearly highlights rhythmic cues of speech through these beats. Hollow log drums—such as the Manguaré drums of Bora/Miraña people of West Amazon—are the most commonly used for telecommunication purposes because of their powerful and very low-frequency signals whose long wavelengths travel relatively far. Spoken-word rhythm is copied in drummed Bora, and different rhythmic units were found as a function of the number of segments (vowels and consonants) between two vowel onsets (V-to-V intervals), word-internally (see Figure 7). Drummed Bora also employs sentence framing and word enphrasing (as described at the end of Section 2), as well as emulating the pitch of vowels (and thus their tone) (Meyer et al. 2012; Seifart et al. 2018; Thiesen 1969). Amplitude levels of drumbeats were found to be fairly constant, and thus do not represent potentially informative amplitude variations of spoken language. However, drummed high and low tones do display two different types of amplitude envelopes, with a quicker amplitude decline for high tone beats (higher pitch is played on the smaller and less resonating log drum), providing redundant information besides frequency heights to identify tones.

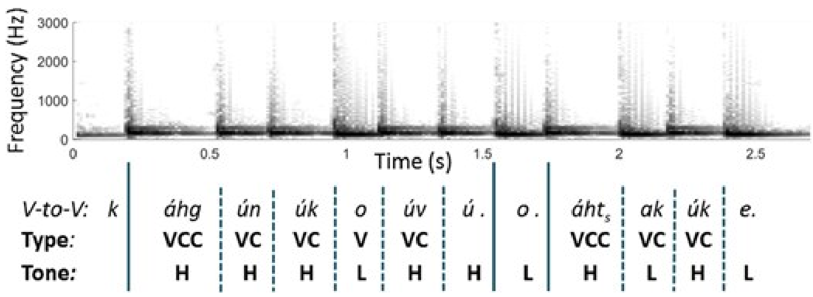

Figure 7 Spectrogram of the drummed Bora phrase /*káʔgúnúkòúβú ò áʔt͡sàkúnè*/ 'I am finishing the cahuana [manioc starch drink]'. The bottom rows provide the transcription—in orthography in use locally—but also the segmentation into vowel-to-vowel (V-to-V) intervals as well as spoken tone levels (in Bora language, a High tone is written with an acute accent and a Low tone is written with a grave accent).

Source: Adapted from Seifart et al. (2018).

3.2 Verbal arts in musical surrogacy

From the wide diversity of musical speech surrogacy systems, we chose to first highlight here drummed traditions of Africa based on proverbial texts in Gùn-gbè and Yorùbá. They are played on different types of skin drum that entail more variability in strokes than log drums used for telecommunication. They offer different techniques by which drummed emulations of speech have adapted to constraints of musical instruments to encode key phonological aspects of languages. Next, we also present how three different musical instruments of the Amazonian Gavião people transform phonetic and phonological features of song lyrics into fixed musical notes. Finally, we explore key challenges of surrogate speech and verbal art emulation via the Ìgbo ọ̀ja flute.

3.2.1 Gùn-gbè drum and its emulation of proverbial texts (Republic of Benin)

In his pioneering analysis of drummed speech emulation in Gùn-gbè, Rouget (1964) observed how different timbres and amplitude envelopes of complex strokes on a skin drum may accurately represent spoken sounds. He proposed 'an articulatory musicology' defined as 'the description of music in terms of the physical motions by which it is produced' (Rouget 1964: 25), and further conjectured that 'drummed utterances can rightly be held to show how speakers sense, if not conceive and in any case interpret the tone system as it operates when they are speaking' (p. 3).

This original hypothesis is testable because a drummer's hands are more easily observed than are the hidden muscles of the larynx (Halle and Stevens 1971; Nissenbaum et al. 2002), and also because many oral traditions already have narratives which explain how instruments are built and how they should be played (e.g. N'Guessan 2020; Niangoran-Bouah 1981; Pacere 1987).

Comparing musical and linguistic properties of Gùn-gbè proverbial texts simulated on a single-membrane drum ('générateur unique'), Rouget concluded:

(i) Contrasts of spoken F0 are not realized as distinct musical notes of different pitches—difficult to produce with a single membrane—but as spectral qualities ('timbres') distinguished in three types of strike. An open hit on the edge has a rich, broad harmonic spectrum and is associated with the M tones. A muted strike on the edge of the drum emphasizes high-frequency partials of the drum (associated with Gùn-gbè H 'tones'), whereas a muted strike on the centre emphasizes low-frequency partials (associated with L 'tones').

(ii) A spoken syllable with H 'tone' or L 'tone' is systematically longer than one with M, but in drumming, the relative duration is reversed as rapid vs gradual decay of resonance are produced by muted vs non-muted manual strikes.

(iii) Of Gbè's three phonetic 'tones' H/M/L, L is least intense—has lowest signal strength in decibels—in both speech and drumming modes. The drummer strikes L on the centre of the membrane, vs the edge for H and M.

Rouget's results are consistent with a theory of acoustic emulation in musical surrogate speech for which the target of emulation need not be restricted to fundamental frequency. Instead, the drummer as a fluent Gbè speaker can also access a set of other acoustic features sufficient to distinguish Gbè H/M/L in both vocal and drummed modes (as summed up in Table 1).

Table 1 Rouget's intensity + duration-based analysis of Gbè oral and drummed auxiliary speech (intensity decay dynamics of the envelope correspond to (ii) and max intensity values correspond to (iii))

	Drum articulation		Drum sound			Speech sound	
	Place	Manner	Intensity	Sleep	decay	Intensity	Duration
H	edge	muted	+	+	+	+	+
M	edge	non-muted	+	–	+	+	–
L	centre	muted	–	+	–	-	+

Most surprising in these parameters is the fact that vocal durations associated with tones are reversed in the drummed form. Perhaps priority was given to emulating frequency and intensity, or perhaps other timing cues are relevant, although Rouget specified that the beat-to-beat duration is not consistent with spoken durations in his data.

Another aspect is asymmetric between vocal and drummed Gbè modalities: the degree of contrast between the three tones. Indeed, in the Gbè language group, the F0 contrast between H and non-H 'tones' is robust, but the M/L distinction fluctuates depending on consonant timbre and syntactic alignment (Ansre 1961; Capo 1991; Gbètɔ 1995 2020; Hyman 1973; Manfredi 2018 2020; Smith 1968; Sprigge 1967; Stahlke 1971). The spoken data presented by Rouget show this effect: spoken H is approximately 60 Hz higher than M, but M is separated from L by 20 Hz at most. However, a different asymmetry appears in Gùn-gbè drumming, as the M/L distinction is the clearest one according to the 'drummed articulation' parameters and to their acoustic expression in frequency, time, and intensity (see Table 21.1).

This might be one of the reasons why Rouget wrote: 'la place qu'on attribue traditionnellement aux intervalles se rélèverait peut-être plus grande qu'elle n'est en réalité' [traditionally more importance is perhaps ascribed to pitch intervals than they actually possess] (Rouget 1964: 26; cf. Rouget 1975: 218, 225). Another reason is probably that, for Rouget, several other aspects could be more fully explored in relation to spoken pitch as transposed on drums:

(i) Spectral qualities of drummed beats could be characterized by the terminology usually employed for spoken phonetic features as a function of sonority such as [compact], [checked], [tense] (cf. Jakobson and Halle 1956).

(ii) The 'tone' distinctions made in terms of only pitch would profit from a more syntagmatic approach in some languages such as Gùn-gbè, that would accommodate other acoustic parameters as drums do and as notions such as 'intonation' or 'accentuation' also do.

A further lesson follows from Rouget's contribution: that it is unreliable to assume a traditional analysis of 'tone' and apply it mechanically to musical speech surrogacy. Instead, auxiliary speech phenomena may highlight in a different way important elements of linguistic pitch analysis.

3.2.2 Yorùbá music and its large diversity of speech surrogates (Nigeria)

From ten Yorùbá-speaking communities, Thieme (1969) recorded surrogate speech by 13 musical instruments spanning 6 formal genera *à la* von Hornbostel and Sachs (1914).[3]

Genus 1: laced hourglass variable-pitch drum *(dùn-dún*, two-headed; *gán-gan*/*àpàlà/àdàmọ̀*, two-headed, smaller than *dùn-dún; kẹríkẹrì*, two-headed, accompanying *bẹ̀mbẹ́; kósó*, single-head).

Genus 2: *bàtá* are conical two head drums, whipped on the small end with a leather flail.

Genus 3: pegged single-head drum *(ìgbìn*, '*keregidi*' [diacritics unknown], *ìgbésẹ̀ lukorigi*/*lukoogi* or *àgẹ̀rẹ̀* (cf. Abraham 1958: 30; Adéoyè 1979: 125).

Genus 4: percussion idiophone *(ọmọ owú*, anvil hammer).

Genus 5: plucked idiophone, metal tongues on wood box or split canes on gourd resonator ('*molo*' [diacritics unknown] or *àgídìgbo*).

Genus 6: wind instruments (*ìpè* 'trumpet', *ekùtù* side-blown antelope horn, *tíokò* bamboo flute). Such multiplicity opens possible comparisons of different instruments conveying identical or similar texts in one speech community. Matching Yorùbá texts have been elicited in parallel on *dùn-dún* and *bàtá*—two mechanically different drums whose virtuosi, however, share initiation in àyàn, an Ọ̀yọ́ palace guild named after the tree from which these instruments are carved (cf. Abraham 1958, p. 86; Àyánkúnlé 1997; González and Olúdáre 2022; Law 1977; Villepastour 2010). Video 21.1 shows parallel renditions of one proverb in two oral and two drummed modalities[4] (source: Àyánkúnlé 1997).

Transcript:

A ò ní sí ǹbẹ̀

We will not be there

níbi wọ́n gbé ń ṣe orí burúkúkú

where they're always manifesting bad luck [lit. 'bad head']

a ò ní sí ǹbẹ̀

We will not be there

[drummed on *dùn-dún*]

[drummed on *bàtá*]

jajà fífí rídà

fífi jí fáfá já ti rí fí

jajà fífí rídà

[drummed on *bàtá*]

[drummed on *dùn-dún*]

Dùn-dún drums. An expert player masters *dùn-dún*'s fine pitch control so that it closely matches vocal F0 contours—including glides (Akínbọ̀ 2019). Yorùbá has three relative pitches H/M/L as well as predictable glides between H and L within a phrase. In contrast with the 'tone language' stereotype and the analogy with musical scales, Yorùbá H and L are positive articulatory gestures and acoustic targets, whereas M only fills in a neutral/unspecified/default value (Akinlabí 1985; Connell and Ladd 1990; Láníran 1992; Siertsema 1958), and this asymmetry of 'markedness' arguably carries over into *dùn-dún* pitch signals. Correlates of oral rhythm (e.g. Dúrójáyé et al. 2021) and amplitude (e.g. Villepastour 2010) are also measurable in the drummed signal of this instrument. For instance, vowel-intrinsic intensity is taken into account when modulating the force with which the membrane is hit; the syllables containing the high vowels /i/ and /u/ are played in a significantly softer manner (lower amplitude), when compared to the surrounding strokes (e.g. González and Olúdáre 2022). Despite their rich acoustic cues, however, *dùn-dún* messages still admit radical ambiguity outside of a finite, learned textual corpus reinforced by context (Adégbọlá 2003); it is not exhaustive, therefore, to assert that '[t]he Yorùbá drum languages [...] do not rely on any hermetic coding process, but on the outstandingly cultivated aural sensibility of both drummers and listeners' (González and Olúdáre 2022: 18). Indeed, it is easy to meet fluent L1 Yorùbá speakers who frankly confess 'N kò gbọ́ ìlù' ('I can't parse the drum', referencing either *dùn-dún* or *bàtá*), revealing an incapacitating lack of familiarity with the repertoire of ritualized utterances. The indeterminacy is illustrated by a famous anecdote (Beier 1970: 12f., transcription adjusted):

> An illiterate drummer was asked to say *This is the Nigerian Broadcasting Service* (in English). The drummer listened to this phrase [...] and he produced the tone sequence presented in Figure 8, which became the Nigerian radio service's signature tune.

