



**HAL**  
open science

# Auxiliary speech by vocal tract modulation and musical surrogacy

Julien Meyer, Victor Manfredi

► **To cite this version:**

Julien Meyer, Victor Manfredi. Auxiliary speech by vocal tract modulation and musical surrogacy. Sammler, D. (ed.). Oxford Handbook of Language and Music, Oxford University Press, In press. hal-04389746v3

**HAL Id: hal-04389746**

**<https://hal.science/hal-04389746v3>**

Submitted on 5 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## **Auxiliary speech by vocal tract modulation and musical surrogacy**

Julien Meyer and Victor Manfredi

Julien Meyer (Université Grenoble Alpes, CNRS, Gipsa-lab, France)  
julien.meyer@cnrs.fr

Victor Manfredi (Boston University, USA)  
manfredi@bu.edu

### **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 & Umiker-Sebeok 1976). Auxiliary speech phenomena attest nonliterate 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, (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.*

## 24.1 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 type. 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<sup>1</sup> 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***—with a large panel of musical instruments + whistled forms (Labouret 1923, Carrington 1949, Pepper 1950, 1954, Nketia 1963, 1971, Sebeok & Umiker-Sebeok 1976, Locke & Agbeli 1981, Zemp & Soro 2004, Kawada 2014, Winter 2014, McPherson 2018).

***East and southern Africa***— with some types of musical instruments + whistled forms (Wedekind 1983, Fournel 2002, Dargie 2007, 2011).

***Europe and North Africa***—whistled forms (Busnel et al 1962, Busnel 1967, 1970, Cañadas & Jimenez 1992, Charalambakis 1994, Meyer et al. 2015, Ridouane et al 2018)

***Central and North America***—whistled forms (Cowan 1948, 1972, Ritzenhaler & Peterson 1954, Voorhis 1971, Busnel 1974, Meyer 2015, Sicoli 2016).

***South American Amazon***— with several types of musical instruments + whistled forms (Whiffen 1915, Izikovitz 1935, Thiesen 1969, Everett 1983, Hill 1993, Beudet 1997, Mindlin et al. 2001, Ermel 2004, Franchetto & Montagnani 2011, Moore & Meyer 2014, van der Voort 2016).

***Southeast Asia***—with a large panel of musical instruments + whistled forms (Stern 1957, Catlin 1982, Busnel et al 1989, Poss 2005, 2012, Meyer 2007)

***Oceania***— with several types of musical instruments + whistled forms (Snyders and Haudricourt 1968, Schefold 1973, Niles 2010, Persoon and Schefold 2017).

---

<sup>1</sup> 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, 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 & Kaufmann 1969).



**Figure 24.1. 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 24.3 and 24.6). Upper row, right: flute used among the Gavião of Rondônia (Brasil) to emulate song lyrics (see also Figures 24.2 and 24.10). 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 24.9). 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 24.7). Lower row, right: Yorùbá dùn-dún ‘talking drum’ used to recite proverbs and pass messages to a large audience in Nigeria. (Photos: Julien Meyer/Laure Dentel. All Rights Reserved).

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 (Nketia 1971, Busnel & Classe 1976, Winter 2014, Sicoli 2016, Ridouane et al 2018, Seifart et al 2018, Dentel & Meyer 2020, Meyer & Moore 2021, James 2021). 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 this way and the degree of precision with which they are rendered by each auxiliary practice depends on various criteria. These include the type of signal used to modify/emulate speech, their production mode, the grammar of the base language, 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 24.3.2.3). The next sections of this article 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 24.2); and next, detailed case studies from languages of Africa, Europe, and Latin America (Section 24.3).

## 24.2 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 24.1). 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, Cowan 1972, Bonhomme 2014, Busnel 1974, Busnel et al 1989, Meyer 2007, 2015, Yegnan-Touré 2008, Niles 2010, Moore & Meyer 2014; Nikolsky 2020, Meyer & Moore 2021).

**Full musical surrogacy.** A non-somatic sound source is coopted 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 (Eboué 1935, Ritzenthaler & Peterson 1954, Pepper 1950, 54, N’Guessan, 2020; Niagoran-Bouah 1980, Catlin 1982, Pacere 1987, Hill 1993, Zemp & Soro 2004; Zemp 2005, Macpherson 2018, N’Guessan 2020, Meyer & Moore 2021, Akínbò 2019, 2021), or a chordophone like the Ethiopian *kar* guitar or the Akan *serepewa* lute (Wedekind 1983, Nketia 1994).

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 esthetics 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 (Ohala 1984, Fitch 2010), 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 of 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 (Sebeok & Umiker-Sebeok 1976, Bagemihl 1988, Seifart et al 2018), although exceptions exist. Indeed, a few drummed practices concern ‘non-tonal’ languages (e.g., Snyders and Haudricourt 1968, Schefold 1973, Winter 2014, Wojtylak 2019), whereas orally modulated jew’s harp practices are rather evenly distributed between ‘tonal’ and ‘non tonal’ languages (Nikolsky 2020) (just as 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 24.2 and 24.3),

whereas several types of drums such as Bendré skin drums, are based on unpitched sounds (Kawada 2014, and see Figure 24.4).

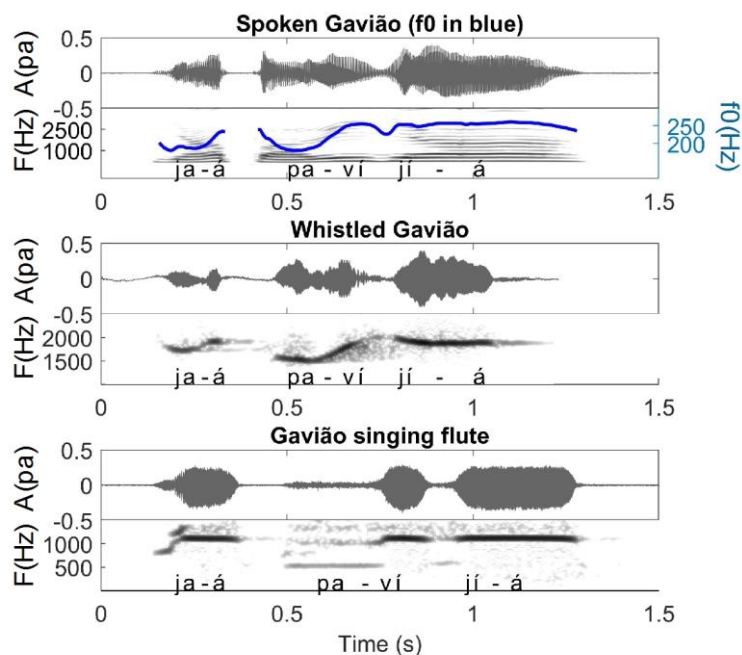
The choice of pitch versus timbre as primary target in the vocal-to-auxiliary transformation is thus mediated by: (1) acoustic limitations of the auxiliary production mode (as just mentioned); but also (2) the role of intonation in the language (for instance, ‘tonal’ languages clearly deploy pitch for lexical meaning more than ‘non-tonal’ ones); and (3) the type of speech that is emulated (sung vs. spoken) because song lyrics emphasise 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 signaled 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 external 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 1952, see Figure 24.5) and the Yorùbá lead *bàtá* drum (Section 24.3.2.2 below).

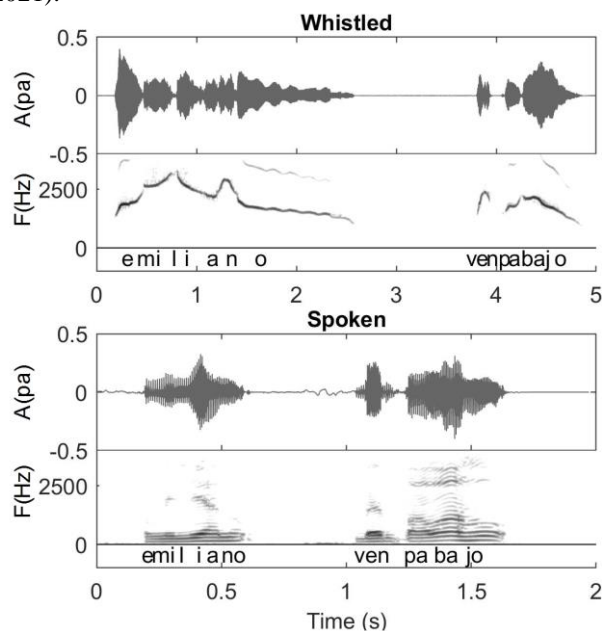
***Vocal pitch => auxiliary timbre.*** Spoken pitch is emulated by varying the timbre of the auxiliary technique. Here, the spectral distribution of the auxiliary timbre mimics the changes of spoken pitch. For example, this occurs when spoken F0 contours in ‘tonal’ languages are imitated with a jew’s-harp by modifying the oral mouth shape (Poss 2012, Nikolsky 2020), or else evoked external to the body by contrasting drum attacks as in Gùn-gbè and Mooré traditions (Rouget 1964, Section 24.3.2.1 below; see also Kawada 2014, Dentel and Meyer 2020, and an example in Figure 24.4).

