Check for updates

Speech and beatboxing cooperate and compromise in beatrhyming

Reed Blaylock[1]*, Ramida Phoolsombat[1] and Kaila Mullady[2]

[1]Department of Linguistics, University of Southern California, Los Angeles, CA, United States, [2]Academy of Noise, New York, NY, United States

Beatrhyming is a type of multi-vocalism in which an individual beatboxes and speaks (i.e., sings or raps) at the same time by interweaving beatboxing sounds and speech sounds within words and phrases. The measurements in this case study of a beatrhyming performance focus on one-to-one sound "replacements" in which a beatboxing sound is integrated into a word by taking the place of a speech sound. The analysis unfolds in two parts: first, a count of how many times beatboxing sounds used in place of speech sounds matched the intended speech sounds for vocal tract constrictor and constriction degree; and second, an assessment of whether the beatboxing sound patterns in beatrhyming (beatboxing with simultaneous lyrics) have the same degree of rhythmic structure as the beatboxing sound patterns in beatboxing (without lyrics). Despite having disparate aims, the separate speech and beatboxing systems work together to create a well-organized combined behavior. Speech tasks (i.e., communicating the linguistic message of the lyrics) are achieved in beatrhyming by replacing some speech sounds with beatboxing sounds that match the speech segment in vocal tract constrictor and in manner/constriction degree. Beatboxing tasks (i.e., establishing a musical rhythm) are achieved through the inviolable use of Outward K Snares {K̚} on the backbeat. Achieving both of these aims in the same performance requires flexibility and compromise between the speech and beatboxing systems. In addition to providing the first scientific description and analysis of beatrhyming, this article shows how beatrhyming offers new insight for phonological theories built to describe spoken language.

1 Introduction

One of many questions in contemporary linguistic research in phonology is how the task of speech interacts with other concurrent motor tasks. Many behaviors may not fall under the purview of speech in the strictest traditional sense but nevertheless collaborate with speech, resulting in a wide variety of spatiotemporally organized speech performance modalities. These include co-speech manual gestures (Krivokapić, 2014; Parrell et al., 2014; Danner et al., 2019), co-speech ticcing produced by speakers with vocal Tourette's disorder (Llorens, 2022), text-setting in singing and chanting (Hayes and Kaun, 1996), tone-tune alignment in the sung music of languages with lexical tone (Schellenberg, 2012; McPherson and Ryan, 2018; Schellenberg and Gick, 2020), and more. Studying these and other multi-task behaviors illuminates the flexibility of speech units and their organization in a way that studying talking alone cannot.

This article introduces beatrhyming, a type of multi-vocalism in which a person produces beatboxing sounds while they are speaking (i.e., singing or rapping). Beatrhyming has scarcely been investigated from a scientific perspective (see Fukuda et al., 2022 for the lone piece of scholarly beatrhyming literature that we know of). This article's case study of one beatrhyming performance from an expert beatboxer provides the first detailed account of how speech and beatboxing can be fluidly combined in beatrhyming. The ensuing analysis contributes a new perspective for how speech units can be flexibly organized with units from otherwise unrelated task systems.

Sections 1.1 introduces beatboxing as precursor to an introduction of beatrhyming in Section 1.2. Section 1.3 lays out hypotheses and predictions for the segmental and temporal organization of beatrhyming.

1.1 Beatboxing

Beatboxing is a type of non-linguistic vocal percussion named for the human beatboxes who created the percussive back bone for hip hop emcees to rap over. Contemporary beatboxers may produce music solo, in ensembles with other beatboxers, or as an accompaniment to linguistic music (e.g., rapping or a cappella singing) performed by a different individual.

The art of beatboxing is shaped by its musical influences and the sound-making potential of the vocal tract. Because of beatboxing's origins as a tool for supporting hip hop emcees, it is generally assumed that a beatboxer's primary intention is to imitate the sounds of a drum kit, electronic beat box, and a variety of other sound effects (Lederer, 2005/2006; Stowell and Plumbley, 2008; Pillot-Loiseau et al., 2020). Hip hop drumming generally features a snare on beats two and four (Greenwald, 2002), so the musical structure of beatboxing utterances (or "beat patterns") almost always includes a sound emulating a snare drum on beats two and four. Likewise, non-continuant stops and affricates that aptly emulate the sounds of a drum kit are fundamental to beatboxing and very common in beat patterns, as described more below. The core tasks of beatboxing thus include production of these percussive sounds in the vocal modality, with certain rhythmic regularities.

Novice beatboxers are often taught to derive their first and most fundamental beatboxing sounds from speech. For example, starting from the English phrase "boots and cats" [buts ænd kæts], devoicing the vowels and intensifying the consonants leads to a basic beat pattern featuring a Kick Drum, Closed Hi-Hat, K Snare, and another Closed Hi-Hat. In the International Phonetic Alphabet, the outcome could be transcribed as something like [p' ts' k͡ʟ̆ ts']. In Standard Beatbox Notation (SBN) (Tyte and Splinter, 2002/2014; Stowell, 2003), the same sounds would be transcribed in curly brackets as {B t K t}. Table 1 lists and describes the five main beatboxing sounds that will be referred to in this article. These beatboxing sounds share major vocal constrictors with their speech originators (i.e., [b] and {B} are both labial) but differ in their articulatory dynamics (Paroni et al., 2022; Blaylock and Narayanan, 2023) and air pressure control strategies.

Beatboxing sounds are produced with a wide range of airstream mechanisms, including some that are not attested as phonologically contrastive in languages of the world—like pulmonic ingressive and lingual (velaric) egressive airstreams (Stowell and Plumbley, 2008, 2010; de Torcy et al., 2014; Blaylock et al., 2017; Pillot-Loiseau et al., 2020; Dehais-Underdown et al., 2021; Paroni et al., 2021a,b). Common beatboxing sounds learned early are often glottalic egressive (ejectives) (Proctor et al., 2013; Blaylock et al., 2017; Dehais-Underdown et al., 2019; Paroni, 2022), though some beatboxers may perform these sounds with use pulmonic egressive airstream instead (Patil et al., 2017). Beatboxers are also known to use multiple airstreams simultaneously, as when humming (pulmonic egressive) and making percussive sounds (lingual egressive and lingual ingressive) at the same time (Paroni et al., 2021a,b; Paroni, 2022).

A guiding theme in beatboxing science is the study of vocal agility and capability (Dehais-Underdown et al., 2021). The complex sounds and patterns of beatboxing offer a chance to build a stronger general phonetic framework for studying the relationship between linguistic tasks, musical tasks, cognitive limitations, physical limitations, and motor constraints in speech production and evolution. The connection between beatboxing and speech via the vocal tract they share has also generated interest in using beatboxing for speech therapy (Pillot-Loiseau et al., 2020) for both adults (Icht, 2018, 2021; Icht and Carl, 2022) and children (Martin and Mullady, n.d.) (see also Himonides et al., 2018; Moors et al., 2020).

A gap in the beatboxing science literature so far is a framework for beatboxing sound organization. A few scholars have suggested phonological analogies for beatboxing sound organization. Evain et al. (2019) and Paroni et al. (2021a,b) posit the existence of a "boxeme" by analogy to the phoneme—an acoustically and articulatorily distinct building block of a beatboxing sequence. Separately, Guinn and Nazarov (2018) argue for sub-segmental features in beatboxing based on evidence from variations in beat patterns and phonotactic place restrictions (an absence of beatboxing coronals in prominent metrical positions), though they do not link features back to larger segment-sized units. But the status of the boxeme as a fundamental combinatorial unit is far from settled, as are other questions about the extent to which beatboxing has its own phonology akin to speech phonology. Likewise, questions also remain about whether or how speech representations and beatboxing representations are connected.

TABLE 1 Sounds of beatboxing used during beatrhyming sections of this performance.

Name	SBN	IPA	Description
Kick Drum	{B}	[p']	Voiceless ejective bilabial stop
PF Snare	{PF}	[p͡f']	Voiceless ejective labio-dental affricate
Closed Hi-Hat	{t}	[t͡s']	Voiceless ejective alveolar affricate
Center K	{k}	[k']	Voiceless ejective velar stop
Outward K Snare	{K}	[k͡ʟ̆]	Voiceless pulmonic egressive lateral velar affricate

These last are particularly relevant for understanding how different specialized cognitive domains may be integrated; they may also lead to more effective ways to use beatboxing as a tool for speech therapy.

This article adds to the nascent work on beatboxing sound organization by investigating how beatboxing and speech are interwoven in beatrhyming. The nature of the relationship between beatboxing and speech found here suggests that the most apt beatboxing representations will have a concrete link to the phonological units of speech via the capabilities of the vocal tract. The question of the nature of those representations is taken up in Section 4.

1.2 Beatrhyming

Beatrhyming is a relatively new variety of music that unites speech and beatboxing within a performance. It is performed by weaving beatboxing sounds into and among the words of a verbal behavior—often singing, but possibly other behaviors like speaking and rapping. Beatrhyming thus combines the musical qualities of beatboxing described above with the goals of a speech act. These combinations can manifest in various ways. For example, Rahzel's beatrhyming performance of "If Your Mother Only Knew"[1] (an adaption of Aaliyah's "If Your Girl Only Knew") uses primarily Kick Drums whereas Kaila Mullady (the third author, whose beatrhyming is analyzed here) more often uses a variety of beatboxing sounds in her beatrhyming. In addition to these and other notable beatrhyming performers like Kid Lucky, more and more beatboxers are taking up beatrhyming as an artistic variation of their beatboxing; we can therefore expect that more and more beatrhyming variety will emerge in the future.

In beatrhyming, speech sounds and beatboxing sounds may be produced in sequence or may overlap temporally. Consider if the word "got" [gat] were produced in conjunction with a beatboxing sound. It could be produced as a sequence of beatboxing and speech sounds as {B}[gat] (a Kick Drum, followed by the word "got"). Or, that Kick Drum might completely overlap with the word-initial [g] resulting in {B}[at]. Or, partial overlap might occur if the Kick Drum appeared part-way through the vowel, resulting in [ga]{B}[at]. Fukuda et al. (2022) even found beatrhyming examples in which beatboxing sounds temporally overlap with entire words. This paper primarily focuses on cases of the second type in which a beatboxing sound fully overlaps a single speech sound, though Section 4 returns to a discussion of partial overlap. Going forward, we refer to these cases of one-for-one overlap as "replacements."

Beatrhyming sound replacement is illustrated by an excerpt from the third author's composition "Dopamine" in Figure 1: in the beatrhymed word "dopamine", the Closed Hi-Hat {t} replaces the intended speech sound [d], and Kick Drum {B} replaces the intended speech sound [p]. In both cases, the [d] and [p] can be seen segmented on the second tier with the same temporal interval as the replacing beatboxing sound labeled on the third tier. Figure 1

also features an example of partial overlap, an Outward K Snare {K} that begins in the middle of an [i].

