HAL
open science

Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication

Frank Seifart, Julien Meyer, Sven Grawunder, Laure Dentel

To cite this version:

Frank Seifart, Julien Meyer, Sven Grawunder, Laure Dentel. Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. Royal Society Open Science, 2018, 5 (4), pp.170354. 10.1098/rsos.170354 . hal-01968842

HAL Id: hal-01968842

https://hal.science/hal-01968842v1

Submitted on 3 Jan 2019

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.

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Seifart F, Meyer J, Grawunder S, Dentel L. 2018 Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *R. Soc. open sci.* **5**: 170354.
http://dx.doi.org/10.1098/rsos.170354

Received: 24 April 2017
Accepted: 15 March 2018

Subject Category:
Psychology and cognitive neuroscience

Subject Areas:
psychology/acoustics/behaviour

Keywords:
drummed speech, speech rhythm, Bora language, Amazonia, talking drums, intervocalic duration

Authors for correspondence:
Frank Seifart
e-mail: f.c.seifart@uva.nl
Julien Meyer
e-mail: julien.meyer@gipsa-lab.fr

†These authors contributed equally to this study.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4058798.

THE ROYAL SOCIETY PUBLISHING

Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication

Frank Seifart[1,2,†], Julien Meyer[3,4,†], Sven Grawunder[5,6] and Laure Dentel[7]

[1]Amsterdam Center for Language and Communication, University of Amsterdam, 1012 CX Amsterdam, The Netherlands
[2]Department of Linguistics, University of Cologne, 50923 Köln, Germany
[3]Université Grenoble Alpes, CNRS, GIPSA-lab, 38000 Grenoble, France
[4]Linguistics Division, Museu Paraense Emilio Goeldi, Belém, Brazil
[5]Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[6]Department of Linguistics, Kiel University, 24118 Kiel, Germany
[7]The World Whistles Research Association, Paris, France

FS, 0000-0001-9909-2088; JM, 0000-0002-4747-1884

Many drum communication systems around the world transmit information by emulating tonal and rhythmic patterns of spoken languages in sequences of drumbeats. Their rhythmic characteristics, in particular, have not been systematically studied so far, although understanding them represents a rare occasion for providing an original insight into the basic units of speech rhythm as selected by natural speech practices directly based on beats. Here, we analyse a corpus of Bora drum communication from the northwest Amazon, which is nowadays endangered with extinction. We show that four rhythmic units are encoded in the length of pauses between beats. We argue that these units correspond to vowel-to-vowel intervals with different numbers of consonants and vowel lengths. By contrast, aligning beats with syllables, mora or only vowel length yields inconsistent results. Moreover, we also show that Bora drummed messages conventionally select rhythmically distinct markers to further distinguish words. The two phonological tones represented in drummed speech encode only few lexical contrasts. Rhythm thus appears to crucially contribute to the intelligibility of drummed Bora.

© 2018 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

Our study provides novel evidence for the role of rhythmic structures composed of vowel-to-vowel intervals in the complex puzzle concerning the redundancy and distinctiveness of acoustic features embedded in speech.

1. Introduction

The human voice can produce rich and varied acoustic signals for transmitting information, but normally only up to a few hundred metres. Drum communication systems like that of the Bora people of the northwest Amazon can extend this range by a factor of up to 100 through emulating speech in sequences of drumbeats. But using acoustically simple signals that propagate well in natural environments across long distances comes with the challenge of being able to represent only a fraction of the acoustic features of spoken language, as produced by the human voice, without rendering speech incomprehensible [1]. Systems of long-distance communication emulating speech with drummed beats or whistles developed by traditional societies thus represent unique natural laboratories for exploring the relative distinctiveness of various linguistically relevant acoustic features. Although the importance of speech emulation for a better understanding of the architecture of human language has long been recognized [2], existing descriptions of drummed speech are still largely impressionistic and lack the methodological rigour of modern phonetics. Previous studies have highlighted two features of drummed speech: (i) 'enphrasing', i.e. elaborating words and sentences to make them longer and less ambiguous, and (ii) the representation of linguistic tone [3–11], but have almost entirely neglected exploring the role of speech rhythm, despite the clearly rhythmic nature of drumming. For Bora, it has even been claimed that rhythm plays no role in drummed speech [12], wrongly as we will argue.

The rhythmic organization of speech is complex and an area of extensive experimental research and lively theoretical debate [13]. Currently debated issues include what should be identified as the carrier of rhythmic beat in spoken languages [14], whether vowel-to-vowel intervals represent basic units of speech rhythm [15] and what is the role of rhythmic structures in language comprehension [16]. Against this background, we ask here which aspects of rhythmic structure are selected in conventionalized drummed speech to optimally contribute to its intelligibility.

To address this question, we carried out a case study on the Bora drum communication system from the northwest Amazon, locally known as *manguaré*, investigating in detail the prosodic features in common between spoken and drummed speech. Drummed Bora represents Bora words with a number of drumbeats that equals the number of its syllables. We focus on word-internal rhythmic units delimited by beats and especially their durations, as explained later in this Introduction. We analyse a corpus of drummed Bora messages showing how drummed speech selects and combines tonal and rhythmic features of the Bora language to communicate a wide variety of messages. In addition to an experimental study on rhythm, we carried out qualitative analyses of tonal, lexical and grammatical properties of drummed and spoken Bora. Taken together, these constitute the most comprehensive linguistic analysis of a drum communication system so far. The present study is part of a larger project that involved an extensive multimedia documentation of this traditional drummed practice and spoken Bora [17], and a methodology to analyse drummed speech that was developed based on a smaller set of *manguaré* data [18].

1.1. Emulated speech systems and drummed speech

Drummed speech systems are attested in Africa, South America, Asia and Oceania [3,4]. They are one type of emulated speech system, all of which operate by transforming spoken language as produced by the human voice into sounds such as whistles or acoustic signals produced with musical instruments. Such systems have also been called 'speech surrogates' [4] or 'speech substitutes' [19] in the older literature. Of these techniques, drummed speech can not only reach the longest distances but also employs the most radical reduction in acoustic complexity in all three main dimensions of acoustic signals: frequency, amplitude and time. By comparison, whistled speech consists of a much less radical acoustic reduction, primarily in the frequency domain, and can still represent, for example, vowel qualities [1,20].

Drummed signals exploit the natural bio-acoustic properties of percussions for optimal sound propagation in natural environments [21]. Their low pitch frequencies are not blocked by large

vegetation [22,23] and their high amplitudes and narrow frequency band reduce further noise-masking effects [24,25].

Drummed emulations of speech should be distinguished from various other types of drummed signalling systems. Firstly, there are systems that use repertoires of drummed codes to symbolically represent meaning, with no iconic relationship to the sound of a spoken language, as often found in Melanesia [26,27]. Secondly, there are musical drumming traditions that use nonsense syllables (so-called vocables) that iconically represent musical sounds for mnemotechnical purposes, but never carry linguistic meaning, as in the North Indian Tabla tradition [28]. By comparison, a purely symbolic association is also made in the occidental solfège to represent notes of the scale by the syllables *do*, *re*, *mi*, *fa*, *sol*, *la*, *ti*. Finally, drummed emulations of speech are also different from systems like the Morse code, which symbolically represents units of a writing system, and does not iconically represent speech sounds.

1.2. Tone, rhythm and enphrasing in drummed speech

The most comprehensive review on drummed speech [4] provides information on 18 different forms of drummed speech, mostly by reprinting older descriptions up to the first half of the twentieth century, on e.g. Ewe, Twi/Akan, Banen (all from West Africa), Chin (from Burma) and an early, short report on Bora [5]. From these descriptions and the few published since on this subject (e.g. [29]), it emerges that such systems most frequently represent two acoustic properties of spoken language: tonal contrasts and rhythmic structures. In fact, drummed speech has been attested almost exclusively for tonal languages, although one rare exception has recently been described [6,30,31]. In tonal languages, drummed speech systems render abstract, underlying phonological tone categories through different drum pitches. Level tones, e.g. high versus low, are represented by single beats on drums with different pitches, and contour tones, i.e. rising or falling, by a succession of two beats of different pitches.

The representation of rhythm in drummed speech has barely been addressed in previous research, beyond the observation that beats correspond to syllables and that longer pauses occur at boundaries between phrases or sentences [7,8]. As far as we know, only two previous studies have analysed rhythmic aspects of drummed speech in some detail: Nketia [9] found that in Akan (Ghana) drummed speech, closed syllables are represented by longer pauses between beats than open syllables followed by a voiced consonant or just a vowel. More recently, Cloarec-Heiss [10] argued that for drummed Banda-Linda (Central African Republic), the liquid consonants [l, r] and the labial fricative [v] are represented by shorter intervals between beats than other consonants in drummed speech. However, both of these studies remain impressionistic in that they do not provide any quantitative phonetic measurements.

Finally, it has been noted that drummed speech has an intrinsically limited semiotic potential due to the reduced acoustic distinctiveness of the drummed signal when compared with human speech. This leads to 'drum homophones', i.e. different spoken words that sound the same in the drummed mode [32]. As far as we know, all drummed speech systems counter this by 'enphrasing', i.e. short words that would come out as homophones in drumming are replaced by longer, less ambiguous expressions [4], often with poetic creativity [33].