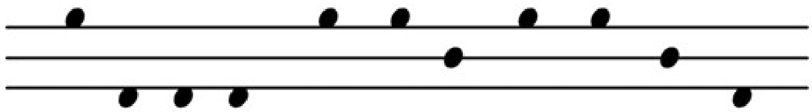

Figure 8 Schematic note annotation showing the rhythmic and frequency variations of the drummed production by a master drummer of the sentence 'This is the Nigerian Broadcasting Service' (in English).

Now, of course the man in the street did not know the origin of the phrase and, looking for a meaning in Yorùbá, he came up with the following:

> *B'Ólú.bàdàn bá kú tal'ó j'oyè?* If the ruler of Ìbàdàn dies, who is his successor?

Other versions were:

> *Ó jẹ̀'gẹ̀dẹ̀ dudu, inú ń ta bọ̀n-ùn.* He ate overripe plantains, his stomach swelled up and burst.
>
> *Kò sí olòṣì n'íbí, lọ sí'lé 'kejì!* There are no paupers here, try the next house!

Nevertheless, when listeners were tested perceptually recently in a more controlled and detailed context (Krowles et al. 2025), examining individual differences in judgement ability when listeners compare vocal and drumming excerpts (short Yorùbá proverbs and idiomatic expressions) with different levels of match, results statistically confirmed that similarity ratings for matching pairs are an additive function of both language familiarity (Yorùbá > English speakers) and music training (Musicians > Nonmusicians).

Bàtá drums. The *bàtá* drums, with fixed membranes, are considered harder to play and interpret, probably because they combine strokes by hand and/or with the leather *bílálà* flail to distinguish two classes of vowel qualities and three Yorùbá tones (Láoyè 1959: 10; Oyèélámì 1991; Rouget 1965: 82). The lead (*ìyá* 'mother') *bàtá* drum, less capable than *dùn-dún* of emulating spoken f0, compensates by controlling different frequency spectral distributions—coding drummed timbre—to emulate both spoken pitch and two distinct vowel qualities. A universal difference in the acoustic resonance of vowels is exploited: F1 (the first spectral formant) is produced much closer to f0 (fundamental frequency) in vowels like [i, u], articulated with strong constriction of the vocal cavity, than in more open vowels like [a, e, o] (Stevens 1998: 262–267; cf. González and Olúdáre 2022: 11f.). Further evidence for *bàtá*'s intentional control of timbre is the obligatory use of an adhesive compound of indigenous latex (*ìda*) applied to the centre of the larger drumhead so as to dampen overtones (M. Kone, p.c.) and reinforce lower partials (this effect can be listened to in Video 21.2; and it resembles the muted strike on the centre of Gùn-gbè drum associated to L 'tones' described by Rouget, see Section 3.2.1). The *bàtá* combinatorics include three types of strikes on this larger head of the drum—open, muted, slap—used to code respectively low, mid, high tones. Video 21.2 illustrates this acoustic effect of *ìda* latex damper on *bàtá* large drumhead[5] (source: Àyánkúnlé 1997).

transcript:

[4 strikes with *ìda* applied]

Ẹ mọ́n-ọn gbọ́ nísinìí

You all listen now

[4 strikes with *ìda* removed]

Iṣẹ́ tí ń [...] tí a bá tún fi le nísin-ìn-ín

The work that's [...] when we stick [ìda] back on

[4 strikes with *ìda* applied]

A simultaneous strike on the other, smaller head of the drum is added to render the distinct spectral quality of open vowels. A strike on the smaller head alone is used to code the mid tone in some specific contexts. Oyèélámì (1991) published the first explicit mapping of tones and vowel timbres onto these seven permutations of bàtá drumstrokes, although that account is idealized, and in performance the drummer may add the secondary enhancement of a whip strike on the smaller drumhead for H tone, even to accompany a non-open vowel quality (González and Olúdáre 2022: 10). There are also quite frequent speech fusions or contractions on the *bàtá* (mostly resembling normal allegro speech effects or performance haplology of adjacent identical phonemes and/or identical pitch). The youngest generation of *bàtá* maestros have gone still further to innovate a transitional type of *bàtá* (Euba 2011), called *omele mẹ́ta* (triple *omele*) bundling three mini *bàtás* together so as to directly emulate f0 to encode tones (H, L, M) in addition to various timbre cues corresponding to vowels (closed vowels are performed with a simultaneous strike on the highest-pitch drum), in this way improving the intelligibility of surrogate speech for popular audiences more accustomed to *dùn-dún*.

To conclude, studies on the different types of Yorùbá drums presented here show different degrees of sound similarities between spoken speech and drum sounds which underlie different speech surrogacy systems for the same language. These different drummed surrogacy forms make tangible the same Yorùbá language structures (tones and vowel quality) in different ways that adapt to specific instrumental constraints. By doing so, they shed light both on the importance of such abstract mental representations of linguistic sounds for this language but also on the fact that the means by which such representations are expressed can vary because they may sometimes be detached from a part of the acoustic correlates on which they rely in spoken modal speech.

3.2.3 Gavião lyrics emulated with three different musical instruments (Amazonia)

Instrumental speech exists among the Gaviaõ people of Rondônia with three different types of musical instrument employed to emulate song lyrics—a flute, an ensemble of three one-note bamboo 'clarinets' (Figure 1) and a mouth arch (Moore and Meyer 2014). The ensemble of three one-note bamboo reed instruments is a specificity of the Amazon, while the flute and the mouth arch are more widespread around the planet. The phonetic and phonological nature of the acoustic iconicity which exists between the words of the songs and the music played is simply based on the cognitive association between, on one side, the tone and syllable length of the Gaviaõ spoken language and, on the other, the notes played with musical instruments. The speech-to-music correspondence is from a spoken pitch line that allows frequency curves to an instrumental music line that is based on fixed notes. The pitch line of the surface phonetic form is replicated, but only with flat notes, which reflect the relative pitch levels of consecutive syllables in phrases. The flattening of the pitch appears to be a response to the constraints of the musical instruments: frequency modulations are technically impossible with the bamboo reed ensemble, and they are hard to produce with the mouth arch and the flute (Meyer and Moore 2021). To teach these flattened melodies of the base spoken form, the players use an intermediate form, sung with the voice, which also flattens the melody. Flattening the phonetics of the linguistic pitch does not, however, hinder the players from rendering the relative height of a syllable in relation to the surrounding syllables (see Figure 2 lower line for flutes and Figure 9 for bamboo clarinets). For example, the musicians manage to emulate processes expressed as rising/falling pitches or downstep tonal effects, which are rendered by relative differences with surrounding flat musical notes (see full explanation in Meyer and Moore 2021, and examples in Figures 2, 9 and 10). Moreover, these instruments provide another perspective on the linguistic rules followed by spoken pitch in the linguistic system of Gavião. An additional surprising aspect of surface speech imitation with Gavião singing musical instruments is that the results of fast speech fusions, which are far from the underlying forms, can be played as flat notes according to the same patterns.

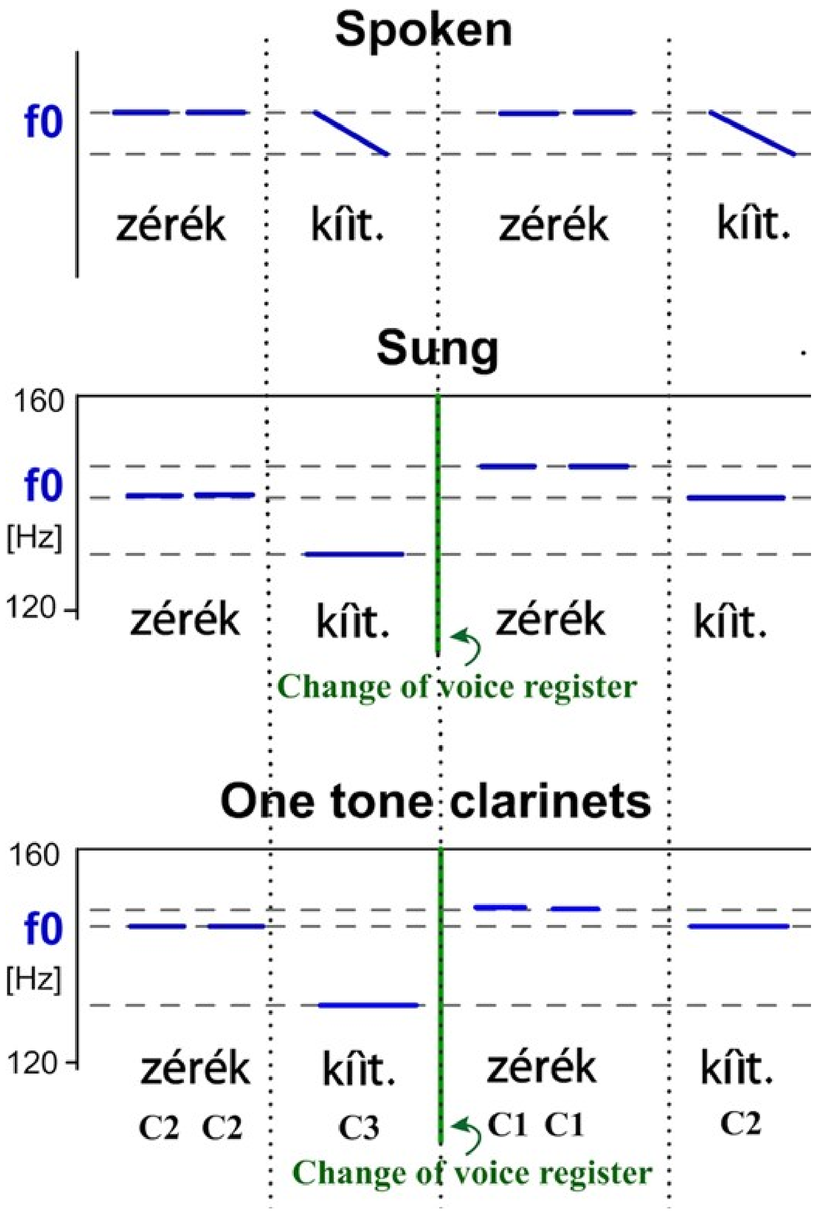

Figure 9 Simplified charts of the spoken, sung, and musical forms of a song played on totoráp clarinets. The same sentence is repeated twice here (meaning 'The white skin. The white skin'). C1, C2, C3 correspond to 'Clarinet' 1, 2, 3, as the instrument is an ensemble of three one-tone bamboo reed instruments (see Figure 1 for a photo). Note here also that the change of voice register in the middle of the sung verse is imitated by the instrumentalists (indicated by a vertical/green line). These are typical musical aesthetic effects. Another difference from the spoken form is the flat *kîît* in the sung and instrumental forms. They still encode the relative pitch relations to surrounding syllables and substitute the falling melody of the spoken form.

Source: Adapted from Meyer and Moore (2021).