***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, 2021, and see explained examples in Figures 24.3 and 24.6).

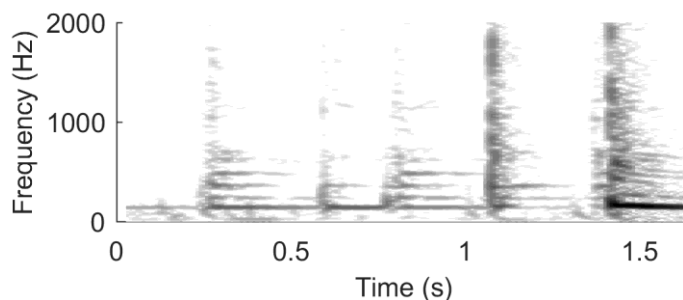
***Vocal pitch => auxiliary pitch.*** This is the most widespread strategy that consists in a direct prosodic correspondence. It is found when pitches produced by musical instruments emulate speech prosody encoded in spoken pitch (e.g., Nketia 1971, Hill 1993, van der Voort 2016, Seifart et al. 2018, Meyer and Moore 2021, and see Figure 24.2, lower row), but also when F0 contours of a ‘tonal’ language are imitated by whistling (Cowan 1948, Moore & Meyer 2014, Figure 24.2, middle row).



**Figure 24.2.** Waveforms and spectrograms of the sentence “*jaá pavíjía*” (meaning “Let’s go to bathe” in the Gavião language of Rondônia, Brasil) uttered in spoken speech (upper row, Audio 24.1), in whistled speech with lip protrusion (middle row, Audio 24.2), and with the flute as in traditional flute songs (lower row, Audio 24.3) (listen also to sound extracts in <https://soundcloud.com/user-28976943/sets/sounds-in-meyer-manfredi-2023>). The blue 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).

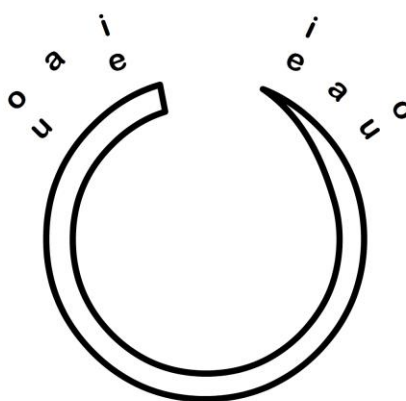


**Figure 24.3.** Waveform and spectrogram of the whistled (upper row) and spoken (lower row) forms of the Spanish sentence “*Emiliano ven pa’ bajo*” [emiliano ben pa βaxo] meaning “Emiliano come down”, where “*pa’ bajo*” is the fast speech fusion of “para bajo”, i.e., an allegro version. The x-axis of 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 is mirroring 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 for example the frequency of /a/ in /ba/ higher than the /a/ in /pa/). Listen to sound extracts in Audio 24.4 or <https://soundcloud.com/user-28976943/sets/sounds-in-meyer-manfredi-2023>.



**Figure 24.4.** Spectrogram of a sentence drummed on the single-membrane Bendré drum of the Mossi people of Burkina Faso (the sentence is “*wënd tēntumd dōlle*”, see Kawada (1985, 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 rather played with an emphasis on low partials to mark the end of the sentence, as often in this tradition (see Kawada 1996). The 1<sup>st</sup> and the 3<sup>rd</sup> 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 2<sup>nd</sup> strike makes an emphasis on low partials and encodes/emulates a Low tone [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 1952: 672, Figure 24.5 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 emphasise 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 24.3 and 24.6).



**Figure 24.5.** 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’) [Adapted from Schneider 1952, 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 1987). 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 (Sicoli 2016, Diaz 2017). 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, Trujillo 1978, Rialland 2005, Diaz 2017, Meyer et al. 2021).

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 a 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, Nikolsky 2020, Meyer & Moore 2021) but also thanks to fine acoustic microstructures (Dúrójayé 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 (Nketia 1963, Thiessen 1969, Meyer 2015, Seifart et al. 2018, MacPherson 2019, Nikolsky 2020, Meyer & Moore 2021).

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 24.9). 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 (Kawada 1985, 2014, Dentel and Meyer 2020), whereas others intersperse syllables without meaning or musical series of notes with no corresponding vocables (Meyer and Moore 2021, Dúrójayé et al. 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, Babáyemí 1980, Ùgonná 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 intensional definition (Cross 2012) — and default instead to metonyms like ‘song’ (Yorùbá *orin*, Ìgbo *ábù*), ‘dance’ (Yorùbá *ijó*), ‘drum’ (Yorùbá *ilù*, Ìgbo *ńkwa*) or ‘play’ (Ìgbo *égwu*), cf. Bǎnjo (1991, 164). These adjacent concepts illuminate the relationships between music and other domains of thought and behavior by referring to circumstances of use, including language.

## 24.3 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 tonemarking in this part of the manuscript is detailed in a footnote<sup>2</sup>.

### 24.3.1 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 meters—is extendable by whistling to a kilometer or more in a favorable environment (Busnel & Classe 1976, Meyer 2015, 2020) and by drumming to several kilometers (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).

#### 24.3.1.1 Whistled speech in Spanish (Canary Islands) and Gavião of Rondônia (Brasil)

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

---

#### <sup>2</sup> Tonemarking in this manuscript

*For the Niger-Congo languages* discussed here, diacritics of ‘tone’ (contrastive F0) are economised by two different conventions according to prosodic type: Binary, e.g., Ìgbo. H [ˈ] and L [ˉ] pitches are written syntagmatically: an unmarked syllable copies the preceding pitch, and successive H marks are cumulatively downstepped (Christaller 1875, 15, Welmers & Welmers 1968, Nwáchukwu 1976, 20). Downstep is an obligatory F0 drop of roughly 10 Hz, phonetically transcribed [!]. Examples: Ìgbo [LL], ógologo osisi [HHHH HHH] ‘tall tree’. ógologo aká [HHHH H!H] ‘long arm’. Ternary, e.g. Yorùbá, Gbè. Every syllable is paradigmatically either H [ˈ], L [ˉ] or unmarked, with the neutral value M (Siertsema 1958, 583, Akinlabí 1985). In Yorùbá, downstep arises at phrasal boundaries and can be marked internal to an orthographic word with the ‘full stop’ punctuation sign [.] (Bámgbòsé 1966, Lání.ran 1992). Examples: Yorùbá [MLH], Lání.ran [HH!M]), oló.dù [MH!L] ‘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).

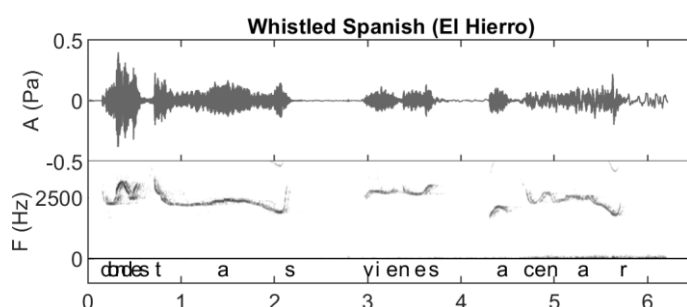
In Gavião of Rondônia, there are contrasts between high (marked by an acute accent), low (unmarked) and rising tones (marked by a circumflex) and also between short and long vowels (transcribed by sequences of two vowels). Also, some long syllables have a floating low tone finally (Moore 1999, Moore 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.

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 Gavião language on Figure 24.2 above), otherwise the target is vocal timbre (aforementioned category *vocal timbre* => *auxiliary pitch*; one example is Spanish, see Figures 24.3 and 24.6). 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 24.3 and 24.6 (details as a function of the vowel type are further explained in Meyer 2015, Ridouane et al 2019, 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 24.2), the articulatory difference is high because the dynamics of spoken sounds produced originally in the larynx are now transposed to the upper oral cavities.

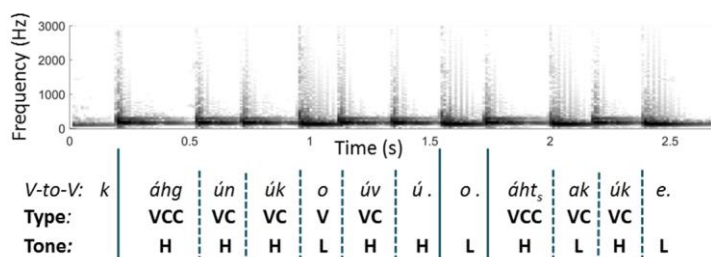
Moreover, in all languages surveyed, secondary cues are also targeted and whistled speech frequency variations may also be partly sensible to stress when it is encoded by intonation (as shown in Figure 24.6, see also Diaz 2017). Finally, in any whistling transposition strategy, phonetic reduction imposes a context of the conversation but whistlers are still able to exchange complex – not stereotyped – sentences in real dialogs (Moles 1970).



**Figure 24.6.** 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 ? Do you come for dinner ?* was recorded with a traditional whistler of El Hierro Canary Island who was targeting long distance communication (whistler: Emiliano Fernandez Armas). 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, such 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 2008). Listen to sound extracts in Audio 24.5 or <https://soundcloud.com/user-28976943/sets/sounds-in-meyer-manfredi-2023>.