1.3. Speech rhythm

Durational patterns contribute to shaping speech rhythm even in languages where contrastive duration is not employed to signal phonological oppositions, e.g. vowel length. Perceptually, rhythm appears to be built on temporal structures that rely on longer intervals, as opposed to shorter ones, and the expectation of regular patterns by listeners results in perceiving sequential events as being grouped together into higher-level patterns [34]. Such grouping is common to both language and music and constitutes an essential step in the interpretation of complex sound sequences. Temporal structure refers to the durational patterning of events, which is measured by the time intervals between successive events. The rhythmic structure of speech involves a temporal organization into intervals of various types, such as syllables or stress [28], which may be emulated in drummed speech. The duration of such intervals and their potentially hierarchical organization are determined by a complex set of factors, some of them language-specific [13]. To investigate speech rhythm in the sense of a sequence of intervals with (potentially) different durations, measurements of syllable (as well as phoneme) durations have been carried out on a range of languages with the aim of characterizing languages as a whole as belonging to distinct rhythmic types [35] as well as to characterize individual speakers [36]. Measures for this purpose include the standard deviation of the duration of vocalic (ΔV) and consonantal (ΔC)

intervals, the percentage to which speech is vocalic (%V) [37] and the pairwise variability index [38], a measure of the degree to which adjacent syllables or vocalic/consonantal intervals contrast in duration. This research revealed, firstly, that languages systematically differ with respect to durational contrasts between syllables, as well as between consonantal and vocalic segments of the syllables. Results of such measurements have also been used to investigate whether languages tend to approximate rhythmic isochrony for syllable durations or inter-stress interval durations [39,40]. But, this classification remains contested [41,42]. Secondly, it was shown that speakers from a very early age on can discriminate languages based on low-pass filtered speech samples preserving rhythmic properties [43].

Rhythmic units associated with syllables also play an important role in grammar in the sense that many languages categorize syllables, treating one class as heavier than another, for purposes such as stress placement and poetic rhyme [15]. Such weight distinctions usually, but not always, correlate with phonetic duration [44]. Syllable weight is often binary and described in terms of moras (light syllables: mono-moraic; heavy syllables: bi-moraic). The moraic model must assume that the weight of coda consonants is subject to language-specific parametrization representations designed to account for languages that treat both CVV and CVC as heavy, while others treat only a subset of these syllable types (usually CVV) as heavy. Note also that consonants in syllable onsets do not contribute to weight in terms of moras [44]. However, the concept of mora cannot easily cope with the fact that weight hierarchies can be highly complex, beyond two or three levels, and that onsets may, in fact, contribute to syllable weight [45], a famous example being C_{voiced} V < $C_{voiceless}$ V < VV < C_{voiced} VV < $C_{voiceless}$ VV in Pirahã [46].

It has recently been proposed that the units of speech rhythm that encode weight for the purpose of stress placement might be vowel-to-vowel intervals (delimited by the beginning of each vowel, see further discussion and illustration below) rather than syllables, so far with mixed results [15,47,48]. In any case, weight distinctions of vowel-to-vowel intervals also correlate with phonetic duration [48]. This proposal is anchored in the strong contribution of the vowel onset as an acoustic correlate for the perception of the beginning of speech events [49,50]. Other studies suggest that the relevant rhythmic units are most accurately delimited by the p-centre [51], roughly, the perceptual 'beat' of the syllable, which approximates the beginning of the vowel but anticipates it slightly for longer onsets [52]. This is supported by findings showing that syllable onsets do contribute to syllable weight, but less so than codas, for the perception of syllable duration and for stress placement [51,53]. Accordingly, probabilistic scales of weight differences between rhythmic units (as a function of onset length) have been proposed that successfully predict stress placement [51].

From the literature reviewed above, we know that speakers are sensitive to fine-grained durational contrasts of intervals associated with syllables for language discrimination and stress placement. We also know that rhythmicity of speech contributes to intelligibility by allowing for predictability of important information contained in syllable nuclei or stressed syllables [54]. However, it has hardly if ever been directly investigated so far to what extent durational contrast of rhythmic units might enable speakers to link acoustic input to meaningful elements, i.e. words and phrases, and if so, which kind of units can perform this function. It is likely that durational contrasts of rhythmic units can fulfil this function for two reasons. Firstly, many languages employ durational contrasts at the level of consonant and vowel phonemes. For instance, two length contrasts in vowels are common [55], but three are also well attested [56]. Secondly, experimentally reduced speech (noise-band vocoding, sine-wave speech and acoustic chimaeras) remains highly intelligible after removing acoustic cues distinguishing consonants and vowels but preserving temporal cues, including the rhythmic contours resulting from the frequency and amplitude pattern of the resonance peaks of a natural utterance [16,57–59].

1.4. Bora *manguaré* drumming

The Boras are an indigenous group of approximately 1500 members residing traditionally in small communities centred around multi-family roundhouses in the Amazonian rainforest of Colombia and Peru [12,60]. *Manguaré* drums are an important element of Bora culture [61] and were traditionally present in every roundhouse. They are used for communication within and across communities, up to 20 km away. Bora *manguaré* is one of the very few ancient drum communication systems attested in South America [62], where it developed independently of potential influence from Africa through the colonial slave trade [63]. In principle, every adult Bora man can use the drums and every community member can understand *manguaré* messages, both of which are acquired without explicit training. Sources from the early twentieth century report that *manguarés* were in daily use at the time, including for conversations about almost any subject and that messages were relayed from one roundhouse to another to reach

Figure 1. *Manguaré* drum pair in Brillo Nuevo, Peru (Photo Julien Meyer/Laure Dentel).

further distances [64]. Today, the approximately 20 *manguaré* drums still in existence among the Boras are gradually falling out of use, just as spoken Bora is giving way to local Spanish.

Bora *manguaré* drums are pairs of wooden slit drums carved from single logs (traditionally by burning, due to lack of metal instruments), each approximately 2 m long (figure 1). The drums are hung in parallel in a wooden structure and the drummer stands in the middle to beat them with two wooden mallets covered with natural rubber. Each drum can produce two pitches (one on each side of the slit) and therefore the pair of drums can produce four pitches. The smaller drum produces higher pitches than the larger one. Not all uses of *manguaré* drums iconically emulate speech. Two uses can be identified in Bora [12] and characterized in terms of an earlier description of Akan drummed speech [9]:

1. The 'musical mode' of *manguaré* is used to perform memorized drum sequences with no or very little variation as part of rituals and festivals. These melodic, musical sequences make use of all four pitches. A subtype of 'musical mode' sequences are associated with syllables of sung lyrics (the 'singing mode'). Other sequences do not correspond to syllables of spoken languages at all, like instrumental music without lyrics. In the 'musical mode', drumbeats follow a fixed, isochronous rhythm.
2. The 'talking mode' of *manguaré* is used to transmit relatively informal messages and public announcements, e.g. to ask someone to bring something or to come to do something (figure 2), to announce the outcome of non-alcoholic drinking competitions (figure 3) and to announce the arrival of visitors or the return of the chief of a roundhouse. In this mode, only two pitches (one of each drum) are used and each beat corresponds to a syllable of a corresponding phrase of spoken Bora. The informative nature of these messages puts pressure on both the emitter and the listener for an effective transmission of meaning and therefore the 'talking mode' is of most interest to the present study. Note that 'talking mode' messages can be preceded by 'musical' sequences, during which both drums may be played simultaneously, thus fulfilling the function of ringtones.

TYPE	ADDRESSEE			MESSAGE CONTENT	END
	CLAN	MARKER	NAME		
ík^{l}òòkáré now *tsà-ʔíɦk^{l}à* come-REP 'Come now!'	*mégʷààhí* hawk	*tùùtáβààbè* male clan member	*ìóóβèhì*	*mè-gʷák^{l}ùté-ʔíɦk^{l}à-kì* we-go_fishing-REP-PURP 'to go fishing'	*ídʒùrè ádʒìíb^{l}è* now liar *nèʔdín^{l}è* don't_say 'now don't say (that I am) a liar'
	ín^{l}èhé palm	*tùùtáβàdʒè* female clan member	*dʒííʔ^{l}ò*	*ííbìì mè-déíɦk^{l}ù-ʔíɦk^{l}à-kì* coca we-chew-REP-PURP 'to chew coca'	
ík^{l}òòkáré now *tsíβà-ʔíɦk^{l}à* bring-REP 'Bring now!'	*núʔbàhé* sun	*tùùtáβáábè-úβù àh^{l}úgʷànííʔ^{l}ò* clan's daughter	*kúgʷáhìbíɦkà*	*úmè-núʔbà-úβù-kè mè-kóβátsò-ʔíɦk^{l}à-kì* salt-sun-DEC-ACC we-burn-REP-PURP 'to burn salt (lit. deceased sun of salt)'	
	òʔgʷámú fox		*dʒíɦk^{l}àù*	*ííbíí-ʔààmí-nè-úβù mè-gʷárí?kò-ʔíɦk^{l}à-kì* coca-leaf-PLURAL-DEC we-toast-REP-PURP 'to toast coca leafs (lit. deceased coca leafs)'	
	bòánámú anaconda		*úút^{l}ágʷà*	*ìámé-tùùtáβààbè bìirí-mú-níígʷákò-úβù-kè mè-dó-ʔíɦk^{l}à-kì* animal-clan_member agouti-PLURAL-hair-DEC-ACC we-eat-REP-PURP 'to eat agouti (lit. deceased hair of agoutis, damaged animal)'	
	pàʔtúgʷám^{l}ú palm		*márímúdʒàʔáhtsì*	*ìámé-tùùtáβààbè néébá-nìiβúgʷà-úβù-kè mè-dó-ʔíɦk^{l}à-kì* animal-clan_member annatto-deer-DEC-ACC we-eat-REP-PURP 'to eat deer (lit. deceased annatto deer, damaged animal)'	
	tóódʒèmé Ocaina		*ròʔdɛ́íbà*	*dááʔìβétsó tsàpàtííjà dááʔìβétsó tsàpàtííjà* dry sneakers dry sneakers 'Dry the sneakers! Dry the sneakers!'	
	(etc.)		(etc.)	(etc.)	

Figure 2. The structure of *manguaré* 'calling messages'. Segmentation into meaningful elements and translation of each of them, in addition to idiomatic translations, are provided. Abbreviations: REP: repetitive; PURP: purposive; DEC: deceased; ACC: accusative case.