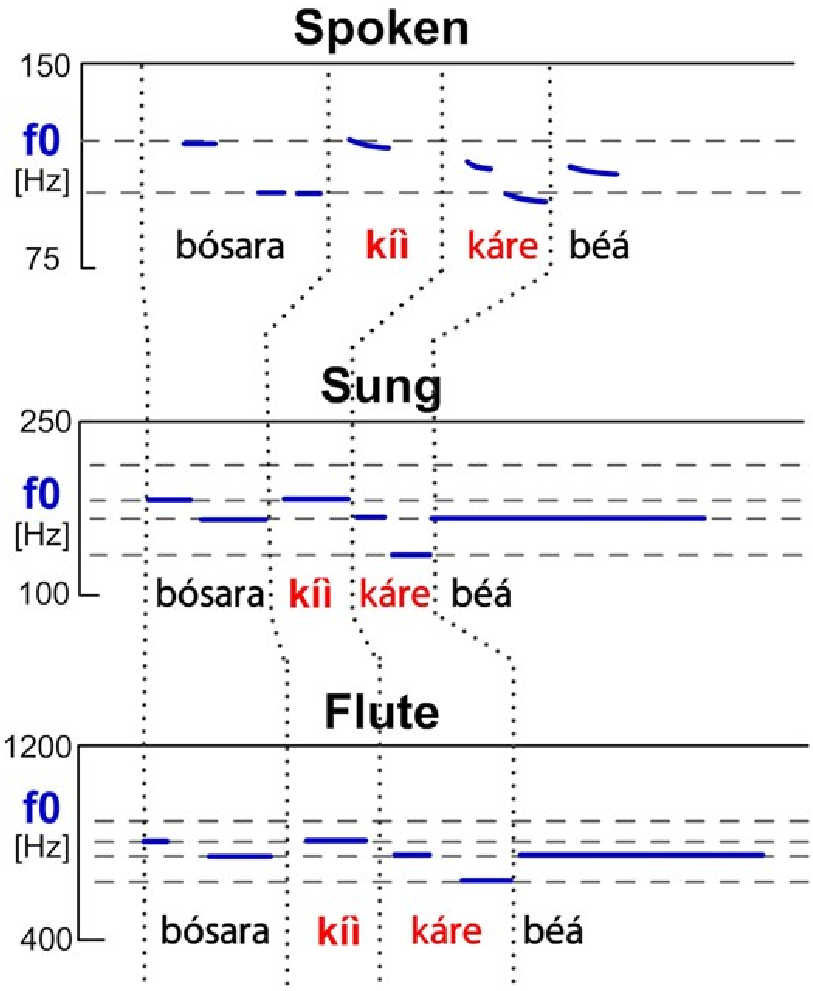

Figure 10 Simplified charts of the spoken, sung, and flute forms of a song extract played on the Gavião kotiráp flute Here we show only the end of a verse meaning 'I will lie down with her daughter again' [see and listen to full sentence in Meyer and Moore 2021, from which the present chart is adapted]. Note the tone downstep phonetically expressed in the three forms after *kîì* in *kîì káre béá* (according to Gavião language descriptions, the syllable *kîì* ends with a floating low tone that is followed by the high tone of the syllable *ká*, and thus attaches to the following high tone: the combination is realized phonetically as a mid-tone in spoken and sung speech, and as a middle note in instrumental speech played on the flute (see Meyer and Moore 2021, for details)). Finally, the verb *bósara*, translated as 'going to bed [retiring to a hammock]' is a lexical item that does not exist in the modern Gavião language.

3.2.4 Ìgbo ọ̀ja *flute (southeast Nigeria)*

The *ọ̀ja* is a small three-hole, cross-shaped wooden flute producing at least five distinct musical pitches (Nwáchuku 1997). It is used to emulate Ìgbo utterances in different contexts such as reciting texts for influential people in the community, during lifecycle celebrations (such as weddings), or to address invisible participants (for example during dance-mask performances). Carter-Ényì et al. (2021) elicited 16 Ìgbo praise expressions (*áhá òtutu*) recorded sequentially in declamatory chanting mode from a male and a female speaker, then emulated in their presence by a player of the *ọ̀ja* flute.[6] The traditional 'chanting mode' of the flute performed during this video documentation is demonstrably more speech-like than the 'singing mode', which also exists (Lo-Bamijoko 1987). This enables a rather straightforward analysis of the 'speech-to-music conversion' of the flute (i.e. flute emulation of speech), avoiding various aesthetic effects added in the singing mode. In order to be able to compare this practice to the other ones described in the preceding paragraphs, we revisited this set of data, exploring how the flute emulates the pitch melodic line of vocal speech.[7] Just as for the Gavião flute (Section 3.2.3), the Ìgbo *ọ̀ja* flute proceeds to some adaptations in the pitch and rhythmic patterns. Those specificities are either due to *instrumental acoustic limitations* (the flute is less capable of frequency modulation than of producing level pitch) or are related to the *base chanting speech form*, which adopts a poetic style. These changes provide the occasion to revisit some important characteristics of the Ìgbo language, and in particular to question why some of them are maintained in the instrumental form despite the acoustic abridgement.

First, the flute versions of the phrases are on average roughly 30% to 40% longer than the vocal ones (male 1.8 seconds, female 1.9, flute 2.8), with a significant part of the flute duration (~40%) due to the extended last note of each phrase played as an emphatic marker of the end of the verses.[8] Moreover, all but two of the texts of this corpus are nominal epithets, i.e. complex noun phrases, a syntactic construction entailing numerous downstep junctures in Ìgbo. This reflects the importance of downstep in this language. In these conditions, we transcribed the spoken Ìgbo text of the 16 phrases in Welmers and Welmers' (1968) tone spelling designed to show downstep economically and transparently, with the convention that a syllable without a tonemark copies the pitch value on the same level of the preceding syllable (see (1) and (2) below for examples on two phrases of the corpus). In the adjacent schematic pitch notation, we marked all the downsteps with the standard phonetic juncture sign '!' (see also Figures 11, 12 and 13, where such pitch annotations correspond to the ones of the spoken form). The flute audibly emulates 14 of the total of 22 downsteps present in the linguistically significant pitch contours of this corpus. However, the material at hand does not allow us to try to predict *which* downsteps are omitted by this flute performer. Downstep has been extensively analysed in linguistic studies of Ìgbo (Éménanjọ 1978a 2015; Liberman et al. 1993; Welmers 1970). Rather than a taxonomic phoneme, it is a syntagmatic *relation* between H tones conditioned by a grammatically imposed prosodic boundary (Clark 1980).[9] It can also be represented analytically as units of syntagmatic 'constituency, such as bracketed grids and metrical trees' (Dilley 2005: 46; cited by Carter-Ényì 2016: 65). Given the significant albeit imperfect registration of downstep cues (14/22) in this *ọ̀ja* corpus, more controlled examples in future research may allow us to discover what combination of structural constraints and performance errors explain the omission of some downsteps, and whether the resulting discrepancies affect intelligibility.

Comparing phrases 2 and 7 from the published recordings of Carter-Ényì's corpus, an instructive difference appears, illustrating what we just explained. Phrase 2 has three syntactic constituents informally shown by the bracketed literal gloss below. As indicated in the pitch transcriptions, the male and female spoken versions have a sequence of six H syllables interrupted by one L syllable (the verb *-zụ̀* 'feed/nurture') and two downsteps, whose abrupt drops (of approximately 10 Hz or more) are clearly visible in the corresponding spectrographic representation (Figure 11). Moreover, the female is slower than the male, and the flute is slower again than the female. The fact that the absolute pitch drops of the downsteps are much greater in the flute than in either of the vocal versions presumably follows automatically from the technical tuning of the instrument. Despite any difference of modality, however, this example shows that the flute can emulate downstep through pitch, and does so reliably (see also Homayounfar 2021).

> (1) *Transcription, gloss, and tonal annotation of phrase 2 (see recordings in* Figure 11*)*:
>
> ‘nurturing person’
>
> [$_{\text{XP1}}$ ala [$_{\text{XP2 relative clause}}$ na-a-zụ [$_{\text{XP3 genitive modifier}}$ nnwa]]]
>
> [$_{\text{XP1}}$ breast [$_{\text{XP2 relative clause}}$ that habitually feeds [$_{\text{XP3 genitive modifier}}$ child]]]
>
> spoken male and female: ála ná-a-zụ̀ ńnwá HH !HHL H!H flute (played on 5 levels of notes): HH !HHL H!H

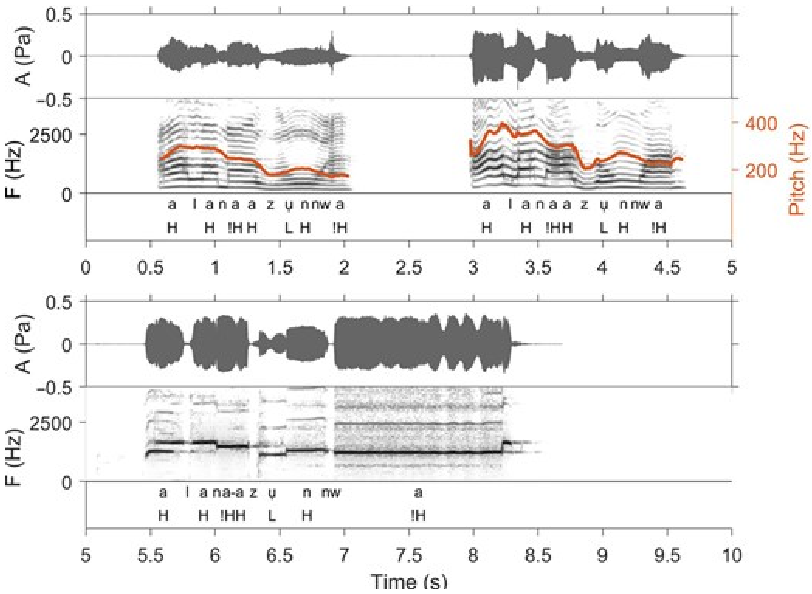

Figure 11 Waveform and spectrogram of three different pronunciations of the same sentence, *ála ná-a-zụ̀ ńnwá*. Spoken by a male (upper left), spoken by a female (upper right), and played with the flute (bottom). The f0 is extracted and shown for the two spoken productions (bold/orange line and pitch scale). The pitch simplified notation—with H, L, !H—refers in the three cases to the spoken reference.

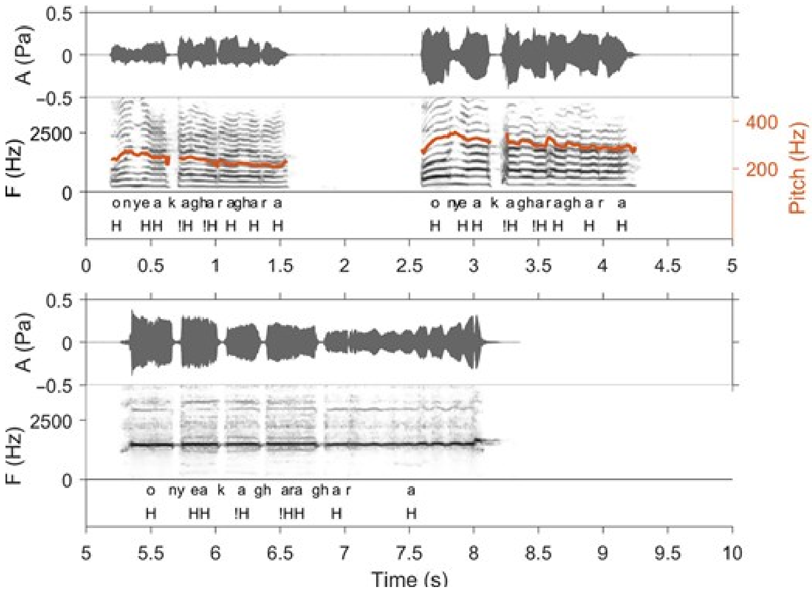

Figure 12 Waveform and spectrogram of three different pronunciations of the same sentence *ónye ak ghároghara*. Spoken by a male (upper left), spoken by a female (upper right), and played with the flute (bottom). The f0 is extracted and shown for the two spoken productions (bold/orange line and pitch scale). The pitch simplified notation—with H, L, !H—refers in the three cases to the spoken reference.