Sections 1.3 presents hypotheses and predictions about how beatboxing and speech may (or may not) cooperate to support the achievement of their respective tasks. Section 2 presents the method used for analysis, and Section 3 describes the results. Finally, Section 4 offers a phonological account of beatrhyming and argues that dynamical action units are more suitable than static symbolic units for such an account.
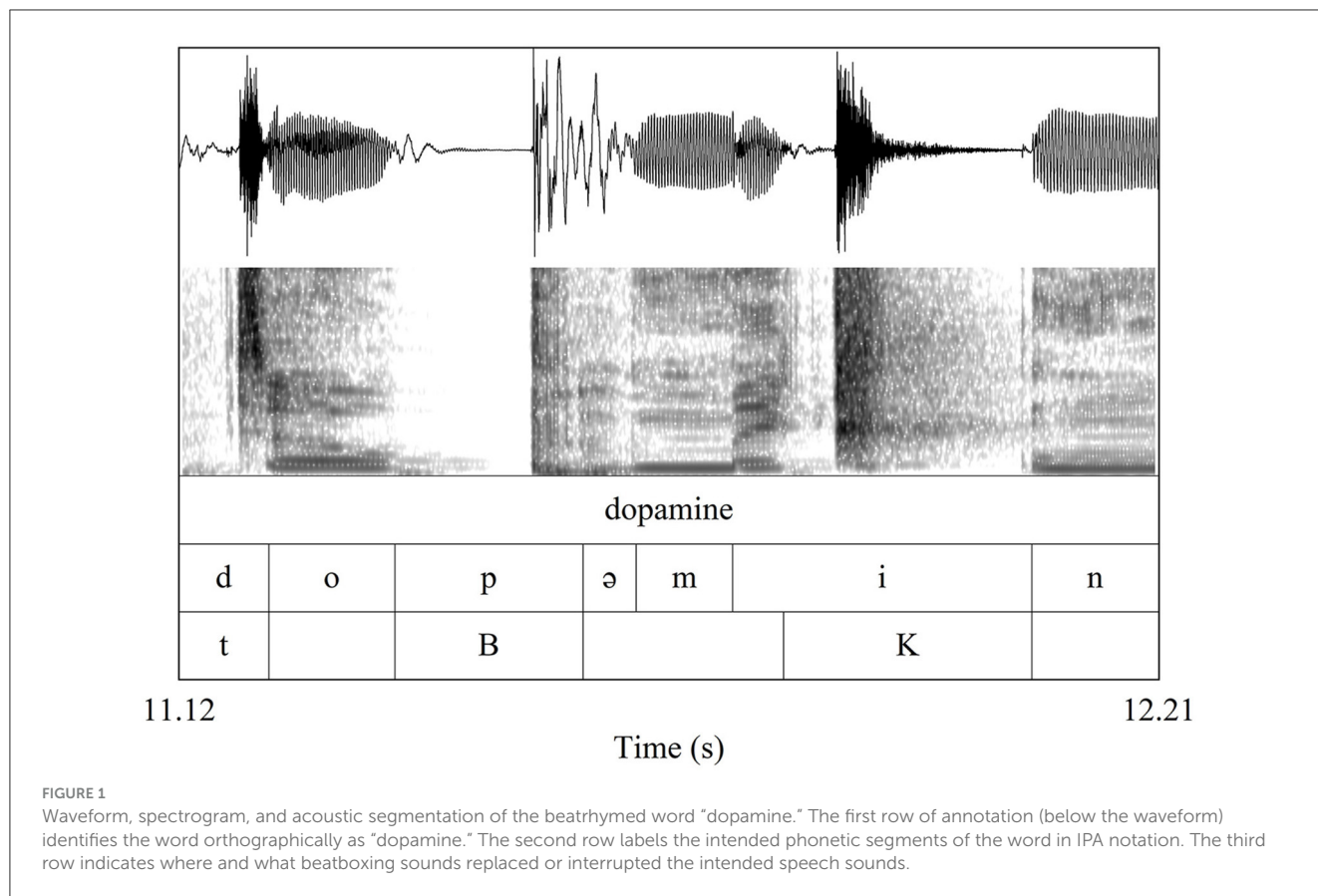
1.3 Hypotheses and predictions

The relationship between language and music—especially tonal music—has been discussed at great length (Lerdahl and Jackendoff, 1983; Feld and Fox, 1994; Patel, 2008). Many similarities between musical and linguistic structure have been noted and modeled, including syntactic structure (Steedman, 1984, 1996; Rohrmeier, 2011; Rohrmeier and Pearce, 2018) and prosodic structure (see Katz, 2022 for a recent overview). Speech and singing are not distinct, unrelated acts; rather, they lie on a continuum of (among other things) rhythmicity and melody (List, 1963; Schellenberg, 2012; Cummins, 2020). The speech-to-song illusion demonstrates this clearly: when a listener hears the same spoken phrase repeated several times, their perception of the phrase changes from hearing speech to hearing song (Deutsch et al., 2011).

These similarities facilitate text settings in which words are arranged to musical meter in ways that resemble the spoken rhythm (Palmer and Kelly, 1992; Halle and Lerdahl, 1993; Hayes and Kaun, 1996; Kiparsky, 2006; Hayes, 2009). In text setting where speech phrases are mapped to musical phrases, the segmental content of the lyrics is somewhat irrelevant. For example, the English nursery rhyme "Twinkle Twinkle Little Star" (Taylor and Taylor, 1806) and the ABC Song have different text set to the same tune (the French "Ah! vous dirai-je, maman"). And although adapting speech rhythm to accommodate musical rhythm may impose minor perturbations on the timing of the actions of the vocal articulators, like when a vowel is sustained for a long note in a song when it would otherwise be shorter in speech, the identity of the vowel does not fundamentally change because no vocal articulations are added or removed or replaced with others.

When beatrhyming is performed as the union of beatboxing and song, as it is in this case study, many of the insights about the relationship between speech and song likely apply. In fact, we take it for granted that the composition discussed here uses song melody and rhythm that reasonably reflect the speech prosody. The focus of this article is instead on the heretofore unknown relationship between speech sounds and beatboxing sounds in beatrhyming.

The mapping between beatboxing and speech is not the same as the text setting cases in which words are aligned to abstract metrical music structure, leaving the words otherwise unperturbed. In beatrhyming, beatboxing sounds often temporally overlap with sung words; because the beatboxing sounds and the speech sounds both require control of the vocal articulators, one result of temporal overlap is the type of replacement described earlier in which a beatboxing sound occurs when a speech sound is expected. This overlap is much more likely to have significant consequences

---

FIGURE 1
Waveform, spectrogram, and acoustic segmentation of the beatrhymed word "dopamine." The first row of annotation (below the waveform) identifies the word orthographically as "dopamine." The second row labels the intended phonetic segments of the word in IPA notation. The third row indicates where and what beatboxing sounds replaced or interrupted the intended speech sounds.

to intelligibility since crucial phonologically contrastive content manifests as a non-contrastive beatboxing sound.

A similar tension between musical aims and language aims has been observed in the literature on melodic text settings of languages with lexical tone. Intelligibility of the lyrics in these tunes could ostensibly be maximized by singing exactly and only the underlying lexical speech tones, but that is usually not what happens. Text settings tend to avoid sung melodies that directly oppose speech melody (i.e., rising musical melody for a falling speech contour), but sung melodies are not necessarily faithful to the spoken melodies either: instead of rising speech melody manifesting as rising song melody, for example, the song melody might stay level (Schellenberg, 2012). In other words, melodies tend to reflect the spoken pitch contour for the same words—but only to a certain extent (Schellenberg, 2012; Schellenberg and Gick, 2020).

The question for beatrhyming is whether speech or beatboxing always accommodates the other, or if they compromise to partially satisfy both their aims. Section 1.3.1 considers what it would look like if beatboxing and speech compromise in the selection of the beatboxing sound that replaces a speech sound—and what a lack of compromise would look like. Rhythm is an important factor to consider in this because beatboxing has its own rhythmic organization unrelated to speech timing. Does beatboxing accommodate to speech by sacrificing its rhythmic organization in service to a well-structured message, or are the lyrics organized around a robust beat pattern? Section 1.3.2 examines these possibilities in greater detail.

1.3.1 Segmental content: constrictor matching and constriction-degree matching

Depending on the nature of the beatboxed replacements, cases like the complete replacement of [d] and [p] in the word "dopamine" from Figure 1 could be detrimental to the tasks of speech production and/or perception. In the production of the word "got" [gat], the [g] is intended to be performed as a dorsal stop. If the [g] were replaced by a beatboxing dorsal stop, perhaps a velar ejective (Center K {k}), there would be some articulatory and acoustic similarity between the intended speech sound and the produced beatboxing sound. In this way, the speech task could be partially achieved while simultaneously beatboxing.

But if the intended [g] were replaced with a labial Kick Drum {B}, perhaps because part of the aesthetic of beatboxing required a Kick Drum at that moment, the resulting replacement would deviate further from the intended speech task. If the difference were great enough, making replacements that do not support the intended speech goals might lead to listeners misperceiving beatrhyming lyrics—in this case, perhaps hearing "pot" [pat] instead of "got."

So then, if the speech task and the beatboxing task influence each other during beatrhyming, it may be optimal for the speech task to have beatrhyming replacements that match the intended speech signal as often as possible and along as many phonetic dimensions as possible. This article investigates whether, and to what extent, beatrhyming replacements support the speech task by

matching the intended speech sounds in active constrictor type and constriction degree, as described below.

The speech consonants[2] under consideration here and the beatboxing sounds that replace them (listed in Table 1) can all be described as having one of three major consonant constrictor types: *labial*, created by bringing together the upper and lower lip or the upper teeth and lower lip; *coronal*, formed by lifting the tongue tip toward the alveolar ridge or teeth; and *dorsal*, achieved by a closure of the tongue body against the velum or palate. Constrictor-matching would be found if speech labials tend to be replaced by beatboxing labials, speech coronals tend to be replaced by beatboxing coronals, and speech dorsals tend to be replaced by beatboxing dorsals.

Constriction degree is approximated by the manner of articulation such as closure or frication. The prediction of constriction degree matching follows a similar logic to the prediction for constrictor matching. If beatrhyming replacements are made with an aim of satisfying speech tasks, then replacements are more likely to occur between speech sounds and beatboxing that have similar constriction degrees. Since the beatboxing sounds in this data set are stops and affricates, the prediction of the main hypothesis is that speech stops and affricates will be replaced more frequently than speech sounds of other manners of articulation.