1.5. Bora phonology

Importantly for the study of *manguaré*, 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 [12], which are expressed by fundamental frequency pitch contours (figure 3*a*). Bora has six phonemic vowels, /*i*, *ɨ*, *u*, *e*, *a*, *o*/ (/*u*/ and /*e*/ are phonetically [*ɯ*] and [*ɛ*]). The Bora plain consonants /*p*, *b*, *t*, *d*, *ts*, *dz*, *tʃ*, *dʒ*, *k*, *g*, *g*^{w}, *ʔ*, *β*, *r*, *m*, *n*, *h*/ almost all have palatal(ized) counterparts, e.g. /*p*^{j}, *b*^{j}, *t*^{j} /. From now on, we will adopt the phonemic transcription without slashes (/). Bora has a (C)V(h/ʔ) syllable structure, i.e. an optional consonant (C) in the syllable onset, an obligatory vowel (V) and an optional consonant in syllable coda position, which can only be *h* or the glottal stop *ʔ*. Therefore, the only consonants preceding other consonants in Bora are the glottal fricative *h* and glottal stop *ʔ*, which also occur in other positions. This restricted distribution is due to a historical development from geminate consonants [66], but, synchronically, *h* and glottal stop *ʔ* behave like other consonants of the language. Additionally, Bora phonologically distinguishes short versus long vowels (written as sequences of two identical vowel symbols), but these long vowels cannot be followed by *h* or *ʔ* within one syllable. Bora phonology distinguishes light, i.e. mono-moraic (C)V syllables from heavy, i.e. bi-moraic (C)VC and (C)VV syllables for various phonotactic and morphophonological purposes [12].

1.6. Aims of this study

To understand the role of rhythm in drummed Bora, we firstly tested which rhythmic units are represented in drummed Bora based on measurements taken in corpora of drummed and spoken Bora. Secondly, we carried out qualitative analyses of the tonal, lexical and grammatical aspects of spoken and drummed Bora that interact with rhythmic structure. Regarding the specific hypothesis tested in the experimental study, it is clear that in *manguaré* messages, the number of beats equals the number of syllables. However, this leaves open various possibilities regarding which consonants preceding or following a vowel are included in the rhythmic units that are represented as intervals between beats and thus determine drummed interbeat durations (IBDs). We test primarily the two main competing theories of rhythmic units, the syllable and the vowel-to-vowel (V-to-V) interval. Through testing the syllable hypothesis, we are also able to test other derived possibilities, namely the 'vowel length-only' hypothesis, derived from a remark in the earliest, brief description of Bora *manguaré* [5], and the 'mora' hypotheses, derived from Bora phonotactics. As mentioned above, V-to-V intervals are delimited by the beginnings of vowels (and so are vowel length-only intervals and mora intervals), i.e. the syllable nuclei that carry the peak of acoustic energy in both spoken language and drummed speech (figure 3). The syllable and V-to-V types of units correspond to different segmentations of all speech segments into rhythmic units.

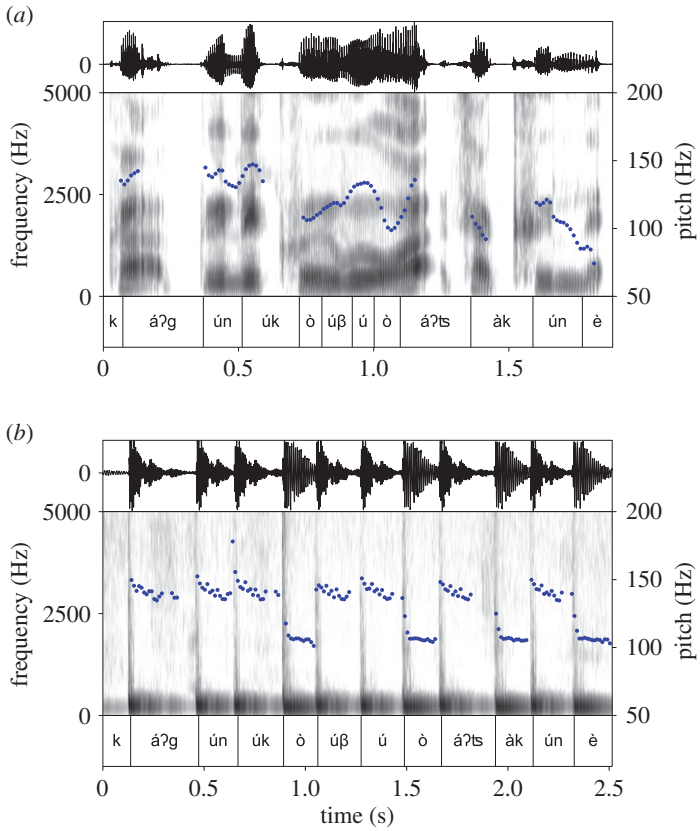

Figure 3. Acoustic properties of the Bora phrase *káʔgúnúkòúβú ò áʔtsàkúnè* 'I am finishing the cahuana (manioc starch drink)', in spoken (*a*) and drummed (*b*) versions. Oscillograms (top rows of *a* and *b*) represent acoustic energy as sound pressure over time. Spectrograms (middle rows) represent the spectrum of frequencies that is interpretable as linguistically distinctive acoustic information. Blue dotted lines indicate pitch. Bottom rows provide transcription and segmentation into vowel-to-vowel intervals (see the text for explanation). Figure created in PRAAT [65]. Corresponding sound files are provided in electronic supplementary material, S3 and S4.

	g^w	á	r	ɨ́	ʔ	k	ò	ʔ	í	h	k^j	à	k	ì
(*a*)	CV		CVC			CV		CVC			CV		CV	
(*b*)		VC		VCC			VC		VCC			VC		

Figure 4. Alternative segmentations of Bora *g^wárɨʔkòʔíhk^{j}àkì* 'in order to toast' into (*a*) syllables and (*b*) V-to-V intervals. Note that word-initial consonants are associated with preceding V-to-V intervals and word-final vowels with the following ones within phrases.

They yield the two following different associations between drummed IBDs and phonemes:

a. *Syllable*: IBDs are associated with the content of corresponding syllables (figure 4, row (*a*)). Possible syllable types in Bora are V, CV, VC (Vh and Vʔ only), CVC (CVh and CVʔ only), VV and CVV (where VV represents long vowels). (Note that the vowel length categories V and VV as well as the moraic categories V and {VV,VC} can be derived from the syllable case.)
b. *V-to-V interval*: IBDs are associated with the phonemes present between the beginning of one vowel up to the beginning of the following vowel (figure 4, row (*b*)). Possible V-to-V interval types in Bora are V, VC, VCC (VhC and VʔC only), VV and VVC. Note that VV intervals, i.e. long vowels followed directly by another vowel with no intervening consonant, are exceedingly rare in Bora. The V-to-V type VV did not occur at all in the drummed corpus and was very rare in the spoken corpus and could therefore not be tested.

We hypothesize, firstly, that Bora drumming represents V-to-V intervals, based on the iconic association of prominent acoustic peaks in drumming and spoken language, rather than syllables, only vowel length or only moras, and that these rhythmic units should be consistently respected across

different drummers. Secondly, we hypothesize that the V-to-V intervals we identified in drumming should also be observable in spoken Bora. Thirdly, we hypothesize that phonological tones play a minor role in the intelligibility of drummed messages when compared with rhythmic units. Finally, we hypothesize that special lexical and grammatical devices are employed in drummed message to support the intelligibility of drummed Bora. Below, we address these hypotheses in roughly reverse order because lexical, grammatical and tonal characteristics provide necessary background for discussing the rhythmic properties.

2. Material and methods

2.1. Materials

The analysed data originate from an extensive corpus of *manguaré* and spoken Bora collected as part of a large-scale project documenting Bora linguistic and cultural heritage [17]. The corpus contains 169 *manguaré* drummed messages that were performed by a total of five native Bora speakers. These were recorded on video and non-compressed audio using a Sony ECM-MS907 microphone. Messages contain on average of 15 words and 60 drumbeats. The majority of the messages were systematically elicited on the basis of a survey of the types of drum messages that are in common use to ensure coverage of message types, typically occurring content and of different drummers. Also included are some messages that were performed spontaneously, e.g. when calling community members for a communal fishing event.

All drummed messages were transcribed, i.e. sequences of drumbeats were annotated with corresponding Bora phrases, and translated together with a native Bora speaker while doing on-site fieldwork using ELAN [67]. This ensured that all messages included in the analysis are interpretable for Bora speakers. We excluded from the analysis drummed data in which the number of drumbeats or their tone category did not match the intended Bora phrase due to 'slips of the drum' that were identified during fieldwork together with a native speaker (approx. 5% of the recorded messages). All data have been archived and are accessible online [17]. They represent, by far, the most comprehensive database accessible online of any drummed speech system so far.

We then selected a set of recordings of spoken Bora for comparison with Bora *manguaré*. These recordings are comparable with drummed messages in terms of, e.g. spontaneity, and were produced by the same Boras who contributed the drummed data. They include spoken versions of some of the phrases and personal names that typically occur in drummed messages, as well as recordings of Bora names of zoological species. These data were then analysed with the same methodology as for the drummed data.

2.2. Design and procedure

For the experimental study, each recorded drum message was analysed as a function of its associated phonemic content (vowels, consonants and phonological tone) as annotated during the field documentation. For each drumbeat, the pitch (high or low) and the exact time position (based on the point with the highest acoustic energy) were automatically identified using PRAAT [65]. This identification was manually double-checked by two researchers. Similarly, a measure point was manually set at the vowel onset in spoken data, again with double-checking. This measure point was set at the third cycle of oscillation of the vocalic segment, in accordance with phonetic standards [68]. In the drummed data, the H or L tone categories of the transcribed vowels matched with approximately 140 Hz (high-pitched drum) and 95 Hz (low-pitched drum), respectively, in the drummed signal, a difference which is easily perceptible by the human ear [21]. The duration of each interval was calculated from the time stamps and each interval was linked to the sequence of vowels and consonants of spoken Bora that it represents, as can be seen in figure 3.

