frontiers in Communication

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fcomm.2021.652690

When Music Speaks: An Acoustic Study of the Speech Surrogacy of the Nigerian Dùndún Talking Drum

Cecilia Durojaye[1,2]*, *Kristina L. Knowles*[3], *K. Jakob Patten*[4], *Mordecai J. Garcia*[3] and *Michael K. McBeath*[1,2]

[1]Psychology Department, Arizona State University, Tempe, AZ, United States, [2]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany, [3]School of Music, Dance and Theatre, Arizona State University, Tempe, AZ, United States, [4]College of Health Solutions, Arizona State University, Tempe, AZ, United States

Yorùbá dùndún drumming is an oral tradition which allows for manipulation of gliding pitch contours in ways that correspond to the differentiation of the Yorùbá linguistic tone levels. This feature enables the drum to be employed as both a musical instrument and a speech surrogate. In this study, we examined four modes of the dùndún talking drum, compared them to vocal singing and talking in the Yorùbá language, and analyzed the extent of microstructural overlap between these categories, making this study one of the first to examine the vocal surrogacy of the drum in song. We compared the fundamental frequency, timing pattern, and intensity contour of syllables from the same sample phrase recorded in the various communicative forms and we correlated each vocalization style with each of the corresponding drumming modes. We analyzed 30 spoken and sung verbal utterances and their corresponding drum and song excerpts collected from three native Yorùbá speakers and three professional dùndún drummers in Nigeria. The findings confirm that the dùndún can very accurately mimic microstructural acoustic temporal, fundamental frequency, and intensity characteristics of Yorùbá vocalization when doing so directly, and that this acoustic match systematically decreases for the drumming modes in which more musical context is specified. Our findings acoustically verify the distinction between four drumming mode categories and confirm their acoustical match to corresponding verbal modes. Understanding how musical and speech aspects interconnect in the dùndún talking drum clarifies acoustical properties that overlap between vocal utterances (speech and song) and corresponding imitations on the drum and verifies the potential functionality of speech surrogacy communications systems.

Keywords: talking drum, speech surrogate, music, acoustics, dundun, Nigeria, pitch, rhythm

OPEN ACCESS

Edited by:
Yoad Winter,
Utrecht University, Netherlands

Reviewed by:
Maarten Mous,
Leiden University, Netherlands
Bruce Connell,
York University, Canada
Meagan Curtis,
SUNY Purchase, United States

****Correspondence:***
Cecilia Durojaye
cdurojay@asu.edu

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

Received: *12 January 2021*
Accepted: *18 June 2021*
Published: *27 July 2021*

Citation:
Durojaye C, Knowles KL, Patten KJ, Garcia MJ and McBeath MK (2021) When Music Speaks: An Acoustic Study of the Speech Surrogacy of the Nigerian Dùndún Talking Drum. Front. Commun. 6:652690. doi: 10.3389/fcomm.2021.652690

INTRODUCTION

Yorùbá *dùndún* drumming is a musical-oral tradition wherein the characteristic of the drum as a variable-pitched membranophone allows the manipulation of intensity and pitch.[1] in ways that can mimic the tones and gliding contours of the Yorùbá, a tonal language, spoken in south-west Nigeria. This unique feature enables the dùndún drum (commonly referred to as the "talking drum") to be

[1]Following Villepastour (2014), we will use pitch to reference the fundamental frequency produced by the drum and the fundamental of a speech utterance, and "speech tone" to refer to the relative pitch of a spoken syllabus in Yorùbá.

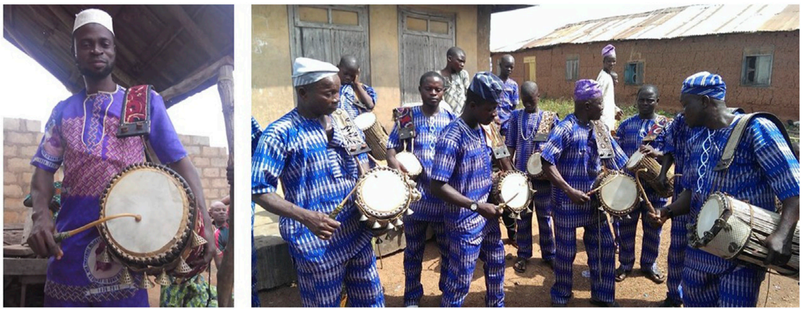

FIGURE 1 | Dùndún percussionists. Drummers can play individually, but more commonly in ensembles.

employed as both a musical instrument and a speech surrogate. The dual function of the drum, its role in the Yorùbá social-cultural milieu, and the belief that it is the most eloquent of Nigerian talking drums (Euba 1990) have thus drawn the attention of linguists, (ethno) musicologists, and anthropologists with a focus on various aspects such as the structure of the drum ensemble (Akpabot 1975; Vidal 2012; Durojaye 2020), the social and religious functions of the drum (Adegbite 1988), principles of dùndún communication (Arewa and Adekola 1980), and more. The dùndún is played both individually and, more commonly, in ensembles, as illustrated in **Figure 1**.

Prominent characteristics of the dùndún include its speech and musical features. The aesthetic and stylistic attributes of drum poetry inescapably share literary and musical space when imitating Yorùbá oral literature (Sotunsa 2009). *Bàtá* drums, a very close relative of the dùndún, use drum strokes as a code that translate into Yorùbá language (Villepastour 2010). Dùndún drummers, however, draw elements from music and speech to communicate emotions on the drum (Durojaye 2019a). The historical and social background around the dùndún tradition, organization and training of drummers, and the three main modes of the dùndún drum are addressed in Akin Euba's monograph *Yorùbá Drumming* (1990), the most extensive work on the dùndún to date.

Of particular importance to the present study is the observation that three drum tones are consistently used to imitate the relative tones of the Yorùbá language. However, when drum speech and human speech are compared, linguistic representations on the drum tend to differ from the original speech utterances (Euba 1990). Acoustical analyses confirm this result, demonstrating that the three distinct tones (Low, Mid, and High) are produced on a global level with three measurably different fundamental frequencies (Connell and Ladd 1990; Akinbo 2019). In a word-level analysis, however, drummers often produce the three tones slightly differently in a way that mimics the spoken language. For instance, a high tone produced before a low tone rises and is significantly higher than one preceding a mid tone (Connell and Ladd 1990; Laniran and Clements 2003). Moreover, Akinbo (2019) noted that the number of drum strikes tend to be correlated with the number of syllables present in either mono- or disyllabic words, However, this research was conducted with single words rather than full speech phrases. Akinbo's research also focused solely on the speech mode of drumming and not the drum as a surrogate for song.

In contrast to the previous work, the precise specific acoustic characteristics and relationships of dùndún drumming modes were examined in two sister articles. Durojaye et al. (2021) provides a *macrostructural* analysis of the overall distributional matches between two dùndún modes for acoustical metrics of intensity, fundamental frequency (f_0), timing, and an entropy measure of timbre. This timbre metric analyzes the slope of the spectral envelope to gauge changes in speech features (Toh et al., 2005). In the present study, we examine the *microstructural* correlations between individual Yorùbá vocalizations at the syllable level and expand the acoustic continuum to four modes of dùndún drumming. Here, we analyze the microtiming patterns, fundamental frequency (f_0) and intensity contours of individual drummers playing the same auditory sequence as one that is vocalized and compare the correlations between the different modes of production. The ongoing acoustic features of f_0 and intensity are fundamentally important to the characterization of music and language (McBeath and Neuhoff 2002; Patten and McBeath 2020; Patten et al., 2019; Yu et al., 2021) and aid the intelligibility of the dùndún communication (Euba 1990; Akinbo 2019). As there are various drums in the general dùndún category, we refer to the *ìyáàlù* (mother of the drum) as the dùndún in this study (see Durojaye 2019b, for detailed description from an emic perspective). Gaining a full knowledge of the communicative mechanisms of the dùndún and its role in the language-music relation benefits from contributions with various perspectives, including acoustical analyses and comparisons of different modes of Yorùbá vocalization and dùndún drumming.

The Yorùbá language, which the drum imitates, uses three relative speech tones: Low, Mid, and High.[2] While the H and L tones are realized as rising and falling intonations, in comparison,

[2] The Yorùbá standard orthography is employed for transcriptions: H tone is marked with an acute accent (´), L tone with the grave (`) and the M tone is usually unmarked. Diacritics such as ẹ/ε/, *ọ*/ɔ/, ṣ/ʃ/, *p*/k͡*p*/, and gb/g͡b/are also used for Yorùbá words.

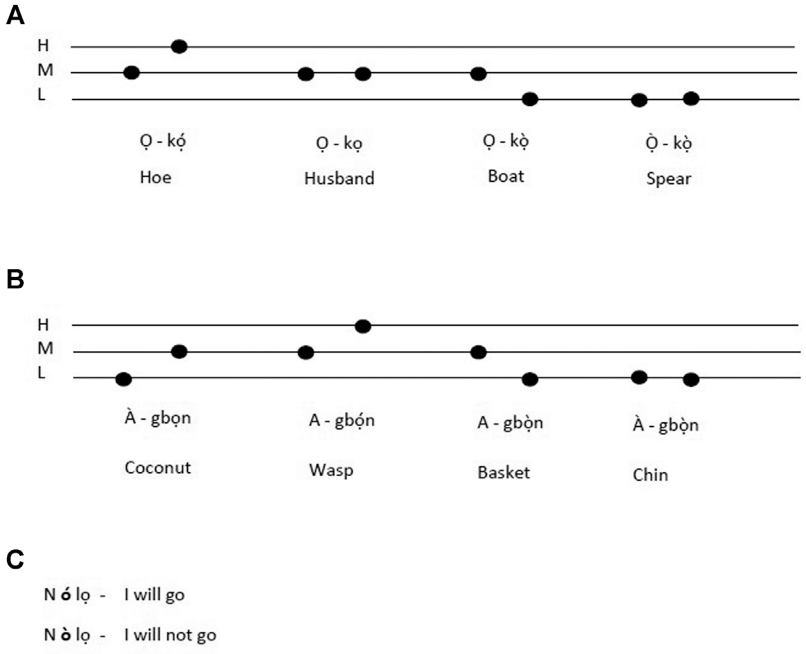

FIGURE 2 | Three examples exemplifying various standard f_0 glide patterns of the Yorùbá language.

the mid tone (M) is flat (its level remains the same and often represents the default tone of a speaker). In other words, the mid tone uses a relaxed laryngeal position, while the H and L tones necessitate muscular tension in the larynx to create a rising (H) or falling (L) tone. The relative speech tones are essential for word and sentence signification and even indicate opposing meaning, i.e., they are lexically and grammatically contrastive, as illustrated in the following examples shown in **Figure 2**.

Unlike monosyllables that have three tonal possibilities, disyllabic words have nine possible tonal patterns (HH, HM, HL, LH, LM, LL, MH, ML, and MM). Although one can employ all possible speech tone combinations in the language, this is, however, dependent on the particular utterance as a word meaning can be distorted if the wrong configurations are employed. (Arewa and Adekola 1980). For example, while different words can be produced from the word *agbọn,* only four tonal combinations are meaningful in the Yorùbá language.

Yorùbá language consists of twelve vowels (seven oral vowels and five nasal vowels), eighteen consonants, and a syllabic nasal/ n/.[3] Every syllable in Yorùbá language contains a vowel. The vowels constitute the Tone Bearing Unit (TBU) of the language and are also the essence of the tonal glides occurring in the language (Eme and Uba 2016). Glides, comprised of both rising and falling tones, occur when a syllable consists of more than one tone or when there are two adjacent vowels, as in *ọlọ́pàá* (the police) or *jọ̀ọ́* (please). Words ending in a vowel are often elided or assimilated when followed by words beginning with a vowel (Pulleyblank 2009). Since the majority of Yorùbá verbs end in a vowel and most nouns begin with a vowel, elision (the deletion of a vowel) is a common occurrence. For example, *gbé odó* (lift a mortar) becomes *gbódó*. Elisions also occur when a preposition is followed by a noun as in *ní ilé* (at home) to *níle*. In vowel assimilation, the assimilated vowel becomes the same as the vowel to which it was added, as in *Mo rí i* (I saw him/her), or *mo pè é* (I called him/her) (Barber and Oyetade 1998). As a result, these glides, elisions, or assimilation of vowels have an effect on the actual speech tone and the syllable duration (Villepastour 2014). Also, unlike the nasalized vowels which do not impact the musical setting, the syllabic nasal, like vowels, adds to the tone length (Villepastour 2014, 31).