To summarize: the main hypothesis is that speech and beatboxing interact with each other in beatrhyming in a way that supports the accomplishment and perception of intended speech tasks. This predicts that beatboxing sounds and the intended speech sounds they replace are likely to match in constrictor and constriction degree and thereby yield an acoustic signal with properties that allow for the recovery of the intended speech sounds, and consequently words, in perception. Conversely, the null hypothesis is that the two systems do not interact or do not interact in a way that supports the intended speech tasks, predicting that beatboxing sounds replace speech sounds with no regard for the intended constrictor or constriction degree.

1.3.2 Timing: beat pattern repetition

Beatboxing beat patterns have their own predictable sound organization. The presence of a snare drum sound on the backbeat (beats 2 and 4) in particular is highly consistent, but beat patterns are also often composed of regular repetition at larger time scales. Speech utterances are highly temporally structured as well, but the sequence of words (and therefore sounds composing those words) is determined less by sound patterns and more by syntax (cf. Shih and Zuraw, 2017). However, artistic speech like poetry and singing leverage the flexibility of language to express similar ideas with a variety of utterance forms to yield different rhythmic aesthetics, alliteration, rhyming, and other specific sound patterns.

------

2   In this article, the beatrhyming sound replacements are limited to complete substitutions of beatboxing sounds for speech consonants. Section 4.1.2 addresses examples of beatboxing sounds partially overlapping with vowels. In principle it is possible for beatboxing sounds to completely replace vowels as well (Fukuda et al., 2022), but that does not happen in the song analyzed here.

There are (at least) two ways beatboxing and speech could temporally interact while maximizing constrictor-matching and constriction-degree matching as hypothesized in Section 1.3.1. First, the lyrics of the song could be planned without any regard for the resulting beat pattern. Any co-speech beatboxing sounds would be planned based on the words of the song, prioritizing faithfulness to the segmental content of the intended spoken utterance. Alternatively, the lyrics could be planned "around" a beatboxing beat pattern, prioritizing the performance of an aesthetically appealing beat pattern. Or of course, both aims could be satisfied to some degree. The basic counts of constrictor matches described in Section 1.3.1 could remain unaffected by this balance, but the two hypotheses do predict that the resulting beat patterns will be structured differently. Specifically, prioritizing the beatboxing beat pattern predicts that beatrhyming will feature highly regular/repetitive beatboxing sound sequences characteristic of beatboxing music, whereas prioritizing the segmental speech structure would lead to irregular (or less regular or non-repeating) beatboxing sound sequences. The rest of this section discusses these predictions in more detail.

A beat pattern is often highly repetitive. Figure 2 shows a transcription of the beatboxing sounds from the first beatboxed (non-lyrical, so not beatrhymed) section of "Dopamine." Each row or "line" or "measure" of music is a temporal interval containing two snare sounds—first an Outward K Snare {K}, then either an Outward K Snare {K} or an inhalation sound effect {in} serving as the snare. The meter of the lines is construed so that the first snare of a line is on beat 2 and the second snare is on beat 4. Labels for the remaining beats (1, 3) and sub-beats (1.5, 2.5, 3.5, 4.5) are filled in based on the rest of the rhythmic structure.

Notice that the first line of the transcription is not unique: it repeats in lines 2, 3, 6, and 7. The same sounds also repeat in line 5, though with a slight change to the rhythmic structure that puts the two Closed Hi-Hats {t} in triplet timing. The fourth line is also not unique because it repeats on line 8 (albeit with another small rhythmic change). There is a great deal of repetition from line to line.

It can be useful to consider the repetition on the smaller time scale of half-measures too: minor rhythmic perturbation in line 5 aside, the first half of line 1 is identical to the first halves of each other line, including lines 4 and 8. Calling that half-measure A, the second half of line 1—half-measure B—repeats on lines 2, 3, 5, 6, and 7. A third half-measure, C, is used only twice on lines 4 and 8.

These eight beatboxing lines can be considered the hierarchical composition of three half-measures organized into larger structures, as depicted in the lower portion of Figure 2. An initial line AB (composed of half-measures A and B) is copied and joined to the first line to make a two-line phrase. A four-line phrase can be created by copying the two-line phrase and joining it with itself. There is room for variation of course, and lines may change based on the artist's musical choices. In this case, the last half-measure of the first four-line phrase is half-measure C instead of half-measure B. Then, the whole four-line phrase is then copied to create an eight-line phrase, resulting in repetition of that deviation to half-measure C at the end of both four-line phrases.

If the rhythmic tasks of beatboxing influence the rhythmic structure of beatrhyming, then the hierarchical composition for beatboxing described above can be used to predict where repeating

	1	1.5	2	2.5	3	3.5	4	4.5
Line 1	B	t t	K		th	B	in	B
Line 2	B	t t	K		th	B	in	B
Line 3	B	t t	K		th	B	in	B
Line 4	B	t t	K		B	h	K	t t
Line 5	B	t t	K		th	B	in	B
Line 6	B	t t	K		th	B	in	B
Line 7	B	t t	K		th	B	in	B
Line 8	B	t t	K		B	h	K	t t

8-line phrase

4-line phrase

2-line phrase

```
1   2   3   4     5   6   7   8          1 line
A B A B A B A C*  A* B A B A B A C     Half-measures
```
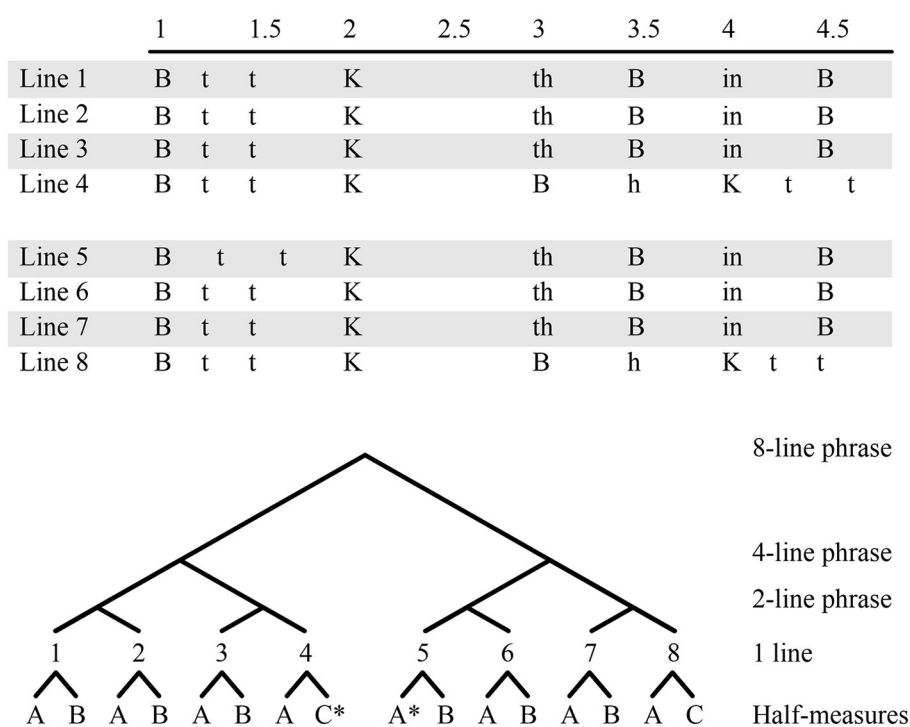
FIGURE 2
Serial and hierarchical representations of an eight-line beatboxing phrase. The asterisks in the hierarchical representation on the eighth half-measure C* and ninth half-measure A* indicate a minor rhythmic deviation.

lines are most likely to be found in an eight-line beatrhyming pattern. The initial repetition of a line to make a two-line phrase predicts that lines 1 and 2 should be similar. Likewise, copying that two-line phrase to make a four-line phrase would predict similarity between lines 3 and 4; at a larger time scale, this would also predict that lines 1 and 3 should be similar to each other, as should lines 2 and 4. In the eight-line phrase composed of two repeating four-line phrases, the repetition relationships from the previous four-line phrase would be copied over (lines pairs 5 and 6, 7 and 8, 5 and 7, and 6 and 8); repetition would also be expected between corresponding lines of these two four-line phrases, predicting similarity between lines 1 and 5, 2 and 6, 3 and 7, and 4 and 8.

Because deviations from the initial line are expected to occur in the interest of musical expression, some two-line phrases are more likely to exhibit clear repetition than others. Consider a two-line phrase composed of two lines with the structures AB and AC—their first half-measures (A) are identical, but their second half-measures (B and C) are different. If this two-line phrase AB-AC is repeated, the resulting four-line phrase would be AB-AC-AB-AC. In this example, lines 1 and 3 match as do lines 2 and 4, but lines 2 and 3 do not match and neither do 1 and 4. For this study, the search for repetition in beatrhyming is limited to just those pairs of lines that are most likely to feature repetition:

- *Adjacent* lines—lines 1 and 2, 3 and 4, 5 and 6, 7 and 8
- *Alternating* lines—lines 1 and 3, 2 and 4, 5 and 7, and 6 and 8
- *Cross-group* lines—lines 1 and 5, 2 and 6, 3 and 7, and 4 and 8

If beatboxing structure is prioritized in beatrhyming—either because beatboxing and speech aren't sensitive to each other at all or because the speech system accommodates beatboxing through lyrical choices that result in an ideal beat pattern—then sequences of co-speech beatboxing sounds should have similarly high repetitiveness compared to beat patterns performed without speech. But if speech structure is prioritized, then the beat pattern is predicted to sacrifice rhythmicity and repeated sound sequences in exchange for supporting the speech task (for example, by matching the intended constrictor and constriction degree of any speech segments being replaced as described in Section 1.3.1).

1.3.3 Summary of hypotheses and predictions

The main hypothesis is that speech and beatboxing interact during beatrhyming to accomplish their respective tasks, and the null hypothesis is that they do not. Support for the first hypothesis could appear in two different forms, or possibly both at the same time. First, if beatrhyming replacements are sensitive to the articulatory goals of the intended speech sound being replaced, then the beatboxing sounds that replace speech sounds are likely to match their targets in constrictor and constriction degree. Second, if beatboxing sequencing patterns are prioritized in beatrhyming, then sequences of beatrhyming sound replacements should exhibit structural repetitiveness akin to the non-lyrical beatboxing sequences. Failing to support either of these predictions would support the null hypothesis and the notion that speech and beatboxing are cognitively independent during beatrhyming.

2 Materials and methods

This section introduces the beatrhymer whose performance was analyzed (Section 2.1), then describes how the data were collected and coded (Section 2.2) and analyzed (Section 2.3).

2.1 The beatrhymer and the data

Kaila Mullady (third author) is an expert beatboxer and two-time female Beatboxing World Champion title holder. At the time of the recording of the piece analyzed here, she had been beatrhyming for 2 years. In those 2 years she practiced beatrhyming by reading aloud from texts and enhancing consonants into beatboxing sounds. During that time, she practiced beatrhyming while busking under the tutelage of expert beatrhymer Kid Lucky. Through this practice she developed the skill to freestyle beatrhyme more or less automatically—quickly, smoothly, and with little effort.