Next consider phrase 7(2), the only example for which Carter-Ényì et al. (2021) provided a pitch track. These authors correctly observed that the flute version is produced with essentially level pitch, whereas the pitch of the spoken versions markedly declines (see Figure 12). This is in contrast to the observations about the beginning of phrase 2, which is also characterized by H tones and a downstep.

> (2) *Transcription, gloss and tonal annotation of phrase* 7 *(see recordings in* Figure 12*)*:
>
> ‘generous person’
>
> [$_{\text{XP1}}$ onye [$_{\text{XP2 genitive modifier}}$ aka [$_{\text{XP3 relative clause}}$ gharaghara]]]
>
> [$_{\text{XP1}}$ person [$_{\text{XP2 genitive modifier}}$ hand [$_{\text{XP3 relative clause}}$ disburses things around]]]
>
> spoken male and female: ónye aká gháraghara HH H!H !HHHH
>
> flute: HH HH HHHH (played on one note, called H here)

As shown in (2), obligatory downsteps occur at the internal phrase boundaries. Phonetically, pitch drops occur in the spoken forms as downsteps in good standing (more clearly in the female than in the male): they are ‘non-automatic’ precisely as defined by Stewart (1964) and are not reducible to the universal vocal trend of ‘back-drop declination’ (cf. Connell and Ladd 1990) as Carter-Ényì et al. (2021: 3) suggest.[10]

A different but notable *ọ̀ja* characteristic clearly visible in these two phrases is that directly adjacent syllable nuclei with the same tone—i.e. vowels or syllabic *n* and *m* in Ìgbo—are often assimilated in a single flute note, even across word boundaries (Figures 11 and 12). The resulting played musical note is not systematically longer than for a one syllable note. Moreover, when tones are different, there are various possibilities: (i) either to play them differently (see ‘*zụ̀ ń*’ of phrase 2, played LH on the flute (Figure 11; see also ‘*mụọ́ ń*’ of phrase 8 in Figure 13, played at three different flute pitches), or (ii) to assimilate all the adjacent syllable nuclei in a same pitch level (as in ‘*rụ ó*’ and ‘*mụọ́ à*’ of phrase 8, see Figure 13). In the latter case, a subtle compromise is made by the flautist as the resulting level pitch is influenced by the different constituent syllables while succeeding in maintaining relative pitch differences with the other neighbouring syllables (as noted also for Gavião flute, see Section 3.2.3). There is also an influence of speech speed, e.g. when sentences are short, the pace of pronunciation is slower, and fewer assimilations occur in the flute form. As in spoken modal speech, assimilation may create allegro phonology—for example, an L that may be converted to a downstep (cf. Green and Ígwè 1963: 158; Ígwè 1999: 844; Williamson 1972).

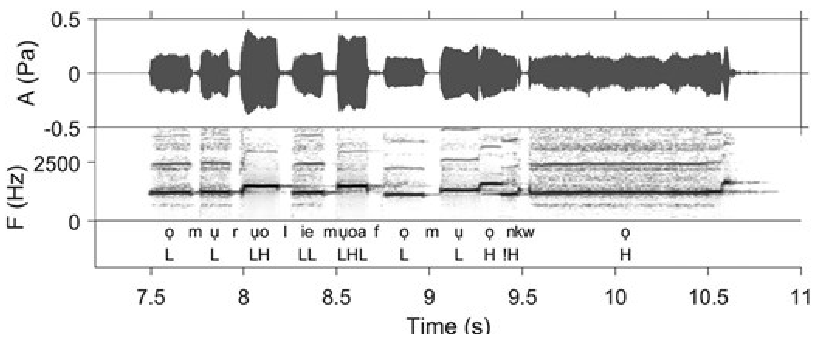

Figure 13 Waveform and spectrogram of the flute speech form of phrase 8, the sentence *ọ̀-mụ-rụ ólìe*, *mụọ́ àfọ*, *mụọ́ ńkwọ* (English literal translation: 'person who gives birth on *ólìe* day, gives birth on *àfọ* day, gives birth on *ńkwọ* day' (meaning 'very fertile woman')). The line with pitch simplified notation—with H, L, !H below the text—refers to the spoken reference.

Other interesting aspects in line with results found in other auxiliary speech forms derive from these observations:

(i) The chanting mode employed with the flute relies on a proverbial style that often enphrases messages with rather long texts (this further illustrates comments in Section 2), while several words also show that the flute is playing a fuller/longer form than what the common speakers produce (e.g. the word for 'child' is played *ńnwa* HH bisyllabic (here pronounced H!H, which is to be expected in the phrase 2 context), a rather archaic and irregular form[11]).

(ii) Consecutive syllable assimilations in *ọ̀ja* provide novel evidence confirming the relevance of structures composed of Vowel-to-Vowel intervals to analyse the complex puzzle concerning features embedded in speech. As shown in Section 3.1.2, such interval types have already been highlighted in drummed auxiliary speech (Dentel and Meyer 2020; Meyer et al. 2012; Seifart et al. 2018), besides being increasingly used to analyse rhythmic units of spoken modal speech (Hirsch 2014; Otterbein et al. 2012; Ryan 2016). Note that previous support for such intervals also comes from phonetic research on tonal alignment (House 1990).

(iii) Third, while most consonants are marked by a silent pause, some can remain continuous (such as the liquid /r/ and the nasal /n/), in line with other results on whistled tonal languages relating to the transposition of formant-like resonant patterns in consonants (Meyer 2015; Rialland 2005). In addition, the *ọ̀ja* flute melody also distinguishes syllabic from non-syllabic nasals more clearly than the spoken form, and thus clearly highlights syllabic nasals of Ìgbo. Here, these syllabic nasals are played as a plain note on the flute, whereas non-syllabic nasals are expressed either with silent pauses or with a short amplitude dip, and eventually a pitch break in the melodic line (see examples for *n* and *m* in Figures 11, 12 and 13, and listen to more audio examples in the corpus for syllabic nasal *m*[12]). *Ọ̀ja* thus offers insights into this important characteristic of the Ìgbo language, giving an example of how the alternative phonetic expression of auxiliary speech forms may contribute to clarify the analysis of some linguistic structures.

The example of *ọ̀ja* surrogacy in Ìgbo finally also illustrates more general problems with attempts to map from 'phonological tones' to pitch targets in non-laryngeal emulation:

(i) The operational concept of 'functional load of tone' is sometimes assumed to be relevant to auxiliary speech studies on tonal language (e.g. Bagemihl 1988) based on the fact that tonal contrasts contribute for a large part to the pitch line resulting from the emulation at play. However, there are several limitations to this approach. First, a functional load quantification requires advanced computing on a large corpus representative of the language use, which is extremely rare in less-described languages (Surendran and Niyogi 2006). Next, traditional phonemic theory (e.g. Pike 1948) extracts 'morphemes' taxonomically—in advance of systemic grammatical analysis—and defaults in practice to establishing minimal pitch contrasts between translated lexical items. In the *ọ̀ja* case, this last shortcut yielded a prima facie implausible inference 'that the functional load of tone in Ìgbo may be higher than that of Yorùbá' (Carter-Ényì et al. 2021: 2), evidently an unintended artefact of also restricting comparison to 'disyllable entries'—effectively, translated nouns—and disregarding confounds inherited from incommensurable lexicographical sources (Crowther 1842/1913, cited as 'Ìbàdàn University 1991' vs Williamson 1972). Finally, several other speech components possibly also influence pitch in speech surrogates of tonal languages, e.g. intonation, consonant coarticulation, or rhythmic cues (see examples in other languages in Foris 2000; Moore and Meyer 2014; Rialland 2005; Seifart et al. 2018).

(ii) Paradigmatic contrasts in natural languages are realized not in the pristine isolation of minimal pairs from which tones are abstracted in structuralist literature, but instead relative to the syntagmatic spans of phrasal prosody (Dilley 2005). This is the frame of downstep juncture, a core property of Ìgbo nominal phrasing and verbal inflection as shown above, whereas it is relatively marginal in other languages such as Yorùbá (Courtenay 1971; Láníran 1992). This difference, and the large but partial manifestation of downstep in the pitch played on *ọ̀ja* flute, extend the current typology of musical surrogacy realization of downstep, showing that it depends not only on acoustic constraints on instruments but also on what kinds of underlying and surface forms are involved in each particular language (Akínbọ̀ 2019; McPherson and James 2021; Meyer and Moore 2021; Nketia 1994; Struthers-Young 2022).[13]

Closer attention to these issues will permit investigation of a question of perennial interest in the literature on instrumental auxiliary speech—whether the emulation target is closer to the raw phonetic output of vocal production or to a more abstract mental representation of sound structure, or more likely to some practical compromise between these idealized extremes, as determined by external performance constraints.

4 Summary

Auxiliary speech practices are the result of the adaptation of human productive and perceptual flexibility to several different environmental, cognitive, and linguistic constraints. These secondary speech codes not only show how much natural language text can be recovered from greatly reduced acoustic signals, but also offer insight into how the sound–meaning correspondences of the primary spoken languages themselves are stored and processed in real time. Altogether, studies of language channelled through different modalities, such as these different auxiliary speech types, shed new light on our general understanding of human language and on its relation to 'music'. They challenge some key concepts of linguistics, such as the phonetics–phonology relation, the notion of prosody (e.g. the purely prosodic melodic line of whistled speech sometimes encodes primarily formants), and the role of the syllable in speech rhythm (to some extent they highlight the role of V-to-V intervals). Crucially, the exploration of this phenomenon and the typological comparison of its expression in a diverse set of sound systems and languages provide a fruitful scientific lens and natural laboratory through which to question classical language descriptions, test linguistic theories, and open new perspectives in language sciences.

The study of musical auxiliary speech was transformed by Rouget's hypothesis that this traditional skill set presents precious evidence for the cognitive status of natural language itself. The same point has been made of other *Kunstsprachen* like Vedic and Homeric poems, developed in non-literate social settings (Kiparsky 1973), and further examples include the onomatopoeia of verbal ideophones (Awóyalé 1981 1989) and the melodic/metric matching of text and song (Éménanjọ 1978b; Zhang and Cross 2021). Each particular case can be expected to combine multiple cognitive connections in indigenous 'theories of mind' in meta-communication and meta-representation (Sperber 2000; Whiten 1991).

Acknowledgements for audiovisual material

Research was conducted in accordance with the Declaration of Helsinki.