Speech perception relies on multiple acoustic features produced by the vocal apparatus. Fundamental frequency has been identified as the major component of tone perception in Yorùbá (Hombert 1976). Intensity of the sound pressure wave of speech (perceived as loudness), which is correlated with syllable accent and emotionality in some non-tonal languages (Sluijter and van Heuven et al., 1996; Laukka et al., 2011), has been identified as a lesser component of tone in tonal languages including Yorùbá and Cantonese (Ọdẹ́jọbí, 2007). Lastly, f_0 and intensity are not perceived independently of one another. Perception of pitch is often correlated with intensity (Neuhoff and McBeath 1996; McBeath and Neuhoff 2002). The reverse is also true in speech production; as speakers raise the intensity of their voice, the f_0 also rises (Scharine and McBeath 2019). Intercorrelations between f_0, intensity, and timbre have also been observed (Patten and McBeath 2020). As mentioned, Yorùbá speech tones are specifically dependent on modulation of the f_0, which is perceived as pitch. As dùndún drummers are attempting to replicate speech, a multiply determined construct,

[3] Yoruba vowels include seven oral vowels a/a/, e/e/, ẹ/ɛ/, I/i/, o/o/, ọ/ɔ/, u/u/; five nasal vowels an/ã/, ẹn/ɛ̃/, in/ĩ/, ọn/ɔ̃/, and un/ũ/.

TABLE 1 | Characteristics of various techniques, execution, and sound types produced on the dùndún drum.

Dùndún sounds	Stick/syllabification technique	Execution
L and M tones	Free stroke (loose wrist movement; stick bounds off the drum head after playing the tone. One drum stroke per syllable.	Moderate pressure on strings (M) Minimal or no tension (L)
H tone	Muted stroke (firm wrist movement; stick does not bounce off the drum head). Muted stroke is also usually used in the speech mode. One drum stroke per syllable.	Maximum pressure on tension strings
Glides (from L to H or vice versa)	One drum stroke for two tone levels/syllables	L–H: Gradual tightening of tension strings immediately after executing the lower tone and maintains the pressure for the glided tone H–L: Gradual release of pressure on tension strings
Vowel assimilation	One drum stroke for two tone levels/syllables	Tightening or releasing of tensioning strings according to the tone contour
Singing	As used for speech as highlighted above. Sometimes the use of hand and stick technique.	Vibration of the arm in addition to loosening and tightening of the tension strings

with an instrument that can most readily vary in f_0 and intensity, it is unclear which elements of the speech signal drummers rely on to create an effective speech surrogate.

Dùndún drummers reproduce the Yorùbá speech tones by loosening or tightening the tension strings (*ọsán*) surrounding the drum's resonator and connecting the two drumheads from one end of the drum to the other. Although there are different dialects spoken among the Yorùbá, the dùndún only imitates the Ọ̀yọ́ dialect (Akpabot 1986; Durojaye 2019b), believed to be the standard Yorùbá (Euba 1990; Villepastour 2014). The *ìyáàlù dùndún* is carried with a shoulder strap which hangs from the player's (usually left) shoulder, wherein the combination of the left fingers, the wrist and hip bone are employed in the manipulation of the strings (*ọsán*). The other hand plays the drum using a stick known as *ọ̀pá* or *kọ̀ngọ́* (a curved stick shaped like lower case 't' without the cross). For the drum to produce the lowest pitch, minimal pressure is applied to the strings, and the more the pressure, the higher the frequency. Thus, the H speech tone is executed with the maximum pressure on the strings (**Table 1**). Given that the drum only captures vowels (the Yorùbá tones) and vowels occur in every syllable of the Yorùbá language, interpretation and analyses of drum rhythm and tone are usually in relation to syllables. The technique of representing syllables can take many forms such as 1) using one drum stroke for each syllable (as for a single tone level and vowel elisions); 2) many strokes for one syllable; 3) one drum stroke for two or more syllables and 4) one drum stroke for a syllable with many speech tone levels as would be the case for some glides, or assimilations (see also, Euba 1990).

Past research (for example, Euba 1990) has shown that the dùndún connects music and language through three modes. First, the "musical speech" mode in which strict or danceable rhythm is employed, which we refer to as *Drum–Dance Rhythm* (*D-DR*). Second, the "song form" where the drum is used to mimic vocal singing of a text, which we refer to as *Drum Singing* (*DS*). Third, "speech mode" in which the drum more strictly imitates speech and follows speech rhythm, which we divided into two sub-categories, referred to as *Drum Talking, Performative* (*DT-P*), and *Drum Talking, Direct* (*DT-D*). We added the subcategory of DT-D in order to test the drumming acoustical match with speech when drummers directly try to maximize their imitation of talking without musical constraints. The categories we compared are shown in **Table 2**. To our knowledge, no prior empirical studies have acoustically compared all of these drum modes to their corresponding vocally spoken or sung forms. The principal goal of this study is to determine the extent of microstructural acoustic representation of speech sounds on the drum, and second to distinguish and acoustically compare the four dùndún mode categories. Here we examine acoustic f_0, intensity, and microtiming patterns and test the microstructural relationship between the various modes of Yorùbá vocalization and dùndún drumming.

Our principal hypothesis, shown in **Table 2**, is that there are successively larger positive correlations between the patterns of (a) drum attack interonset intervals (i.e., duration from one attack to the next) and interonset intervals from one syllable to the next, and between the (b) drum and speech f_0, as the drum sequentially increases in extent of speech surrogacy from drum as song (DS) to direct talking (DT-D). Such a pattern validates the functional variance represented by the different drum modes. Our secondary hypothesis is that there is a progression toward musical rhythms across adjacent drumming modes (e.g., DT-D to DT-P to DS to DD-R) with the more musical modes featuring greater similarity in the average length of IOIs than more speech-like modes, which are more likely to feature non-isochronous rhythms. Together, the hypothesized pattern of findings would acoustically confirm that the dùndún drum spans the range of surrogacy communication systems between speech and music at a microstructural (syllable-by-syllable) level.

METHODS

Recording and Data Collection Procedure

Three Yorùbá vocal performers produced 10 examples each of both speech and singing, for a total of 60 vocalizations. Three independent professional drummers from three Yorùbá towns were randomly assigned 10 each of the previously recorded speech and singing examples and were asked to represent their assigned examples on the *dùndún*. Drummers performed 10 examples each of Drum-Dance Rhythm (D-DR), Drum Singing (DS), Drum Talking, Performative (DT-P), and Drum Talking, Direct (DT-D) for a total of 120 performances. Drum

TABLE 2 | Categories of speech and drumming that were compared.

Vocalizing	Drumming	
Vocal talking (*VT*)	Drum talking–Direct (*DT-D*)	↑Predicted larger vocal vs drum correlation
	Drum talking–Performative (*DT-P*)	
Vocal singing (*VS*)	Drum singing (*DS*)	
	Drum–Dance rhythm (*D-DR*)	

and vocal performances were matched such that there were 30 song/DS, 30 speech/DT-P, and 30 speech/DT-D pairs. D-DR performances did not have a matched vocal recording. All the performers were male and monolingual Yorùbá. Drummer A and C have approximately 35 years of experience of dùndún drumming, while drummer B has been playing for 28 years. The Yorùbá speakers also doubled as singers as they are well conversant with traditional Yorùbá songs.

Sample phrases were recorded at a professional studio in Ibadan, South-West Nigeria. After listening to each of the spoken utterances, the drummers were asked to replicate the same sequence in the corresponding drum language. The same procedure was followed for the songs and their drum representation. Each drummer, with their personal drums, reproduced ten samples each of the previously recorded spoken utterances and songs. For the song samples, performers were asked to choose from *dùndún* popular repertoire and were thus at liberty to perform similar pieces. For each of the speech samples, performers were asked to create two recordings, one of the drum as it is typically used to represent speech (DT-P) and one where the drummer tries to maximize their imitation of the speech example (DT-D). The recordings ranged between 5 and 10 s in length. The drum data were recorded in a soundproofed room with a SHURE SM57 dynamic microphone at a 3-inch distance from the drum. Spoken utterances and songs were recorded with Audio Technica AT 2035 cardioid condenser microphones. All recordings were made at a sampling rate of 44.1 KHz in WAV format and saved as separate files.[4] The recordings were then analyzed independently for timing, and f_0 and intensity information. In the initial analysis stage, recording errors resulting in missing data were discovered for two of the 180 recordings. As recordings were paired for vocal utterances and corresponding drum modes, we removed the data that was paired with the incomplete recordings, resulting in a total of five recordings omitted from analysis (one each of VT, DT-D, DT-P, VS, and DS), leaving a total of 175 recording samples in our final analysis.

Analysis of Timing in Speech and Dùndún Drumming

In order to allow for comparison of timing patterns between the drum excerpts and speech and song excerpts we measured interonset intervals (IOIs). An IOI is defined as the duration between successive event onsets, a standard measurement within studies on sensorimotor synchronization (Madison 2001; Repp 2005), expressive timing in music (Benadon 2006; Goldberg 2015; Ohriner 2018), and cognitive processes like free recall of items from memory (e.g., Rhodes and Turvey 2007; Patten et al., 2020). In keeping with standard methods for measuring IOIs in prerecorded music (Repp 1992; Ashley 2002), drum samples were uploaded into the sound processing program *Audacity* (Mazzoni 2021) and were then analyzed in two stages. In the first stage, the time point for each attack on the drum was marked in the recording. Since the physical onset of the sound envelope for a drum typically corresponds to the perceptual attack time, IOIs were calculated based on the physical onset of the sound which was verified in the sound file both visually and aurally. In the second stage analysis, onsets were coded for whether they were produced by a single strike on the head of the drum, or by utilizing rhythmic embellishments such as flams (short rapid note on the drum) or *àfikún* (additions), a pair of rapid 16th notes. Following musical conventions for both Western and Nigerian drumming, the second attack of a flam was marked as the onset for the drum stroke.

Excerpts containing verbal utterances (recordings of speech and song excerpts by performers) underwent a related process for analyzing timing information. Unlike drum attacks, the perception of which corresponds to the onset of the sound event, perceptual attack times in language are correlated with the onset of vowels, which are considered the syllable nucleus (Peterson and Lehiste 1960; Allen 1972; Greenberg et al., 2003). In order to calculate onsets of vowels in the speech (VT) and song (VS) samples, recordings were uploaded into Praat (Boersma and Weenink, 2021) and then parsed first for word boundaries, then for syllables, and finally vowels. Vowels were identified by the onset of characteristic vowel formants. Following onset identification, attention was given to the mid-point of the amplitude rise (approxmiately 50%) at the beginning of the vowel, which is broadly considered to be the "perceptual center," or p-center, of a vowel. (Pompino-Marschall 1991; Harsin 1997; Ohriner 2019). The p-center was used when calculating IOIs for speech and song excerpts.