For the analysis of the drummed corpus, we first selected the recordings of the drummer who had contributed the most data (our 'first drummer'), who is also the one with the highest proficiency in drumming according to the Bora population in Peru. This subcorpus contains 2030 word tokens and 7459 intra-word intervals, i.e. pairs of drumbeats. To investigate consistency of the results across drummers, we selected the data of the drummer (our 'second drummer') who had contributed the second-largest amount of data in our corpus (1886 intra-word intervals). We identified a high rate of repetition in drummed messages, e.g. only 231 different word types in total in the subcorpus of the first drummer. The corpus of spoken Bora, produced by the same two drummers/speakers, consists of 557 word tokens with 147 different word types and 1278 intra-word intervals for the first drummer/speaker and 333 word tokens with 299 different word types and 731 intra-word intervals for the second drummer/speaker. To

study some aspects of spoken–drummed correspondence in more detail, we identified 14 different word types that occurred in both our spoken and drummed corpora.

We statistically analysed the variations in the qualitative IBD variable as a function of several explanatory variables, including their interactions. Details of each model are provided in the next section. Additionally, durational differences between some classes of words were tested by applying non-parametric tests, because the normality assumption for a *t*-test was violated in these sub-samples. Details of these tests are provided in electronic supplementary material, S1-d.

3. Results and discussion

3.1. Grammar and lexicon of *manguaré* messages

In our analysis of the 'talking mode' of *manguaré* drumming, we found that messages follow a fairly rigid structure (or 'syntax'), with a fixed order of building blocks of various types. The structure of the most frequent message type, 'calling messages', is represented in figure 2. Within this structure, the elements '(message) type', 'marker' and 'end' in particular are more or less invariable, formulaic elements signalling the structure of a message. New information is principally communicated in the slots corresponding to the addressee's clan and proper name and especially in the 'message content', in which a potentially infinite number of different messages can occur.

We found that Bora *manguaré* messages apply three types of enphrasing devices, i.e. conventional ways of making words and phrases longer and less ambiguous. Firstly, with few exceptions, nouns and verbs are marked with special disyllabic markers. On nouns, the marker *-úβù* is used (which in spoken Bora is used on nouns referring to deceased humans), for instance in *ííbíí-ʔààmɨ-nè-úβù* (coca-leaf-PLURAL-DECEASED) 'coca leaves' (figure 2). For verbs, the marker *-ʔíhkʲà* is used (which in spoken Bora signifies repeated action), for instance in *tsíβà-ʔíhkʲà* (bring-REPETITIVE) 'Bring!' In drummed messages, these markers do not carry any semantic value, but function purely to identify the preceding sequences of beats as representing nouns or verbs. Such marking of grammatical information through enphrasing in drumming has, to our knowledge, so far not been described.

Secondly, the 'marker' element *tùùtáβà-àbè/-ʤè* (literally 'the damaged one') functions to identify preceding sequences as clan names and following sequences as proper names. Note that Bora speakers have no intuitions why elements with the literal meanings 'deceased', 'repeated' and 'damaged' should be used in *manguaré* message. This supports that they have been selected because of their rhythmic properties, as discussed below.

Thirdly, there are conventional long forms for words that occur frequently in *manguaré* messages to render them less ambiguous, as has been described for other drummed speech systems before [4,32]: for instance, the Bora noun referring to a commonly hunted deer species *nííβúgʷà* is replaced in *manguaré* messages with *ìámé-tùùtáββààbè néébá-nííβúgʷà-úβù*, literarily 'deceased annatto deer, damaged animal'.

3.2. The representation of phonological tones

We found that, as expected, *manguaré* messages represent Bora high (H) and low (L) tones by beats on the small versus large drum, but not using the different tongues on each drum. The blue dotted lines representing pitch contours in figure 3*a* show that the realization of underlying tones in spoken Bora is modulated by an intonation contour with a characteristic 'downdrift' marking phrase boundaries. This information is not represented in drummed Bora, but only the two underlying tone categories, high and low (figure 3*b*). Note also that maximum amplitude levels of drumbeats are fairly constant and thus do not represent potentially informative amplitude variations of spoken language. Drummed high and low tones do display two different types of amplitude envelopes, namely a more rapid amplitude decline for high tone beats. But, this is intrinsically linked to the pitch properties of the smaller and less resonating hollow log used to produce high tone beats, which cannot be varied.

Our analysis of spoken Bora also confirmed that tone patterns carry primarily grammatical information, e.g. nominalization and phrase boundaries, and only very rarely differentiate words with different lexical meanings from each other. In addition, the occurrence of tone patterns is highly restricted by phonotactic rules which (i) prohibit sequences of two low tones except at the end of tone phrases (which may encompass various words) and (ii) impose word-final high tone in phrase-medial position and word-final low tone in phrase-final position [12]. Consequently, a trisyllabic Bora word can typically only have three out of eight logically possible tone patterns: HHH, HLH and LHH phrase-medially and HLL, HHL and LHL phrase-finally, i.e. a sequence of three tones in a given position in a phrase

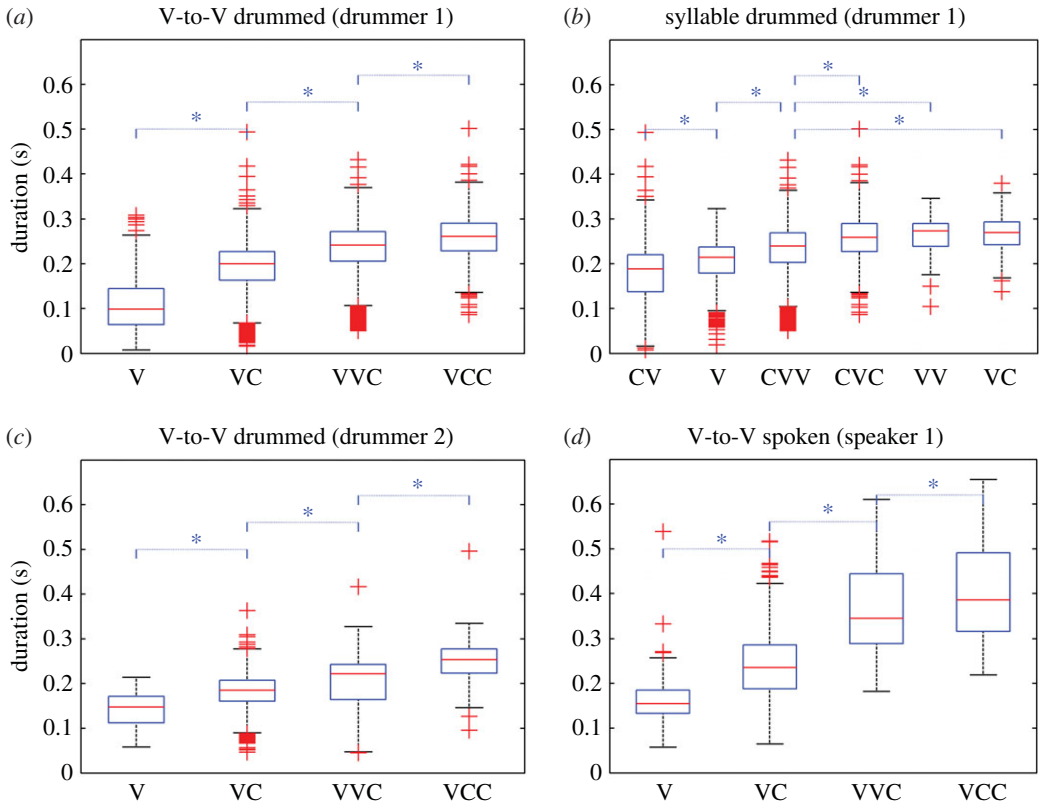

Figure 5. Distribution of drummed IBDs as a function of V-to-V types (*a*) and syllable types (*b*) for the first drummer, drummed IBDs as a function of V-to-V types for the second drummer (*c*) and durations of spoken V-to-V types (*d*) (VV, long vowel; CC, hC or ʔC). Further results are provided in electronic supplementary material, S1.

may represent any one out of one-third of all trisyllabic Bora words. For decoding drum messages, tone patterns thus facilitate parsing by indicating phrase boundaries and possibly encoding some other grammatical meanings. But for communicating lexical content, the distinctiveness of tone patterns is very limited.

3.3. Rhythmic structure

To find out which rhythmic units are represented in drummed Bora interbeat durations, we evaluated statistically which kind of hypothesis (syllable versus V-to-V; figure 5) explains best the IBD temporal distribution by constructing two different linear models of IBD (models 1 and 2, see tables 1 and 2) in R [69] and using the *lm* function to perform linear analyses. In each of these models, we included two variables as fixed factors: the variable DRUMMER and alternatively the variables corresponding to each one of the studied hypothesis (SYLLABLETYPE, V-TO-VTYPE). As there were only two drummers in this study, we included the variable DRUMMER as a fixed effect [70]. Using nested modelling techniques and *F*-tests, we performed a backward selection and we verified for each model that both DRUMMER and SYLLABLE (in model 1) and both DRUMMER and V-TO-VTYPE (in model 2) and their interactions had a significant effect on the IBD variations. Visual inspection of residual plots of both models did not reveal any obvious deviations from homoscedasticity or normality, nor the presence of influential outliers. Next, we performed multiple comparisons. For each of the models for which it was needed, the multiple comparisons were done with Hothorn *et al.*'s method [71] using the function *glht* from the package *multcomp* of R. This method allows us to perform multiple comparisons of means after building appropriate matrices of contrasts (using the function *lsmeans* from the package *lsmeans*). This ensures that the overall type I error associated with the simultaneous decisions does not exceed the pre-specified significance level (here 0.05, as usual) by adjusting *p*-values [69,71,72]. Finally, we compared the models by using the adjusted R^2 values obtained for each model, as they indicate the part of variation in the IBD variable that is explained by the explanatory variables of each model, taking into account their number of parameters.