In Brazil (for Audios 1, 2 and 3), there was at that time no standing ethics committee at the Museu Goeldi, our host institution in Belém. Ethical questions, if they arose, were addressed by an internal investigative committee called a *sindicância*. In Brazilian law and practice, the participating indigenous community indicates, either orally or in writing, its informed consent to the proposed research to the local office of the National Indian Foundation (FUNAI), which in turn transmits that consent, in the form of a document, to the central FUNAI office in the national capital. This office issues written research permits. Our research followed these established procedures. Native local authorities authorized our work in all of the visited communities. Permits were obtained from the National Indian Foundation (FUNAI) and the National Research Council (CNPq). Copies of the recordings and the research results were given to the community.

In El Hierro Island (Spain) (Audios 4 and 5), informed consent was obtained from the whistlers, and the research was authorized by the Council of the Government of El Hierro (Expte 1997/2021, Canary Islands, Spain).

Videos are used courtesy of Àyàn Làmídì Àyánkúnlé.

Appendix 1

Tone-marking in this manuscript: (a) For the Niger-Congo languages discussed here, diacritics of 'tone' (contrastive F0) are economized by two different conventions according to prosodic type. Binary, e.g. Ìgbo, H [´] and L [`] pitches are written syntagmatically: an unmarked syllable copies the preceding pitch, and successive H marks are cumulatively downstepped (Christaller 1875: 15; Nwáchukwu 1976: 20; Welmers and Welmers 1968). Downstep is an obligatory F0 drop of roughly 10 Hz, phonetically transcribed [ꜜ]. Examples: Ìgbo [LL], ógologo osisi [HHHH HHH] 'tall tree'. ógologo aká [HHHH HꜜH] 'long arm'. Ternary, e.g. Yorùbá, Gbè. Every syllable is paradigmatically either H [´], L [`] or unmarked, with the neutral value M (Akinlabí 1985; Siertsema 1958, p. 583). In Yorùbá, downstep arises at phrasal boundaries and can be marked internal to an orthographic word with the 'full stop' punctuation sign [.] (Bám̄gbóṣé 1966; Láníran 1992). Examples: Yorùbá [MLH], Lání.ran [HHꜜM]), oló.dù [MHꜜL] 'holder of an *òdù* pot'.

For the Amazonian languages: Bora is a two-tone language, i.e. every syllable is associated with either high (H, e.g. *bá*) or low (L, e.g. *bà*) tone, which are expressed by fundamental frequency pitch contours (Seifart et al. 2018). In Gavião of Rondônia, there are contrasts between high (marked by an acute accent), low (unmarked), and rising tones (marked by a circumflex) and also between short and long vowels (transcribed by sequences of two vowels). Also, some long syllables have a floating low tone finally (Moore 1999; Moore and Meyer 2014). In this transcription a grave accent is used on the second vowel to indicate long tones which have final floating low tones. Those syllables with an underlying long syllable with a final floating low tone trigger a fall in the phonetic level of any immediately following high tone to a mid-level. This lowers the register of all following high tones (downstep). If nothing follows the syllable, the floating low tone attaches to the end of the syllable and the result is a falling tone.

References

Abraham, R. (1958), *Dictionary of Modern Yorùbá* (London: University of London Press).
Google Scholar Google Preview WorldCat COPAC

Adédéjì, J. A. (1972), 'The Origin and Form of the Yorùbá Masque Theatre', *Cahiers d'études africaines*, 12/46, 254–276.
https://doi.org/10.3406/cea.1972.2763
Google Scholar WorldCat

Adégbọlá, T. (2003), 'Probabilistically Speaking: A Quantitative Exploration of Yorùbá Speech Surrogacy', in A. Akinlabí and O. Adéṣọlá, eds, *Proceedings of the 4th World Congress of African Linguistics* (Cologne: Rüdiger Köppe), 363–373.
Google Scholar Google Preview WorldCat COPAC

Adéoyè, C. L. (1979), *Àṣà àti ìṣe Yorùbá* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Akínbọ̀, S. (2019), 'Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis', *Linguistique et langues africaines*, 5, 11–23.
Google Scholar WorldCat

Akínbọ̀, S. (2021), 'The Language of Gángan, a Yorùbá Talking Drum', *Frontiers in Communication*, 6, Article 650382.
https://doi.org/10.3389/fcomm.2021.650382
Google Scholar WorldCat

Akinlabí, A. (1985), *Tonal Underspecification and Yorùbá Tone*, doctoral dissertation, University of Ibadan.

Alexandre, P. (1967), *Langues et langage en Afrique noire* (Paris: Payot).
Google Scholar Google Preview WorldCat COPAC

Alvar M. (1975), *Atlas lingüístico y etnográfico de las Islas Canarias* (Las Palmas: Cabildo Insular de Gran Canaria).
Google Scholar Google Preview WorldCat COPAC

Ansre, G. (1961), *The Tonal Structure of Èwè*, MA thesis, Hartford Seminary.
Google Scholar Google Preview WorldCat COPAC

Ányaanwụ́, R.-J. (1998), *Aspects of Ìgbo Grammar: Phonetics, Phonology, Morphology and the Tonology of Nouns* (Münster: Lit Verlag).
Google Scholar Google Preview WorldCat COPAC

Arom, S. (1970), *L'arc musical ngbaka, République de Centrafrique* (Paris: CNRS Comité du Film Ethnographique).
Google Scholar Google Preview WorldCat COPAC

Awóyalé, Y. (1981), 'Nominal Compound Formation in Yorùbá Ideophones', *Journal of African Languages and Linguistics*, 3, 139–157.
Google Scholar WorldCat

Awóyalé, Y. (1989), 'Reduplication and the Status of Ideophones in Yorùbá', *Journal of West African Languages*, 19, 15–34.
Google Scholar WorldCat

Àyánkúnlé, L. (1997), *Ọ̀sẹ̀ bàtá* [Video]. Department of Performing Arts, University of Ìlọrin.
manfredi.mayfirst.org/AyanLamidiSunUnReOOO.html

Babáyẹmí, S. O. (1980), *Egúngún among the Ọ̀yọ́ Yoruba* (Ọ̀yọ́ State Council for Arts and Culture).
Google Scholar Google Preview WorldCat COPAC

Bagemihl, B. (1988), *Alternate Phonologies and Morphologies*, doctoral dissertation, University of British Columbia. https://open.library.ubc.ca/collections/831/items/1.0097954
Google Scholar Google Preview WorldCat COPAC

Bám̄gbóṣé, A. (1966), 'The Assimilated Low Tone in Yorùbá', *Lingua*, 16, 1–13. https://doi.org/10.1016/0024-3841(66)90014-3
Google Scholar WorldCat

Beaudet, J. M. (1997), *Souffles d'Amazonie: les orchestres tule des Wayãpi* (Nanterre: Société d'Ethnologie).
Google Scholar Google Preview WorldCat COPAC

Beier, U. (1970), *Yorùbá Poetry: An Anthology of Traditional Poems* (Cambridge: Cambridge University Press).
Google Scholar Google Preview WorldCat COPAC

Bonhomme, J. (2014), 'Lavoix du mongongo ou comment faire parler un arc musical', *Cahiers d'anthropologie sociale*, 10/1, 93–111. https://doi.org/10.3917/cas.010.0093
Google Scholar WorldCat

Busnel, R.-G. (1967), *Documents sur une langue sifflée turque* [Video]. https://www.canal-u.tv/chaines/cerimes/ethnologie-civilisations/documents-sur-une-langue-sifflee-turque

Busnel, R.-G. (1970), 'Recherches expérimentales sur la langue sifflée de Kusköy', *Revue de phonétique appliquée*, 14–15, 41–75.
Google Scholar WorldCat

Busnel, R.-G. (1974), 'Bio-acoustique de la langue sifflée mazathèque', *Revue d'acoustique*, 29, 94–100.
Google Scholar WorldCat

Busnel, R.-G., Alcuri, G., Gautheron, B., and Rialland, A. (1989), 'Sur quelques aspects physiques de la langue à ton sifflée du peuple H'mong', *Cahiers de l'Asie du Sud-Est*, 26, 39–52.
Google Scholar WorldCat

Busnel, R.-G., and Classe, A. (1976), *Whistled Languages* (New York: Springer).
Google Scholar Google Preview WorldCat COPAC

Busnel, R.-G., Moles, A., and Vallancien, B. (1962), 'Sur l'aspect phonétique d'une langue sifflée des Pyrénées françaises', in A. Sovijärvi and P. Aalto, eds, *Proceedings of the Fourth International Congress of Phonetic Sciences* (The Hague: Mouton), 533–546.
Google Scholar Google Preview WorldCat COPAC

Cañadas, A. S., and Jiménez, M. I. (1996), 'Instrumentos musicales de barro en Andalucía (i): aerófonos', *Música oral del Sur*, 2, 85–186.
Google Scholar WorldCat

Capo, H. B. C. (1991), *A Comparative Phonology of Gbè* (Dordrecht: Foris).
Google Scholar Google Preview WorldCat COPAC

Carrington, J. F. (1949), *Talking Drums of Africa* (Carey Kingsgate Press).
Google Scholar Google Preview WorldCat COPAC

Carter-Ényì, A. (2016), *Contour Levels: An Abstraction of Pitch Space Based on African Tone Systems*, doctoral dissertation, Ohio State University. http://rave.ohiolink.edu/etdc/view?acc_num=osu1461029477
Google Scholar Google Preview WorldCat COPAC

Carter-Ényì, A., Amadi, N. C., Carter-Ényì, Q., Chukwudozie, C., Nwankwo, J., and Omodoro, E. (2021), 'Ìgbo Speech Surrogacy: Preliminary Findings Based on the Ọ̀*ja* Flute', *Frontiers in Psychology*, 12, Article 653068. https://doi.org/10.3389/fpsyg.2021.653068

Catlin, R. C. (1982), 'Speech Surrogate Systems of the Hmong: From Singing Voices to Talking Reeds', in B. T. Downing and D. P. Olney, eds, *The Hmong in the West: Observations and Reports* (Minneapolis: University of Minnesota Press), 170–197.
Google Scholar Google Preview WorldCat COPAC

Charalambakis, C. (1994), 'A Case of Whistled Speech from Greece', *Current Issues in Linguistic Theory*, 117, 389–396.
https://doi.org/10.1075/cilt.117.56cha
Google Scholar WorldCat

Christaller, J. G. (1875), A Grammar of the Asante and Fante Language Called Tshi [chwee, twi] (Basel: Evangelical Missionary Society).