### 24.3.1.2 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 frequencies 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 24.7). Drummed Bora also employs sentence framing and word enphrasing (as described in the end of Section 2) as well as emulating the pitch of vowels (and thus their tone) (Thiesen 1969, Meyer et al, 2012, Seifart et al, 2018). 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 24.7.** Spectrogram of the drummed Bora phrase /káʔgúnúkòúβú ò áʔʔàkúnè/ ‘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). [Adapted from Seifart et al 2018].

### 24.3.2 Verbal arts in musical surrogacy

Among the large 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 drums 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 òja flute.

### 24.3.2.1 *Gùn-gbè drum and its emulation of proverbial texts (Rép. 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” (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.” (1964, 3).

This original hypothesis is testable because a drummer’s hands are more easily observed than are the hidden muscles of the larynx (Halle & Stevens 1971, Nissenbaum & 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., Niagoran-Bouah 1980, Pacere 1987, N’Guessan 2020).

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

- (i) Contrasts of spoken F0 are not realised as distinct musical notes of different pitches - difficult to produce with a single membrane - but as spectral qualities (“timbres”) distinguished in three types of strikes. An open hit on the edge has a rich broad harmonic spectrum and is associated to the M tones. A muted strike on the edge of the drum emphasises high frequency partials of the drum (associated to *Gùn-gbè* H ‘tones’), whereas a muted strike on the center emphasises low frequency partials (associated to 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 center of the membrane, versus 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 24.1).

**Table 24.1.** 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))

	<i>drum articulation</i>		<i>drum sound</i>		<i>speech sound</i>	
	<i>place</i>	<i>manner</i>	<i>intensity</i>	<i>steep decay</i>	<i>intensity</i>	<i>duration</i>
H	edge	muted	+	+	+	+
M	edge	non-muted	+	–	+	–
L	center	muted	–	+	–	+

Most surprising in these parameters is the fact that vocal durations associated to 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, Sprigge 1967, Smith 1968, Stahlke 1971, Hyman 1973, Capo 1990, Gbètɔ 1995, 2020, Manfredi 2018, 2020). 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 24.1).

This might be one of the reasons why Rouget wrote: “... la place qu’on attribue traditionnellement aux intervalles se relèverait peut-être plus grande qu’elle n’est en réalité.” [...traditionally more importance is perhaps ascribed to pitch intervals than they actually possess.] (1964, 26, cf. 1975, 218, 225).

Another reason is probably that for him several other aspects could be more fully explored in relation to spoken pitch as transposed on drums. He suggested that:

- 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 & Halle 1956).
- 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, namely 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.

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

From ten Yorùbá-speaking communities, Thieme (1969) recorded surrogate speech by thirteen musical instruments spanning six formal genera à la Hornbostel & Sachs (1914)<sup>3</sup>.

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

**Genus 2:** *bátá* 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 *àgèrẹ̀* (cf. Adéoyè 1979, 125, Abraham 1958, 30)

**Genus 4:** percussion idiophone (*omọ 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 (*ìpẹ̀* ‘trumpet’, *ekùtù* side-blown antelope horn, *tìokò* bamboo flute)

<sup>3</sup> Each local example is identified with the reel number of its audiotape in the UCLA archive [catalogeth.library.ucla.edu/vwebv/holdingsInfo?bibId=11178](http://catalogeth.library.ucla.edu/vwebv/holdingsInfo?bibId=11178): *dùn-dún*: Ede 021, 024, 025, 026, Ìbàdàn 029, Ilé-Ifè 044, 046, Isundunrin (diacritics unknown) 067, 070, 072, Ìwó 078, 080, Èkó (Lagos) 088, 089, 090; *gán-gan/àpàlà/àdàmò*: Adó Àwáyè 017, Ilé-Ifè 046, 062, Isundunrin (diacritics unknown) 061, 066, 067, 072; *kerikeri*: Abèòkúta 007; *kósó*: Èjìgbò 061, 027; Adó Àwáyè 010; *bátá*: Ìbàdàn 029, 031, Ìwó 074, Èkó (Lagos) 088; *ìgbìn*: Isundunrin (diacritics unknown) 064; *ìgbésẹ̀*: Ilé-Ifè 047; *lukorigi/lukoogi* or *àgèrẹ̀*: Ìgàngàn 053, 061; *omọ owú*: Isundunrin (diacritics unknown) 068; *molo* (diacritics unknown) or *àgídìgbo*: Èjìgbò 027, Ísẹ̀yìn 058, Ìwó 080, Mowolowo (diacritics unknown) 087; *ìpẹ̀*: Ilé-Ifè 043, 045, 049; *ekùtù*: Isundunrin (diacritics unknown) 062; Isundunrin (diacritics unknown) 062.

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 Òyó palace guild named after the tree from which these instruments are carved (Àyánkúnlé & al. 1997, Villepastour 2010, González & Olúdáre 2022, cf. Abraham 1958, 86, Law 1977). Video 24.1 shows parallel renditions of one proverb in two oral and two drummed modalities<sup>4</sup> (source: Àyánkúnlé et al. 1997).

transcript:

*A ò ní sí ñbè*  
*We will not be there*  
*níbi wón gbé ñ se orí burúkúkú*  
*where they're always manifesting bad luck [lit. bad head]*  
*a ò ní sí ñbè*  
*We will not be there*  
*[drummed on dùn-dún]*  
*[drummed on bàtá]*

*jajà fífí rídà*  
*fífí 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 (Akínbò 2019). Yorùbá has three relative pitches H/M/L as well as predictable glides between H and L within a phrase. Contra 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 (Siertsema 1958, Akinlabi 1985, Connell & Ladd 1990, Láni.ran 1992), 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 (Adégbolá 2003), therefore it's mistaken 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 & Olúdáre 2022, 18). On the contrary, it's easy to meet fluent L1 Yorùbá speakers who frankly confess "N kò gbó ilù" (I can't parse the drum, referencing either *dùn-dún* or *bàtá*), revealing in this way an incapacitating lack of familiarity with the repertoire of ritualized utterances. The indeterminacy is illustrated by a famous anecdote (Beier 1970, 12f., transcription adjusted):

---

<sup>4</sup> <https://people.bu.edu/manfredi/Lamidi2.mp4>

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 24.8, which became the Nigerian radio's signature tune.



**Figure 24.8.** 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 jè 'gèdè dudu, inú n' ta bòn-ùn.*  
He ate overripe plantains, his stomach swelled up and burst.

*Kò sí olòṣì n'ìbí, lẹ sí'lé 'kejì!*  
There are no paupers here, try the next house!

**Bàtá drums.** The *bàtá* drums, with fixed membranes, are considered as harder to play and interpret, probably because they combine strokes by hand and/or with leather *bílálà* flail to distinguish two classes of vowel qualities and three Yorùbá tones (Láoyè 1959, 10, Rouget 1965, 82, Oyèélámì 1991). The lead (*iyá*, mother) *bàtá* drum, less capable than *dùn-dùn* to emulate 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, 262-67, cf. González & Olúdàré 2022, 11f.). Further evidence for *bàtá*'s intentional control of timbre is the obligatory use of an adhesive compound of indigenous latex (*ida*) applied to the center 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 24.2; and it resembles to muted strike on the center of Gùn-gbè drum associated to L ‘tones’ described by Rouget, see 24.3.2.1). 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 24.2 illustrates this acoustic effect of *ida* latex damper on *bàtá* large drumhead<sup>5</sup> (source: Àyánkúnlé et al. 1997).

<sup>5</sup> <https://people.bu.edu/manfredi/Lamidi1.mp4>



transcript:

[4 strikes with *ìda* applied]

*È mọn-on gbó nísinií*

*You all listen now*

[4 strikes with *ìda* removed]

*Iṣé tí n... tí a bá tún fì le nísin-ìn-ín*

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

[4 strikes with *ìda* applied]

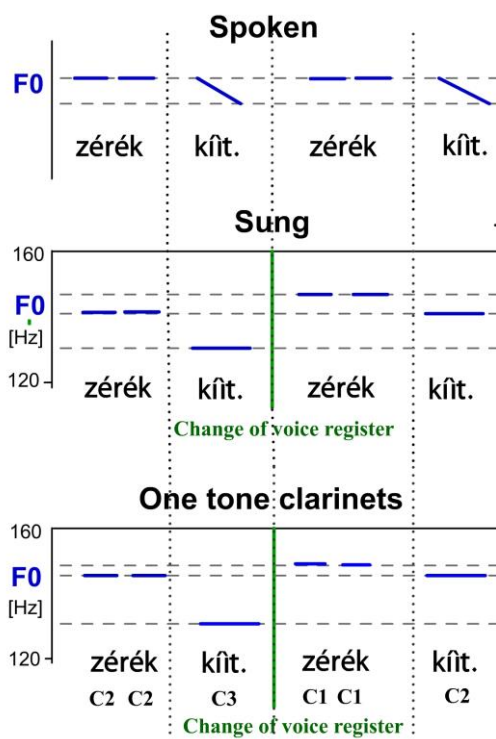
A simultaneous strike on the other smaller head of the drum is added to render the distinct spectral quality of open vowels. A strike on the smaller head alone is used to code the mid tone in some specific contexts. Oyèlámì (1991) published the first explicit mapping of tones and vowel timbres onto these seven permutations of *bàtá* drumstrokes, although that account is idealised, 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 & Olúdàré 2022, 10). There are also quite frequent speech fusions or contractions on the *bàtá* (mostly resembling normal allegro speech effects or performance hapology 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 méta* (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-dùn*.