Table 1. Results for predictors and interactions present in model 1 (SYLLABLETYPE (V, CV, CVV, CVC, VV, VC) for each DRUMMER (1 and 2)).

variable	d.f.	sum square	mean square	*F*-value	Pr(>*F*)
DRUMMER	1	0.1265	0.12647	40.2258	2.367×10^{-10}
SYLLABLETYPE	5	8.4501	1.69002	537.5405	$<2.2 \times 10^{-16}$
DRUMMER:SYLLABLETYPE	5	0.1248	0.02497	7.9418	1.772×10^{-7}

Table 2. Results for predictors and interactions present in model 2 (V-TO-VTYPE (V, VC, VVC, VCC) for each DRUMMER (1 and 2)).

variable	d.f.	sum square	mean square	*F*-value	Pr(>*F*)
DRUMMER	1	0.1249	0.1249	42.427	7.716×10^{-11}
V-TO-VTYPE	3	10.2952	3.4317	1165.560	$<2.2 \times 10^{-16}$
DRUMMER:V-TO-VTYPE	3	0.1763	0.0588	19.964	6.801×10^{-13}

In the model testing the syllabic segmentation, the factors were SYLLABLETYPE (six levels: CV, VC, CVC, CVV, V and VV) and DRUMMER (two levels, 1 and 2). Nested modelling tests lead us to keep the interaction of second order between SYLLABLETYPE and DRUMMER ($F_{59\,326} = 7.94$, $p < 0.001$). The adjusted R^2 value was 0.2279.

Multiple comparisons between different levels of SYLLABLETYPE (see electronic supplementary material, table S1-4) showed that V, CV and CVV were significantly different in duration, while VV, VC and CVC were not significantly different from each other, although these three types were all significantly longer than V, CV and CVV (figure 5). As a consequence, the IBD scale for syllable types is assumed as follows: CV < V < CVV < {CVC, VV, VC}, with undetermined relative durations of the last three. While also distinguishing a total of four types of durational units, this scale does not match our expectations regarding the representation of consonants and vowels. In particular, consonants in onsets of syllables influence IBDs in a non-expected way: CV is significantly shorter than V, and CVV shorter than VV. In both of these cases, this would mean that withdrawing an onset consonant would add more duration to the syllable. An additional consequence of this syllable hypothesis is that drummed CVV syllables would be shorter than VC syllables, which is unexpected [73,74].

The IBD scale for syllable types suggests that heavy, that is, bi-moraic (C)VV and (C)VC syllables always correspond to longer IBDs when compared with light mono-moraic (C)V syllables. However, among the 30 comparisons between different syllable types (see electronic supplementary material, table S1-4), seven are not coherent with the hypothesis that only the contrast between mono-moraic and bi-moraic syllables would be transposed by the drums in IBDs. For instance, VC and VV syllables (both bi-moraic) are significantly different but should not be according to this hypothesis, just as CVV and VV syllables. In data from drummer 2, mono-moraic V syllables are not different from bi-moraic VV syllables. As a consequence, the binary opposition mono-moraic (C)V versus bi-moraic (C)VV and (C)VC syllables disregards the contrast between VC and VV sequences that we find in IBDs, as well as the contribution of syllable onsets to the duration of the preceding interval.

The hypothesis that only phonological vowel length is represented in drummed IBDs could also be tested using these results. By exploring the multiple comparisons between different levels of SYLLABLETYPE (see electronic supplementary material, table S1-4), we found that among the 30 comparisons, 21 are not consistent with this hypothesis. Some are inconsistent because the intervals compared should be different according to this hypothesis (e.g. VC versus VV) but are not, others because they are different but should not be (e.g. V versus VC) and yet others because they display an effect in the unexpected direction (V is longer than VV).

For testing the V-to-V hypothesis, the variables were V-TO-VTYPE (four levels: V, VC, VCC and VVC) and DRUMMER (two levels: 1 and 2). We added in the initial model their interaction of second order. The nested modelling test led us to keep the interaction between V-TO-VTYPE and DRUMMER ($F_{3,9336} = 19.97$, $p < 0.001$). The adjusted R^2 value was 0.2777.

Multiple comparisons between different levels of V-TO-VTYPE showed that differences between all categories were statistically significant for both drummers (see figure 5 and the electronic supplementary material, table S1-5) resulting in a scale of IBDs as follows: V < VC < VVC < VCC. This scale reflects effects of both vowel length (VVC longer than VC) and of the number of consonants (VCC longer than VC and VC longer than V).

Taken together, these findings constitute evidence that the rhythmic units encoded in drummed Bora correspond more closely to V-to-V intervals (as annotated in figure 3) rather than to syllables, vowel lengths or moras. They show that consonants preceding a vowel are irrelevant for the IBD of the rhythmic unit associated with this vowel, while the number of consonants (including consonants that syllabify with following syllables) and the vowel length within V-to-V intervals are associated with significantly different IBDs, which are ordered in a coherent scale.

3.4. Consistency across drummers

Each Bora drummer has his own drumming style and rate, which may depend on factors such as arm length and body height. Therefore, we analysed data from different drummers separately. Comparison of results from both drummers shows that IBDs are consistently represented across drummers, except in one case for the syllable hypothesis. This confirms that rhythmic structure is represented consistently in drumming, and even more so for the V-to-V hypothesis of drummed Bora. (See data from the second drummer in figure 5*c* for V-to-V types and electronic supplementary material, S1 for syllable types.) Multiple comparisons showed that the same four V-to-V types were significantly different across drummers, and therefore that these followed the same scale, despite clearly observable individual differences (compare figure 5*a* with figure 5*c*). For syllable segmentation, the data of the second drummer displayed the same relative distribution and inconsistencies as for drummer 1, except that in data from drummer 2 V and CVV categories were not statistically different (see electronic supplementary material, figure S1-1 and table S1-4).

3.5. Rhythmic disambiguation in drumming

The standard procedure in linguistics to prove the distinctiveness of a phonetic feature is the identification of so-called minimal pairs, i.e. words that are distinguished only by this single feature. For example, *pet* versus *bet* proves that voicing of the consonants *b* versus *p* is distinctive in English because these two words have separate meanings. Following this procedure, we found many examples that prove that rhythmic structure is distinctive in drummed Bora. For instance, the three proper names *(ʤ)ì.íʔʲ.ò*, *(n)èèp.áhʲ.ù* and *(r)òʔʤ.ɨ́b.à* are all trisyllabic and have the tone pattern LHL. However, they contrast in the first V-to-V interval type they contain, V (*.ì.*) versus VVC (*.èèp.*) versus VCC (*.òʔʤ.*), which allows Boras to identify these names in drummed messages. Note that proper names occur in the same slot in the message scheme (figure 2) and thus there is no contextual information for further disambiguation except for the preceding clan name, which narrows down possible proper names. The distinctiveness of rhythmic structures compared to tonal patterns can be illustrated with the set of 18 trisyllabic proper names that occur in our drummed data (see electronic supplementary material, S1-c). In this set, seven rhythmically distinct sequences of intervals occur (e.g. V.VC.V versus VC.VCC.V), narrowing down the choice to on average 2.5 intended proper names. By contrast, only two distinct tonal patterns occur in this set, corresponding to on average nine distinct proper names each.

Quasi-experimental situations during transcription of drummed data, when short sequences of beats were presented in isolation, further show that Bora listeners reliably differentiate elements that are distinguished only by minimal rhythmic contrasts. An example here is *mègʷákʲùtéʔíhkʲàkì* 'to go fishing' versus *mèkóóβàtéʔíhkʲàkì* 'to bring fire wood', which is tonally identical but rhythmically differentiated by the second interval (*ákʲ* VC versus *óóβ* VVC).

Maybe the most striking examples of rhythmic contrasts are the special markers for verbs (*-ʔíhkʲà/á*) and nouns (*-úβù/ú*). As they have identical tone patterns (HH phrase-medially and HL phrase-finally), they must have been selected in the development of drummed Bora because of the distinctiveness of the rhythmic interval types involved, namely V (preceding *-úβù/ú*) versus VC (preceding *-(ʔ)íhkʲà/á*) and VCC (*íhkʲ*) versus VC (*úβ*). Both of these interval pairs are indeed significantly different in mean duration in our data (see electronic supplementary material, S1-d).

Finally, Bora 'drum homophones' (i.e. words with identical tone patterns and rhythmic structures) that appear in similar contexts are conventionally elaborated in 'enphrasing'. For instance, names referring to the commonly hunted animals *bɨ̀ɨ̀rɨ́mù* 'agouti' and *nìì̀βúgʷà* 'deer' (both LHL and

Table 3. Results for predictors and interactions present in model 3: V-TO-VTYPE (V, VC, VVC, VCC) for each DRUMMER (1, 2) and SPEECHTYPE (drummed versus spoken).

variable	d.f.	sum square	mean square	*F*-value	Pr(>*F*)
SPEECHTYPE	1	8.247	8.2474	2365.8540	$<2.2 \times 10^{-16}$
DRUMMER	1	0.434	0.4336	124.3909	$<2.2 \times 10^{-16}$
V-TOVTYPE	3	18.550	6.1833	1773.7490	$<2.2 \times 10^{-16}$
SPEECHTYPE:DRUMMER	1	0.162	0.1621	46.5061	9.599×10^{-12}
SPEECHTYPE:V-TOVTYPE	3	2.791	0.9305	266.9219	$<2.2 \times 10^{-16}$
DRUMMER:V-TO-VTYPE	3	0.150	0.0501	14.3702	2.424×10^{-9}
SPEECHTYPE:DRUMMER:V-TO-VTYPE	3	0.037	0.0123	3.5294	0.0142

(C).VVC.VC.V) are replaced in Bora drummed messages with longer expressions that include contrastive rhythmic structures and tone patterns (figure 2).