Although Yorùbá is a language without diphthongs (Bamgbose 2000), assimilation (the approximation of two successive phonemes toward each other and away from their isolated pronunciation), and elision (the deletion of a phoneme) present similar challenges to phonetic segmentation. Some scholars have attributed an independent attack to each vowel in an assimilated or elided phrase; however, this refers to a descriptive, phonological end and not a perceptual one (Ola Orie and Pulleyblank 2002). The current paper relies on perceptual work regarding diphthongs to segment speech into

[4] Sample recordings provided in Results section.

syllabic intervals. While longer in duration, English diphthong vowels are not perceived as categorically different from their monophthong counterparts (Lehiste and Petersen 1961; Fox 1983). Diphthongs are also acoustically distinct from glides between two monophthong vowels in that diphthongs do not reach the formant range of their ending vowel portion but have a trajectory approaching it. This formant trajectory difference serves to differentiate a diphthong from a monophthong rather than cue a link between two vowels (Gay 1970). As such, diphthongs are typically treated as a single vowel in the speech timing analysis. In the current work, formant frequencies were used to identify vowel onsets and assimilations; clear vowel differences were marked as independent onsets, but formant signatures more indicative of a trajectory toward another vowel (a signature common to diphthongs and elisions) were treated as a single onset (Byrd 1992; Jun 2004). For example, the phrase "bi a" (if we) is two distinct words but is phonetically produced as [bjɑ]. As a result, when analyzing the timing of syllable onsets, elisions and assimilations were treated as a single onset. The duration of syllables, an analog of the drum IOI, was then calculated as the duration between successive vowels between syllables. We estimate that our method for identifying vowel onsets produces a degree of variability in the order of 5–10 ms, though this window of error may be marginally higher in the case of vowel onsets that are preceded by nasal or liquid consonants (e.g., [m] or [l]), as the vowel amplitude rise is not readily apparent. In either case, this variation is likely in the realm of an imperceptibly small change (Huggins 1972). Timing data for speech and song were then compared to their matched drum modes to determine correlations in timing profiles.

In order to test for the occurrence of a continuum of acoustic features spanning the four drumming modes, the difference between successive IOIs was computed and compared as a measure of rhythmicity over sample phrases. Our hypothesis here is that the average difference between IOIs increases linearly across modes between drumming as a purely musical instrument (featuring a greater incidence of isochronous rhythms and even rhythmic ratios) up to directly mimicking speech, which features largely non-isochronous rhythms and is most prone to longer breaks. Specifically, the average IOI differences of the four drumming modes are predicted to be ordered D-DR < DS < DT-P < DT-D. We also gathered data to perform a post-hoc descriptive analysis of the relationship between the IOIs of speech samples and those of their matched drumming modes, but we do not have an a priori hypothesis regarding any relationship differences between the modes.

Analysis of Fundamental Frequency (f_0) and Intensity

Fundamental frequency (f_0) and intensity were extracted from vocal talking (VT), vocal singing (VS), drum as song (DS), drum as performative speech (DT-P), and drum as direct speech (DT-D) samples with Praat (Boersma and Weenink 2021). The tempo of recorded performances differed both between vocal and drum phrases and between sub-phrases within a particular song. Acoustic information was extracted and compared on a sub-phrase basis, identified by long gaps of silence in both vocal and drum recordings. Occasionally, some quiet time-keeping rhythms that occurred between sub-phrases and had no f_0 information were omitted. These rhythms are discussed further in the microtiming analysis results. Tempo still differed in the resulting sub-phrases, so longer sample phrases were downsampled to match the shorter one. To demonstrate similarity between vocal and drum performances, f_0 and intensity correlations were calculated. However, pitch - the perception of the f_0 of speech - is not solely dependent on fundamental frequency; pitch is also influenced by intensity (e.g., Patten and McBeath 2020). Thus, correlations are not only computed between drum and vocal fundamental frequencies and drum and vocal intensities, but also across acoustic features. In other words, this analysis attempts to understand how dùndún speech surrogacy uses acoustic features other than just f_0 to convey pitch information. The hypothesis for this analysis is similar to that of the timing analysis, that correlations between all acoustic features will increase as there are fewer musical features (i.e., DS < DT-P < DT-D). Similarly, it is expected that f_0 and intensity will significantly correlate between drum and speech phrases and that cross correlations between features will also be significant.

RESULTS

Timing Analysis

The timing analysis consisted of three parts, each examining and comparing the interonset intervals (IOIs) of the various vocal and drumming modes. The initial, principal analysis tested temporal correlations of IOIs between corresponding vocal and drumming modes in order to confirm the various levels of instrumental speech surrogacy. The secondary analysis compared the average IOI rates for each of the drumming modes to confirm the distinctiveness of the drumming mode categories. The third analysis compared average IOI rates between corresponding vocal and drumming modes to explore overall systematic acoustic trends regarding vocal vs. drumming modes.

Table 3 summarizes the results of the principal temporal analysis comparing corresponding vocal vs. drumming modes across all samples. As shown in the third column, our principal hypotheses were supported, with significant correlations occurring between each of the corresponding vocal and drum modes, and systematically decreasing correlations when moving from drum mimicking direct talking (DT-D) through talking in performances (DT-P) and drum singing (DS). The fourth column is included to verify that the drumming IOIs of the different drumming modes, which are composed of different performances and should not show a relationship, do not correlate with each other. The only exception to this is the case of the DT-D and DT-P modes, which replicate the same speech sample and exhibit a small correlation of $r(28) = 0.24$.

The overall average correlation between the interonset intervals for Vocal Talking (VT) and Drum Talking, Direct (DT-D) was $r = 0.72$, a large correlation that was backed up by a significant one-sample t-test, $t(28) = 13.95$, $p < 0.001$ with a very large effect size of Cohen's $d = 2.59$. To put this in perspective, Cohen (1988) norms

TABLE 3 | Average timing correlations of syllabic occurrences between different modes of vocalizing and drumming on the same verbal-musical sequence. One-sample *t*-tests comparing correlation coefficients to a null set are included below average correlations.

Vocalizing	Drumming	Vocal and drum	Drum and drum
Vocal talking (*VT*)	Drum talking–Direct (*DT-D*)	$r_{\mathrm{VT,DTD}}$ = 0.72** *d* = 2.59	$r_{\mathrm{DTD,DTP}}$ = 0.35** *d* = 1.00)
	Drum talking–Performative (*DT-P*)	$r_{\mathrm{VT,DTP}}$ = 0.37** *d* = 0.93	—
Vocal singing (*VS*)	Drum singing (*DS*)	$r_{\mathrm{VS,DS}}$ = 0.25** *d* = 0.75	$r_{\mathrm{DS,DTD}}$ = 0.07 $r_{\mathrm{DS,DTP}}$ = 0.10
	Drum–Dance rhythm (*D-DR*)	—	$r_{\mathrm{DDR,DTD}}$ = −0.05 $r_{\mathrm{DDR,DTP}}$ = 0.06 $r_{\mathrm{DDR,DS}}$ = 0.09

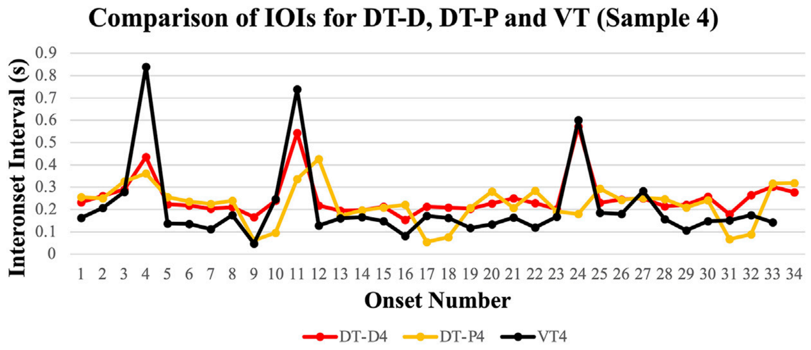

FIGURE 3 | Interonset intervals (IOIs) for a typical vocal talking (VT) sample and corresponding imitations of the speech sequence on the drum, both direct imitation (DT-D) and while musically performing the speech sequence (DT-P). Syllabic and drum onsets are presented as a monotonic temporal series on the *X*-axis, and the duration of each onset is indicated on the *Y*-axis. Thus, shorter IOIs correspond to lower values and longer durations correspond to higher values. Significant peaks represent gaps between sub-phrases. Note how the DT-D timing profile closely resembles VT, aligning robustly at the IOI peaks between sub-phrases, while the use of rhythmic embellishments in DT-P results in a somewhat less related timing profile than that of VT and DT-D (see sound samples **Supplementary Samples 1A–C**).

suggest a *d* over 0.8 is a large effect. This confirms that drummers are able to mimic the timing patterns of Yorùbá speech very robustly when tasked to do so directly without the need to add musical context. When playing in performance mode with some musical context, the average correlation between VT and DrumTalking, Performative (DT-P) drops to $r = 0.37$ while maintaining a significant one-sample *t*-test, $t(28) = 5.39$, $p < 0.001$, Cohen's $d = 0.93$. One possible reason for this lower correlation is the increased use of rhythmic embellishments in DT-P. However, removing flams and the first attack of every paired 16th note that makes up an àfikún does not have a significant effect on the overall correlation ($r_{\mathrm{original}} = 0.37$, $r_{\mathrm{corrected}} = 0.43$, an increase of only 4% variance explained). **Figure 3** illustrates the match between VT, DT-D, and DT-P for a typical representative sequence. Peaks in the graphs of duration onsets (IOIs) indicate breaks between sub-phrases. Note the relative alignment of IOI peaks between vocal and drumming modes showing alignment in phrasing. We also include a link to corresponding auditory recordings of this sequence to allow readers to experience the different levels of drumming IOI and vocal sequence synchronization.

Figure 4 illustrates the match between Vocal Singing (VS) and Drum Singing (DS), again with auditory recordings of the sequences. Here, the correlation between vocal and drumming modes drops further to $r = 0.25$, though still maintains a significant one-sample *t*-test, $t(28) = 4.02$, $p < 0.001$, Cohen's $d = 0.75$, a medium effect size. A contributing factor to the lower correlation between VS and DS is the increased use of rhythmic embellishments of some syllable onsets in the form of flams and àfikún (Euba 1990) along with the insertion of short rhythmic patterns that may serve a time-keeping function (Panteleoni 1972). Removing flams and the second onset from each pair of 16th notes that make up the àfikún ranges from a minor positive impact on correlations to a significant improvement on correlations for individual performers, but did not produce a significant change in the overall pattern of correlations (r_{original} = 0.25, $t(28) = 4.02$, p = <0.001, $d = 0.75$, $r_{\mathrm{corrected}} = 0.40$, $t(28) = 5.53$, $p < 0.001$, $d = 2.59$, a 10% increase in variance accounted for).

Figure 5 illustrates the Drum-Dance Rhythm (D-DR) IOI pattern with corresponding auditory recording. Here, the sequence is not correlated with any particular vocal pattern so none are shown. This IOI pattern shows the most musical rhythmic characteristics, with repeating patterns of durations that relate via even ratios (e.g., 2:1) which follow the same patterns of musical beats and their subdivisions, alternating with series of periodic durations. As with the other samples, large peaks represent breaks between sub-phrases.

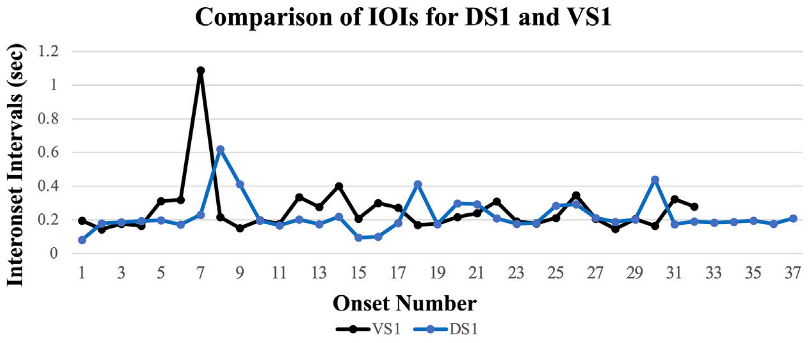

FIGURE 4 | Interonset intervals (IOIs) for a typical vocal singing sample (VS) and the corresponding imitation of the song on the drum (DS). Present in the graph of DS are rhythmic embellishments that shift the timing profiles of DS and VS out of alignment (see sound samples **Supplementary Samples 2A,B**).