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 on both 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 variate and may sometimes be detached from the acoustic correlates on which they rely in spoken modal speech.

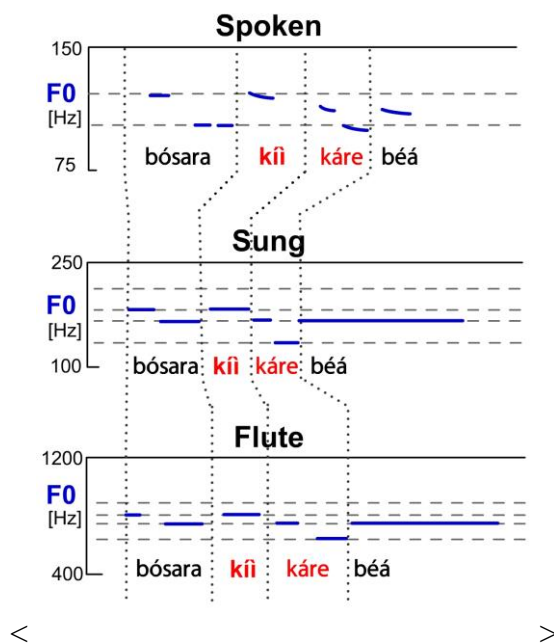
### 24.3.2.3 *Gavião* lyrics emulated with three different musical instruments (Amazonia)

Instrumental speech exists among the *Gavião* people of Rondônia with three different types of musical instruments employed to emulate song lyrics—a flute, an ensemble of three one-note bamboo ‘clarinets’ (Figure 24.1) 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 *Gavião* spoken language and, on the other side, 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 hinder, however, the players to render the relative height of a syllable in relation to the surrounding syllables (see Figure 24.2 lower line for flutes and Figure 24.9 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 24.2, 24.9, and 24.10). 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 24.9.** Simplified charts of the spoken, sung and musical forms of a song played on the 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 24.1 for a photo). Note here also that the change of voice register in the middle of the sung verse is imitated by the instrumentists (delimited by a green line). These are typical musical esthetic effects. Another difference with the spoken form is the flat kîit 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 [Adapted from Meyer and Moore, 2021].



**Figure 24.10.** 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 lay 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 *kîi* in *kîi káre béá* (according to Gavião language descriptions, the syllable *kîi* 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 with 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.

#### 24.3.2.4 Ìgbo òja flute (southeast Nigeria)

The *òja* 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 life-cycle celebrations (such as weddings), or to address invisible participants (for example during dance-mask performances). Carter-Ényi 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 *òja* flute<sup>6</sup>. The traditional “chanting mode” of the flute performed during this video documentation is known to be closer to speaking 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<sup>7</sup>. Just as for the Gavião flute (Section 24.3.2.3 above), the Ìgbo *òja* 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 are 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 abridgment.

First, the flute versions of the phrases are on average roughly 30 to 40% longer than the vocal ones

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

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

(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<sup>8</sup>. 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’ (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 24.11, 24.12, and 24.13, 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 (Welmers 1970, Émènanjọ 1978a, 2015, Liberman & al 1993). Rather than a taxonomic phoneme, it is a syntagmatic *relation* between H tones conditioned by a grammatically imposed prosodic boundary (Clark 1980)<sup>9</sup>. It is also representable analytically as units of syntagmatic “constituency, such as bracketed grids and metrical trees” (Dilley 2005, 46 cited by Carter-Ényi 2016, 65). Given the significant though imperfect registration of downstep cues (14/22) in this òja corpus, more controlled examples in future research may hope 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-Ényi’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 *-zù* ‘feed/nurture’) and two downsteps, whose abrupt drops (of approximately 10Hz or more) are clearly visible in the corresponding spectrographic representation (Figure 24.11). 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 Himayounfar 2021).

(1) *Transcription, gloss and tonal annotation of phrase 2 (see recordings in Figure 24.11):*

‘nurturing person’

[XP1 *ala* [XP2 *relative clause* *na-a-zù* [XP3 *genitive modifier* *nnwa* ] ] ]

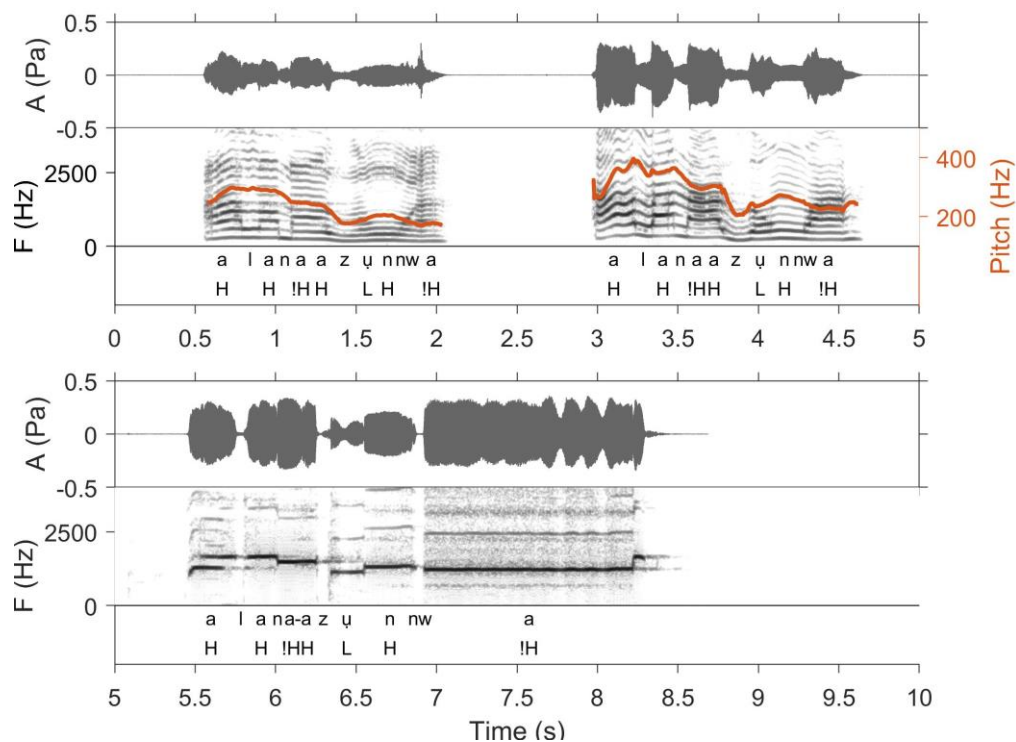
[XP1 *breast* [XP2 *relative clause* *that habitually feeds* [XP3 *genitive modifier* *child* ] ] ]

spoken male and female:                      *ála ná-a-zù nnwá*                      HH !HHL H!H

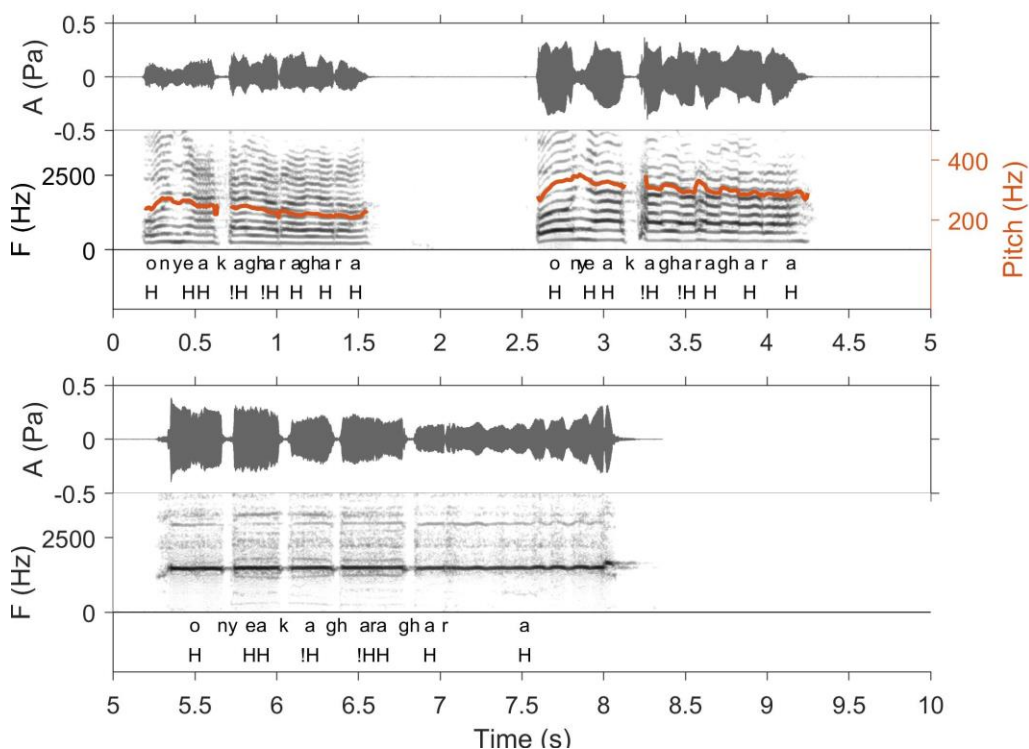
flute (played on 5 levels of notes):                      HH !HHL H!H

<sup>8</sup> 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 Fig 24.4 in this article), either for phatic communication or to simply parse the message.