3.6. Comparison to spoken Bora

The tone patterns of drummed Bora represent underlying, abstract tone categories, which clearly have categorical status in the grammatical system of the language [12]. Our results on durational drummed interval types suggest that these similarly may represent abstract units, defined by different numbers of consonants and vowel length. But unlike for tones, the structural status of these intervals in the language is less clear. For instance, they are not relevant units for phonotactic rules, which are determined by syllables. To find out whether they have phonetic reality in spoken Bora, we applied the same analysis as for drummed Bora (see figure 5*d* for speaker 1 and electronic supplementary material, figure S1-1 for speaker 2). For this verification of V-to-V type for both spoken and drummed speech and their interactions (model 3, table 3), the variables were DRUMMER (two levels: 1 and 2), SPEECHTYPE (two levels: Drummed (D) and Oral (O)) and V-to-VTYPE (four levels: V, VC, VVC and VCC). The nested modelling test led us to keep the interaction between V-TO-VTYPE, DRUMMER and SPEECHTYPE ($F_{3,11\,305} = 3.52$, $p = 0.0142$). Multiple comparisons between different levels of V-TO-VTYPE showed that differences in durations between all four categories were statistically significant for both drummers/speakers also in spoken data (see electronic supplementary material, table S1-6) and that these followed the same scale of IBDs observed in drummed data.

Another way to approach this question is to compare rhythmic properties of words that occurred in both the spoken and the drummed corpora (14 items for the first speaker/drummer). Here, we found that the relative durations between consecutive V-to-V intervals in the drummed form always corresponded to those in the spoken form in each individual word. Results for three words are given in figure 6. For instance, the word *múútsúʔòhtsɨ́ne* contains the V-to-V intervals *úúts* (VVC), *úʔ* (VC), *òhts* (VCC) and *ɨ́n* (VC), among which the expected relative durations VC < VVC < VCC were realized in both spoken and drummed forms.

We thus conclude that the rhythmic structure of drummed Bora not only represents abstract, underlying units but also matches surface realizations of spoken Bora. To distinguish four interval types, drummed Bora makes use of the possibility of gradual distinctions in interval length, as against the strictly two-way distinction in pitch that *manguaré* drums are restricted to.

4. Conclusion

Conventionalized systems for emulating speech using drums are natural speech registers that have developed independently across the world and are produced and comprehended with great ease. However, how exactly they encode speech has so far remained poorly understood and has led to various misconceptions [12]. Here, we show for drummed Bora the complex interplay between the representations of linguistic tone, speech rhythm and enphrasing, and highlight the crucial role of durational rhythmic units. Regarding enphrasing, we show that this is not only a device for the simple

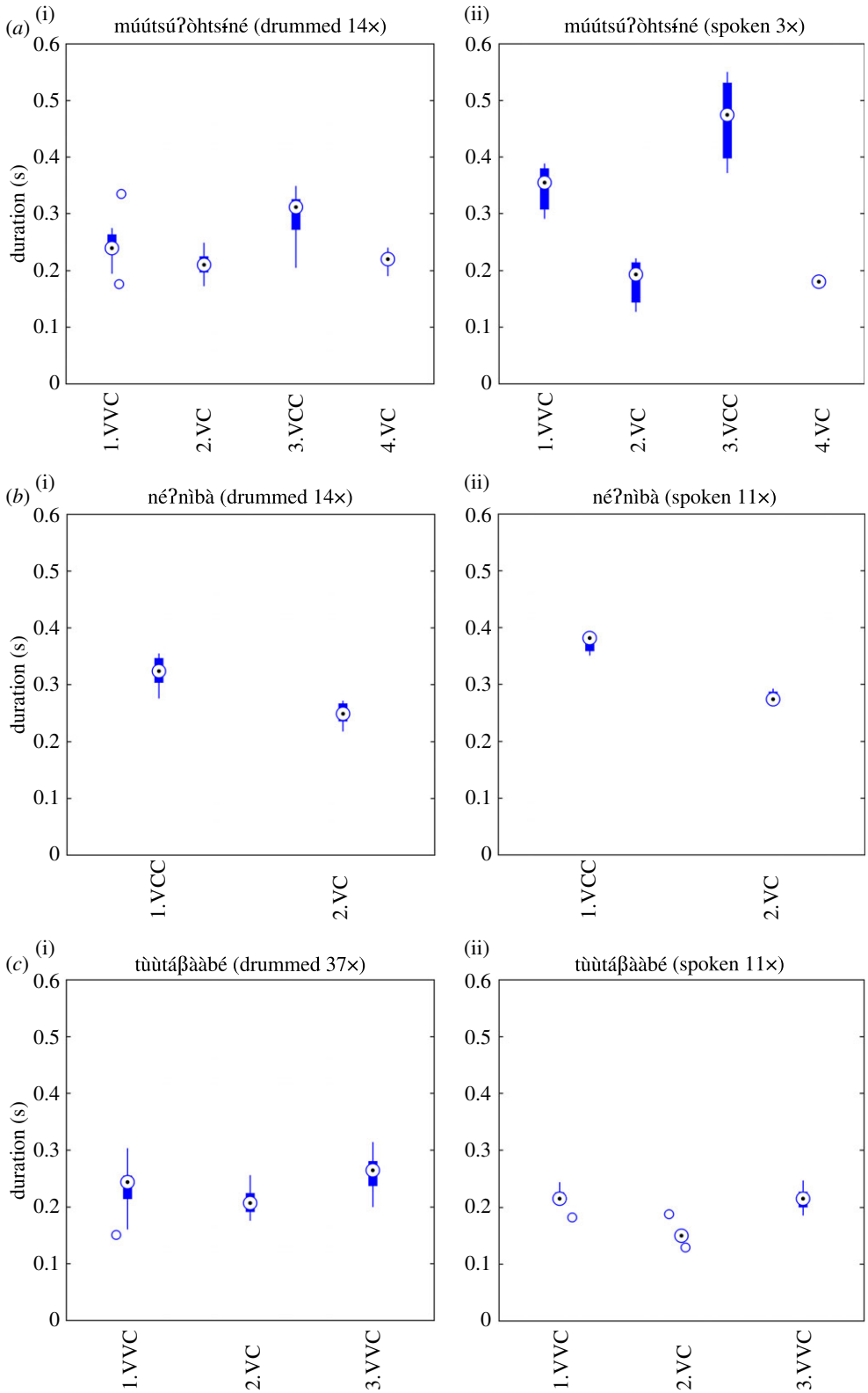

Figure 6. V-to-V durations as a function of V-to-V types for three Bora words (*a*–*c*) in their drummed (i) and spoken (ii) forms (VV: long vowel; CC: hC or ʔC).

replacement of words for longer, less ambiguous expressions [32], but that it can also be used to mark grammatical categories such as nouns versus verbs.

Regarding tone, our current knowledge of systems emulating speech with drums suggests that such systems overwhelmingly develop in tone languages. Interestingly, Bora's neighbouring, non-tonal

languages Ocaina and Witoto also use *manguaré* drums, but only for restricted, symbolic signalling or in a 'musical mode'. This suggests that phonological tone may be a feature that strongly favours the development of such systems. However, our results suggest that representing tone is not sufficient for an efficient system with the expressive range of Bora *manguaré*, because tone patterns alone are not particularly distinctive in this language. This appears to be the case also in other languages with drum communication systems. The typology of linguistic tone suggests that in languages with two or three contrastive tones (as against four or more), tone tends to have a low functional load and rarely distinguishes lexical items [75], also in West African languages [76] for which drummed speech is attested.

Regarding rhythm, our study highlights the importance of non-isochronous rhythmic structure embedded in speech, as against isochronous rhythmic structures that aid parsing of utterances into words, syllables and phonemes [28,77,78]. We found that Bora *manguaré* exploits the semiotic potential of non-isochronous rhythmic structure in speech for linking acoustic input to meaningful elements, i.e. for word recognition, and is in line with previous experimental evidence for the actual exploitation of such dynamic temporal patterns in speech comprehension [16,58,79].

We resolved two crucial questions regarding the nature of such rhythmic patterns in speech, as represented in drumming. Firstly, we provided evidence that the rhythmic units encoded in drummed Bora match vowel-to-vowel (V-to-V) intervals more closely than syllables, i.e. that drumbeats align with vowel nuclei irrespective of onset consonants within the same syllable, which are associated with preceding intervals. The fact that drummed Bora, as a naturally evolved practice, selects to represent these units lends credibility to the concept of V-to-V intervals in the rhythmic structure of spoken languages. Previous support for vowel-to-vowel (V-to-V) intervals comes, for instance, from phonetic research on tonal alignment [80], on the importance of vowel onsets as an acoustic correlate of speech rhythm [50] and research dealing with syllable weight as a unit measure of quantity (e.g. [81]), but also from psycholinguistic studies on the notion of perceptual centre (p-centre)—the perceived moment of the occurrence of an acoustic event—which is commonly non-congruent with the physical signal onset [52,82] and is more closely aligned to the beginning of the vowel than the beginning of the syllable [51,83]. Moreover, the vowel-to-vowel interval appears to explain other aspects related to speech rhythm such as poetic rhyme in different languages and can possibly also account for stress placement [47].

Secondly, we found four distinct rhythmic units in data from two Bora speakers, both in spoken and drummed forms of speech, representing different numbers of consonants and vowel length: V < VC < VVC < VCC. As shown by our data, these units are differentiated by approximately 20 ms, which is within the resolution of differentiation for the human perceptual and cognitive systems for e.g. acoustic events for beats in experimental situations [84,85] and for phonetic events in spoken signals [25,53,86].