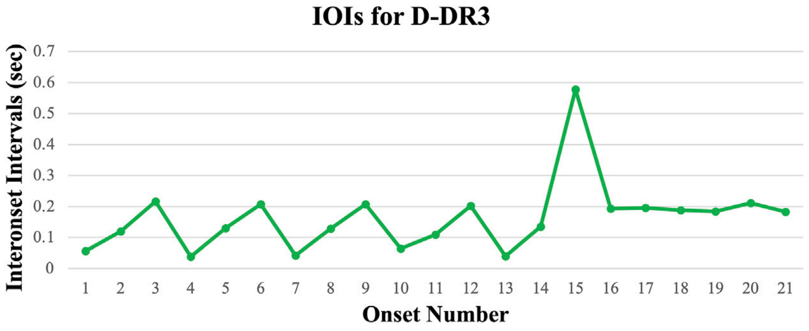

FIGURE 5 | Interonset intervals (IOIs) for a typical drumming pattern in the Drum Dance Rhythm condition (D-DR). The repeating pattern of durations present in the opening of the sample roughly follows a 4:2:1 ratio, which translates into durations of 200, 100, and 50 ms, respectively, with microtiming variation in each category. The sample ends with an isochronous series of durations (see sound **Supplementary Samples 3**).

A three-way, omnibus ANOVA comparing the effects of and interactions between drum and vocal correlations across mode (DT-D and VT, DT-P and VT, and DS and VS), performer, and correction for rhythmic embellishments revealed only a significant main effect of mode, $F(2{,}156) = 21.29$, $p < 0.001$, partial $\eta^2 = 0.21$, a medium–though verging on large–effect according to conventional norms (Cohen et al., 2003). There was a significant linear contrast between individual modes with VT and DT-D yielding the highest average correlation coefficients (average $r = 0.72$), VT and DT-P yielding a moderate amount (average $r = 0.40$), and VS and DS yielding the lowest average correlation coefficients (average $r = 0.33$), $F(1{,}171) = 31.57$, $p < 0.001$, $\eta^2 = 0.20$, a large effect according to Cohen (1988) norms. The discrepancy between the effect sizes of the omnibus test and the linear contrast is likely due to the difference in degrees of freedom. Either way, there is between a medium and large effect of mode. These analyses (for original data only) can be seen graphically in **Figure 6**.

In our secondary timing analysis, we compared the average difference between successive IOI times of the four drumming modes. Timing differences between IOIs differed as a function of drum mode type, as can be seen in **Figure 7**. As predicted, there was a significant linear trend indicating that differences between successive IOIs increased as drumming became more constrained to speechlike characteristics (e.g., from pure musical rhythm (D-DR) to direct speech imitation (DT-D)), $F(1{,}116) = 43.38$, $p < 0.001$, $\eta^2 = 0.27$, which is a large effect according to Cohen (1988) norms. **Figure 7A** illustrates four typical IOI patterns for comparison, and **Figure 7B** illustrates the linear trend of increasing IOI mean as the drumming mode has more speechlike qualities and fewer musical rhythms.

As a third and final timing analysis, we also performed a post-hoc comparison of timing data from speech excerpts and their correlated drum samples. Flams and àfikún differed across mode and drummer. A two-way ANOVA of flam rates indicated a significant main effect of drummer, but not mode or an interaction between the two, $F(2{,}80) = 5.44$, $p < 0.01$, partial $\eta^2 = 0.12$, a medium effect. Pairwise comparison revealed that this effect was driven by Drummer A’s higher average use of flams ($M = 0.40$ flams per phrase) than either Drummer B ($M = 0$) or C ($M = 0.10$). A two-way ANOVA of àfikún rates, however, revealed a significant main effect of mode, drummer, and an interaction between the two. Drum/vocal mode yielded a difference, $F(2{,}80) = 9.10$, $p < 0.001$, partial $\eta^2 = 0.19$, a medium effect, that pairwise comparisons indicated was due to the significantly lower rate of àfikún in DT-D ($M = 0.17$ per

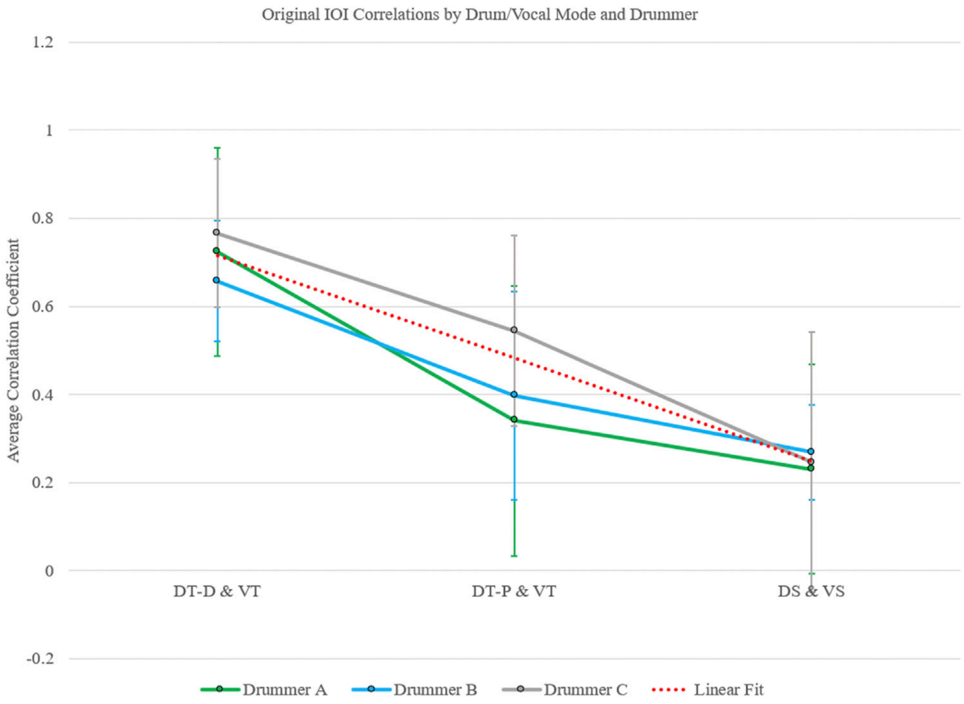

FIGURE 6 | Original IOI correlations between modes and drummers. The data closely follows the linear fit (dotted red line), showing a successive, incremental decrease in average correlation coefficient as the drum/vocal gains musical rhythms. As discussed later, this is likely due to the presence of increasing inter-phrase rhythms as modes become more musical. Error bars are ±2 SE.

phrase) compared to DT-P (M = 1.13) and DS (M = 1.09). Drummer yielded a difference, $F(2,80) = 12.67$, $p < 0.001$, η^2 = 0.24, a large effect, that pairwise comparisons indicated was due to Drummer C's higher usage (M = 1.52) compared to either Drummer A (M = 0.60) or Drummer B (M = 0.27). Finally, the correlation between mode and drummer was also significant, $F(4.80) = 3.39$, p <0 .05, $\eta^2 = 0.15$, a medium effect. According to pairwise comparisons, Drummer A rarely used àfikún at all, Drummer B more evenly recruited the technique across their modes of playing, and Drummer C relied most heavily on àfikún in DS.

Similarly, inter-phrase attacks (drum attacks without pitch information that appear to serve a time-keeping role) differed greatly among the recorded phrases. A two-way ANOVA comparing the amount of inter-phrase rhythms between drum/vocal modes and performers yielded a significant main effect of drummer, $F(2,80) = 16.06$, $p < 0.001$, $t(28) = 3.71$, $p < 0.001$, partial η^2 = 0.28, a large effect. Pairwise comparisons revealed that Drummer B drove this difference, $p < 0.001$ compared to both other drummers. There was also a significant main effect of drum/vocal mode, $F(2,80) = 26.32$, $p < 0.001$, partial η^2 = 0.39, a large effect. Pairwise comparisons revealed that this difference was driven by the DS and VS mode, $p < 0.001$ compared to both other modes. Finally, the interaction between drummer and mode was significant, $F(4,80) = 17.04$, $p < 0.001$, partial η^2 = 0.46, also a large effect. Drummer B produced, on average, 7.6 attacks per phrase in the DS mode, but zero inter-phrase attacks in both other modes. Drummer A produced, on average, 1.4 attacks in the DS mode, 0.4 in the DT-P mode, and zero in the DT-D mode. Drummer C produced no inter-phrase attacks in any mode. An example of these inter-phrase rhythms in the use of the drum as song surrogate is shown in **Figure 8**, with the inter-phrase rhythms resulting in more drum onsets than syllabic onsets in the corresponding song sample. It appears that in some instances the drummer is simultaneously recreating the song sample phrases and playing a pacemaker rhythm to keep time (Anku 1997).

Discussion of Timing Results

The results confirmed our hypotheses, with the highest correlations for timing occurring between DT-D and VT and getting progressively smaller as drum samples become increasingly more music-like, with the lowest correlation occurring between DS and VS. DT-D samples typically present a near-direct mapping of syllables to drum onsets (r = 0.72), following the same phrasing patterns with gaps between sub-phrases as seen in **Figure 3**. Most DT-D samples do feature between 1–3 additional attacks beyond the number of syllables present in the speech samples, though this is likely due to differences in the representation of elisions on the drum vs. in speech (Euba 1990). In contrast, DT-P samples typically incorporate a greater number of attacks, including the use of significantly more rhythmic embellishments such as flams and àfikún. The inclusion of these rhythmic embellishments varies across the samples collected, resulting in variances in correlation of the timing profiles between DT-P and VT. In most cases, the

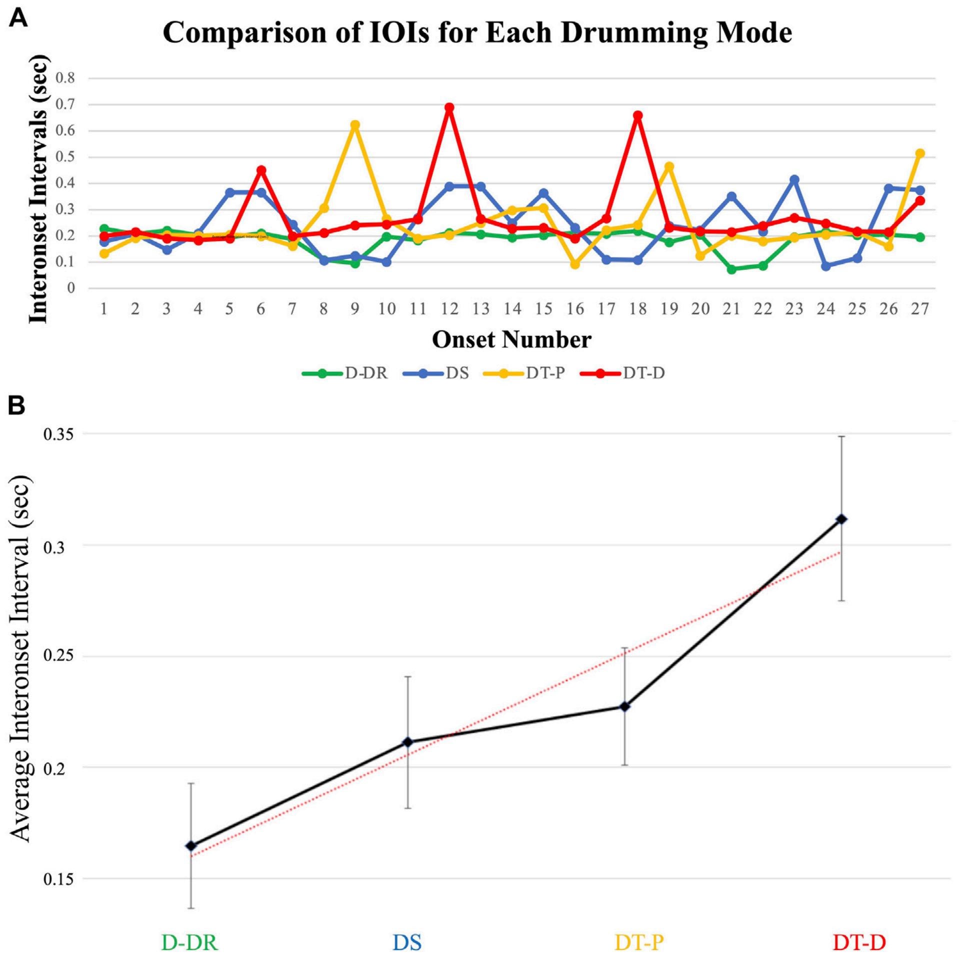

FIGURE 7 | A comparison of IOIs in seconds for representative samples of the four drumming modes. **(A)** Top graph illustrates the temporal IOI pattern for each sample (see sound samples **Supplementary Samples 4A–D**). **(B)** The bottom graph summarizes the average IOIs across all samples for each drumming mode (with error bars indicating ±2 SE). Note the increase in average duration length as samples move from musical rhythms (D-DR, sound sample **Supplementary Samples 4A**) towards the direct imitation of speech (DT-D, sound sample 4d) where more speechlike constraints are introduced to drumming rhythms. Dotted line illustrates a significant linear trend.