<sup>9</sup> 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 & Ígwè (1963) has failed to note its fundamental importance (albeit under the misleading label of “mid tone”; Íkekeonwú 1982, Ányaanwú 1998; 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 & Ígwè 1963, 6f.).



**Figure 24.11.** Waveform and spectrogram of three different pronunciations of the same sentence *ála ná-a-zù nńwá*. 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 (orange line and scale). The pitch simplified notation – with H, L, !H – refers in the three cases to the spoken reference.



**Figure 24.12.** Waveform and spectrogram of three different pronunciations of the same sentence *ónye ak gháraghara*. 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 (orange line and 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-Ényi 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 24.12). 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 24.12):*

‘generous person’

[XP1 onye [XP2 genitive modifier aka [XP3 relative clause gharaghara] ] ]

[XP1 person [XP2 genitive modifier hand [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), there are two obligatory downsteps occurring 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 (1965), they are not reducible to the universal vocal trend of “back-drop declination” (cf. Connell & Ladd 1990) as Carter-Ényi et al. (2021, 3) wrongly suggest<sup>10</sup>.

A different but notable other *òja* 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 24.11 and 24.12). 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 ‘zù n’ of phrase 2, played LH on the flute (Figure 24.11; see also ‘mùó n’ of phrase 8 in Figure 24.13, played at three different flute pitches), or (ii) to assimilate all the adjacent syllable nuclei in a same pitch level (as in ‘ru ó’ and ‘mùó à’ of phrase 8, see Figure 24.13). In the latter case, a subtle compromise is made by the flutist as the resulting level pitch is influenced by the different constituent syllables while succeeding to maintain relative pitch differences with the other neighboring syllables (as noted also for Gavião flute, see Section 24.3.2.3 above). There is also an influence of speech speed as when sentences are short, the pace of pronunciation is slower and less 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 (Williamson 1972, Ígwè 1999, 844, cf. Green & Ígwè 1963, 158).

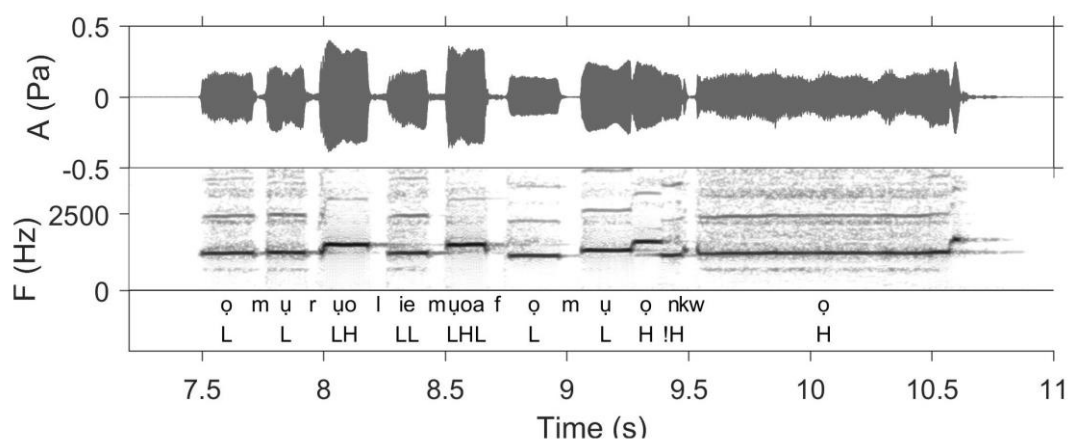
### **Other interesting aspects in line with results found in other auxiliary speech forms derive from these observations:**

- First, the chanting mode employed with the flute relies on a proverbial style that often enphrases messages with rather long texts (this further illustrates mentions made in Section 24.2), while several words also show that the flute is playing a fuller/longer form than what the common speakers produce (for example, the word for child is played *n̄nwa* HH bisyllabic (here pronounced H!H which

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

is to be expected in the phrase 2 context), a rather archaic and irregular form<sup>11</sup>).

- Secondly, consecutive syllable assimilations in *òja* provide novel evidence confirming the relevance of structures composed of Vowel-to-Vowel intervals to analyze the complex puzzle concerning features embedded in speech. As shown in Section 24.3.1.2, such interval types have already been highlighted in drummed auxiliary speech (Meyer et al 2012, Seifart et al 2018, Dentel and Meyer 2020) besides being increasingly used to analyze rhythmic units of spoken modal speech (Ryan 2016, Otterbeim et al 2012, Hirsh 2014). Note that previous support for such intervals comes, also, from phonetic research on tonal alignment (House 1990).
- Third, while most consonants are marked by a silent pause, some of them can remain continuous in some cases (such as the liquid /r/ and the nasal /n/), in line with other results on whistled tonal languages about the transposition of formant-like resonant patterns in consonants (Rialland 2005, Meyer 2015). In addition, the *òja* flute melody also distinguishes syllabic versus 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 either expressed 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 24.11, 24.12 and 24.13, and listen to more audio examples in the corpus for syllabic nasal *m*<sup>12</sup>). As such, *òja* lends 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.



**Figure 24.13.** Waveform and spectrogram of the flute speech form of phrase 8, the sentence *ò-mu-ru ólie, muó àfò, muó ñkwò* (English literal translation: ‘person who gives birth on *ólie* day, gives birth on *àfò* day, gives birth on *ñkwò* day’ (meaning ‘very fertile woman’)). The line with pitch simplified notation– with H, L, !H below the text – refers to the spoken reference.

### The example of *òja* surrogacy in Ìgbo finally also illustrates more general problems with attempts to map from “phonological tones” to pitch targets in nonlaryngeal emulation :

(i) The operational concept of ‘functional load of tone’ is sometimes assumed as 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 lesser described languages (Surendran and Niyogi 2006). Next, traditional phonemic theory (e.g., Pike 1948) extracts ‘morphemes’ taxonomically – in

<sup>11</sup> At least not the most common form found in the dialect base of Williamson’s Ìgbo dictionary.

<sup>12</sup> 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 Higs)

advance of systemic grammatical analysis – and defaults in practice to establishing minimal pitch contrasts between translated lexical items. In the *òja* case, this last shortcut yielded a *přimā faciē* implausible inference “that the functional load of tone in *Ìgbo* may be higher than that of *Yorùbá*” (Carter-Ényi 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, like, for instance, 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 realised 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, for example, *Yorùbá* (Courtenay 1971, Lání.ran 1992). This difference and the large but partial manifestation of downstep in the pitch played on *òja* flute extends 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 (Nketia 1994, Akinbo 2019, McPherson and James 2021, Meyer and Moore 2021, Struthers-Young 2021)<sup>13</sup>.

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 idealised extremes, as determined by external performance constraints.

## 24.4 Summary

Auxiliary speech practices are the result of the adaptation of human productive and perceptive 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, conversely, they also lend insight into how the sound-meaning correspondences of the primary spoken languages themselves are stored and processed in real time. All together, studies of language channeled 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 (for instance, the purely prosodic melodic line of whistled speech sometimes encodes primarily formants), and the role of the syllable in speech rhythm (they rather 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 the Vedic and Homeric poems, developed in nonliterate social

---

<sup>13</sup> 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* (Akinbo 2019), Northern Toussain balafon (Struthers-Young 2021) and balafon forms of sung *Seenku* (McPherson and James 2021).



[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

settings (Kiparsky 1973), and further examples can be considered: the onomatopoeia of verbal ideophones (Awóyalé 1981, 1989) and the melodic/metric matching of text and song (Éménanjo 1978b, Zhang & Cross 2021). Each particular case can be expected to combine multiple cognitive connections in indigenous ‘theories of mind’ in meta-communication and meta-representation (Whiten ed. 1991, Sperber ed. 2000).

## References

Abraham, R. (1958). *Dictionary of Modern Yorùbá*. University of London Press.

Adédéjì, J. (1972). The origin and form of the Yorùbá masquerade theater. *Cahiers d’Études africaines* **12**, 254-76.

Adégbolá, ’T. (2003). Probabilistically speaking; a quantitative exploration of Yorùbá speech surrogacy. *4th World Congress of African Linguistics* (= ACAL 34), edited by A. Akinlabí & O. Adéṣolá, 363-73. Köppe, Köln.

Adéoyè, C. (1979). *Àṣà àti Ìṣe Yorùbá*. Oxford University Press, Ìbàdàn.

Akínbò, S. (2021). The language of *gangan*, a Yorùbá talking drum. *Frontiers in Communication* **6**, 650382.

Akínbò, S. (2019). Representation of Yorùbá Tones by a Talking Drum: An Acoustic Analysis. *Linguistique et Langues Africaines* **5**, 11–23.

Akinlabí, A. (1985). Tonal underspecification & Yorùbá *tone*. Dissertation, University of Ìbàdàn.

Alexandre, P. (1967). *Langues & langage en Afrique noire*. Payot, Paris.

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

Ansre, G. (1961). *The tonal structure of Èwè*. M.A. thesis, Hartford Seminary, Connecticut.

Ányaanwú, R-J. (1998). *Aspects of Ìgbo Grammar; phonetics, phonology, morphology & the tonology of nouns*. Lit Verlag, Hamburg.