We thus submit that the representation of this kind of rhythmic structure plays a crucial role in drummed speech to reliably differentiate elements that are only distinguished by rhythmic structure, and to successfully communicate entirely novel and unexpected messages, e.g. the one in the bottom row of figure 2.

Language is a redundant system, which simultaneously provides many cues to enable the comprehension of the message encoded by the speaker, including contextual, grammatical and phonetic cues. For example, detailed qualitative properties of phonological segments are perceived and processed in language comprehension [87], such as the resonances differentiating vowels or the noise associated with fricative consonants, as visualized in the spectrogram in figure 3*a*. The current study shows that under certain circumstances, such qualitative distinctions can be entirely dispensed with. This is indicated by the empty spaces between beats in the spectrogram in figure 3*b*, where speech is reduced to rhythm combined with two pitch levels. To render drummed messages less ambiguous, drummed speech systems then elaborate lexicon, morphology and syntax. Such naturally evolved systems thus provide living models for a deeper understanding of the complexity of human language, regarding key questions in language sciences such as the interaction of redundancy, reduction and elaboration at different levels of linguistic structure to efficiently communicate linguistic meaning.

A number of directions of future research emerge from our study. Firstly, it seems promising to compare our results with those that could be obtained by adopting a dynamic approach to rhythmic units in speech, which accounts for potential (slight) mismatches between vowel onsets and p-centres [51]. The rhythmic units of spoken Bora could thus be defined dynamically by applying heuristic estimations of p-centres as including varying fractions of syllable onsets. Secondly, a better understanding of the limits of intelligibility of drummed Bora remains a desideratum and should be studied by controlled

comprehension experiments. Finally, the present study developed a methodology for the analysis of rhythmic structures of drummed speech, which can now be applied to other languages, particularly the several African languages that are known to have developed systems of drummed speech. We thus provide the opportunity for further exciting research on the role of rhythmic structures in drummed and spoken forms of language.

Ethics. The permission to work with Bora collaborators and collect data on the language and drum tradition was granted to F.S. by the Federación de Comunidades Nativas del Ampiyacu (FECONA), the indigenous association representing Bora people which has the authority for providing working permits and ethical controls of such activities in this area of Peru (permit number 04-Psdnte~FECONA-03, 7.11.2003). Data collection took place in 2003–2005 and no other specific authorization was required at that time in Peru. All informants provided informed consent and the study was reviewed and approved in accordance with the DOBES Code of Conduct V2 (http://dobes.mpi.nl/ethical_legal_aspects/DOBES-coc-v2.pdf) and the Declaration of Helsinki.

Data accessibility. Our data are provided in open access in the documents electronic supplementary material S1.pdf and S2.xlsx (electronic supplementary material, files 1 and 2). These data correspond to measurements carried out on audio files and associated annotations of drummed data from the Bora language documentation online archive at https://hdl.handle.net/1839/00-0000-0000-000C-1B20-E@view (see detail in electronic supplementary material, S1-a).

Authors' contributions. F.S. proposed the study and also collected and annotated the spoken and drummed data with the Bora expert drummers and analysed grammatical and lexical properties. J.M. designed the research methods for the phonetic analysis and proposed to test the V-to-V hypothesis. L.D. and J.M. reviewed the drummed data to correct temporal alignment between linguistic annotations and drummed beats and processed the data analysis and developed analytic tools. L.D. and S.G. implemented detection of drummed beats. F.S and S.G. segmented spoken data. J.M., L.D. and S.G. performed statistical analyses. J.M., L.D. and F.S. analysed data. F.S., J.M., S.G. and L.D. wrote the paper and prepared electronic supplementary material files.

Competing interests. We declare we have no competing interests.

Funding. Financial support came from Volkswagen Foundation (DoBes grants II/80 110 and II/83 522), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Pós Doutorado Junior Fellowship no. 159747/2010-4) and the 7th research program of the European Community FP7 2007-2013 (Marie Curie IIF Fellowship 630076 and a 2013–2014 MSCA-COFUND Fellowship of the European Institutes for Advanced Study (EURIAS) at the Collegium of Lyon (Réseau Français des Instituts d'Etudes Avancées)).

Acknowledgements. The authors thank the Bora drummers and speakers, and particularly the curaca Manuel Ruiz Miveco. F.S. and S.G. thank Bernard Comrie and the Max Planck Society for the support during the project. J.M. thanks Silvain Gerber of the Gipsa-Lab for the support and guidance during the implementation of linear models. We are grateful for the helpful comments from Anne Cutler, Anita Wagner, Paul Boersma, Roger Mundry, Christian F. Hempelmann, Nathalie Vallée and especially two anonymous reviewers and the editors of RSOS.

References

1. Meyer J. 2015 *Whistled languages*. Berlin, Germany: Springer.
2. Sapir E. 1921 *Language. An introduction to the study of speech*. New York, NY: Harcourt, Brace & Co.
3. Stern T. 1957 Drum and whistle 'languages': an analysis of speech surrogates. *Am. Anthropol.* **59**, 487–506. (doi:10.1525/aa.1957.59.3.02a00070)
4. Sebeok TA, Umiker-Sebeok DJ (eds). 1976 *Speech surrogates: drum and whistle systems*. The Hague, The Netherlands: Mouton.
5. Thiesen W. 1969 The Bora signal drums. *Lore* **19**, 101–103.
6. Tang P. 2007 *Masters of the sabar: Wolof griot percussionists of Senegal*. Philadelphia, PA: Temple University Press.
7. Carrington JF. 1949 *Talking drums of Africa*. London, UK: Carey Kingsgate Press.
8. Agawu VK. 1995 *African rhythm: a Northern Ewe perspective*. Cambridge, UK: Cambridge University Press.
9. Nketia JHK. 1976 Surrogate languages of Africa. In *Speech surrogates: drum and whistle systems* (eds TA Sebeok, DJ Umiker-Sebeok), pp. 825–864. The Hague, The Netherlands: Mouton.
10. Cloarec-Heiss F. 1999 From natural language to drum language: an economical encoding procedure in Banda-Linda, Central African Republic. In *Language diversity and cognitive representations* (eds C Fuchs, S Robert), pp. 145–157. Amsterdam, The Netherlands: John Benjamins Publishing Company.
11. Ames DW, Gregersen EA, Neugebauer T. 1971 Taaken sàmàarii: a drum language of Hausa youth. *Africa* **41**, 12–31. (doi:10.2307/1159675)
12. Thiesen W, Weber D. 2012 *A grammar of Bora with special attention to tone*. Dallas, TX: SIL International.
13. Fletcher J. 2010 The prosody of speech: timing and rhythm. In *The handbook of phonetic sciences* (eds WJ Hardcastle, J Laver, FE Gibbon), pp. 521–602, 2nd edn. Chichester, UK: Blackwell.
14. Mairano P. 2010 *Rhythm typology: acoustic and perceptive studies*. Torino, Italy: Università degli studi di Torino. See https://tel.archives-ouvertes.fr/tel-00654261.
15. Ryan KM. 2016 Phonological weight. *Lang. Linguist. Compass* **10**, 720–733. (doi:10.1111/lnc3.12229)
16. Xu L. 2016 Temporal envelopes in sine-wave speech recognition. In *Interspeech 2016, San Francisco, CA, 8–12 September*, pp. 1682–1686. Grenoble, France: International Speech Communication Association (ISCA). (doi:10.21437/Interspeech.2016-171)
17. Seifart F, Fagua D, Gasché J, Echeverri JA (eds). 2009 *A multimedia documentation of the languages of the people of the center. Online publication of transcribed and translated Bora, Ocaina, Nonuya, Resígaro, and Witoto audio and video recordings with linguistic and ethnographic annotations and descriptions*. Nijmegen, The Netherlands: The Language Archive. See https://hdl.handle.net/1839/00-0000-0000-001C-7D64-2@view.
18. Meyer J, Dentel L, Seifart F. 2012 A methodology for the study of rhythm in drummed forms of languages: application to Bora Manguaré of Amazon. In *Interspeech-2012, Portland, OR, 9–13 September*, pp. 687–690. Grenoble, France: International Speech Communication Association (ISCA).
19. Nöth W. 1995 *Handbook of semiotics*. Bloomington, IN: Indiana University Press.
20. Meyer J. 2016 'Whistled languages' reveal how the brain processes information. *Sci. Am.* **316**, 61–67.
21. Rossing TD. 2000 *Science of percussion instruments*. Singapore: World Scientific.
22. Bass HE, Sutherland LC, Zuckerwar AJ. 1990 Atmospheric absorption of sound: update. *J. Acoust. Soc. Am.* **88**, 2019–2021. (doi:10.1121/1.400176)