general phrase structure is preserved, but more attacks are added into each sample phrase, altering the timing profile while still retaining the same general shape. In keeping with this pattern of increasing musicality changing the drum signal, differences between IOIs linearly increased from D-DR to DS to DT-P to DT-D. Significant differences in IOIs, particularly those that are not in an even ratio with one another (e.g., 2:1) can be understood as either breaks or deviations from a prevailing rhythmic pattern, or as a representation of the more irregular timing patterns of speech. In essence, this result simply shows that DS more closely adheres to rhythmic synchrony than DT-D, which shows little evidence of rhythmic synchrony. The occurrence of inter-phrase rhythms follows a similar pattern.

Though the majority of the DT-P samples feature a greater, though not significantly greater, number of onsets than their DT-D and VT counterparts, the overall rhythmic profile is still significantly more irregular than what is seen in the samples when the drum is being used as a musical instrument (D-DR). Samples where the drum is imitating specific songs (DS) represents a middle ground between the rhythmic profiles of D-DR and DT-P samples, combining some aspects of the irregular timing patterns seen in the DT-P and DT-D samples with periods of isochrony or repeating rhythmic patterns reminiscent of the rhythmic regularity of the D-DR samples. Like the DT-P samples, some of the DS samples also use flams and àfikún, and these rhythmic embellishments can at times obscure the similarities in timing profile that do exist between the DS samples and their corresponding song excerpts (VS). Removing flams and the second attack from the paired 16th notes that make up the àfikún gestures does not significantly improve the overall correlations across all samples, though it does have a differential effect on individual correlations for some performers. Filtering these embellishments out of the data set does not address all of the differences between DS and VS samples that can be attributed to rhythmic variation. For example, **Figure 8** (top) shows the timing profiles for a DS sample that

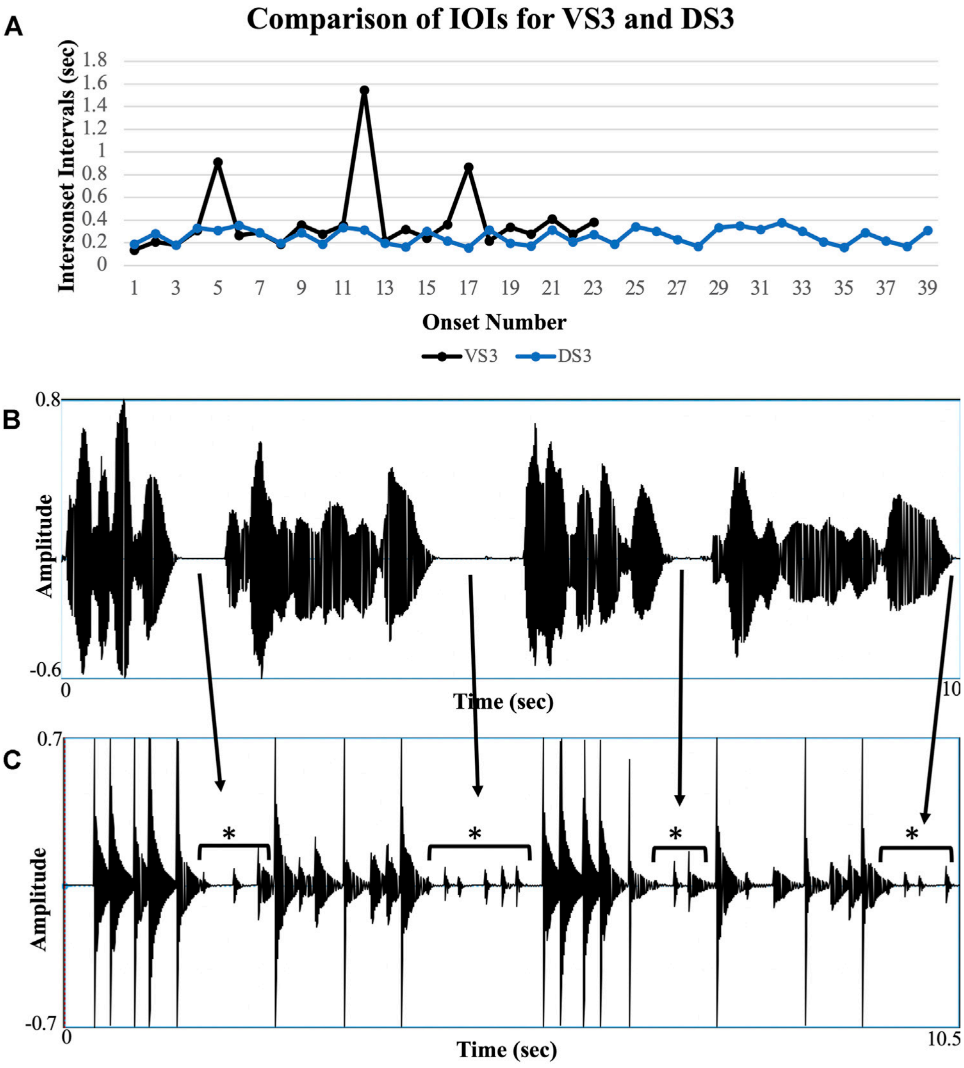

FIGURE 8 | Example of the temporal acoustic correspondence between a typical VS and DS sample phrase. Shown are sound pressure levels during a voice-drum sequence of about 10 s **(A)** IOI timing pattern of a matched set of VS and DS sample phrases (see sound samples **Supplementary Samples 5A,B**). **(B)** Temporal acoustic intensity graph for the VS sample. **(C)** Temporal acoustic intensity graph for the DS sample. Note how the two samples have approximately the same discrete onsets (syllables or drum strikes, respectively) and rate, but that the DS sample has more onset, particularly between sub-phrases consisting of quieter, repeating patterns of attacks that maintain a rhythmic musical time-keeping structure (exemplars marked with an asterisk).

uses neither flams nor àfikún. While similarities in the timing profiles between this DS sample and its paired VS sample are visually apparent, the timing profile of DS is altered by the use of inter-phrase rhythms that shift it away from the timing profile of VS. The overall regularity and repetitive nature of most of the inter-phrase rhythms is suggestive of a time-keeping function. It may be that the ìyáàlù dùndún is simultaneously representing the song sample phrases and providing a pacemaker rhythm to keep time (Locke 1982; Anku 1997). Differences in performers regarding the use of inter-phrase rhythms results in differing effects on correlations when removing rhythmic embellishments. Performer C uses no inter-phrase rhythms, so the removal of rhythmic embellishments from Performer C's DS samples significantly increases correlations between DS and VS (r_{original} = 0.22, $r_{\text{corrected}}$ = 0.73), an increase from three of nine significant individual correlations to eight of 9, and an increase of 49% variance accounted for. In contrast, Performer B uses inter-phrase rhythms in all of his DS samples, with an average of 7.6 inter-phrase rhythm onsets per sample. Consequently, removing rhythmic embellishment has a negligible impact on correlations between DS and VS (r_{original} = 0.26, $r_{\text{corrected}}$ = 0.27), with no changes in the significance of individual correlations. Performer A represents a middle ground between Performers C and B, using inter-phrase rhythms in some (but not all) samples, along with a greater use of rhythmic embellishments than Performer B, and so the removal of rhythmic embellishment from his DS samples has a minor positive effect on correlations between VS and DS (r_{original} = 0.17, $r_{\text{corrected}}$ = 0.23), with an

TABLE 4 | Rhythm-matched fundamental frequency and intensity correlations between different modes of vocalizing and their matched drumming modes.

Vocalizing	Drumming	Vocal f_0 and drum f_0	Vocal intensity and drum intensity	Vocal f_0 and drum intensity	Vocal intensity and drum f_0
Vocal talking (*VT*)	Drum talking–Direct (*DT-D*)	$r_{\mathrm{VT,DTD}} = 0.32^{**}$ $d = 1.46$	$r_{\mathrm{VT,DTD}} = 0.13^{**}$ $d = 0.97$	$r_{\mathrm{VT,DTD}} = 0.20^{**}$ $d = 1.27$	$r_{\mathrm{VT,DTD}} = 0.05$ $d = 0.29$
	Drum talking–Performative (*DT-P*)	$r_{\mathrm{VT,DTP}} = 0.25^{**}$ $d = 1.31)$	$r_{\mathrm{VT,DTP}} = 0.08^{*}$ $d = 0.47$	$r_{\mathrm{VT,DTP}} = 0.17^{**}$ $d = 0.87$	$r_{\mathrm{VT,DTP}} = 0.04$ $d = 0.19$
Vocal singing (*VS*)	Drum Singing(*DS*)	$r_{\mathrm{VS,DS}} = 0.58^{**}$ $d = 3.61$	$r_{\mathrm{VS,DS}} = 0.29^{**}$ $d = 1.83$	$r_{\mathrm{VT,DS}} = 0.27^{**}$ $d = 1.85$	$r_{\mathrm{VT,DS}} = 0.16^{**}$ $d = 0.78$

increase from 0 of 10 significant individual correlations to three of 10.

Results of Fundamental Frequency (f_0) and Intensity Analysis

Table 4 summarizes the correlational results comparing the f_0 and intensity patterns between vocal and corresponding drumming modes. Fundamental frequency (f_0) of non-singing speech (VT) correlated significantly with the f_0 of drumming as direct speech (DT-D) and drumming as performative speech (DT-P), though the latter to a lesser extent. Twenty-three of twenty-nine individual samples (average $df = 248$) in which we compared f_0 of speech and DT-D produced significant correlations, with an average group correlation of $r = 0.32$. A single-sample t-test comparing individual correlations to a null set was significant, $t(28) = 7.88$, $p < 0.001$, Cohen's $d = 1.46$, a large effect according to Cohen (1988) norms. Similarly, twenty-four of twenty-nine individual samples (average $df = 236$) comparing the f_0 of speech and DT-P correlated significantly, producing an average group correlation of $r = 0.25$. A single-sample t-test of these correlations also was significant, $t(28) = 7.07$, $p < 0.001$, Cohen's $d = 1.31$, a large effect. DT-D produced significantly larger f_0 correlations with drum ($M = 0.32$) than DT-P ($M = 0.25$), paired $t(28) = 2.50$, $p < 0.05$, Cohen's $d = 0.46$, nearly a medium effect. Like the pattern found with the f_0 correlations, the correlation in intensity patterns between VT and DT-D was larger than that between VT and DT-P (with Cohen's d indicating a small effect size), however the difference did not reach statistical significance, $t(27) = 1.53$, $p = 0.14$, Cohen's $d = 0.29$.