The data in this study come from a beatrhyming performance called "Dopamine," created and performed by the third author (publicly available on YouTube[3] This performance was recorded in January 2017 with a Shure SM7B condenser microphone in a relatively small room of an apartment. "Dopamine" includes sections of beatboxing in isolation (beatboxing sections) and sections of beatboxing interwoven with speech (beatrhyming sections). These beatrhyming sections were composed by first applying her freestyle beatrhyming skill to the lyrics, then making artistic changes as needed.

2.2 Annotation

The third author provided the lyrics of the piece to enable comparison between the intended speech sounds and their beatboxing sound replacements. The second author performed manual acoustic segmentation of "Dopamine" using Praat (Boersma, 2001); the first author verified the segmentation, and any disagreements about the annotation were discussed until consensus was reached. The third author later confirmed that the annotation was correct in identifying the five beatboxing sounds used during the beatrhymed sections of "Dopamine": Kick Drum {B}, Closed Hi-Hat {t}, PF Snare {PF}, Center K {k}, and Outward K Snare {K}. Each beatboxing sound was coded by its major constrictor as follows: {B} and {PF} were coded as "labial," {t} was coded as "coronal" (tongue tip), and {k} and {K} were coded as "dorsal" (tongue body). Full phonetic descriptions are listed in Table 1.

The annotation was further verified by examining spectral descriptors of the release for each beatboxing sound, a strategy that has been used to distinguish place of articulation among voiceless obstruents in speech (Forrest et al., 1988; Jongman et al., 2000). The spectrum for each beatboxing sound was created using Praat's default settings with a 40 ms Gaussian

---

3 https://www.youtube.com/watch?v=4BCcydkZqUo

---

window the left edge of which was manually positioned 10 ms before the release burst. Visual inspection of a three-dimensional plot of center of gravity, standard deviation, and skewness for all the beatboxing sounds revealed fairly robust clustering of the sounds of our annotation (e.g., everything marked as a Kick Drum {B} clustered together), indicating that we had correctly identified the beatboxing sounds. The measurements are available as part of the Supplementary material.

Acoustic segmentation was performed at the level of words, intended and actualized speech sounds, beatboxing sounds, and the musical beat on which beatboxing sounds were performed. For complete one-to-one sound replacements, the start and end of the annotation for the interval of an intended speech sound were the same as the start and end of the replacement beatboxing sound interval (partial sound replacements and replacements of two speech sounds by a single beatboxing sound occurred occasionally and are not included in this analysis). Beatboxing sounds replacing speech sounds preceding sonorant sounds were segmented as ending at the start of voicing for the following sonorant sound. Segment boundaries were placed manually by inspection of the waveform and narrowband spectrogram. Beatboxing sounds not followed by sonorants were determined to end where the high-amplitude energy of their release gave way to the amplitude of the ambient silence in the video. A beatboxing sound was coded as replacing a speech sound when the beatboxing sound fully masked any audible trace of the underlying intended sound.

During annotation, each beatboxing sound was manually aligned to a musical meter based on numerous iterations of listening to different portions of the audio signal. By convention, the meter was assumed to be common time—four main metrical divisions, noted as beats 1, 2, 3, and 4, within each measure. The Outward K Snare {K} was identified as regularly occurring on the backbeat, so Outward K Snares participating in setting the backbeat groove were annotated as occurring on beat 2 or beat 4. The tempo of the piece was perceived to be consistent throughout, so having established the locations of beats 2 and 4 and by keeping the pulse of the rhythm while listening (i.e., by finger tapping), it was possible to mark any beatboxing sounds that occurred on beats 1 or 3. The metrical position of sounds occurring on finer divisions of the measure (sub-beats) was determined through further iterative listening and was annotated using decimal notation (e.g., a sound occurring half-way between beat 1 and beat 2 was marked as occurring on beat 1.5).

Subsequent visual review of the Praat PointTier confirmed that sounds noted as occurring on beats 1 or 3 were roughly equidistant from beats 2 and 4 in real time. Likewise, sounds produced on sub-beats were roughly equidistant from their nearest beat neighbors. The rhythm of "Dopamine" is swung, so sounds noted as occurring on a sub-beat ending in 0.25 or 0.75 occurred closer in real time to the following sub-beat ending in 0.5 or 0.0 than to the preceding 0.0 or 0.5. Sounds produced with triplet timing (noted as 0.33 or 0.67) were performed straight; for example, in a triplet occurring over beats 2, 2.33, and 2.67, the sound at 2.33 is roughly equidistant from beats 2 and 2.67, and likewise the sound at 2.67 is roughly equidistant from beats 2.33 and 3.

2.3 Measurements

2.3.1 Constrictor-matching assessment

The mPraat software (Boril and Skarnitzl, 2016) for MATLAB was used to count the number of complete one-to-one replacements made in the entire performance (n = 88) (excluding partial replacements or cases in which one beatboxing sound replaced two speech sounds). The constrictor of the originally intended speech sound was then compared against the constrictor for the replacing beatboxing sound, noting whether the constrictors were the same (matching) or different (mismatching). Constriction degree matching was likewise measured by counting how many speech sounds of varying constriction degree (i.e., "manner;" namely, stops, affricates, nasals, fricatives, approximants) were replaced. All the beatboxing sounds that served as replacements were either stops {B} or affricates {PF, t, k, K}. A higher propensity for constriction degree matching would be found if the speech sounds being replaced were likely to also be stops and affricates rather than nasals, fricatives, or approximants.

2.3.2 Repetition assessment

Four eight-line sections (labeled B, C, D, and E) were chosen for repetition analysis ("dopamine" begins with a refrain, section A, that was not analyzed for repetition because it has repeated lyrics that would inflate the repetition measurements—the intent is to assess whether beat patterns in beatrhyming are as repetitive as beat patterns without lyrics, not how many times the same lyrical phrase was repeated). Sections B and D were non-lyrical beat patterns (just beatboxing, no words) between the refrain and the first verse and between the first and second verses, respectively. Sections C and E were the beatrhymed (beatboxing with words) first and second verses, respectively. The second verse E was 12 measures long but was truncated to the first eight measures for the analysis in order to match the eight-measure durations of the other sections. Text-based transcriptions of these sections are available in the Supplementary material.

Repetitiveness was assessed using two different metrics. The first metric counted how many unique half-measure sequences of beatboxing sounds were performed as part of a section of music. The greater the number of unique half-measures found, the less repetition there is. Rhythmic variations within a measure were ignored for this metric so as to accommodate artistic flexibility in timing. For example, Figure 3 contains two lines, each of which constitutes a single measure or equivalently two half-measures; of the four half-measures in Figure 3, this metric would count three of them as unique: {B t t K}, {t PF K B}, and {B K B}. The first half-measures of each line would be counted as the same because the sequence of sounds in the half-measure is the same despite use of triplet timing on the lower line (using beats 1.33 and 1.67 instead of beats 1.25 and 1.5). This uniqueness metric provides an integer value representing how much repetition there is over an eight-line section; if beatrhyming beat patterns resemble non-lyrical beatboxing patterns, each section should have roughly the same number of unique half-measures.

The second metric is a proportion here called the repetition ratio. For a given pair of lines, the number of beats that had

1		1.5	2		2.5	3		3.5	4		4.5
B	t	t	K				t	PF	K		B
B	t	t	K					B	K		B

FIGURE 3
Example of a two-line beat pattern.

matching beatboxing sounds was divided by the total number of beats that hosted a beatboxing sound across that pair of lines. This provides the proportion of beats in the two lines that were the same, normalized by the number of beats that could have been the same, excluding beats for which neither line had a beatboxing sound.

For example, the two lines in Figure 3 have a repetition ratio of 0.4. The first half-measures of each line have the same sounds on beats 1 and 2, but sounds which occur on beats 1.25 and 1.5 of the first line are performed with triplet timing on beats 1.33 and 1.67. Therefore in the first half-measure, six beats have a beatboxing sound in either line—beats 1, 1.25, 1.33, 1.5, 1.67, and 2—but only two of those six beats have matching sounds. In the second half-measure, four beats have a beatboxing sound in either phrase—beats 3, 3.5, 4, and 4.5. The sound on beat 3 of the first line does not have a mate in the second line, and the sounds on beat 3.5 of the two lines do not match. However, the sounds on beats 4 and 4.5 are the same from line to line. In the first half-measure, two of six beats match; in the second half-measure, two of four beats match. Taken together, this brings the total number of repeated sounds on matching beats in these two lines to 4/10 for a repetition ratio of 0.4.

This calculation penalizes cases like the first half of the example in Figure 3 in which the patterns are identical except for a slightly different rhythm. The rhythmic sensitivity of this repetition ratio measurement was useful to complement the rhythmic insensitivity of the previous technique for counting how many unique half-measures were in a beat pattern. In practice, this penalty happened to only lower the repetition ratio for phrases that were beatboxed without lyrics; that is because co-speech beat patterns rarely had patterns with the same sounds but different rhythms, so there were few opportunities to be penalized in this way. Despite this, we will see that the repetition ratios for beatrhymed patterns were still lower than the repetition ratios for beatboxed patterns in the same song (see Section 3.3.2 for more details).

Within each section, the repetition ratio was calculated for adjacent pairs of lines (lines 1 and 2, 3 and 4, 5 and 6, 7, and 8), alternating pairs of lines (lines 1 and 3, 2 and 4, 5 and 7, 6 and 8), and cross-group pairs of lines (lines 1 and 5, 2 and 6, 3 and 7, 4 and 8), as described in Section 1.3.2. Additionally, repetition ratio was calculated between Sections B and D and between sections C and E to see if musically related sections used the same beat pattern. Repetition ratios measured for the beatboxed and beatrhymed sections were then compared pairwise to assess whether the beatrhymed sections were as repetitive as the beatboxed sections.[4]

---

4 A transcription of the beatboxing sounds of "Dopamine" was used for both measurement techniques. This transcription excluded phonation and

3 Result

Section 3.1 measures the extent to which the beatrhyming replacements were constrictor-matched, and Section 3.2 does likewise for manner of articulation; both assess whether the selection of beatboxing sounds accommodates the speech task. Section 3.3 quantifies the degree of repetition during beatrhyming to determine whether the selection of lyrics accommodated the beatboxing task. A full summary of how many times each intended speech sound was replaced with each beatboxing sound can be found in the Supplementary material.

3.1 Constrictor-matching

Section 3.1.1 finds that replacements are constrictor-matched overall. Section 3.1.2 considers replacements on and off the backbeat separately: off the backbeat (everywhere except beats 2 and 4), a high degree of constrictor-matching supports the main hypothesis that there is an interaction between speech and beatboxing in beatrhyming; but on the backbeat, constrictor-matching at near chance levels matches the prediction of the null hypothesis. Section 3.1.3 offers possible explanations for the few exceptional replacements that were off the backbeat and not constrictor-matched.