Clark, M. (1980), 'On the Treatment of Syntactically-Distributed Downstep', *Studies in African Linguistics*, 11, 101–137.
Google Scholar WorldCat

Cloarec-Heiss, F. (1999), 'From Natural Language to Drum Language: An Economical Encoding Procedure in Banda-Linda (Central African Republic)', in C. Fuchs and S. Robert, eds, *Language Diversity and Cognitive Representations* (Amsterdam: John Benjamins), 145–157.
Google Scholar Google Preview WorldCat COPAC

Connell, B., and Ladd, D. R. (1990), 'Aspects of Pitch Realisation in Yorùbá', *Phonology*, 71, 1–29.
Google Scholar WorldCat

Cross, I. (2012), 'Music as a Social and Cognitive Process' in P. Rebuschat, M. Rohmeier, J. A. Hawkins, and I. Cross, eds, *Language and Music as Cognitive Systems* (Oxford: Oxford University Press), 315–328.
Google Scholar Google Preview WorldCat COPAC

Courtenay, K. (1971), 'Yorùbá, a "Terraced-Level" Language with Three Tonemes', *Studies in African Linguistics*, 2, 239–255.
Google Scholar WorldCat

Cowan, G. M. (1948), 'Mazateco Whistle Speech', *Language*, 243, 280–286. https://doi.org/10.2307/410362
Google Scholar WorldCat

Cowan, G. M. (1972), 'Segmental Features of Tepehua Whistle Speech', in R. André and C. René, eds, *Proceedings of the Seventh International Congress of Phonetic Sciences* (Berlin: De Gruyter Mouton), 695–698. https://doi.org/10.1515/9783110814750-083
Google Scholar Google Preview WorldCat COPAC

Crowther, A. and Wakeman, C. (1843–1852 1913 and reprints), *A Dictionary of the Yorùbá Language* (Lagos/Oxford: Church Missionary Society).
Google Scholar Google Preview WorldCat COPAC

Dargie, D. (2007), 'Umakhweyane: A Musical Bow and its Contribution to Zulu Music', *African Music*, 8/1, 60–81.
https://doi.org/10.21504/amj.v8i1.1712
Google Scholar WorldCat

Dargie, D. (2011), 'The Xhosa *Umrhubhe* Mouthbow: An Extraordinary Musical Instrument', *African Music*, 91, 33–55.
https://doi.org/10.21504/amj.v9i1.1757

Dentel, L., and Meyer, J. (2020), 'How Rhythm and Timbre Encode Mooré Language in Bendré Drummed Speech', in *Proceedings Interspeech 2020* 1889–1902. https://doi.org/10.21437/Interspeech.2020-2532
Google Scholar WorldCat

Diaz, R. D. (2017), *El Lenguaje Silbado en la Isla de El Hierro*, 2nd edn (Las Palmas: LeCanarien).

Dilley, L. C. (2005), *The Phonetics and Phonology of Tonal Systems*, doctoral dissertation, Massachusetts Institute of Technology.
Google Scholar Google Preview WorldCat COPAC

Dúrójáyé, C., Fink, L., Roeske, T., Wald-Fuhrmann, M., and Larrouy-Maestri, P. (2021), 'Perception of Nigerian Dùndún Talking Drum Performances as Speech-Like vs. Music-Like: The Role of Familiarity and Acoustic Cues', *Frontiers in Psychology*, 12, Article 652673. https://doi.org/10.3389/fpsyg.2021.652673
Google Scholar WorldCat

Eboué, F. (1935), 'La clef musicale des langages tambourines et sifflés: essai de linguistique musicale', *Bulletin du Comité d'études historiques et scientifiques de l'Afrique occidentale française*, 18, 356–360.
Google Scholar WorldCat

Éménanjọ, N. (1978a), *Elements of Modern Ìgbo Grammar* (Ìbàdàn: Ìbàdàn University Press).
Google Scholar Google Preview WorldCat COPAC

Éménanjọ, N. (1978b), 'An Ìgbo Folk-Song: A Linguistic Study', *African Languages/Langues Africaines*, 4, 96–109.
Google Scholar WorldCat

Éménanjọ, N. (2015), *A Grammar of Contemporary Ìgbo: Constituents, Features and Processes* (Lagos: M and J Grand Orbit Communications).
Google Scholar Google Preview WorldCat COPAC

Ermel, P. B. (2004), *L'arc et la lire: proposition d'un narratif visuel et sonore* (Bry-sur-Marne: Collège Iconique de l'Institut National de l'Audiovisuel).
Google Scholar Google Preview WorldCat COPAC

Euba, A. (2011), 'Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code', *Ethnomusicology*, 55/3, 516–520. https://doi.org/10.5406/ethnomusicology.55.3.0516
Google Scholar WorldCat

Everett, D. L. (1983), *A Língua Pirahã e a teoria da sintaxe: descricao, perspectivas e teoria* (Campinas: Universidade Estuadual de Campinas).
Google Scholar Google Preview WorldCat COPAC

Fant, G. (1960), *Acoustic Theory of Speech Production* (The Hague: Mouton).
Google Scholar Google Preview WorldCat COPAC

Fitch, W. T. (2010), *The Evolution of Language* (Cambridge: Cambridge University Press). https://doi.org/10.1017/CBO9780511817779
Google Scholar Google Preview WorldCat COPAC

Foris, D. P. (2000), *A Grammar of Sochiapam Chinantec* (Arlington, TX: SIL International).
Google Scholar Google Preview WorldCat COPAC

Fournel, T. (2002), *Polyphonies Ari*. [Music] Collection Ocora/Radio France.
Google Scholar Google Preview WorldCat COPAC

Franchetto, B., and Montagnani, T. (2011), 'Flûtes des hommes, chants des femmes: images et relations sonores chez les Kuikuro du haut-Xingu', *Gradhiva*, 13, 94–111. https://doi.org/10.4000/gradhiva.2052
Google Scholar WorldCat

Gbètɔ, F. (1995), 'Quelques aspects comparatifs et diachroniques de la tonologie du Gbè', *Afrika und Übersee*, 78/1, 73–99.
Google Scholar WorldCat

Gbètɔ, F. (2020), *Les tons en Gbè, langue Kwa de l'Afrique de l'ouest: étude comparative, typologique and diachronique*, MS, Université d'Abomey-Calavi.

González, M., and Olúdáre, O. (2022), 'The Speech Surrogacy Systems of the Yorùbá Dùndún and Bàtá Drums: On the Interface Between Organology and Phonology', *Frontiers in Communication*, 6, Article 652542. https://doi.org/10.3389/fcomm.2021.652542
Google Scholar WorldCat

Green, M., and Ígwè, G. (1963), *A Descriptive Grammar of Ìgbo* (Akademie Verlag, for Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Halle, M., and Stevens, K. N. (1971), 'A Note on Laryngeal Features', *Quarterly Progress Report of the Research Laboratory of Electronics*, 101, 198–213.
Google Scholar WorldCat

Hill, J. D. (1993), *Keepers of the Sacred Chants: The Poetics of Ritual Power in an Amazonian Society* (Tucson: University of Arizona Press).
Google Scholar Google Preview WorldCat COPAC

Hirsch, A. (2014), 'What Is the Domain for Weight Computation: The Syllable or the Interval?' in J. Kingston, C. Moore-Cantwell, J. Pater, and R. Staubs, eds, *Proceedings of the 2013 Annual Meeting on Phonology* (Linguistics Society of America), 1–12.
Google Scholar Google Preview WorldCat COPAC

Homayounfar, R. (2021), 'Exploring a Bigram System in the Analysis of Ìgbò Ọ̀jà Flute Surrogacy' *[Blog]* https://speechsurrogates.surge.sh/

House, D. (1990), *Tonal Perception in Speech* (Lund: Lund University Press).
Google Scholar Google Preview WorldCat COPAC

Hyman, L. M. (1973), 'The Role of Consonant Types in Natural Tonal Assimilation', in *Consonant Types and Tone*: *Southern California Occasional Papers in Linguistics*, 1, 152–179.
Google Scholar WorldCat

Ígwè, G. (1985/1999), *Ìgbo–English Dictionary* (Ibadan: University Press PLC).
Google Scholar Google Preview WorldCat COPAC

Íkekeọnwụ́, C. (1982), 'Tones in Ìgbo: A Classification', 15th West African Languages Congress, Port Harcourt.

Izikowitz, K. G. (1935), 'Idiophones', in K. G. Izikovitz, ed, *Musical and Other Sound Instruments of the South American Indians: A Comparative Ethnographical Study* (Mölnlycke: Elanders), 7–33.
Google Scholar Google Preview WorldCat COPAC

Jakobson, R. (1960), 'Linguistics and Poetics', in T. A. Sebeok, ed, *Style in Language* (Cambridge, MA: MIT Press), 350–377.
Google Scholar Google Preview WorldCat COPAC

Jakobson, R., and Halle, M. (1956), *Fundamentals of Language* (The Hague: Mouton).
Google Scholar Google Preview WorldCat COPAC

James, L. (2021), 'Systems of Communication: Aspects of Culture and Structure in Speech Surrogates', *Frontiers in Communication,* 6, Article 653268. https://doi.org/10.3389/fcomm.2021.653268
Google Scholar WorldCat

Kawada, J. (1985), *Textes historiques oraux des Mosi méridionaux (Burkina-Faso)* (Tokyo: Institut de recherches sur les langues et les cultures d'Asie et d'Afrique).
Google Scholar Google Preview WorldCat COPAC

Kawada, J. (2014), ‘Notes on the Drum Language of the Ancient Mossi Kingdoms (Burkina Faso)’, *Japanese Review of Cultural Anthropology*, 15, 199–215.
Google Scholar WorldCat

Kawada, J., and Jun-Ichi, O. (1996), *A Study of a Drum Language as a Communication System in an African Society* (Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa).
Google Scholar Google Preview WorldCat COPAC

Kiparsky, P. (1973), ‘Phonological Representations’, in O. Fujimara, ed, *Three Dimensions of Linguistic Theory* (Pontyclun: TEC Co.), 5–86.
Google Scholar Google Preview WorldCat COPAC

Krowles, K. L., Patten, K. J., Durojaye, C., Dada, B. O., and MacBeath, M. K. (2025), ‘Yorùbá Speech Surrogacy with the Dùndún Talking Drum: Recognition Accuracy Is a Function of Language Familiarity and Music Training’, *Music Perception*, 42/4, 287–310. https://doi.org/10.1525/mp.2025.2327968
Google Scholar WorldCat

Labouret, L. (1923), ‘Langage Tambouriné et Sifflé’, *Bulletin du Comité de l’Afrique Occidentale Française*, 6, 120–158.
Google Scholar WorldCat

Láníran, Y. (1992), *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yorùbá*, doctoral dissertation, Cornell University.

Láoyè, T. Ẹ. (1959), ‘Yorùbá Drums’, *Odu: A Journal of Yoruba and Related Studies*, 7, 5–14.
Google Scholar WorldCat

Law, R. (1977), *The Ọ̀yọ́ Empire, c.1600–c.1836: A West African Imperialism in the Era of the Atlantic Slave Trade* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Liberman, M., Schultz, J. M., Hong, S., and Okeke, V. (1993), ‘The Phonetic Interpretation of Tone in Ìgbo’, *Phonetica*, 50/3, 147–160. https://doi.org/10.1159/000261935
Google Scholar WorldCat

Lo-Bamijoko, J. N. (1987), ‘Classification of Ìgbo Musical Instruments, Nigeria’, *African Music*, 6/4, 19–41. https://doi.org/10.21504/amj.v6i4.1259
Google Scholar WorldCat

Locke, D., and Agbeli, G. (1981), ‘Drum Language in Adzogbo’, *The Black Perspective in Music*, 9/1, 25–50. https://doi.org/10.2307/1214103
Google Scholar WorldCat

Manfredi, V. (2018), ‘Cyclic Accentuation in Yorùbá’, in O. Adéṣọlá, A. Akinlabi, and O. O. Orie, eds, *Data-Rich Linguistics: Papers in honor of ‘Yíwọlá Awóyalé’* (Newcastle upon Tyne: Cambridge Scholars), 211–236.