Similarly, speech intensity was correlated with drum intensity for both DT-D and DT-P, albeit with smaller effect sizes. Sixteen of twenty-nine individual samples (average $df = 328$) correlated significantly, producing an average group correlation of $r = 0.13$, a small effect. A single-sample t-test yielded a significant result, $t(28) = 5.23$, $p < 0.001$, Cohen's $d = 0.97$, a large effect. Fourteen of twenty-eight individual correlations between speech and DT-P intensity were significant (average $df = 324$) correlated significantly, producing an average group correlation of $r = 0.08$, a null effect. A single-sample t-test indicated that these correlations significantly differed from a null set, $t(27) = 2.51$, $p < 0.05$, Cohen's $d = 0.47$, a small effect. These correlations are also presented in **Table 4**. Though these intensity correlations differ slightly in both coefficient and effect size, a paired t-test indicated no significant difference between DT-D and DT-P, $t(27) = 1.53$, $p = 0.14$, Cohen's $d = 0.29$, a small effect.

Because perception of f_0 and intensity are not one-to-one correlations, cross-feature correlations were also examined. Drum f_0 and speech intensity were correlated for neither DT-D (average $r = 0.05$, one-sample $t(28) = 1.55$, $p = 0.13$, Cohen's $d = 0.29$) nor DT-P (average $r = 0.04$, one-sample $t(28) = 1.0$, $p = 0.33$, Cohen's $d = 0.19$). Drum intensity and speech f_0, however, were both significantly correlated. The average correlation for DT-D intensity and speech f_0 ($r = 0.20$, one-sample $t(28) = 6.86$, $p < 0.001$, Cohen's $d = 1.27$, a large effect) was slightly larger than DT-P intensity and speech $f0$ ($r = 0.17$, one-sample $t(28) = 4.70$, $p < 0.001$, Cohen's $d = 0.87$), though not significantly so (paired $t(28) = 1.10$, $p = 0.28$, Cohen's $d = 0.20$). An exemplar showing the similar f_0 and intensity contours can be seen in **Figure 9**.

Vocal singing (VS) and the corresponding drum imitation (DS) were also significantly correlated in f_0. Twenty-nine of twenty-nine individual samples (average $df = 472$) correlated significantly, producing an average group correlation of $r = 0.58$. A single-sample t-test comparing individual correlations to a null set was significant, $t(28) = 19.42$, $p < 0.001$, Cohen's $d = 3.61$, a large effect. Similarly, the intensity of singing and the matched drum rhythm was significantly correlated. Twenty-eight of twenty-nine individual samples (average $df = 611$) correlated significantly, producing an average group correlation of $r = 0.29$. A single-sample t-test comparing individual correlations to a null set was significant, $t(28) = 9.85$, $p < 0.001$, Cohen's $d = 1.83$, a large effect. Cross-acoustic correlations were also significant. Twenty-two of twenty-nine drum f_0 and singing intensity correlations were significant, producing an average correlation of $r = 0.16$ (average $df = 472$), a small effect, and a significant one-sample t-test, $t(28) = 4.08$, $p < 0.001$, Cohen's $d = 0.76$, a medium effect. Twenty-eight of twenty-nine drum intensity and singing f_0 correlations were significant, producing an average correlation of $r = 0.27$ average $df = 566$), a nearly medium effect, and a significant one-sample t-test, $t(28) = 9.94$, $p < 0.001$, Cohen's $d = 1.85$, a large effect. These correlations are also summarized in **Table 4**.

Differences in average correlation across drum/vocal modes and the acoustic features correlated were examined in an omnibus ANOVA. There was a significant difference of average correlation between DT-D, DT-P, and DS, $F(2{,}335) = 36.75$, $p < 0.001$, $\eta^2 = 0.07$, a medium effect according to Cohen (1988) norms. Pairwise comparisons reveal that this effect is carried by the much larger average correlation for DS ($r = 0.33$) compared to DT-D ($r = 0.18$)

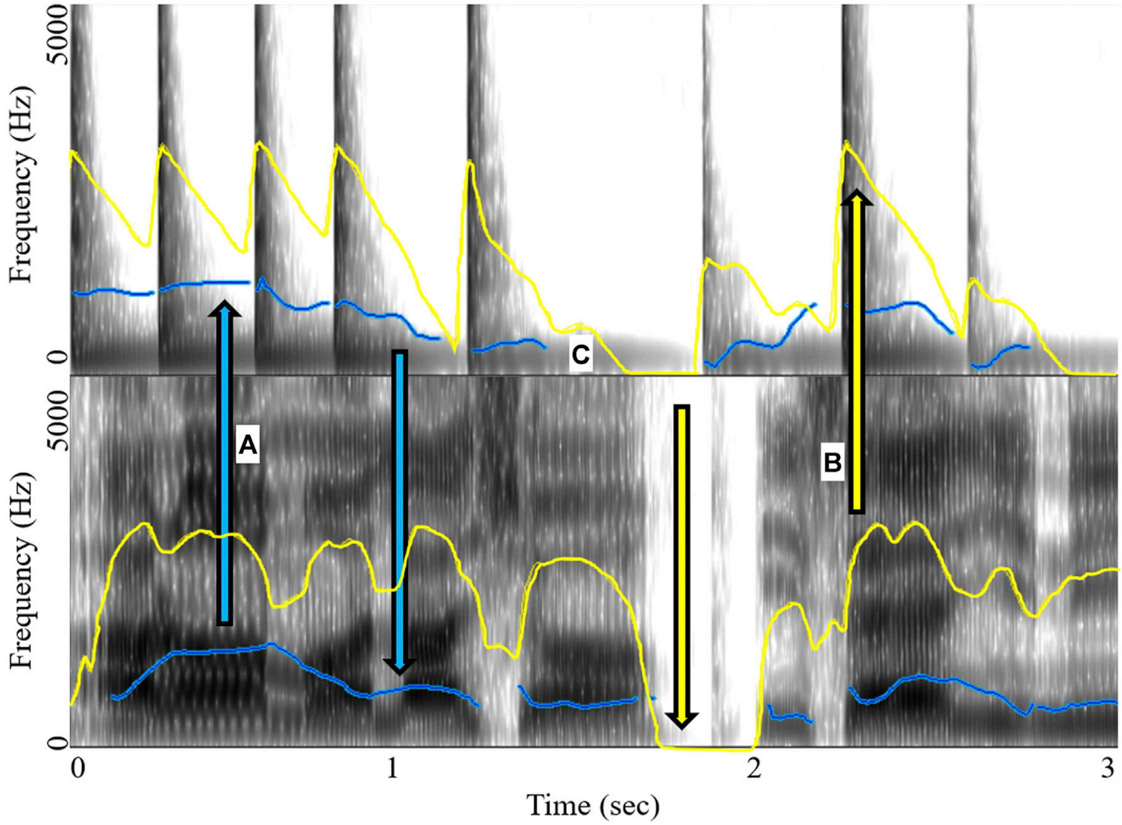

FIGURE 9 | Fundamental frequency (f_0) and intensity contours for drum and speech. This exemplar of the f_0 and intensity data (pulled from Praat for Drummer B's Song 5, which exhibited an f_0 correlation of $r(429) = 0.80$ and intensity $r(664) = 0.50$) shows the correlated movements between **(A)** drum and speech f_0 and **(B)** drum and speech intensity. This graphic also shows **(C)** gaps that appear in the data (most often in f_0), which are often caused by an actual lack of data at the moment or a dearth of data points that does not allow for an accurate estimate of the acoustic feature. Finally, the temporal asymmetry between samples is also evident. This is what necessitates the downsampling procedure to equalize the length of each sample phrase.

and DT-P (r = 0.13). There was also a significant effect of the acoustic features (f_0 or intensity) correlated with one another, $F(3,335) = 42.96$, $p < 0.001$, $\eta^2 = 0.13$, nearly a large effect. Pairwise comparisons show that this effect is due to the significantly higher f_0/f_0 correlations (r = 0.38) compared to others, and the significantly lower drum intensity/speech f_0 correlations (r = 0.08) compared to others. The intensity/intensity (r = 0.17) and drum f_0/speech intensity (r = 0.21) correlations are not statistically different from one another. There was also a statistically significant interaction between drum/voice mode and correlated acoustics, $F(6,335) = 2.86$, p <0 .05, $\eta^2 = 0.01$, a small effect, though very close to the benchmark minimum. This interaction is likely spurious as it seems to indicate no meaningful difference. These average correlations are depicted graphically in **Figure 10**. A post hoc one-way ANOVA to identify the nature of these differences investigated the standard deviation of f_0 differences across all modes independently (DT-D, DT-P, VT, DS, and VS), which revealed a significant effect, $F(4,140) = 11.91$, $p < 0.001$, $\eta^2 = 0.25$, a large effect. Pairwise comparisons revealed that this result was driven only by a significantly lower f_0 standard deviation for VT (M = 17.83) compared to DT-D (M = 23.50), DT-P (M = 27.37), DS (M = 26.00), and VS (M = 23.53).

Discussion of Fundamental Frequency (f_0) and Intensity Results

Overall, significantly higher correlations are observed Drum Singing and Vocal Singing (DS and VS), which is in line with neither our hypothesis nor the results of the timing analysis. This result is notable because this is the first study to show the relationship between dùndún drumming as song and vocal singing. The DS and VS correlations are larger for both f_0 and intensity as well as their cross correlations, which may be due, in part, to the more restricted range of the f_0 for Vocal Talking (VT). It is typical for the frequency range of singing to be larger than that of talking (Hacki 1996), so this is not an abnormal result. While singing, however, the variance in f_0 increases to match the exhibited f_0 range of the drum. DS and VS were the only correlates to show a significant average correlation coefficient for all acoustic feature correlations. This indicates that in addition to the similar f_0 range, drummers may be using different information from the VS phrases to produce their DS phrases

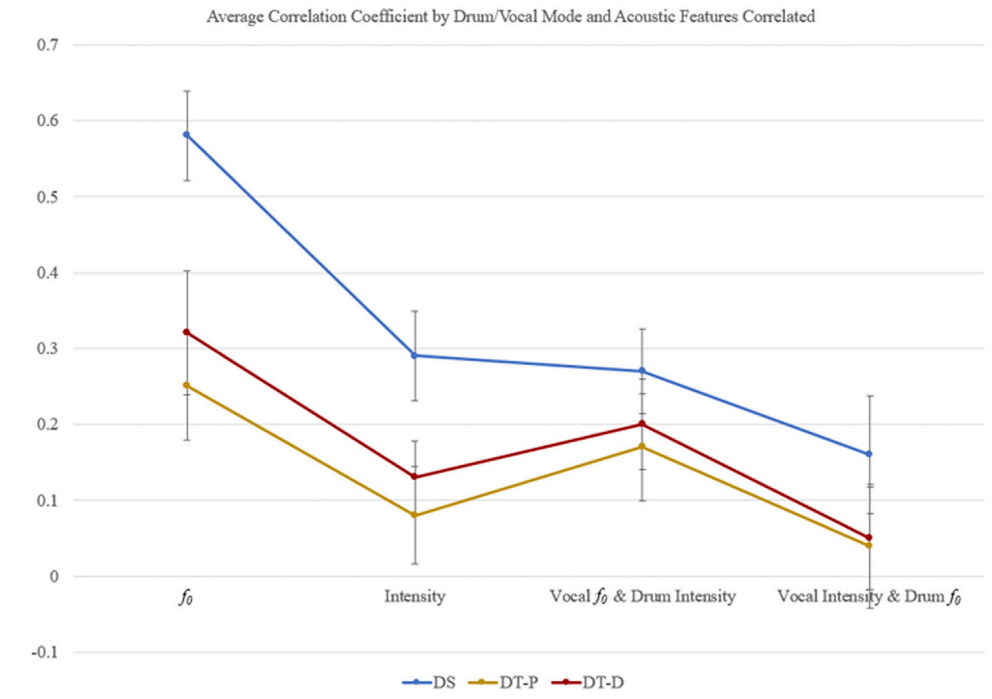

FIGURE 10 | Average correlation coefficients for all songs broken down by drum/vocal mode (DS, DT-P, and DT-D) and acoustic features used to obtain the correlation (drum f_0/vocal f_0, drum intensity/vocal intensity, drum intensity/vocal f_0, and drum f_0/vocal intensity). The significant difference of DS compared to both DT-P and DTD is apparent, as is the large effect of f_0/f_0 correlations, small-to-medium effect of intensity/intensity and drum intensity/vocal f_0, and the insignificant effect of drum f_0/vocal intensity. Error bars are ±2 SE.

than when listening to VT and performing either DT-D or DT-P. While the DT-D and VT and DT-P and VT correlations were not as high and did not produce a significant correlation between all acoustic features, the correlations were significant between vocal and drum f_0, vocal and drum intensity, and vocal f_0 and drum intensity.