3.1.1 All replacements

Table 2 shows the contingency table of replacements by constrictor. Cells along the upper-left-to-bottom-right diagonal represent constrictor matches; all other cells are constrictor mismatches. Reading across each row reveals how many times an intended speech constriction was replaced by each beatboxing constrictor. For example, intended speech labials were replaced by beatboxing labials 19 times, by beatboxing coronals 0 times, and by beatboxing dorsals 10 times. A chi-squared test over this table rejects the null hypothesis that beatboxing sounds replace intended speech sounds at random ($\chi^2 = 79.59$, df = 4, p < 0.0001).

Overall, beatboxing sounds and the intended speech sounds they replaced tended to share the same constrictor. The intended speech dorsals were almost exclusively replaced by beatboxing dorsals, with only two exceptions. The majorities of intended labials and intended coronals were also replaced by beatboxing sounds with matching labial or coronal constrictors, though there was still a notable number of mismatches for each (10/29 mismatches for labials, 10/31 mismatches for coronals). However, this degree of mismatching is less than the levels of chance predicted by a lack of interaction between beatboxing and speech. These results support the hypothesis that speech and beatboxing interact in beatrhyming.

3.1.2 Replacements on and off the backbeat

The 10 cases in which a labial speech sound was replaced by a non-labial beatboxing sound all featured an Outward K Snare on the backbeat, as did 8/10 cases in which a coronal speech sound was replaced by a non-coronal beatboxing sound. This is consistent with beatboxing musical structure which almost always features a snare sound on the backbeat. This conspiracy of so many dorsal replacements being made on the backbeat suggests that it would be more informative to consider replacements on the backbeat ($n = 29$) separately from the other replacements ($n = 59$).

Figure 4 shows the Table 2 values graphically, highlighting the difference between replacements on and off the backbeat. Blue bars represent replacements off the backbeat: dark blue bars represent the number of times a beatboxing sound replaced an intended speech sound of the same constrictor, while light blue bars represent the number of times a beatboxing sound replaced an intended speech sound of a different constrictor. Red bars represent backbeat replacements, all of which involved replacement by an Outward K Snare {K}. Focusing on all the replacements not on the backbeat, 55 of 59 replacements were made with matching constrictor. This distribution closely matches the prediction of main hypothesis, indicating that beatboxing sounds are sensitive to the constrictor of the intended speech sound they replace in beatrhyming.

On the other hand, the distribution of replacements made on the backbeat appears to support the null hypothesis. Beatboxing sounds on the backbeat in "Dopamine" are restricted to the dorsal constrictor for the Outward K Snare {K}. The replacements are fairly evenly distributed across all intended speech constrictors, congruent with there being no interaction between beatboxing constrictions and intended speech constrictors. Taking this result with the previous, this provides evidence for the speech task being achieved during replacements under most circumstances but not on the backbeat.
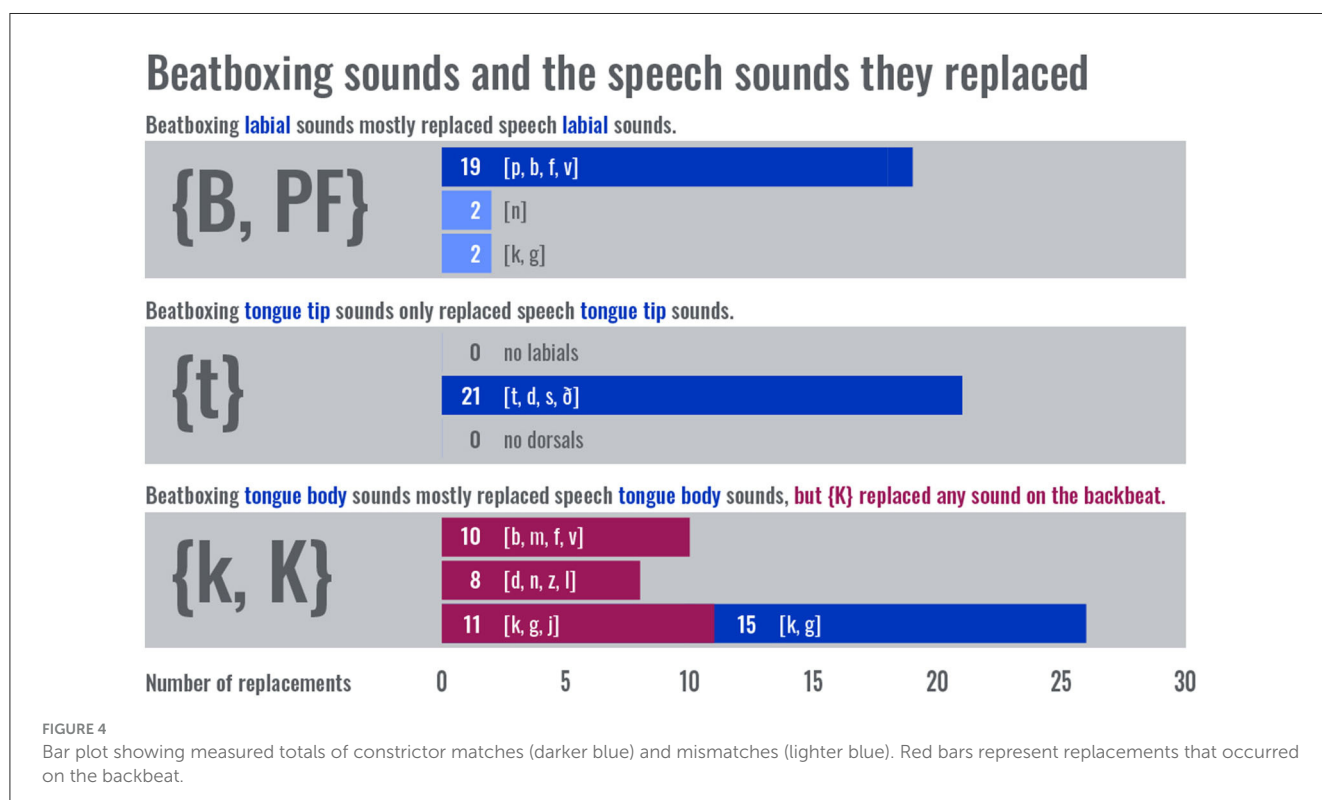
One more granular finding obfuscated by the constrictor type analysis above is that speech labials tended to be constrictor-matched to the labial Kick Drum {B} (ejective bilabial stop) while the speech labiodentals were constrictor matched to the PF Snare {PF} (ejective labiodental affricate). Thus we see constriction location matching even within the (labial) constrictor-type matching. Specifically, PF Snares only ever replaced [f]s, and 5 out of 7 replaced [f]s were replaced by PF Snares (the other two were on the backbeat, and so replaced by Outward K Snares). There were two [v]s off the backbeat, both of which were in the same metrical position and in the word "of," and both of which were replaced by Kick Drums. In the main constrictor-matching analysis, the (relatively small number of) labiodentals were grouped with the rest of the labials. However, for future beatrhyming analyses, it may be useful to separate bilabial and labio-dental articulations into separate groups rather than covering them with "labial."

---

trill sounds during the beatboxing patterns because they extend over multiple beats and would inflate the number of beats counted in the calculation of the repetition ratio. The excluded beatboxing sounds were repeated as consistently as the other sounds in the beatboxing section, and they did not occur during the beatrhymed sections.

3.1.3 Examining mismatches more closely

There are four constrictor mismatches not on the backbeat: two in which a labial beatboxing sound replaces an intended speech coronal and two in which a labial beatboxing sound replaces an intended speech dorsal.

TABLE 2   Contingency table of beatboxing sound constrictors (top) and the constrictors of the speech sounds they replace (left). Speech labial sounds were [p b f v m]; speech coronal sounds were [t d s z ð n l]; speech dorsal sounds were [k g ŋ j].

Intended speech constrictor	Replacing beatboxing sound			Total
	Labial {B, PF}	Coronal {t}	Dorsal {k, K}	
Labial	19	0	10	29
Coronal	2	21	8	31
Dorsal	2	0	26	28
Total	23	21	44	88

FIGURE 4
Bar plot showing measured totals of constrictor matches (darker blue) and mismatches (lighter blue). Red bars represent replacements that occurred on the backbeat.

Both labial-for-coronal cases are of a Kick Drum replacing the word "and," which we assume (based on the style of the performance) would be pronounced in a reduced form like [n]. Acoustically, the low frequency burst of a labial Kick Drum {B} is probably a better match to the nasal murmur of the intended [n] (and thus the manner of the nasal) than the higher frequency bursts of a Closed Hi-Hat {t}, Outward K Snare {K}, or Center K {k}. All the other nasals replaced by beatboxing sounds were on the backbeat and therefore replaced by the Outward K Snare {K}.

The two cases where a Kick Drum {B} replaced a dorsal sound can both be found in the first four lines of the second verse (see the transcriptions in the Supplementary material). In one case, a {B} replaced the [g] in "got" on beat 1 of line 3. Beat 1 in this performance was usually marked by a Kick Drum, and all replacements that occurred on beat 1 involved a Kick Drum, so this mismatch may be the result of a preference for placing Kick Drums on beat 1 akin to the preference for placing Outward K Snares

on the backbeat. However, there were few beat 1 replacements in general, due in part to the musical arrangement placing relatively few words on beat 1, so there is not enough evidence in this performance to determine whether this type of mismatch reflects a systematic pattern. The other mismatch also involved a Kick Drum {B} replacing a dorsal, this time the [k] in the word "come." The replacing {B} in this instance was part of a recurring beatboxing sequence {B B} that didn't otherwise overlap with speech. The mismatch between {B} and [k] may have resulted from a brief musical prioritization of the {B B} sequence that overruled the usual constrictor-matching pattern.

In short, tentative explanations are available for the few constrictor mismatches that occur off the backbeat: two mismatches could be because intended nasal murmur likely matches the low frequency burst of a Kick Drum better than the burst of the other beatboxing sounds available, and the other two could be due to established musical patterns specific to this performance.

TABLE 3 Counts of replacements by beatboxing sounds (top) against the manner of articulation of the speech sound they replace (left). Most replacements occurred on sounds that were intended to be stops. The beatboxing sounds are stops {B, k} and affricates {PF, t, K}.