Manfredi, V. (2020), ‘A Toneless Theory of 2-and-a-Half Tonemes in Gbè’, in E.-A. Urua, F. Egbokhare, O. Adesola, and H. Adeniyi, eds, *African Languages in Time and Space: A Festschrift in honour of Professor Akinbiyi Akinlabi* (Ibadan: Zenith BookHouse), 65–97.
Google Scholar Google Preview WorldCat COPAC

Marsh, N. (2013), *The Decline of Speech Surrogacy in Bata Cuban Drumming*, dissertation, University of Auckland.
Google Scholar Google Preview WorldCat COPAC

McPherson, L. (2018), ‘The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications’, *Anthropological Linguistics*, *60/*3, 255–294. https://doi.org/10.1353/anl.2019.0006

Google Scholar WorldCat

McPherson, L. (2019), 'The Role of Music in Documenting Phonological Grammar: Two Case Studies from West Africa', *Proceedings of the 2018 Annual Meeting on Phonology*. https://doi.org/10.3765/amp.v7i0.4493
WorldCat

McPherson, L., and James, L. (2021), 'Artistic Adaptation of Seenku Tone: Musical Surrogates vs. Vocal Music', in A. Akinlabi, L. Bickmore, M. Cahill, M. Diercks, L. J. Downing, J. Essegbey, K. Franich, L. McPherson, and S. Rose, eds, *Celebrating 50 Years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics* (Berlin: Language Science Press), 203–223. https://doi.org/10.5281/zenodo5578834
Google Scholar Google Preview WorldCat COPAC

Meyer, J. (2007), 'Acoustic Features and Perceptive Cues of Songs and Dialogues in Whistled Speech: Convergences with Sung Speech', *Proceedings of the International Symposium on Musical Acoustics* 2007, 1–8. https://doi.org/10.48550/arXiv.0711.3704
Google Scholar WorldCat

Meyer, J. (2015), *Whistled Languages: A Worldwide Inquiry on Human Whistled Speech* (New York: Springer).
Google Scholar Google Preview WorldCat COPAC

Meyer, J. (2020), 'Coding Human Languages for Long-Range Communication in Natural Ecological Environments: Shouting, Whistling, and Drumming', in T. Aubin and N. Mathevon, eds, *Coding Strategies in Vertebrate Acoustic Communication* (New York: Springer International), 91–113. https://doi.org/10.1007/978-3-030-39200-0_4
Google Scholar Google Preview WorldCat COPAC

Meyer, J. (2021), 'Environmental and Linguistic Typology of Whistled Languages', *Annual Review of Linguistics*, 7/1, 493–510. https://doi.org/10.1146/annurev-linguistics-011619-030444
Google Scholar WorldCat

Meyer, J., Dentel, L., and Seifart, F. (2012), 'A Methodology for the Study of Rhythm in Drummed Forms of Languages: Application to Bora Manguaré of Amazon', *Proceedings of Interspeech 2012*, 687–690. https://doi.org/10.21437/Interspeech.2012-210
Google Scholar WorldCat

Meyer, J., and Moore, D. (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, Article 674289. https://doi.org/10.3389/fpsyg.2021.674289
Google Scholar WorldCat

Mindlin, B., Tsoraba, D., and Sebirop, C. (2001), *Couro dos espíritos: namoro, pajés e cura entre os índios Gavião-Ikolen de Rondônia* (São Paulo: Terceiro Nome).
Google Scholar Google Preview WorldCat COPAC

Moles, A. (1970), 'Étude sociolinguistique de la langue sifflée de Kuskoy', *Revue de phonétique appliquée*, 14/15, 78–118.
Google Scholar WorldCat

Moore, D. (1999), 'Tonal System of the Gavião Language of Rondônia, Brazil', in K. Shigeki, ed, *Tupian Perspective: Cross-linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics* (Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa), 297–310.
Google Scholar Google Preview WorldCat COPAC

Moore, D., and Meyer, J. (2014), 'The Study of Tone and Related Phenomena in an Amazonian Tone Language: Gavião of Rondônia', *Language Documentation and Conservation*, 8, 613–636.
Google Scholar WorldCat

Neeley, P. (1999), *People of the Drum of God—Come!* (Tucson, AZ: SIL International).
Google Scholar Google Preview WorldCat COPAC

N'Guessan, K. (2020), *Drummologie: la connaissance aux sons du tambour et du balafon* (Éditions Jets d'encre).
Google Scholar Google Preview WorldCat COPAC

Niangoran-Bouah, G. (1981), *Introduction à la drummologie* (Abidjan: Nationale de Côte d'Ivoire, Institut d'Ethno-sociologie).
Google Scholar Google Preview WorldCat COPAC

Nigerian Educational Research Development Council (1991), *Quadrilingual Glossary of Legislative Terms (English, Hausa, Igbo, Yoruba),* ed. A. Bán̄jọ (Lagos: Spectrum Books).
Google Scholar Google Preview WorldCat COPAC

Nikolsky, A. (2020), '"Talking Jew's Harp" and Its Relation to Vowel Harmony as a Paradigm of Formative Influence of Music on Language', in N. Masataka, ed, *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism* (New York: Springer), 217–322.
Google Scholar Google Preview WorldCat COPAC

Niles, D. (2010), 'Editor's Introduction to Zemp and Kaufmann', *Kulele 4: Occasional Papers on Pacific Music and Dance*, 1–50.
Google Scholar WorldCat

Nissenbaum, J., Hillman, R. E., Kobler, J. B., Curtin, H. D., Halle, M., and Kirsch, J. E. (2002), 'High-Speed MRI of Laryngeal Gestures during Speech Production', *Journal of the Acoustical Society of America*, 111/5_Supplement 2479–2480.
https://doi.org/10.1121/1.4778611
Google Scholar WorldCat

Nketia, J. H. K. (1963), *Drumming in Akan Communities of Ghana* (Edinburgh: Thomas Nelson).
Google Scholar Google Preview WorldCat COPAC

Nketia, J. H. K. (1971), 'Surrogate Languages of Africa', in T. A. Sebeok and D. J. Umiker-Sebeok, eds, *Speech Surrogates: Drum and Whistle Systems* (The Hague: Mouton), 825–864.
Google Scholar Google Preview WorldCat COPAC

Nketia, J. H. K. (1994), 'Generative Processes in Seperawa Music', in E. C. Leichtman, ed, *To the Four Corners: A Festschrift in honor of Rose Brandel* (Sterling Heights, MI: Harmonie Park Press), 127–151.
Google Scholar Google Preview WorldCat COPAC

Nwáchuku, M. A. (1997), 'On the Resonant Frequencies of the Oja', *Nigerian Journal of Technology*, 18/1, 1–21.
Google Scholar WorldCat

Nwáchukwu, P. A. (1976), *Noun Phrase Sentential Complementation in Ìgbo*, doctoral dissertation, University of London.

Ohala, J. J. (1984), 'An Ethological Perspective on Common Cross-language Utilization of f_0 of Voice', *Phonetica*, *41*/1, 1–16.
https://doi.org/10.1159/000261706
Google Scholar WorldCat

Otterbein, S., Abel, C., Heinemann, L. V., Kaiser, J., and Schmidt-Kassow, M. (2012), 'P3b Reflects Periodicity in Linguistic Sequences', *PLOS One*, 7, Article e51419. https://doi.org/10.1371/journal.pone.0051419
Google Scholar WorldCat

Oyèélámì, M. (1991), *Yorùbá Bàtá Music: A New Notation with Basic Exercises and Ensemble Pieces* (Bayreuth: Ìwàlẹwà Haus).
Google Scholar Google Preview WorldCat COPAC

Pacere, T. (1987), *Bendrologie et littérature culturelle* (Burkina Faso: Pacéré Titinga).
Google Scholar Google Preview WorldCat COPAC

Pepper, H. (1950), 'Musique Centre-Africaine', *Encyclopédie Coloniale et Maritime* (Paris).

Pepper, H. (1954), *Considérations sur le langage tambouriné et autres langages musicaux d'Afrique centrale, sur la pensée musicale Africaine* (Bioko: CIAO, Fernando Pô Guinée Espagnole).
Google Scholar Google Preview WorldCat COPAC

Persoon, G. A., and Schefold, R. (2017), 'Frightened by the Eagle: Recording Songs and Music from the Island of Siberut, Mentawai Islands', *Wacana*, 18/3, 581–613. https://doi.org/10.17510/wacana.v18i3.629
Google Scholar WorldCat

Pike, K. (1948), *Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion* (Ann Arbor: University of Michigan Press).
Google Scholar Google Preview WorldCat COPAC

Poss, N. F. (2005), *The Communication of Verbal Content on the Hmong Raj: An Ethnographic Analysis of Performance Practice*, thesis, Ohio State University.

Poss, N. F. (2012), *Hmong Music and Language Cognition: An Interdisciplinary Investigation*, dissertation, Ohio State University.

Rialland, A. (2005), 'Phonological and Phonetic Aspects of Whistled Languages', *Phonology*, 22/2, 237–271. https://doi.org/10.1017/S0952675705000552
Google Scholar WorldCat

Ridouane, R., Turco, G., and Meyer, J. (2018), 'Length Contrast and Covarying Features: Whistled Speech as a Case Study', *Proceedings of Interspeech* 2018 1843–1847. https://doi.org/10.21437/Interspeech.2018-1060
Google Scholar WorldCat

Risset, J.-C. (1968), 'Sur certains aspects fonctionnels de l'audition', *Annales des Télécommunications*, 23/3, 91–120. https://doi.org/10.1007/BF02996236
Google Scholar WorldCat

Ritzenthaler, R. E., and Peterson, F. A. (1954), 'Courtship Whistling of the Mexican Kickapoo Indians', *American Anthropologist*, 56/6 1088–1089. https://doi.org/10.1525/aa.1954.56.6.02a00110
Google Scholar WorldCat

Ros, S. (2021), 'Rhythm–Speech Correlations in a Corpus of Senegalese Drum Language', *Frontiers in Communication*, 6, Article 643683. https://doi.org/10.3389/fcomm.2021.643683
Google Scholar WorldCat

Rouget, G. (1964), 'Tons de la langue, en gùn (Dahomey), et tons du tambour', *Revue de musicologie*, 50/128, 3–29. https://doi.org/10.2307/927529
Google Scholar WorldCat

Rouget, G. (1965), 'Notes et documents pour servir à l'étude de la musique Yorùbá', *Journal des Africanistes*, 35/1, 67–108.
Google Scholar WorldCat

Rouget, G. (1975), 'Tons, intonation, accentuation; problèmes d'identification', in G. Manessy and J. Thomas, eds, *Langues sans tradition écrite: méthodes d'enquête et de description* (Kampala: SELAF, 217–240.
Google Scholar Google Preview WorldCat COPAC

Ryan, K. M. (2016), 'Phonological Weight', *Language and Linguistics Compass*, 10, 720–733. https://doi.org/10.1111/lnc3.12229
Google Scholar WorldCat

Schefold, R. (1973), 'Schlitztrommeln und Trommelsprache in Mentawai', *Zeitschrift für Ethnologie*, 98/1, 36–73.
Google Scholar WorldCat

Schneider, M. (1976), 'Zur Trommelsprache der Duala', in T. Sebeok and D. J. Umiker-Sebeok, eds, *Speech Surrogates: Drum and*

Whistle Systems (The Hague: Mouton), 669–678.
Google Scholar Google Preview WorldCat COPAC

Sebeok, T. A., and Umiker-Sebeok, D. J. (1976), *Speech Surrogates: Drum and Whistle Systems* (Berlin: De Gruyter Mouton).
Google Scholar Google Preview WorldCat COPAC

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', *Royal Society Open Science*, 5/4, Article 170354.
https://doi.org/10.1098/rsos.170354
Google Scholar WorldCat

Shannon, C. E. (1948), 'A Mathematical Theory of Communication', *Bell System Technical Journal*, 27/3, 379–423.
https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Google Scholar WorldCat

Sicoli, M. A. (2016), 'Repair Organization in Chinantec Whistled Speech', *Language*, 92/2, 411–432.
Google Scholar WorldCat

Siertsema, B. (1958), 'Some Notes on Yorùbá Phonetics and Spelling', *Bulletin de l'Institut Fondamental d'Afrique Noire, Série B: Sciences humaines*, 20/3, 576–592.
Google Scholar WorldCat

Smith, N. (1968), 'Tone in Èwè', in *Quarterly Progress Report* 88 (Cambridge, MA: Research Laboratory in Electronics, MIT).