DS and VS may also have exhibited stronger correlations due to differing perceptions of f_0 in speech and music. While f_0 is an important component of speech perception, it is not always directly perceived. Interactions with intensity and spectral centroid, as mentioned, can deviate pitch perception from frequency (things). Similar pitch changes are also perceived differently based on their location in speech (at the beginning or ending of a syllable) and their location relative to silence (House 1990, House 1995; Mertens 2004). Very quiet f_0s are also not often perceived when occluded by other speech sounds (Mertens 2013, Mertens 2014). While speech is characterized by formant frequencies that vary greatly between speech sounds, this variation is often less pronounced in singing–possibly in an effort to maintain consistent vocal quality for the duration of the sung vocalization (Bloothoofft and Plomp 1986). In the case of music, both vocal and instrumental songs have been found to exhibit similar correlations between changes in f_0 and intensity, so the current music correlations are not unexpected (Scharine and McBeath, 2019; Patten and McBeath, 2020). There are three levels of correlations for acoustic features. The highest correlation, across all modes, is between vocal f_0 and drum f_0. As Yorùbá is a tonal language and f_0 change captures most of the variance of changes between High, Mid, and Low tones (Ọdẹ́jọbí 2007), this result is not surprising. Across all modes, this correlation explains 14% of the variance and, for DS and VS, 34% of the variance. Correlations between vocal intensity and drum intensity, and those between vocal f_0 and drum intensity do not differ from one another significantly. As with the f_0/f_0 correlations, the intensity/intensity correlations are not surprising and simply confirm the dùndún directly imitates speech. The cross correlation between vocal f_0 and drum intensity, however, is surprising. This correlation may indicate that drummers are using intensity variations to perceptually change the pitch of the notes they play and further enhance the dynamics of the drum. While this correlation only explains 4% of the variance overall, it explains 7% for DS and VS. Lastly, the correlation between drum f_0 and vocal intensity was significantly lower than all other correlations. In fact, for all modes except DS and VS, it was significantly different from a null set of correlations. This indicates that, while some acoustic features inform others, the intensity changes in voice are not features used by dùndún drummers to inform the f_0 of their performances (Oyetade et al., 2003). While this correlation is significant for DS and VS and may indicate a difference in how DS is performed to imitate singing, the correlation is small according to Cohen (1988) norms and the significance of the one-sample t-test could be due to a large sample size.

GENERAL DISCUSSION

Yorùbá dùndún drumming is a classic example of speech surrogacy in which a musical instrument represents pitch and rhythmic characteristics of vocal utterances. The primary purpose of this research was to examine the speech-music relationship in the dùndún, thereby laying the groundwork for the understanding of the functioning of speech surrogate systems. Our goal was to determine the extent of acoustic representation of speech sounds on the drum, as well as to compare four dùndún performance modes. Here, we examine microtiming, f_0, and intensity patterns and provide a microstructural acoustic analysis that verifies the acoustic correspondence between talking and singing vocalization modes as well as their corresponding drumming modes. Analysis of 29 spoken and sung verbal utterances (n = 58) and corresponding drum modes (n = 87), along with samples of the drum performing as a musical instrument (n = 30) demonstrated that the four distinct drumming modes reflect a variety of imitation styles, ranging from purely rhythmic to direct speech imitation. Our microtiming findings confirmed our general hypotheses that comparisons between vocal and corresponding drumming modes are highest for when the drum is directly imitating speech (DT-D vs VT), next when the drum is performatively imitating speech (DT-P vs VT), and weakest when the drum is imitating song (DS vs VS). As anticipated, the purely rhythmic drumming mode (D-DR) differed significantly in comparison to other drumming modes, demonstrating that the dùndún can embody multiple distinct modes (e.g., rhythmic instrument, song surrogate, and speech surrogate). Correlations between f_0 and intensity, however, revealed a somewhat contradictory finding in that DS and VS yielded significantly better coefficients across all acoustic feature correlations, while DT-D and VT and DT-P and VT did not differ. It is possible this difference is, in part, due to the removal of inter-phrase rhythms from the f_0 and intensity analysis, a necessary change to equalize the duration of sub-phrases and accurately model correlations. The significantly greater number of rhythmic embellishments in DT-P and DS likely hindered IOI correlations but not those of acoustic features, as many did not produce f_0 information, were produced with low intensities, or occurred quickly enough that variations between correlates were negligible.

While previous studies have demonstrated frequency correlation between drum and word-level utterances (Akinbo 2019), we showed that this frequency correlation also exists at the phrase level. Novel to this study is our finding of significantly higher frequency correlations between the DS mode and song excerpts when compared to speech, demonstrating that the use of the drum as a speech surrogate can also be extended to the imitation of song as well. Previous research has often treated drum patterns as either representations of pure linguistic utterances (e.g., poetic phrases) or has treated those same drum patterns as musical patterns depending on the disciplinary focus of the researcher. Following Nketia (1963), most researchers discuss the signal, dance, and speech modes of the dùndún talking drum, with most acoustical analysis taking place on the speech mode of the drum (Akinbo 2019). To our knowledge, Euba (1990) is the only scholar to acknowledge the singing mode of the drum, and to date, no one has studied the acoustical features of this mode, making our study the first to do so. One reason scholars may have traditionally omitted the singing mode from discussion is due to difficulties in distinguishing between the speech and song modes of the dùndún drum (Euba 1990). However, our results show a consistent microstructural difference between speech (DT-D and DT-P) modes and singing modes (DS) on the drum, offering new evidence that can assist in acoustically differentiating between these modes in future studies. One of the differences we found between the speech and song drum modes was the inclusion of inter-phrase rhythms in several of the DS samples. These inter-phrase rhythms helped to communicate a clear and consistent pulse through the addition of repetitive rhythms between sub-phrases or in the use of one- or two-timed attacks. Research on West African drumming consistently cites the importance of a steady beat, often established and maintained via the bell pattern, a recurrent timeline played by one or more instruments in a drum ensemble (Locke 1982; Anku 1997; Oludare 2016). The inter-phrase rhythms in the DS samples may serve a similar purpose, providing a pacemaker rhythm that keeps time underneath the song sample phrases that the drum is performing. Indeed, it was found that the inter-phrase rhythm of performer B tends to follow the Long-short-short (L-S-S) pattern, known as “konkolo” rhythm among the Yoruba (Oludare 2016), and is also consistent with the accompaniment pattern in Segu Bamana drumming (Polak and London 2014). It is worth noting that the use of inter-phrase rhythms varied on a performer basis, with Performer B consistently using inter-phrase rhythms in all of his imitations of song on the drum, while Performer C used no inter-phrase rhythms. This variation between performers in our study parallels similar inter-performer differences shown by Polak and London (2014) in Malian drumming and also shows a characteristic element of dùndún performance in which drummers employ individual forms of expression while also maintaining a more general time-keeping function.

Also of note is our finding of significant cross-feature correlations between drum and vocal phrases. Correlations between the f_0 of drum and vocal phrases, and those between intensity, demonstrate that the drum can be used as a surrogate in both talking and singing modes, as mentioned. However, the significant correlation between vocal f_0 and drum intensity may indicate a perceptual correlation between f_0 and intensity in speech surrogacy drumming. Similarly, the use of drum intensity to indicate f_0 changes in speech suggests that drummers are collapsing the highly multidimensional qualities of speech into a few dimensions that can be captured by dùndún drumming.

Future research is needed to continue to deepen our understanding of the acoustical differences between the speech and singing modes of the ìyáàlù dùndún. While our study is the first to demonstrate clear microstructural acoustical differences between these modes, further study is warranted. Additionally, further examination of the different modes of the dùndún could compare them to modes of other percussion instruments in the talking drum family that do not seem to mimic speech as directly. Such analyses could complement our current findings, especially how other drumming modes are similar to or differ from the dùndún's speech and song surrogacy modes. Such a focus promises to reveal the deeper relations between speech and its imitation on musical instruments, but also aid further understanding of general percussive mechanisms of speech surrogates.

In conclusion, our findings confirm that the dùndún can very accurately mimic microstructural acoustic temporal characteristics of Yorùbá vocalization when doing so directly, and that this acoustic match systematically decreases as more musical context is specified. This pattern of imitation is maintained to a high degree for both speech and song in talking drum performances and is largely absent when used principally as a dance rhythm instrument. Frequency and intensity characteristics, on the other hand, are more closely matched for song than talking, which may be due to the constrained frequency range in the vocal talking phrases. Our findings acoustically verify the distinction between the four drumming mode categories DT-D, DT-P, DS, and D-DR, and their acoustical match to corresponding verbal modes. Understanding the process of music and speech interconnectivity in the dùndún talking drum helps clarify acoustical properties that overlap between these modes of communication and verifies the potential functionality of speech surrogacy communication systems.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CD, KK, and MM designed the study. CD collected the data. KK, JP, and MG analyzed the data. CD, KK, and JP wrote the article. MM contributed to writing. All authors reviewed, edited and approved the content of the article.

ACKNOWLEDGMENTS

Special thanks to Mitchell Ohriner, who offered advice on methods for analyzing timing of syllable onsets in speech, and to Robert Ghilduta for assistance. We also thank the performers and the ELYON studio staff for the time spent on recording the samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.652690/full#supplementary-material

REFERENCES

Adegbite, A. (1988). The Drum and its Role in Yoruba Religion. *J. Religion Afr.* 18, 15–26. doi:10.1163/157006688X00207

Akinbo, S. (2019). Representation of Yorùbá Tones by a Talking Drum: an Acoustic Analysis. *Ling. Lang. Afric.* 5, 11–23. doi:10.31234/osf.io/43gf6

Akpabot, S. (1975). The Talking Drums of Nigeria. *Afric. Mus.* 5, 36–40. doi:10.21504/amj.v5i4.1616

Akpabot, S. (1986). *Foundations of Nigerian Traditional Music*. Ibadan: Spectrum Books.

Allen, G. D. (1972). The Location of Rhythmic Stress Beats in English: an Experimental Study I. *Lang. Speech* 15, 72–100. doi:10.1177/002383097201500110

Anku, W. (1997). Principles of Rhythm Integration in African Drumming. *Black Music Res. J.* 17, 211–238. doi:10.2307/779369

Arewa, O., and Adekola, N. (1980). Redundancy Principles of Statistical Communications as Applied to Yorùbátalking-Drum. *Anthropos* 75, 185–202.

Ashley, R. (2002). Do[n't] Change a Hair for Me: The Art of Jazz Rubato. *Music Percept.* 19, 311–332. doi:10.1525/mp.2002.19.3.311

Bamgbose, A. (2000). *A Grammar of Yoruba*. Cambridge: Cambridge University Press, Vol. 5.

Barber, K., and Oyetade, A. (1998). *Yoruba Wuyi. Iwe Kinni (Book One): A Beginners' Course in Yoruba*. London: Hakuna Matata Press.

Benadon, F. (2006). Slicing the Beat: Jazz Eighth-Notes as Expressive Microrhythm. *Ethnomusicology* 50, 73–98.

Bloothofft, G., and Plomp, R. (1986). The Sound Level of the Singer's Formant in Professional Singing. *The J. Acoust. Soc. America* 79, 2028–2033. doi:10.1121/1.393211

Boersma, P., and Weenink, D. (2021). Praat: Doing Phonetics by Computer. Version 6.1.37. Available at: http://www.praat.org/(Accessed January 1, 2021).