Arom, S. (1970). *L’arc Musical Ngbaka, République de Centrafrique*. Paris: CNRS Comité du Film Ethnographique.

Awóyalé, ’Y. (1981). Nominal compound formation in Yorùbá ideophones. *Journal of African Languages & Linguistics* **3**, 139-57.

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Awóyalé, 'Y. (1989). Reduplication and the status of ideophones in Yorùbá. *Journal of West African Languages* **19**, 15-34.

Àyánkúnlé, L. & al. (1997). Òsè bàtá 1-4, 7-9 October. Video 294 minutes. Department of Performing Arts, University of Ilorin. [manfredi.mayfirst.org/AyanLamidiSunUnReOOO.html](http://manfredi.mayfirst.org/AyanLamidiSunUnReOOO.html).

Babáyemí, S. (1980). *Egúngún among the Òyó Yorùbá*. Òyó State Council for Arts and Culture, Ìbàdàn.

Bagemihl, B. (1988). *Alternate Phonologies and Morphologies*. Vancouver, BC: University of British Columbia.

Bámgbóṣé, A. (1966). The assimilated low tone in Yorùbá. *Lingua* **16**, 1-13.

Bánjò, A. & al. (eds., 1991). *Quadrilingual Glossary of Legislative Terms: English, Hausa, Ìgbo, Yorùbá*. Nigerian Educational Research & Development Council, Lagos.

Beudet, J. M. (1997). *Souffles d'Amazonie, Les Orchestres tute Des Wayãpi*. Nanterre: Société d'Ethnologie.

Beier, U. (1970). *Yorùbá Poetry; an anthology of traditional poems*. Cambridge University Press.

Bonhomme, J. (2014). La voix du mongongo ou comment faire parler un arc musical. *Cahiers d'anthropologie sociale* **10**, *L'Image rituelle*, edited by C. Fausto & C. Severi, 93-111. L'Herne, Paris.

Busnel R-G. (1967). *Documents sur une langue sifflée turque*. SFRS, Paris.

Busnel R-G. (1970). Recherches expérimentales sur la langue sifflée de Kusköy. *Revue de Phonétique Appliquée* 14/15: 41-57.

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

Busnel RG, Classe A. (1976). *Whistled Languages*. Berlin: Springer-Verlag

Busnel, R. G., Moles, A., and Vallencien, B. (1962). "Sur l'aspect phonétique d'une langue sifflée dans les Pyrénées françaises," in *Proc. Int. Congress of Phon. Sci. (Helsinki, Finland)*, 533–546.

Busnel, R-G., Gustavo, A., Gautheron, B., 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.

Cañadas A.S. & Jiménez M. I. (1992) *Instrumentos musicales de barro en Andalucía, Granada, Centro de*

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Documentación Musical de Andalucía, Junta de Andalucía.

Capo, H. (1991). *A Comparative Phonology of Gbè*. Foris, Dordrecht.

Carrington, J. F. (1949). *Talking Drums of Africa*. London: Carey-Kingstate Press

Carter-Ényi, A. (2016). *Contour levels; an abstraction of pitch space based on African tone systems*.

Dissertation, Ohio State University, Columbus.

Carter-Ényi, A. & al. (2021). Ìgbo speech surrogacy, preliminary findings based on the òja flute. *Frontiers in Psychology* **12**, article 653068.

Catlin, A. (1982). “Speech Surrogate Systems of the Hmong: from singing voices to talking reeds” in *The Hmong in the West: Observations and Reports*, eds B. T. Downing and D. P. Olney (Minneapolis, MI: University of Minnesota Press). 170–197.

Charalambakis C. (1994). A case of whistled speech from Greece, *Current Theory in linguistics*, 117, 389-396.

Christaller, J. (1875). *A Grammar of the Asante & Fante Language (...)*. Missionsbuchhandlung, Basel.

Clark, M. (1980). On the treatment of syntactically-distributed downstep. *Studies in African Linguistics* 11, 101-37.

Cloarrec-Heiss (1999). From natural language to drum language : An economical encoding procedure in Banda-Linda (Central African Republic). In *Language diversity and cognitive representations / C. Fuchs, S. Robert (Eds.) Philadelphia : J. Benjamins Publ.,145-157*

Connell, B. & D. Ladd. (1990). Aspects of pitch realisation in Yorùbá. *Phonology* **7**, 1-29.

Cross, I. (2012). Music as a social and cognitive process. *Language & Music as Cognitive Systems*, edited by P. Rebuschat & al., 315-28. Oxford University Press.

Courtenay, K. (1971). Yorùbá, a terraced-level language with three tonemes. *Studies in African Linguistics* 2, 239-55.

Cowan G. (1948). Mazateco whistle speech. *Language* 24, 280-286.

Cowan, G. (1972). Segmental features of whistled Tepehua. *Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal, 695-98*.

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Crowther, A. & C. Wakeman. (1843-1852, 1913 and reprints). *A Dictionary of the Yorùbá Language*.

Church Missionary Society, Lagos/Oxford University Press, Ìbàdàn.

Dargie, D. (2011). The Xhosa umrhubhe mouthbow: An extraordinary musical instrument. *African Music:*

*Journal of the International Library of African Music*, 9(1), 33–55.

Dargie, D. (2007). Umakhweyane: A musical bow and its contribution to Zulu music. *African Music:*

*Journal of the International Library of African Music*, 8(1), 60–81.

Dentel, L., and Meyer, J. (2020). “How Rhythm and Timbre Encode Mooré Language in Bendré Drummed

Speech” in *Proceedings of Interspeech 2020- 21st Annual Conference of the International Speech*

*Communication Association (Germany: DBLP)*, 1898–1902. doi: 10.21437/Interspeech.2020- 2532

Dilley, L. (2005). *The phonetics & phonology of tonal systems*. Dissertation, M.I.T., Cambridge Mass.

Diaz, R. D. (2017). *El Lenguaje Silbado en la Isla de El Hierro (The Silbo Language on the Island of El*

*Hierro)*, 2nd ed. (Tenerife, LeCanarien Ediciones, La Orotava, Spain).

Dúrójayé, 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. *Front. Psychol.* 12:652673. doi: 10.3389/fpsyg.2021.652673

Eboué F. (1935). La clef musicale des langages tambourinés et sifflés. *Bulletin du Comité d'étude*

*historiques et scientifiques de l'Afrique occidentale française*, 18, 353-360.

Émènanjò, 'N. (1978a). *Elements of Modern Ìgbo Grammar*. Ìbàdàn. University Press.

Émènanjò, 'N. (1978b). An Ìgbo folksong—a linguistic study. *African Languages/Langues africaines*

[London], 4, 96-109.

Émènanjò, 'N. (2015). *A Grammar of Contemporary Ìgbo; constituents, features & processes*. M & J Grand

Orbit Communications, Port Harcourt.

Ermel, P. B. (2004). *L'Arc et la Lire: proposition d'un Narratif Visuel et Sonore*. Paris: Collège Iconique de

l'Institut National de l'Audiovisuel (INA).

Euba, A. (2011). Review: Villepastour, Amanda. *Ethnomusicology* 55, 516–520.

doi:10.5406/ethnomusicology.55.3.0516

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Everett, D. (1983). *A língua Pirahã e a Teoria da Sintaxe: descrição, Perspectivas e Teoria*. Campinas: Universidade Estadual de Campinas.

Fant G. (1960). *Acoustic theory of vowel production*. Mouton, s'Gravenhage, The Netherlands.

Fitch, T. (2010). *The Evolution of Language*. Cambridge University Press.

Foris, D.P. (2000). *A grammar of Sochiapam Chinantec*. Dallas: SIL International and UT Arlington.

Fournel, T. (2002). *Polyphonies Ari*. Paris: Collection Ocora/Radio France.

Franchetto, B., and Montagnani, T. (2011). Flutes des hommes, chant des femmes. Images et relations sonores chez les Kuikuro du haut Xingu. *Gradhiva* 13, 94–111. doi: 10.4000/gradhiva.2052

Gbè̀tò, F. (1995). Quelques aspects comparatifs et diachroniques de la tonologie du Gbè̀. *Afrika & Übersee* **78**, 73-99.

Gbè̀tò, F. (2020). *Les tons en gbè̀, langue kwa (sic) de l'Afrique de l'Ouest; étude comparative, typologique & diachronique*. Ms., Université d'Abomey-Calavi, Bénin.

González M and Oludare O (2022) The Speech Surrogacy Systems of the Yoruba Dùndún and Bàtá Drums. On the Interface Between Organology and Phonology. *Front. Commun.* 6:652542. doi: 10.3389/fcomm.2021.652542

Green, M. & G. Ígwè̀. (1963). *A Descriptive Grammar of Ígbo*. Ostberlin: Akademie Verlag, for Oxford University Press.

Halle, M. & K. Stevens. (1971). A note on laryngeal features. *Quarterly Progress Report* **101**, 198-213. Research Laboratory of Electronics, M.I.T., Cambridge, Mass.