23. Padgham M. 2004 Reverberation and frequency attenuation in forests — implications for acoustic communication in animals. *J. Acoust. Soc. Am.* **115**, 402–410. (doi:10.1121/1.1629304)
24. Zwicker E. 1982 *Psychoakustik*. Berlin, Germany: Springer.
25. Green DM. 1985 Temporal factors in psychoacoustics. In *Time resolution in auditory systems* (ed. A Michelsen), pp. 122–140. Berlin, Germany: Springer.
26. Zemp H, Kaufmann C. 1969 Pour une transcription automatique des 'langues Tambourinés' mélanésiens. *L'Homme* **9**, 38–88. (doi:10.3406/hom.1969.367047)
27. Niles D. 2010 Editor's introduction to Zemp & Kaufmann. *Kulele Occas. Pap. Pac. Music Dance* **4**, 1–50.
28. Patel AD. 2008 *Music, language, and the brain*. Oxford, NY: Oxford University Press.
29. Kawada J. 1998 *La voix: étude d'ethno-linguistique comparative*. Paris, France: Éd. de l'École des hautes études en sciences sociales.
30. Tang P. 2008 Rhythmic transformations in Senegalese sabar. *Ethnomusicology* **52**, 85–97. (doi:10.2307/20174567)
31. Winter Y. 2014 On the grammar of a Senegalese drum language. *Language* **90**, 644–668. (doi:10.1353/lan.2014.0069)
32. Gleick J. 2011 *The information: a history, a theory, a flood*. New York, NY: Pantheon Books.
33. Finnegan RH. 1970 *Oral literature in Africa*. Oxford, UK: Clarendon P.
34. Fraisse P. 1956 *Les structures rythmiques: étude psychologique*. Louvain, Belgium: Publications Universitaires de Louvain, Erasme.
35. Ramus F, Nespor M, Mehler J. 2000 Correlates of linguistic rhythm in the speech signal. *Cognition* **75**, 265–292. (doi:10.1016/S0010-0277(00)00101-3)
36. Dellwo V, Leemann A, Kolly M-J. 2015 Rhythmic variability between speakers: articulatory, prosodic, and linguistic factors. *J. Acoust. Soc. Am.* **137**, 1513–1528. (doi:10.1121/1.4906837)
37. Ramus F, Mehler J. 1999 Language identification with suprasegmental cues: a study based on speech resynthesis. *J. Acoust. Soc. Am.* **105**, 512–521. (doi:10.1121/1.424522)
38. Grabe E, Low EL. 2002 Durational variability in speech and the Rhythm Class Hypothesis. In *Laboratory phonology 7* (eds C Gussenhoven, N Warner), pp. 515–546. Berlin, Germany: De Gruyter Mouton.
39. Pike KL. 1945 *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.
40. Abercrombie D. 1967 *Elements of general phonetics*. Chicago, IL: Aldine.
41. Bertinetto PM, Bertini C. 2008 On modeling the rhythm of natural languages. In *Proc. 4th Int. Conf. on Speech Prosody, SP 2008*, pp. 427–430.
42. Arvaniti A. 2012 The usefulness of metrics in the quantification of speech rhythm. *J. Phon.* **40**, 351–373. (doi:10.1016/j.wocn.2012.02.003)
43. Nazzi T, Bertoncini J, Mehler J. 1998 Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 756–766. (doi:10.1037/0096-1523.24.3.756)
44. Gordon M. 2006 *Syllable weight: phonetics, phonology, typology*. New York, NY: Routledge.
45. Topintzi N. 2010 *Onsets: suprasegmental and prosodic behaviour*. Cambridge, UK: Cambridge University Press.
46. Everett DL. 1988 On metrical constituent structure in Pirahã phonology. *Nat. Lang. Linguist. Theory* **6**, 207–246. (doi:10.1007/BF00134230)
47. 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.
48. Lunden A. 2017 Syllable weight and duration: a rhyme/intervals comparison. *Proc. Linguist. Soc. Am.* **2**, 33: 1–12. (doi:10.3765/plsa.v2i0.4084)
49. Allen GD. 1973 Segmental timing control in speech production. *J. Phon.* **1**, 219–237.
50. Allen GD. 1975 Speech rhythm: its relation to performance universals and articulatory timing. *J. Phon.* **3**, 75–86.
51. Ryan KM. 2014 Onsets contribute to syllable weight: statistical evidence from stress and meter. *Language* **90**, 309–341. (doi:10.1353/lan.2014.0029)
52. Morton J, Marcus S, Frankish C. 1976 Perceptual centers (P-centers). *Psychol. Rev.* **83**, 405–408. (doi:10.1037/0033-295X.83.5.405)
53. Goedemans RWN. 1998 *Weightless segments — a phonetic and phonological study concerning the metrical irrelevance of syllable onsets*. The Hague, The Netherlands: Holland Academic Graphics.
54. Lehiste I. 1977 Isochrony reconsidered. *J. Phon.* **5**, 253–263.
55. Donohue M, Hetherington R, McElvenny J, Dawson V. 2013 *World phonotactics database*. Canberra, Australia: Department of Linguistics, The Australian National University. See http://phonotactics.anu.edu.au.
56. Prehn M. 2012 *Vowel quantity and the fortis-lenis distinction in North Low Saxon*. Utrecht, The Netherlands: LOT. See: http://hdl.handle.net/11245/1.358624.
57. Remez RE, Rubin PE, Pisoni DB, Carrell TD. 1981 Speech perception without traditional speech cues. *Science* **212**, 947–949. (doi:10.1126/science.7233191)
58. Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. 1995 Speech recognition with primarily temporal cues. *Science* **270**, 303–304. (doi:10.1126/science.270.5234.303)
59. Smith ZM, Delgutte B, Oxenham AJ. 2002 Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87–90. (doi:10.1038/416087a)
60. Seifart F. 2005 *The structure and use of shape-based noun classes in Miraña* (*North West Amazon*). Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.
61. Gasché J. 2009 La sociedad de la 'gente del centro'. In *A multimedia documentation of the languages of the people of the center. Online publication of transcribed and translated Bora, Ocaina, Nonuya, Resígaro, and Witoto audio and video recordings with linguistic and ethnographic annotations and descriptions* (eds F Seifart, D Fagua, J Gasché, JA Echeverri), Nijmegen, The Netherlands: The Language Archive.
62. Izikowitz KG. 1935 *Musical and other sound instruments of the South American Indians: a comparative ethnographical study*. Göteborg, Sweden: Elander.
63. Nordenskiöld E. 1930 *Modifications in Indian culture through inventions and loans*. Göteborg, Sweden: Elanders boktryckeri aktiebolag.
64. Whiffen T. 1915 *The North-West Amazons: notes of some months spent among cannibal tribes*. London, UK: Constable.
65. Boersma P. 2001 Praat, a system for doing phonetics by computer. *Glot Int.* **5**, 341–345.
66. Seifart F, Echeverri JA. 2015 Proto Bora-Muinane. *Liames. Línguas Indígenas Am.* **15**, 279–311. (doi:10.20396/liames.v15i2.8642303)
67. Brugman H, Russel A. 2004 Annotating multimedia/multi-modal resources with ELAN. In *Proc. 4th Int. Conf. on Language Resources and Language Evaluation* (eds M Lino, M Xavier, F Ferreira, R Costa, R Silva), pp. 2065–2068. Paris, France: European Language Resources Association.
68. Turk A, Nakai S, Sugahara M. 2006 Acoustic segment durations in prosodic research: a practical guide. In *Methods in empirical prosody research* (eds S Sudhoff, D Lenertova, R Meyer, S Pappert, P Augurzky, I Mleinek, N Richter, J Schließer), pp. 1–28. Berlin, Germany: De Gruyter.
69. R Core Team. 2016 *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See http://www.R-project.org.
70. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin C, Robinson B, Hodgson DJ, Inger R. 2017 Best practice in mixed effects modelling and multi-model inference in ecology. *PeerJ Preprints* **5**, e3113v1. (doi:10.7287/peerj.preprints.3113v1)
71. Hothorn T, Bretz F, Westfall P. 2008 Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363. (doi:10.1002/bimj.200810425)
72. Lenth RV. 2016 Least-squares means: the R package lsmeans. *J. Stat. Softw.* **69**, 1–33. (doi:10.18637/jss.v069.i01)
73. Goldsmith JA. 1990 *Autosegmental and metrical phonology*. Oxford, UK: Basil Blackwell.
74. Blevins J. 1995 The syllable in phonological theory. In *The handbook of phonological theory* (ed. JA Goldsmith), pp. 206–244. Oxford, UK: Blackwell.
75. Ratliff M. 1992 Form and function in tone languages. In *18th Annual Meeting of the Berkeley Linguistics Society* (ed. LA Buszard-Welcher), pp. 134–144. Berkeley, CA: Berkeley Linguistics Society.
76. Bird S. 1999 Strategies for representing tone in African writing systems. *Written Lang. Lit.* **2**, 1–44. (doi:10.1075/wll.2.1.02bir)
77. Cutler A, Otake T. 2002 Rhythmic categories in spoken-word recognition. *J. Mem. Lang.* **46**, 296–322. (doi:10.1006/jmla.2001.2814)
78. Jadoul Y, Ravignani A, Thompson B, Filippi P, de Boer B. 2016 Seeking temporal predictability in speech: comparing statistical approaches on 18 world languages. *Front. Hum. Neurosci.* **10**, 568. (doi:10.3389/fnhum.2016.00586)
79. Salverda AP, Dahan D, McQueen JM. 2003 The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* **90**, 51–89. (doi:10.1016/S0010-0277(03)00139-2)
80. House D. 1990 *Tonal perception in speech*. Lund, Sweden: Lund University Press. See http://lup.lub.lu.se/record/528914.
81. Kato H, Tsuzaki M, Sagisaka Y. 2003 Functional differences between vowel onsets and offsets in

temporal perception of speech: local-change detection and speaking-rate discrimination. *J. Acoust. Soc. Am.* **113**, 3379–3389. (doi:10.1121/1.1568760)

82. Scott SK. 1998 The point of P-centres. *Psychol. Res. Psychol. Forsch.* **61**, 4–11. (doi:10.1007/PL00008162)
83. Otterbein S, Abel C, Heinemann LV, Kaiser J, Schmidt-Kassow M. 2012 P3b reflects periodicity in linguistic sequences. *PLoS ONE* **7**, e51419. (doi:10.1371/journal.pone.0051419)
84. Rammsayer TH, Lima SD. 1991 Duration discrimination of filled and empty auditory intervals: cognitive and perceptual factors. *Percept. Psychophys.* **50**, 565–574. (doi:10.3758/BF03207541)
85. Seifert U, Olk F, Schneider A. 1995 On rhythm perception: theoretical issues, empirical findings. *J. New Music Res.* **24**, 164–195. (doi:10.1080/09298219508570681)
86. Meyer J. 2005 *Description typologique et intelligibilité des langues sifflées, approche linguistique et bioacoustique*. Lyon, France: Cyberthèse.
87. Cutler A. 2012 *Native listening: language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.