Snyders, J., and Haudricourt, A.-G. (1968), 'Le langage par tambours à San Cristoval, British Solomon Islands', *Journal de la Société des Océanistes*, 24, 133–138.
Google Scholar WorldCat

Sperber, D., ed. (2000), *Metarepresentations: A Multidisciplinary Perspective* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Sprigge, R. G. S. (1967), 'Collected Field Reports on Tone in the Àdángbè Dialect of Èwè', *Collected Language Notes 8* (Accra: Institute of African Studies, University of Ghana).
Google Scholar Google Preview WorldCat COPAC

Stahlke, H. F. W. (1971), *Topics in Èwè Phonology*, doctoral dissertation, University of California Los Angeles.

Stern, T. (1957), 'Drum and Whistle "Languages": An Analysis of Speech Surrogates', *American Anthropologist*, 59/3, 487–506.
https://doi.org/10.1525/aa.1957.59.3.02a00070
Google Scholar WorldCat

Stevens, K. N. (1998), *Acoustic Phonetics* (Cambridge, MA: MIT Press).
Google Scholar Google Preview WorldCat COPAC

Stewart, J. M. (1964), *The Typology of the Twì Tone System* (Accra: Institute of African Studies, University of Ghana).
Google Scholar Google Preview WorldCat COPAC

Struthers-Young, A. (2022), 'A Preliminary Account of the Northern Toussian Balafon Surrogate Language', *Frontiers in Communication*, 6, Article 652740. https://doi.org/10.3389/fcomm.2021.652740
Google Scholar WorldCat

Surendran, D., and Niyogi, P. (2006), 'Quantifying the Functional Load of Phonemic Oppositions, Distinctive Features, and Suprasegmentals', in O. N. Thomsen, ed, *Competing Models of Linguistic Change: Evolution and Beyond* (Amsterdam: John Benjamins), 43–58.
Google Scholar Google Preview WorldCat COPAC

Thieme, D. L. (1969), *A Descriptive Catalogue of Yorùbá Musical Instruments*, doctoral dissertation, Catholic University of America.
Google Scholar Google Preview WorldCat COPAC

Thiesen, W. (1969), 'The Bora Signal Drums', *Lore*, 19, 101–103.
Google Scholar WorldCat

Tran Ngoc, A., Meunier, F., and Meyer, J. (2022), 'Testing Perceptual Flexibility in Speech through the Categorization of Whistled Spanish Consonants by French Speakers', *JASA Express Letters*, 2/9, Article 095201. https://doi.org/10.1121/10.0013900
Google Scholar WorldCat

Trujillo, R. (1978), *El silbo gomero: análisis lingüístico* (Las Palmas: Editorial Interinsular Canaria).
Google Scholar Google Preview WorldCat COPAC

Ùgoṅnà, N. (1984), *Mmọnwụ: A Dramatic Tradition of the Ìgbo* (Lagos: Lagos University Press).
Google Scholar Google Preview WorldCat COPAC

van der Voort, H. (2016), 'Aikanã and Kwaza: Their Ethno-historical and Sociolinguistic Context in Rondônia, Brazil', in L. Filipović and M. Pütz, eds, *Endangered Languages and Languages in Danger: Issues of Documentation, Policy, and Language Rights* (Amsterdam: John Benjamins), 203–230. https://doi.org/10.1075/impact.42.09van
Google Scholar Google Preview WorldCat COPAC

Villepastour, A. (2010), *Ancient Text Messages of the Yorùbá Bàtá Drum: Cracking the Code* (Abingdon: Routledge).
Google Scholar Google Preview WorldCat COPAC

von Hornbostel, E. M., and Sachs, C. (1914), 'Systematik der Musikinstrumente: Ein Versuch', *Zeitschrift für Ethnologie*, 46/4–5, 553–590.
Google Scholar WorldCat

Voorhis, P. H. (1971), 'Notes on Kickapoo Whistle Speech', *International Journal of American Linguistics*, 37, 238–243.
Google Scholar WorldCat

Wedekind, K. (1983), 'A Six-Tone Language in Ethiopia: Tonal Analysis of Benchnon', *Journal of Ethiopian Studies*, 16, 129–156.
Google Scholar WorldCat

Welmers, B. F., and Welmers, W. E. (1968), *Ìgbo: A Learner's Dictionary* (Berkeley: African Studies Center, University of California).
Google Scholar Google Preview WorldCat COPAC

Welmers, W. (1970), 'Ìgbo Tonology', *Studies in African Linguistics*, 1, 255–278.
Google Scholar WorldCat

Whiffen, T. (1915), *The North-West Amazons: Notes of Some Months Spent among Cannibal Tribes* (New York: Duffield and Co.).
Google Scholar Google Preview WorldCat COPAC

Whiten, A., ed (1991), *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading* (Oxford: Blackwell).
Google Scholar Google Preview WorldCat COPAC

Williamson, K. (1972), *Dictionary of Ònìchà Igbo* (Addis Ababa: Ethiope Press).
Google Scholar Google Preview WorldCat COPAC

Winter, Y. (2014), 'On the Grammar of a Senegalese Drum *Language*', *Language*, 90/3, 644–668.
https://doi.org/10.1353/lan.2014.0061

Wojtylak, K. (2019), 'Traversing Language Barriers: "Witoto" Signal Drums from Northwest Amazonia', *International Journal of*

Language and Culture, 6/1, 196–217. https://doi.org/10.1075/ijolc.00022.woj
Google Scholar WorldCat

Yegnan-Touré, A. (2008), ‘La technique et le jeu de l’arc musical’, *Cahiers d’ethnomusicologie*, 21, 205–223.
Google Scholar WorldCat

Zemp, H. (2005), *Siaka, an African Musician* [DVD]. Documentary Educational Resources. www.der.org/siaka-an-african-musician

Zemp, H., and Kaufmann, C. (1969), ‘Pour une transcription automatique des “langages tambourinés” mélanésiens (un exemple Kwoma, Nouvelle-Guinée)’, *L’Homme*, 9/2, 38–88.
Google Scholar WorldCat

Zemp, H., and Soro, S. (2004), ‘Paroles de balafon’, *L’Homme*, *17*, 313–331. https://doi.org/10.4000/lhomme.24930
Google Scholar WorldCat

Zhang, X., and Cross, I. (2021), ‘Analysing the Relationship between Tone and Melody in Chaozhou Songs’, *Journal of New Music Research*, 50/4, 299–311. https://doi.org/10.1080/09298215.2021.1974490
Google Scholar WorldCat

Notes

1 Excluded from consideration here are speech enhancements that keep primary acoustic features intact: traditional voice modifiers in which spoken and sung texts are still produced with vocal cords and are further transformed by the use of external instruments—e.g. gourd or wooden instruments and coconut or conch shells (see Niles 2010). Excluded also are secondary representations like Morse code or other ‘lexical ideographs’ that lack phonetic resemblance to spoken language (Stern 1957; Zemp and Kaufmann 1969).

2 The tonemarking conventions adopted in this chapter are given in Appendix 1.

3 Each local example is identified with the reel number of its audiotape in the UCLA archive catalogeth.library.ucla.edu/vwebv/holdingsInfo?bibId=11178: *dùn-dún*: Ẹdẹ 021, 024, 025, 026, Ìbàdàn 029, Ilé-Ifẹ̀ 044, 046, Isundunrin (diacritics unknown] 067, 070, 072, Ìwó 078, 080, Èkó (Lagos) 088, 089, 090; *gán-gan/àpàlà/àdàmọ̀*: Adó Àwáyè 017, Ilé-Ifẹ̀ 046, 062, Isundunrin (diacritics unknown) 061, 066, 067, 072; kẹríkẹrì:Abẹ́òkúta 007; *kósó*: Èjìgbò 061, 027; Adó Àwáyè 010; *bàtá*: Ìbàdàn 029, 031, Ìwó 074, Èkó (Lagos] 088; *ìgbìn*: Isundunrin (diacritics unknown) 064; *ìgbésẹ̀*: Ilé-Ifẹ̀ 047; *lukorigi/lukoogi* or *àgẹ̀rẹ̀*: Ìgàngàn 053, 061; *ọmọ owú*: Isundunrin (diacritics unknown) 068; *molo* (diacritics unknown) or *àgídìgbo*: Èjìgbò 027, Ìsẹ́yìn 058, Ìwó 080, Mowolowo (diacritics unknown) 087; *ìpè*: Ilé-Ifẹ̀ 043, 045, 049; ekùtù: Isundunrin (diacritics unknown) 062: Isundunrin (diacritics unknown) 062.

4 https://people.bu.edu/manfredi/Lamidi2.mp4.

5 https://people.bu.edu/manfredi/Lamidi1.mp4.

6 https://www.youtube.com/watch?v=H5Kzk4YJuCY (from the Africana Digital Ethnography Project, www.africanadept.org).

7 Given that the tone transcriptions of the corpus in Carter-Ényì et al. (2021) are garbled overall to the point of being unusable, we analysed this excellent set *dē novō*.

8 Note that initial and/or final syllables behave often differently than the rest of the phrase in several other auxiliary speech forms such as those based on whistling (Meyer 2015; Sicoli 2016) and drumming (see also Kawada and Jun-Ichi 1996, or Dentel and Meyer 2020, and corresponding Figure 4 in this chapter), either for phatic communication or to simply parse the message.

9 Depending on the abstractness of the framework, downstep is either predominantly or entirely a phenomenon of syntactic phrasing, and no credible description of Ìgbo from Green and Ígwè (1963) has failed to note its fundamental importance (albeit under the misleading label of ‘mid tone’; (Ányaanwụ́ 1998; Íkekeọnwụ́ 1982); the label ‘mid tone’ was

indeed used in the pioneering formal Ìgbo grammar but the concept of downstep was present in the rule that “a tone following a mid-tone on the same [pitch] level is a high tone”; Green and Ígwè 1963: 6f.).

10 We are constrained to note that the error of Carter-Ényì et al. (2021) is in describing the linguistic form of this text as “all high tone level” (Carter-Ényì et al. 2021: 3), perhaps influenced by the special flat flute line which here omits to mark downsteps with frequency. Note also that simple all-H sequences in Ìgbo actually rise in pitch, with slope inversely proportional to the length of the span (Liberman et al. 1993: 154) and the same is true in Yorùbá (Connell and Ladd 1990; Láníran 1992).

11 At least not the most common form found in the dialect base of Williamson’s (1972) Ìgbo dictionary.

12 For example, the ninth sentence of this corpus (https://www.youtube.com/watch?v=H5Kzk4YJuCY)—available between 2min36s and 2min45s—ends with /mma/ (HH bisyllabic and here played H!H, where the first m is syllabic and the second is not and thus marked by a silence between the two Highs).

13 Other examples of phrasal downstep in instrumental speech surrogacy include Serepewa lute in Akan (Nketia 1994), Gavião instrumental speech forms (Meyer and Moore 2021), Yorùbá dùn-dún (Akínbọ̀ 2019), Northern Toussain balafon (Struthers-Young 2022) and balafon forms of sung Seenku (McPherson and James 2021).