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

Himayounfar R. (2021) Exploring a Bigram System in the Analysis of Ìgbò Òjà Flute Surrogacy. <https://speechsurrogates.surge.sh/>

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

- Hirsch A. (2014) What is the domain for weight computation: the syllable or the interval? In *Proc. 2013 Annual Meeting on Phonology*, Amherst, MA, 8–10 November (eds J Kingston, C Moore-Cantwell, J Pater, R Staubs), pp. 1–12. Washington, DC: Linguistic Society of America.
- v. Hornbostel, E. & C. Sachs. (1914). *Systematik der Musikinstrumente, ein Versuch*. *Zeitschrift für Ethnologie* 46, .553-90.
- House D. (1990) *Tonal perception in speech*. Lund, Sweden: Lund University Press.
- Hyman, L. (1973). The role of consonant types in natural tone assimilations. *Consonant Types & Tone*. *Southern California Occasional Papers in Linguistics* 1, 152-79.
- Ígwè, G. (1985/1999). *Ìgbo-English Dictionary*. University Press, Ibadan.
- Íkekeonwú, C. (1982). Tones in Ìgbo, a classification. 15th West African Languages Congress, Port Harcourt, 4-10 April.
- Izikovitz, K. G. (1935). “Idiophones” in Musical and Other Sound Instruments of the South American Indians: a Comparative Ethnographical Study, ed. K. G. Izikovitz (Göteborg: Elanders Boktryckeri Aktiebolag), 7–33.
- Jakobson, R. (1960). “Linguistics and poetics,” in *Style in Language*. ed. T. Sebeok (Cambridge: MIT Press), 350–377.
- Jakobson, R. & M. Halle. (1956). *Fundamentals of Language*. Mouton, Den Haag.
- James, L. (2021) *Systems of Communication: Aspects of Culture and Structure in Speech Surrogates*. *Front. Commun.* 6:653268. doi: 10.3389/fcomm.2021.653268
- Kawada, J. (1985). *Textes historiques oraux des Mosi méridionaux (Burkina-Faso)*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA).
- Kawada, J. (2014). Notes on the Drum Language of the Ancient Mossi Kingdoms (Burkina Faso). *Japanese Rev. Cult. Anthropol.* 15, 199–215.
- Kawada J., 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 (ILCAA).
- Kiparsky, P. (1973). Phonological representations. *Three Dimensions of Linguistic Theory*, edited by O.

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Fujimura, 5-86. T.E.C., Tokyo.

Labouret H. (1923). Langage tambouriné et sifflé. *Bulletin du comité d'études historiques et scientifiques de l'Afrique Occidentale Française* 6, 120-158.

Láníran, 'Y. (1992). *Intonation in tone languages; the phonetic implementation of tones in Yorùbá*. Dissertation, Cornell University, Ithaca New York.

Láoyè T. E. (1959). Yorùbá drums. *Odù [Ìbàdàn]* 7, 5-14.

Law, R. (1977). *The Òyó Empire c.1600-c.1836, a West African imperialism in the era of the Atlantic slave trade*. Oxford University Press.

Lieberman, M. & al. (1993). The phonetic interpretation of tone in Ìgbo. *Phonetica* 50, 147-60.

Lo-Bamijoko, J. N. (1987). Classification of Igbo musical instruments, Nigeria. *Afr. Music J. Int. Libr. Afr. Music* 6, 19–41. doi: 10.21504/amj.v6i4.1259

Locke, D., and Agbeli, G. (1981). *Drum language in Adzogbo*. Black Perspect. Music 9, 25–50. doi: 10.2307/1214103

McPherson, L. (2018). The Talking Balafon of the Sambla: grammatical Principles and Documentary Implications. *Anthropol. Linguist.* 60, 255–294. doi: 10.1353/ anl.2019.0006

McPherson, L. (2019). The role of music in documenting phonological grammar: two case studies from West Africa. *Proceedings of the Annual Meeting on Phonology 2018* (New York, NY: Linguistic Society of America). doi: 10.3765/ amp.v7i0.4493

McPherson, L., and James, L. (2021). Artistic adaptation of Seenku tone: musical surrogates vs. vocal music, *Celebrating 50 Years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, eds A. Akinlabi., L. Bickmore., M. Cahill., M. Diercks., L. J. Downing., J. Essegbey., K. Franich., et al. (Berlin: Language Science Press).

Manfredi, V. (2018). Cyclic accentuation in Yorùbá. *Data-rich Linguistics; papers in honour of 'Yíwólá Awóyalé*, edited by O. Adéşolá & al., 211-36. Cambridge Scholars Publishing, Newcastle-upon-Tyne, England.

Manfredi, V. (2020). A toneless theory of 2-and-a-half tonemes in Gbè. *African Languages in Time &*

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

*Space; papers in honour of Professor Akínbiyí Akinlabí*, edited by E. Urua & al., 40-64. Zenith BookHouse, Ìbàdàn.

Marsh, N. (2013). *The decline of speech surrogacy in Bata Cuban drumming*. M.Sc. Dissertation University of Auckland. New Zealand.

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* (Spain: Universitat Politecnica de Catalunya). 1-S4-4, 1–8.

Meyer, J. (2015). *Whistled Languages: A Worldwide Inquiry About Whistled Speech*. Berlin: Springer.

Meyer, J. (2020) Coding Human Languages for Long-Range Communication in Natural Ecological Environments: Shouting, Whistling, and Drumming. In: Aubin T., Mathevon N. (eds) *Coding Strategies in Vertebrate Acoustic Communication. Animal Signals and Communication*, vol 7. pp.91-113, Springer, Cham. [https://doi.org/10.1007/978-3-030-39200-0\\_4](https://doi.org/10.1007/978-3-030-39200-0_4)

Meyer J (2021) Environmental and linguistic typology of whistled languages. *Annual Review of Linguistics*, 7, 493-510.

Meyer, J., Dentel, L., and Seifart, F. (2012). A methodology for the study of rhythm in drummed forms of languages: bora of Amazon. *Proceedings of Interspeech 2012 Annual Conference of the International Speech Communication Association* , 686–690. doi: 10.21437/Interspeech. 2012-210

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. *Front. Psychol.* 12:674289. doi: 10.3389/fpsyg.2021.674289

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.

Moles, A. (1970). Etude sociolinguistique de la langue sifflée de Kuskoy. *Revue de Phonétique Appliquée*, 14(15), 78-118

Moore, D. (1999). “Tonal System of the Gavião Language of Rondônia, Brazil” in *Tupian Perspective Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics*. ed. K.



[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Shigeki (Tokyo: ILCAA/Tokyo University of Foreign Studies), 297–310.

Moore, D., and Meyer, J. (2014). The study of tone and related phenomena in an Amazonian tone language, Gavião of Rondônia. *Lang. Document. Conserv.* 8, 613–636.

Neeley, P. (1999). *People of the Drum of God—Come!* Dallas, TX: Summer Inst of Linguistics.

Niagoran Bouah, G. (1980) *Introduction à la drummologie*. Abidjan, Université nationale de Côte d’Ivoire, Institut d’ethnosociologie.

Nikolsky, A. (2020). Talking jew’s harp and its relation to vowel harmony as a paradigm of formative influence of music on language. *The Origins of Language Revisited; differentiation from music and the emergence of neurodiversity & autism*, edited by N . Masakata, 217-322. Springer, Singapore.

Niles, D. (2010). Editor’s Introduction to Zemp and Kaufmann. *Kulele Occasion. Pap. Pacif. Music Dance* 4, 1–50.

Nissenbaum, J. & al. (2002). High speed MRI of laryngeal gestures during speech production. *Journal of the Acoustical Society of America* **111**, 2479-80.

N’Guessan, K. (2020). *Drummologie : la connaissance aux sons du tambour et du balafon*. Editions Jets d’encre. Paris.

Nketia, J. H. K. (1963). *Drumming in Akan Communities of Ghana*. Edinburgh: Thomas Nelson and sons.

Nketia, J. H. K. (1971). “Surrogate languages of Africa” in *Speech Surrogates: drum and Whistle Systems*, eds T. A. Sebeok and D. J. Umiker-Sebeok (1976, Mouton: La Hague and Paris), 825–864.

Nketia, J. H. K. (1994). “Generative processes in Seperawa music” in *To the Four Corners: a Festschrift in Honor of Rose Brandel*, ed. E. C. Leichtman (Michigan: Harmonie Park Press), 127–151.

Nwáchuku, M. A. (1997). On the resonant frequencies of the oja. *Nigerian J. Technol.* 18, 1–21.

Nwáchukwu, A. (1976). *Noun Phrase sentential complementation in Ìgbo*. Dissertation, University of London.

Nwáchukwu, A. (1995). *Tone in Ìgbo Syntax*. Ìgbo Language Association, Department of Linguistics & Nigerian Languages, University of Nigeria, Nsùkà.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica*

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

41, 1-16.

Otterbein S, Abel C, Heinemann LV, Kaiser J, Schmidt-Kassow M. (2012). P3b reflects periodicity in linguistic sequences. *PLoS ONE* 7, e51419.

Oyèélámì, M. (1989). *Yorùbá Dùn-dún Music; a new notation with basic exercises & five Yorùbá drum repertoires*. Ìwàlẹ̀wà Haus, Bayreuth.

Oyèélámì, M. (1991). *Yorùbá Bàtá Music; a new notation with basic exercises & ensemble pieces*. Ìwàlẹ̀wà Haus, Bayreuth.