Byrd, D. (1992). Perception of Assimilation in Consonants Clusters:A Gestural Model. *Phonetica* 49 (1), 1–24. doi:10.1159/000261900

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed.. Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd ed.. New York: Routledge.

Connell, B., and Ladd, D. R. (1990). Aspects of Pitch Realisation in Yoruba. *Phonology* 7, 1–29. doi:10.1017/s095267570000110x

Durojaye, C. (2019a). *Evoked Emotional Responses in the Performance Practices of Selected Yorùbá Dùndún Ensembles [dissertation]*. [South Africa]: University of Cape Town.

Durojaye, C. (2019b). Born a Musician: the Making of a Dundun Drummer Among the Yoruba People of Nigeria. *J. Arts. Hum.* 8, 43–55. doi:10.18533/journal.v8i1.1551

Durojaye, C. (2020). The Dùndún Talking Drum of the Yorùbá Community in South-West Nigeria. *J. Arts. Hum.* 9, 11–19. doi:10.18533/journal.v9i7.1902

Durojaye, C., Fink, L., Roeske, T., Wald-Fuhrmann, M., and Larrouy-Maestri, P.. (2021). Perception of Nigerian Talking Drum Performances as Speech-like vs Music-like: the Role of Familiarity and Acoustic Cues. Manuscript in Preparation.doi:10.4324/9781003175049

Eme, C. A., and Uba, E. D. (2016). A Contrastive Study of the Phonology of Igbo and Yoruba. *Ujah J. Art Hum.* 17, 65–84. doi:10.4314/ujah.v17i1.4

Euba, A. (1990). *Yorùbá Drumming: The Dùndún Tradition*. Bayreuth: Bayreuth African Studies.

Fox, R. A. (1983). Perceptual Structure of Monophthongs and Diphthongs in English. *Lang. Speech* 26, 21–60. doi:10.1177/002383098302600103

Gay, T. (1970). A Perceptual Study of American English Diphthongs. *Lang. Speech* 13, 65–88. doi:10.1177/002383097001300201

Goldberg, D. (2015). Timing Variations in Two Balkan Percussion Performances. *Emp. Mus. Rev.* 10, 305–328. doi:10.25148/lawrev.10.2.8

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal Properties of Spontaneous Speech-A Syllable-Centric Perspective. *J. Phonetics* 31, 465–485. doi:10.1016/j.wocn.2003.09.005

Hacki, T. (1996). Comparative Speaking, Shouting and Singing Voice Range Profile Measurement: Physiological and Pathological Aspects. *Logopedics Phoniatrics Vocology* 21, 123–129. doi:10.3109/14015439609098879

Harsin, C. A. (1997). Perceptual-center Modeling Is Affected by Including Acoustic Rate-Of-Change Modulations. *Perception & Psychophysics* 59 (2), 243–251. doi:10.3758/bf03211892

Hombert, J. M. (1976). Consonant Types, Vowel Height, and Tone in Yorùbá. *UCLA Working Pap. Phonetics* 33, 40–54.

House, D. (1990). *Tonal Perception in Speech*. Lund: Lund University Press.

House, D. (1995). "Perception of Prepausal Tonal Contours: Implication for Automatic Stylization of Intonation," in Proceedings of the 4th European Conference on Speech Communication and Technology, 4 (1), 949–952. Available at: https://www.speech.kth.se/prod/publications/files/3131.pdf.

Huggins, W. F. (1972). Just Noticeable Differences for Segment Duration in Natural Speech. *J. Acous. Soc. Am.* 51, 270. doi:10.1121/1.1912971

Jun, J. (2004). "Place Assimilation," in *Phonetically Based Phonology*. Editors B. Hayes, R. Kirchner, and D. Steriade (Cambridge: Cambridge University Press), 58–86. doi:10.1017/cbo9780511486401.003

Laniran, Y. O., and Clements, G. N. (2003). Downstep and High Raising: Interacting Factors in Yoruba Tone Production. *J. Phonetics* 31, 203–250. doi:10.1016/S0095-4470(02)00098-0

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., and Elenius, K. (2011). Expression of Affect in Spontaneous Speech: Acoustic Correlates and Automatic Detection of Irritation and Resignation. *Computer Speech Lang.* 25, 84–104. doi:10.1016/j.csl.2010.03.004

Lehiste, I., and Peterson, G. E. (1961). Transitions, Glides, and Diphthongs. *J. Acoust. Soc. America* 33, 268–277. doi:10.1121/1.1908638

Locke, D. (1982). Principles of Offbeat Timing and Cross-Rhythm in Southern Eve Dance Drumming. *Ethnomusicology* 26, 217–246. doi:10.2307/851524

Madison, G. (2001). Variability in Isochronous Tapping: Higher Order Dependencies as a Function of Intertap Interval. *J. Exp. Psychol. Hum. Perception Perform.* 27, 411–422. doi:10.1037/0096-1523.27.2.411

Mazzoni, D. (2013). AudacityI: Free Audio Editor and Recorder (Version 2.0.3) [Mac OS X]. Available at: https://audacityteam.org (Accessed January 1, 2021).

McBeath, M. K., and Neuhoff, J. G. (2002). The Doppler Effect Is Not what You Think it Is: Dramatic Pitch Change Due to Dynamic Intensity Change. *Psychon. Bull. Rev.* 9, 306–313. doi:10.3758/BF03196286

Mertens, P. (2004). "The Prosogram: Semi-automatic Transcription of Prosody Based on Tonal Perception Model" in Proceedings of Speech Prosody 2004, Nara. *Japan*, 23–26. doi:10.20396/joss.v4i2.15053

Mertens, P. (2013). "Automatic Labelling of Pitch Levels and Pitch Movements in Speech Corpora," in Proceedings of Tools and Resources for the Analysis of Speech Prosody 2013. Editors B. Bigi and D. Hirst (Aix-en-Provence: Laboratoire Parole et Language), 42–46. ISBN: 978-2-7466-6443-2.

Mertens, P. (2014). Polytonia: a System for the Automatic Transcription of Tonal Aspects in Speech Corpora. *J. Speech Sci.* 4, 17–57.

Neuhoff, J. G., and McBeath, M. K. (1996). The Doppler Illusion: the Influence of Dynamic Intensity Change on Perceived Pitch. *J. Exp. Psychol. Hum. Perception Perform.* 22, 970–985. doi:10.1037/0096-1523.22.4.970

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

Ohriner, M. (2018). *"Expressive Timing,"* in *The Oxford Handbook of Critical Concepts in Music Theory*. Editors A. Rehding and R. Steven (New York: Oxford University Press), 369–396.

Ohriner, M. (2019). Lyric, Rhythm, and Non-alignment in the Second Verse of Kendrick Lamar's "Momma". *Mus. Theo. Online* 25, 1. doi:10.30535/mto.25.1.10

Ola Orie, O., and Pulleyblank, D. (2002). Yoruba Vowel Elision: Minimality Effects. *Nat. Lang. Ling. Theo.* 20, 101–156. doi:10.1023/a:1014266228375

Oludare, O. (2016). An Analysis of the Two Forms of the 'Kónkónkóló' Rhythm in Sakara Music. *J. Assoc. Nig. Mus.* 10, 186–196.

Oyetade, A., Hayward, K., and Watkins, J. (2003). "The Phonetic Interpretation of Register: Evidence from Yoruba," in *Phonetic Interpretation. Papers in Laboratory Phonology IV*. Editors J. Local, R. Ogden, and R. Temple (Cambridge: Cambridge University Press), 305–321.

Ọdéjọbí, Ọ. À. (2007). A Quantitative Model of Yorùbá Speech Intonation Using Stem-ML. *INFOCOMP J. Computer Sci.* 6 (3), 47–55.

Panteleoni, H. (1972). Three Principles of Timing in Anlo Dance Drumming. *Afr. Music* 5, 50–63. doi:10.21504/amj.v5i2.1419

Patten, K. J., Greer, K., Likens, A. D., Amazeen, E. L., and Amazeen, P. G. (2020). The Trajectory of Thought: Heavy-Tailed Distributions in Memory Foraging Promote Efficiency. *Mem. Cogn.* 48, 772–787. doi:10.3758/s13421-020-01015-7

Patten, K. J., and McBeath, M. K. (2020). "The Difference between Shrieks and Shrugs: Spectral Envelope Correlates with Changes in Pitch and Loudness," in Proceedings of the 2nd International Conference on Timbre (Timbre 2020), September 2020, 3–4. Thessaloniki (online), Greece

Patten, K. J., McBeath, M. K., and Baxter, L. C. (2019). Harmonicity: Behavioral and Neural Evidence for Functionality in Auditory Scene Analysis. *Aud. Percept. Cogn.* 1, 158–172. doi:10.1080/25742442.2019.1609307

Peterson, G. E., and Lehiste, I. (1960). Duration of Syllable Nuclei in English. *J. Acoust. Soc. America* 32, 693–703. doi:10.1121/1.1908183

Polak, R., and London, J. (2014). Timing and Meter in Mande Drumming from Mali. *Music Theor. Online* 20, 1-27. doi:10.30535/mto.20.1.1

Pompino-Marschall, B. (1991). The Syllable as a Unit and the So-Called P-center Effect. *Forschungberichte des Instituts für Phonetik und Sprachliche Kommunikationder Universitdt München (FIPMK)* 29, 65–123.

Pulleyblank, D. (2009). Vowel Deletion in Yoruba. *J. Afri. Lang. Ling.* 10, 117–136. doi:10.1515/jall.1988.10.2.117

Repp, B. H. (1992). Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann's "Träumerei". *J. Acoust. Soc. America* 92, 2546–2568. doi:10.1121/1.404425

Repp, B. H. (2005). Sensorimotor Synchronization: a Review of the Tapping Literature. *Psychon. Bull. Rev.* 12, 969–992. doi:10.3758/BF03206433

Rhodes, T., and Turvey, M. T. (2007). Human Memory Retrieval as Lévy Foraging. *Physica A: Stat. Mech. its Appl.* 385, 255–260. doi:10.1016/j.physa.2007.07.001

Scharine, A. A., and McBeath, M. K. (2019). Natural Regularity of Correlated Acoustic Frequency and Intensity in Music and Speech: Auditory Scene Analysis Mechanisms Account for Integrality of Pitch and Loudness. *Aud. Percept. Cogn.* 1, 205–228. doi:10.1080/25742442.2019.1600935

Sluijter, A. M. C., and van Heuven, V. J. (1996). "Acoustic Correlates of Linguistic Stress and Accent in Dutch and American English," in Proceedings of Fourth International Conference on Spoken Language Processing (USA: ICSLP '96, Philadelphia, PA), Vol. 2, 630–633.

Sotunsa, M. (2009). *Yorùbá Drum Poetry*. London: Stillwatersstudios.

Toh, A. M., Togneri, R., and Nordholm, S. (2005). Spectral Entropy as Speech Features for Speech Recognition. *Proc. PEECS* 1, 92.

Vidal, A. (2012). ""Traditional Music Instruments of the South-West Nigeria: Forms and Distribution," in *Selected Topics on Nigerian Music*. Editors F. Adedeji (Ile-Ife: Ife: Obafemi Awolowo University Press),", 43–53.

Villepastour, A. (2010). *Ancient Text Messages of the Yorùbá Bata Drum*. Burlington, VT: Ashgate.

Villepastour, A. (2014). "Talking Tones and Singing Speech Among the Yorùbá of Southwest Nigeria," in *Jahrbuch des Phonogrammarchivs der Österreichischen Akademie der Wissenschaften*, 29–47.

Yu, C. S.-P., McBeath, M. K., and Glenberg, A. M. (2021). The Gleam-Glum Effect: Phonemes Generically Carry Emotional Valence. *J. Exp. Psychol. Learn. Mem. Cogn.* 47. doi:10.1037/xlm0001017

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