Manner of intended speech sound	Replacing beatboxing sound					Total
	{B}	{PF}	{t}	{k}	{K}	
Stop	14	0	17	12	16	59
Fricative	2	5	4	0	8	19
Nasal	2	0	0	0	6	8
Approximant	0	0	0	0	2	2
Total	18	5	21	12	32	88

3.2 Constriction degree (manner of articulation)

Table 3 shows that the sounds that were reliably used for constrictor-matching replacements—the stop Kick Drum {B} and affricates PF Snare {PF}, Closed Hi-Hat {t}, and Center K {k}—collectively replaced 43 stops but replaced only 11 fricatives, 2 nasals, and 0 approximants (no sounds with other manners of articulation than these were replaced at all). In general, the beatboxing sounds tended to match the constriction degree of the intended speech sounds they replaced. A chi-squared test over this table rejects the null hypothesis that beatboxing sounds replace intended speech sounds at random ($\chi^2 = 35.89$, df = 12, $p < 0.001$), and supports the main hypothesis that speech and beatboxing interact in beatrhyming.

The Outward K Snare {K} likewise replaced its own share of stops (16), but on its own also replaced almost as many fricatives (8) and more nasals (6) and approximants (2) than the other four beatboxing sounds combined. As in Section 3.1.2, the broader distribution of replacements by the Outward K Snare breaks from the pattern and more closely resembles the prediction of the null hypothesis.

If many stops were in positions to be replaced by a beatboxing sound but were not replaced, this finding of stops in lyric words as a target for beatboxing replacement would carry less weight. It may be feasible in future work to determine whether non-replaced "beatboxable" sounds have a uniform distribution.

3.3 Repetition

Section 3.3.1 finds that there was little repetition of half-measure beatboxing sequences during beatrhyming, especially compared to the high degree of repetition of half-measure beatboxing sequences during non-lyrical beatboxed sections of the song. Section 3.3.2 reinforces this finding by showing that there is comparatively little repetition of beatboxing rhythm within and between lyrical sections. Taken together, these results indicate that the interaction between speech and beatboxing in beatrhyming is somewhat one-sided: the intended speech lyrics determine which beatboxing sounds will be used as described in Sections 3.1 and 3.2, but the repetitious structure of beatboxing patterns does not influence or constrain the selection of the lyrics.

Note however that the beatboxing structure is strong on the backbeat: Outward K Snares occur regularly on every beat 2 and 4 in the beatrhyming sections (with one exception on the last beat 4 of section E).

3.3.1 Analysis 1: unique measure count

The number of unique half-measure beatboxing sound sequences in an eight-line phrase indicates how much overall repetition there is in that phrase. Sections B and D of the song, the two eight-line phrases without lyrics (just beatboxing), had a combined total of just 3 unique half-measure beatboxing sound sequences: the same three sound sequences were used over and over again, as described in Section 1.3.2 and depicted in Figure 2. Section C, the first beatrhymed verse, had 16 unique half-measures (no repeated measures), and Section E, the second beatrhymed verse, had 13 unique half-measures (3 half-measures were repeated once each). The beatrhymed sections of the song therefore had far less repetition of half-measures than the beatboxed sections, indicating that the lyrics were not arranged in a way that serves a highly structured beat pattern.

At a more granular level, some sound sequences smaller than a half-measure were repeated during the beatrhyming sections. For example, the sequence {B t t K} from the beatboxing sections is also found as part of half-measures {B t t K K}, {B t t K B}, and {B t t K k} from the first beatrhyming verse (section C). On its face, this appears to suggest that beatrhyming could have repetitious structure at a smaller time scale than the unique half-measure count analysis can account for. However, it turns out that these recycled subsequences are mostly non-lyrical chunks within the beatrhyming sections: the {B t t} portions are not actually beatrhymed because they are not coproduced with any lyrics. The fact that these and other smaller sequences repeat within the beatrhyming sections of the song cannot serve as evidence for an interaction between speech and beatboxing because the repeating sound sequences are not coordinated with words.

3.3.2 Analysis 2: repetition ratio

Summary Table 4 shows the repetition ratios calculated within each section. The mean repetition ratios of beatrhyming sections C and E (0.35 and 0.3, respectively) were much lower than the mean repetition ratios for beatboxing sections B and D (0.68 and 0.70, respectively). The bottom two sections of Table 4 additionally show

TABLE 4 Summary of all repetition ratios calculated for this data set. For any pair of lines, the denominator in a fraction indicates how many beats had any beatboxing event on one or both lines, and the numerator indicates how many times a beat had the same beatboxing sound on both lines. The mean repetition ratio for a section is given as a decimal; values closer to 1 indicate a higher degree of repetition. Each ratio is given as a fraction.

Within-section comparisons of adjacent, alternating, and cross-group pairs of lines														
Section	Mean	Adjacent pairs				Alternating pairs				Cross-group pairs			
		1 & 2	3 & 4	5 & 6	7 & 8	1 & 3	2 & 4	5 & 7	6 & 8	1 & 5	2 & 6	3 & 7	4 & 7
B	0.68	8/8	4/10	6/10	4/9	8/8	4/10	6/10	4/9	6/10	8/8	8/8	7/11
C	0.35	3/10	5/11	6/12	3/11	5/12	3/11	3/12	3/11	5/10	5/12	3/13	3/10
D	0.70	8/8	4/10	6/10	4/10	8/8	4/10	6/10	4/10	6/10	8/8	8/8	9/9
E	0.30	5/10	2/11	3/11	2/7	5/12	2/10	3/10	2/9	4/12	3/9	3/10	2/9
Line-by-line comparisons of beatboxing sections B & D and beatrhyming sections C & E									
Section	Mean	1 & 1	2 & 2	3 & 3	4 & 4	4 & 5	6 & 6	7 & 7	8 & 8
B & D	0.96	8/8	8/8	8/8	9/9	8/8	8/8	8/8	7/11
C & E	0.29	5/9	5/9	4/15	2/10	2/12	3/12	2/10	1/9

repetition ratios calculated from line-by-line comparisons across sections: beatboxing sections B and D were nearly identical with a mean cross-section repetition ratio of 0.96, whereas beatrhyming sections C and E have a low mean cross-section repetition ratio of 0.29. The finding of low repetition ratios for beatrhymed sections corroborates the observation from Section 3.3.1 that there is relatively little repetition of beatboxing sound sequences during beatrhyming. The repetition/rhythmicity associated with beatboxing does not carry much weight in beatrhyming.

4 Discussion

The analysis above tested the hypothesis that beatboxing and speech interact during beatrhyming in a way that supports both speech and beatboxing tasks being achieved. The results provide evidence in support of such an interaction hypothesis. Speech tasks are achieved, in a local sense, by generally selecting replacement beatboxing sounds that match the intended speech sound in vocal tract constrictor and in manner/constriction degree. This is understood to serve the global task of communicating the linguistic message of the lyrics. The task demands of beatboxing can also affect beatboxing sound selection: the inviolable use of Outward K Snares {K} on beats 2 and 4 of each measure establishes the fundamental musical rhythm. The goals of each system—linguistic and musical—are reached, to an extent.

Achieving both speech goals and beatboxing goals in the same performance requires compromise. Optimizing the speech task comes at the cost of inconsistent beat patterns during lyrical beatrhyming. In some world, both the speech task and the beatboxing repetition regularity task could have been achieved by careful selection of lexical items whose speech-compatible beatboxing replacement sound would also satisfy repetition, but this did not happen. Instead, beatboxing sounds were selected in such a way as to optimize speech task achievement, but lexical items were not selected so as to optimize beatboxing repetition. On the other side of the compromise, the baseline musical structure established by Outward K Snares comes at the cost of its dorsal

beatboxing constriction sometimes not matching the constrictor or constriction degree of the intended speech sound it replaces. In sum, the speech task and the beatboxing task do interact such that one or the other receives priority at different moments in time during the artistic performance.

Beatrhyming is the union of a beatboxing system and a speech system. Each system is goal-oriented, defined by aesthetic tasks related to the musical genre, communicative needs, motor efficiency, and many other considerations. These tasks act as forces that shape the organization of the sounds of speech and of beatboxing in beatrhyming. Despite having disparate aims, the tasks of the speech and beatboxing systems work together flexibly (after much practice!) to create a well-organized combined behavior.

Beatrhyming is not the only behavior to exhibit this type of interplay between speech and music. Referring to the relationship between speech melody and sung melody in text settings for tone languages, Schellenberg (2012, p. 266) summarizes: "music accommodates language when it is convenient but is perfectly willing and able to override linguistic requirements." Beatrhyming fits this description well: beatboxing accommodates speech through constrictor-matched replacements and looser rhythmic constraints on beatboxing sounds, but overrides speech with highly regular snares that pay no heed to the linguistic content they replace.

4.1 Beatrhyming phonology in Optimality Theory

Ultimately, a core interest in the study of speech sounds is to understand the forces at work shaping the within- and across- word patterning of spoken language. Phonological theories accounting for why sounds in a language pattern a particular way turn to explanations of effective message transmission and motor efficiency almost axiomatically. But until we understand how these tasks manifest under a wider variety of linguistic behaviors,

we will not have a full sense of the speech system's flexibility or its limitations. To that end, we have examined how the goal of optimizing message transmission in beatrhyming via sound replacements is balanced with aesthetic beatboxing tasks. We turn now to a simple phonological model of beatrhyming, then highlight two advantages of an articulatory gestural (Browman and Goldstein, 1986, 1989) approach to beatrhyming analysis over a segmental/symbolic approach.

The analysis in Section 3 demonstrated that when speech and beatboxing are interwoven in beatrhyming, the selection of beatboxing sounds to replace a speech sound is generally constrained by the intended speech task and overrides the constraints of beatboxing task, except in one environment (the backbeat) in which the opposite is true. Given that the selection of lexical items does not appear to be sensitive to location in the beatboxing structure, the achievement of both tasks simultaneously is not possible. The resulting optimization can therefore be modeled by ranking the speech and beatboxing tasks differently in different lyrical environments. Optimality Theory (Prince and Smolensky, 1993/2004) is well suited to such an approach; ranked/weighted constraints guide the prediction of a surface output representation from an underlying input representation (we do not mean to suggest that Optimality Theory is necessarily the best paradigm in which to model beatrhyming. We use Optimality Theory because it is a convenient starting point for a descriptive model of joint beatboxing and speech sound organization: it is relatively simple, it has a decades-long history of use in phonological theory, and it is apt for the findings reported above).

We assume that beatrhyming is a highly learned skill as is beatboxing and of course speech. Further, we assume that this artist's beatrhyming can be formalized as having its own specialized Optimality Theory grammar that draws on the representations from both speech phonology and beatboxing phonology but with different constraints relevant specifically to the tasks of beatrhyming (We do not assume that this grammar necessarily generalizes to other artists' beatrhyming, though the finding by Fukuda et al. (2022) that other beatrhymers make one-for-one sound replacements hints that it may). Based on this article's interpretation that beatboxing sounds replace speech sounds in beatrhyming, the performer's internal grammar takes speech representations as inputs and returns surface forms composed of combined beatboxing and speech units as outputs. For the purposes of this simple illustration, the computations are restricted to the selection of a single beatboxing sound that replaces a single speech segment (Presumably there are higher-ranking constraints that determine which input speech segment representations should be replaced by a beatboxing sound in the output). We begin with a consideration of traditional segmental representations (i.e., phonemes and boxemes), and then note some shortcomings that can be ameliorated by using gestural representations instead.