Pacere T. (1987) *Bendrologie et littérature culturelle*. Maison Pusga. Ouagadougou

Pepper H. (1950) *Musique centre-africaine*. Paris (FR) : Encyclopédie Coloniale et Maritime, 1950

Pepper, H. (1954) *Considérations sur le langage tambouriné et autres langages musicaux d'Afrique centrale, sur la pensée musicale Africaine*. Ivo C.I.A.O., Fernando Pô Guinée Espagnole

Persoon G.A., Schefold R. (2017) Frightened by the eagle Recording songs and music from the Island of Siberut, Mentawai Island, *Wacana* 18(3), 581-613.

Pike, K. (1948). *Tone languages; a technique for determining the number & type of pitch contrasts in a language, with studies in tonemic substitution & fusion*. Ann Arbor: University of Michigan Press.

Poss, N. (2012). *Hmong Music and Language Cognition: An Interdisciplinary Investigation*. PhD Thesis (Columbus, OH: Ohio State University).

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

Rialland, A. (2005). Phonological and phonetic aspects of whistled languages. *Phonology* 22, 237–271. doi: 10.1017/S0952675705000552

Ridouane R., Turco G., Meyer J. (2018). Length Contrast and Covarying Features: Whistled Speech as a Case Study. *Proceedings of Interspeech 2018*, Sep 2018, Hyderabad, India, 1843-1847

Risset, J. C. (1968). Sur certains aspects fonctionnels de l'audition. *Ann. Télécommun.* 23, 91–120. doi: 10.1007/BF02996236

Ritzenthaler, R. E., and Peterson, F. A. (1954). Courtship Whistling of the Mexican Kickapoo Indians. *Am.*

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Anthropol. 56, 1088–1089. doi: 10.1525/aa.1954.56.6. 02a00110

Ros, S. (2021). Rhythm-speech correlations in a corpus of Senegalese drum language. *Frontiers in Communication* 6, article 643683.

Rouget, G. (1964). Tons de la langue en gùn (Dahomey) et tons du tambour. *Revue de musicologie* 50 (128), 3-29.

Rouget, G. (1965). Notes et documents pour servir a l'étude de la musique yorùbá. *Journal de la Société des Africanistes* 35, 67-107.

Rouget, G. (1975). Tons, intonation, accentuation; problèmes d'identification. *Langues sans tradition écrite; méthodes d'enquête et de description*, edited by G. Manessy & J. Thomas, 217-40. SELAF, Paris.

Ryan, KM. 2016 Phonological weight. *Lang. Linguist. Compass* 10, 720–733.

Schefold, R. (1973) Schlitztrommeln und Trommelsprache in Mentawai, *Zeitschrift für Ethnologie* 98(1), 36-73.

Schneider M. (1952). Zur trommelsprache der duala. In Sebeok T. & Umiker-Sebeok D.J. (Eds). *Speech surrogates: drum and whistle systems*. 1976 La Haye and Paris, Mouton, 669-678

Sebeok, T. A., and Umiker-Sebeok, D. J. (1976). *Speech Surrogates: drum and Whistle Systems*. Mouton: La Hague and Paris. doi: 10.1515/9783110804416

Seifart\*, F., Meyer\*, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: amazonian Bora drummed language exploits speech rhythm for long-distance communication. *R. Soc. Open Sci.* 5:170354. doi: 10.1098/rsos.170354 (\* equivalent contributions)

Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, Vol. 27, 379–423.

Sicoli MA. 2016. Repair organization in Chinantec whistled speech. *Language* 92(2):411–32

Siertsema, B. (1958). Some notes on Yorùbá phonetics and spelling. *Bulletin de l'Institut Fondamental d'Afrique Noire, Série B, Sciences humaines* 20, 576-92.

Smith, N. (1968). Tone in Èwè. *Quarterly Progress Report* 88, Research Laboratory in Electronics, M.I.T., Cambridge Mass.

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Snyders J., Haudricourt A. (1968). Le langage par tambours à San Cristoval, British Solomon Islands.

*Journal de la Société des océanistes* 24, 133-138

Sperber, D. ed. (2000). *Metarepresentations; a multidisciplinary perspective*. Oxford University Press.

Sprigge, R. (1967). Collected field reports on tone in the Àdángbè dialect of Èwè. *Collected Language*

*Notes* 8. Institute of African Studies, University of Ghana, Legon.

Stahlke, H. (1971). *Topics in Èwè phonology*. Dissertation, U.C.L.A., Los Angeles.

Stern, T. (1957). Drum and whistle languages: an analysis of speech surrogates. *Am. Anthropol.* 59, 487–

506. doi: 10.1525/aa.1957.59.3.02a0 0070

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stewart, J. (1965). The typology of the Twì tone system, with comments by P. Schachter & W.E.

Welmers. Preprint from the *Bulletin of the Institute of African Studies* [Legon] 1, 1-67.

Struthers-Young, A. (2021). A preliminary account of the Northern Toussian balafon surrogate language.

*Front. Commun.* doi: 10.3389/fcomm.2021.652740

Surendran, D., and Niyogi, P. (2006). “Quantifying the functional load of phonemic oppositions, distinctive

features, and suprasegmentals,” in *Competing Models of Linguistic Change: Evolution and Beyond*,

ed O. N. Thomsen (Amsterdam; Philadelphia, PA: John Benjamins), 43–58.

Thieme, D. (1969). *A descriptive catalogue of Yorùbá musical instruments*. Dissertation, Catholic University

of America, Washington D.C.

Thiesen, W. (1969). The Bora signal drums. *Lore* 19, 101–103.

Tran Ngoc, A., Meunier, F., Meyer, J. (2022). Testing perceptual flexibility in speech through the

categorization of whistled Spanish consonants by French speakers. *JASA Express Lett.* 2(9):095201

Trujillo, R. (1978). *El Silbo Gomero : análisis lingüístico*. Editorial Interinsular Canaria. Santa Cruz de

Tenerife

Ùgoónà, N. (1984). *M̀m̀ónwù, a Dramatic Tradition of the Ìgbo*. Lagos University Press.

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bàtá Drum; cracking the code*. Ashgate,

Farnham England.

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

- van der Voort, H. (2016). “Aikanã and Kwaza: their ethno-historical and sociolinguistic context in Rondônia, Brazil” in *Endangered Languages and Languages in Danger: Issues of Documentation, Policy, and Language Rights*, eds L. Filipoviæ and M. Pütz (Amsterdam: John Benjamins), 203–230. doi: 10.1075/impact.42.09van
- Voorhis P.H. (1971). Notes on the Kickapoo whistle speech. *International Journal of American Linguistics*, 37/4.
- Wedekind, K. (1983). A Six-Tone language in Ethiopia: tonal Analysis of Benchnon. *J. Ethiop. Stud.* 16, 129–156.
- Welmers, W. (1970). Ìgbo tonology. *Studies in African Linguistics* 1, 255-78.
- Welmers, W. & B. Welmers. (1968). *Ìgbo, a learner's dictionary*. Department of Linguistics, University of California, Los Angeles.
- Whiffen, T. (1915). *The North-West Amazons: Notes of Some Months Spent Among Cannibal Tribes*. New York, NY: Duffield and Co.
- Whiten, A. ed. (1991). *Natural Theories of Mind; evolution, development & simulation of everyday mindreading*. Blackwell, Oxford.
- Williamson, K. (1972) *Dictionary of Ònìchà Igbo*. Ethiope press. Benin City
- Winter, Y. (2014). On the Grammar of a Senegalese Drum Language. *Language* 90, 644–668. doi: 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). pp. 196-217
- Yegnan-Touré, A. (2008). La technique et le jeu de l'arc musical. *Cahiers Ethnomusicol.* 21, 205–223.
- Zemp, H. (2005). *Siaka, an African musician*. Video 79 minutes. Documentary Educational Resources, Watertown Mass. [www.der.org/siaka-an-african-musician](http://www.der.org/siaka-an-african-musician).
- Zemp, H. and Kaufman, C. (1969). Pour une transcription automatique des « langues tambourinés » mélanésiens (Un exemple kwoma, Nouvelle-Guinée). *L'Homme* 9(2), 38-88
- Zemp, H., and Soro, S. (2004). Paroles de balafons. *L'Homme* 17, 313–332. doi: 10.4000/lhomme.24930
- Zhang, X. & I. Cross. (2021). Analysing the relationships between tone and melody in Chaozhou songs. *Journal of New Music Research* 50, 299-311.

## Biographical paragraphs

**Julien Meyer** is a researcher in linguistics and bioacoustics at the CNRS (Gipsa-Lab/Université Grenoble Alpes, France). With inquiries on whistled, drummed and other instrumental forms of languages based on first hand fieldwork, documentation, and study in 15 different language communities, he developed an expertise on natural human telecommunication systems and musical surrogacy in human language. His book *Whistled Language. A Worldwide Inquiry on Human Whistled Speech* is one of his publications available at this page : [https://hal.science/search/index/?q=%2A&rows=30&authIdHal\\_s=julien-meyer](https://hal.science/search/index/?q=%2A&rows=30&authIdHal_s=julien-meyer)

[Will be published as: Meyer, J. & Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

**Victor Manfredi** is currently a research affiliate in the African Studies Center of Boston University. He works on the comparative grammar of the Benue Kwa languages of the Niger Congo family and the social anthropology of southern Nigeria and adjacent regions, and he has taught about these subjects since 1980 at a dozen universities in Nigeria, Europe and North America. His publications can be consulted at this page: <https://people.bu.edu/manfredi/writing.html>.