Because the analysis below requires reference to the metrical position of a sound, input representations are graphically tagged with the associated beat number as a subscript. The input $/b_2/$, for example, symbolizes a speech representation for a voiced bilabial stop on the second beat of a measure. Output candidates are marked with the same beat number as the corresponding input; the input-output pairs $/b_2/ \sim$ { B_2 } and $/b_2/ \sim$ { K_2 } are both possible

TABLE 5 Optimality Theory tableau in which a speech labial stop is replaced by an Outward K Snare on the backbeat. An asterisk indicates that a candidate violates a constraint of the grammar, and an exclamation point indicates that the violation is severe enough to disqualify the candidate as an output.

	$/b_2/$	*BACKBEATWITHOUTSNARE	*PLACEMISMATCH
a.	{ B_2 }	*!	
b. ☞	{ K_2 }		*

TABLE 6 Optimality Theory tableau in which a speech labial stop is replaced by a Kick Drum off the backbeat.

	$/b_1/$	*BACKBEATWITHOUTSNARE	*PLACEMISMATCH
a. ☞	{ B_1 }		
b.	{ K_1 }		*!

in the system because the share the same subscript, but the input-output pair $/b_2/ \sim$ { B_3 } is never generated as an option because the input and output have different subscripts. Two initial Optimality Theory constraints merit consideration:

*BACKBEATWITHOUTSNARE—assign a violation to outputs on beats two and four ({X_2} or {X_4}) that are not snares.
*PLACEMISMATCH—assign a violation to an output whose Place does not match the Place of the corresponding input.

Within this phonological framework, "Place" is typically viewed as a set of three (or more) abstract phonological features: [labial], [coronal], and [dorsal] (where the square brackets [] indicate an abstract feature). For a theory of beatboxing phonology compatible with Optimality Theory and these speech phonological features, we assume the existence of a corresponding set of place features {labial}, {coronal}, and {dorsal}, where curly brackets {} indicate an abstract beatboxing-phonological feature.

The tableaux in Tables 5, 6 demonstrate how different possible input-output pairs might be selected by the grammar depending on the beat associated with the input sound. In an Optimality Theory tableau, an asterisk (*) indicates that a candidate violates a constraint and an exclamation point (!) indicates which violation caused a candidate to be ruled out as an optimal output. The pointing hand (☞) denotes the most optimal output candidate for that tableau. We postulate that *BACKBEATWITHOUTSNARE is ranked above *PLACEMISMATCH so as to ensure that beats 2 and 4 always have an Outward K Snare. Given an input voiced bilabial stop on beat 2 $/b_2/$ in Table 5, the output candidate {B_2} is constrictor-matched to the input and satisfies *PLACEMISMATCH but violates high-ranking *BACKBEATWITHOUTSNARE; the alternative output {K_2} violates *PLACEMISMATCH but is a more optimal candidate than {B_2} based on this constraint ranking. On the other hand, for an input $/b_1/$ which represents a voiced bilabial stop on beat 1, the constrictor-matched candidate {B_1} violates no constraints and therefore will always be selected over {K_1} which violates *PLACEMISMATCH (Table 6).

This phonological formalism is simple but effective; just these two constraints produce the desired outcome for 84 of the 88 replacements in this data set. The remaining replacements described in Section 4.1.3 may be possible to account for either by additional constraints designed to fit more specific conditions, by a related but more intricate phonological constraint model like MaxEnt (Hayes and Wilson, 2008), or by gradient symbolic representations (Smolensky et al., 2014) that permit more flexibility in the input-output relationships.

That said, we consider below two reasons *not* to use symbolic phonological representations in models of beatrhyming: the arbitrariness of speech-beatboxing feature mappings and the impossibility of splitting a segment.

4.1.1 Arbitrary mappings

The purpose of an abstract phonological place feature like [labial] is to encode linguistic information, and that information is defined by the feature's contrastive relationship to other features within the same phonological system. Different theories of phonology propose stronger or weaker associations between an abstract mental representation, such as the feature [labial], and its physical effector, such as the lips, but there nevertheless exists an inherent duality separating abstract phonological representations from the concrete phonetic effectors that implement them (Fowler, 1980).

For the beatrhyming phonology example above, we postulated the existence of both linguistic and beatboxing place features—- [labial], [coronal], [dorsal] and {labial}, {coronal}, {dorsal}. Each feature is associated with the constrictor it maps to; that is, [labial] and {labial} are both associated with physical lip constriction. But the {labial} beatboxing feature and [labial] phonological feature have no inherent correspondence or connection because they are defined by completely different information-bearing roles within their distinct systems. Mapping the abstract Place feature [labial] to the abstract beatboxing {labial} feature would all else equal be arbitrary and no more or less computationally efficient as (poor) mappings such as [labial] to {dorsal}. The logical reason to map [labial] with {labial} is of course because they ultimately in motor execution share an association to the physical lips, but the crux of the mapping—the only property shared by both units— is a phonetic referent that is not integrated into either abstract underlying system (the phonological or the beatboxing). That is a failing of this approach where there should exist a strength.

4.1.2 Splitting a segment

The second issue with symbolic features is that they are notoriously static, with no internal temporal component intrinsic to their representation. When timing is encoded in symbolic approaches, the representations are laid out either in a sequence of representational units associated with timing slots or moras, or possibly organized into hierarchical tiers, such as autosegmental tiers (Goldsmith, 1976). This encodes precedence relations, and only precedence relations, among the phonological units. A consequence of atemporality is that segments are temporally indivisible—they cannot start at one time, pause for a bit, then pick up again where they left off. This is a known challenge

for analyzing glottalized vowels in Vietnamese (Kirby, 2011) and rearticulated vowels in Yucatec Maya (Pike, 1946; Bricker et al., 1998; Bennett, 2016), and it is also a challenge for beatrhyming. Figure 5 illustrates the split-segment example of the word "sky" pronounced as [ska]{K}[a], with a K Snare interrupting and splitting in two what would otherwise be considered a single [a] (the canonical diphthongal closure to [i] is not acoustically realized). This phenomenon occurs several times in "Dopamine," and not always with K Snares; for another example, the word "move" [muv] is pronounced as [mu]{B}[uv] with a Kick Drum splitting the [u]. These cases of beatboxing sounds that interrupt speech segments are challenging to represent in a serial symbolic phonological model because they require splitting an indivisible representation into two parts to achieve the appropriate relative timing in the output representation.[5]
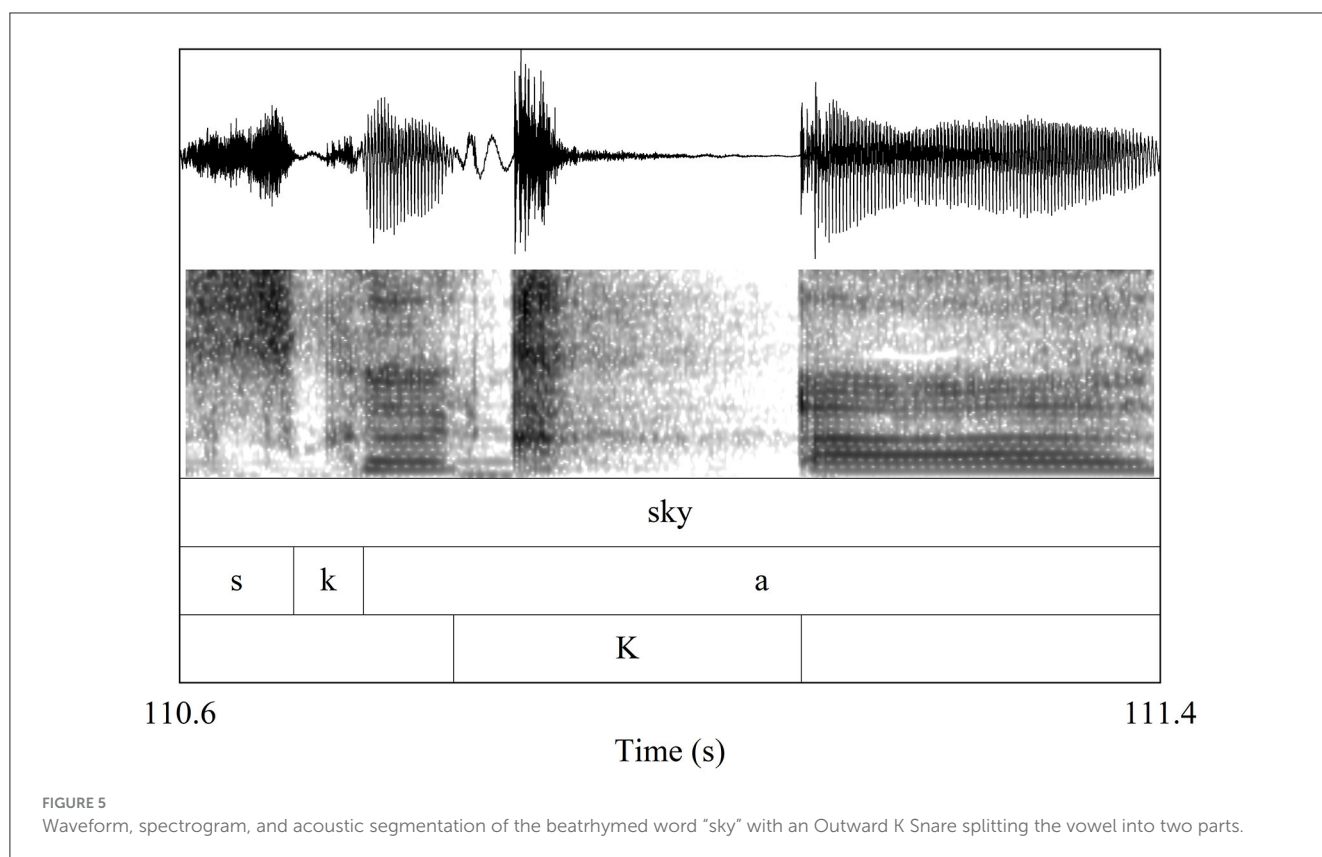
We consider both these interruptions and the speech-beatboxing constrictor mapping discussed earlier as evidence against symbolic units. In Section 4.2, we entertain an alternative theoretical account relying on gestural action units as the atomic units of representation.

4.2 Beatrhyming phonology in Articulatory Phonology

Articulatory Phonology is a theory of phonological representation leveraging the hypothesis that the fundamental units of spoken language are action units, called "gestures" (Browman and Goldstein, 1986, 1989). Unlike symbolic features which are static and which only reference the physical vocal tract abstractly (if at all), gestures as phonological units are dynamic spatiotemporal entities with deterministic and directly observable physical consequences in the vocal tract. Phonological phenomena that are stipulated through computational "interface" processes in other models emerge in Articulatory Phonology from the coordination of dynamic gestures in an utterance. Gestures are defined as dynamical systems that characterize the spatiotemporal unfolding of the vocal tract articulators toward their goals, as computationally implemented for example in the framework of Task Dynamics (Saltzman and Munhall, 1989). Rather than being simply present or absent as a symbol would be, a gesture has an activation interval during which its activation waxes and wanes. While a gesture is active, it exerts control over a vocal tract variable (e.g., lip aperture) to accomplish some linguistic task (e.g., a complete labial closure for the production of a labial stop) as specified by its intrinsic parameter settings.

Constrictor-matching between phonological and beatboxing atomic units emerges naturally within a gestural framework because gestures are defined on the specific vocal tract variable— and ultimately, the constrictor—they control. Gestures are motor plans that leverage and tune the movement potential of the vocal

---

5 Some theories of abstract symbolic segmental representations like Aperture Theory (Steriade, 1993, 1994) and Q-Theory (Shih and Inkelas, 2014, 2018) permit a certain amount of intra-segment temporal flexibility. Even in these it is not clear how to solve the problem of inserting an entire beatboxing segment into a sub-segment-sized slot.

FIGURE 5
Waveform, spectrogram, and acoustic segmentation of the beatrhymed word "sky" with an Outward K Snare splitting the vowel into two parts.

articulators for speech-specific purposes, but speech gestures are not the only action units that can control the vocal tract. The vocal tract task variables used for speech purposes are available to any other system of motor control deploying these effectors in a goal-oriented behavior, including beatboxing. This is the fundamental insight that allows for a non-arbitrary relationship between the atomic phonological units of speech and of beatboxing: a speech unit and a beatboxing unit that both control lip aperture are inherently linked because they control the same vocal tract variable.

The cases in which a beatboxing sound temporarily interrupts a vowel could be modeled in Task Dynamics with a parameter called gestural blending strength. When two gestures that use the same constrictor overlap temporally, the movement plan during that time interval becomes the average of the two gestures' spatial targets (and their time constants "stiffness") weighted by their relative blending strengths. A stronger gesture exerts more influence, and a gesture with very high relative blending strength will effectively override any co-active gestures. For beatrhyming, the interrupting beatboxing sounds could be modeled as having sufficiently high blending strength that the phonological units they co-occur with are overridden by the beatboxing gesture; when the gestures for the beatboxing sound end, control of the vocal tract returns solely to the speech phonological gesture. The gestural score in Figure 6 depicts this for the case of [ska]{K}[a] from Figure 5, with the Outward K Snare {K} temporarily overlapping with and overriding the vowel [a].
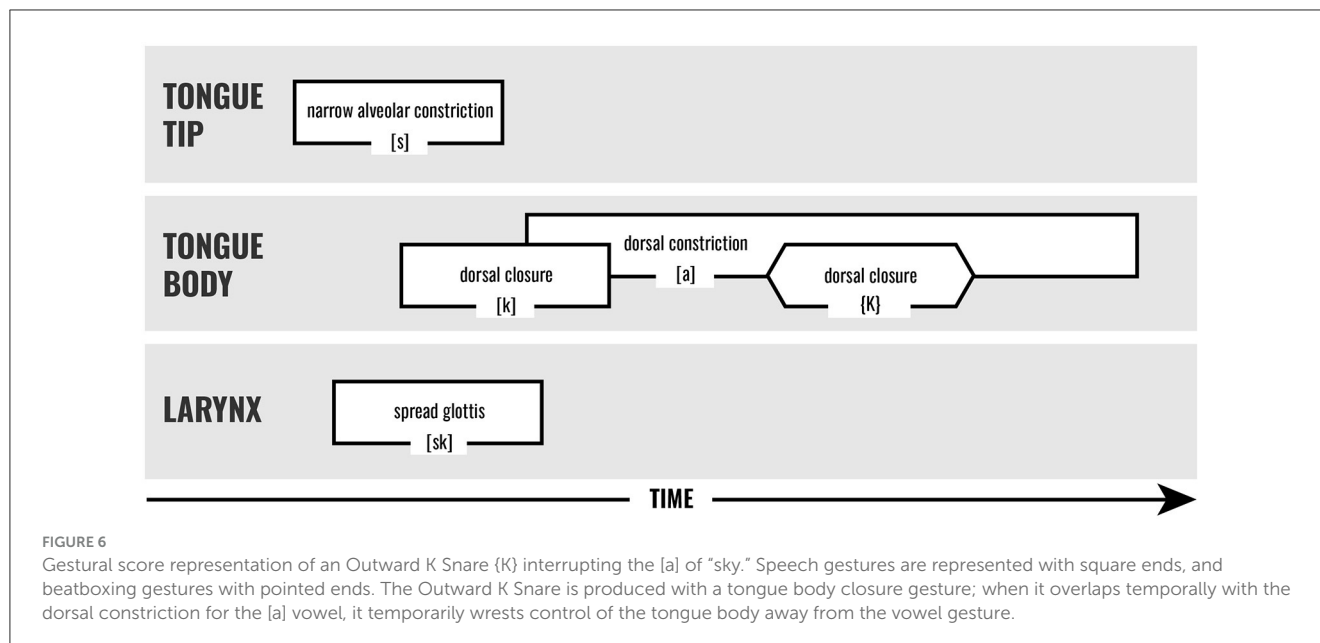
The gestures of Articulatory Phonology are compatible with Optimality Theory and can be used as the atomic units in place of phonological and beatboxing symbolic features (see Smith, 2018

for an example of Optimality Theory operating over gestures). The *PLACEMISMATCH constraint is then improved because it can map beatboxing gestures directly onto speech gestures via their constrictor tract variables. *BACKBEATWITHOUTSNARE would remain roughly the same. The use of speech and beatboxing gestures in place of abstract representations allows the beatrhyming grammar to map speech and beatboxing constrictors onto each other non-arbitrarily; it also provides the basis of explanation, in a more complex model, for how beatboxing sounds can partially interrupt or overlap with speech sounds.

5 Conclusion

The primary aim of this paper has been to introduce the scientific community to the art of beatrhyming, a musical speech behavior that offers a fresh way to examine how the fundamental units of speech can be related to similar musical units (in this case, beatboxing sounds). The methods used here to quantify one-to-one sound replacements in beatrhyming showed that the speech task guides the placement and selection of beatboxing sounds to an extent, but the beatboxing aesthetic takes over in key and predictable places (the backbeat). For an account of beatrhyming sound organization within the world of phonological theory, we have suggested that dynamical units with intrinsic connection to the body are preferable to atemporal, fully abstract units.

Vocal music is a powerful lens through which to study speech, offering insights about speech that may not be accessible from studies of talking. Beatrhyming in particular demonstrates how

FIGURE 6
Gestural score representation of an Outward K Snare {K} interrupting the [a] of "sky." Speech gestures are represented with square ends, and beatboxing gestures with pointed ends. The Outward K Snare is produced with a tongue body closure gesture; when it overlaps temporally with the dorsal constriction for the [a] vowel, it temporarily wrests control of the tongue body away from the vowel gesture.

the fundamental units of speech can interact with the units of a completely different behavior—beatboxing—in a complex yet organized way. When joined with speech, the aesthetic goals of musical performance offer a liminal zone in which to reconsider our understanding of the cognitive representations available for vocal behavior in humans.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving human participation in accordance with the local legislation and institutional requirements. 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

RB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing—original draft, Writing—review and editing. RP: Conceptualization, Data curation, Investigation, Methodology, Writing—review and editing. KM: Resources, Writing—review and editing.

Acknowledgments

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2023.1253817/full#supplementary-material

References

Bennett, R. (2016). Mayan phonology. *Lang. Linguist. Compass* 10, 469–514. doi: 10.1111/lnc3.12148

Blaylock, R. (2022). *Beatboxing Phonology* (Dissertation). Los Angeles (CA): University of Southern California.

Blaylock, R., and Narayanan, S. (2023). Beatboxing kick drum kinematics. *Interspeech* 2023, 2583–2587. doi: 10.21437/Interspeech.2023-1880

Blaylock, R., Patil, N., Greer, T., and Narayanan, S. S. (2017). Sounds of the human vocal tract. *Interspeech* 2017, 2287–2291. doi: 10.21437/Interspeech.2017-1631

Blaylock, R., and Phoolsombat, R. (2019). Beatrhyming probes the nature of the interface between phonology and beatboxing. *J. Acoust. Soc. Am.* 146, 3081. doi: 10.1121/1.5137696

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot Int.* 5, 341–345.

Boril, T., and Skarnitzl, R. (2016). "Tools rPraat and mPraat," in *Text, Speech, and Dialogue,* eds. P. Sojka, A. Horák, I. Kopeček, and K. Pala (Cham: Springer International Publishing), 367–374.

Bricker, V. R., Po?ot Yah, E., and Dzul de Po?ot, O. (1998). *A Dictionary of the Maya Language: As Spoken in Hocabá, Yucatán.* Salt Lake City, UT: University of Utah Press.

Browman, C. P., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251. doi: 10.1017/S0952675700001019

Browman, C. P., and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonol. Yearbook* 3, 219–252. doi: 10.1017/S0952675700000658

Cummins, F. (2020). The territory between speech and song. *Music Percept.* 37, 347–358. doi: 10.1525/mp.2020.37.4.347

Danner, S. G., Krivokapić, J., and Byrd, D. (2019). Co-speech movement behavior in conversational turn-taking. *J. Acoust. Soc. Am.* 146, 3082–3082. doi: 10.1121/1.5137700

de Torcy, T., Clouet, A., Pillot-Loiseau, C., Vaissière, J., Brasnu, D., and Crevier-Buchman, L. (2014). A video-fiberscopic study of laryngopharyngeal behaviour in the human beatbox. *Logoped. Phoniatr. Vocol.* 39, 38–48. doi: 10.3109/14015439.2013.784801

Dehais-Underdown, A., Buchman, L., and Demolin, D. (2019). "Acoustico-Physiological coordination in the Human Beatbox: A pilot study on the beatboxed Classic Kick Drum," in *19th International Congress of Phonetic Sciences* (Melbourne, VIC). Available online at: https://hal.archives-ouvertes.fr/hal-02284132

Dehais-Underdown, A., Vignes, P., Crevier-Buchman, L., and Demolin, D. (2021). In and out: Production mechanisms in Human Beatboxing. *Proc. Meet. Acoust.* 45. doi: 10.1121/2.0001543

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252. doi: 10.1121/1.3562174

Evain, S., Contesse, A., Pinchaud, A., Schwab, D., Lecouteux, B., and Henrich Bernardoni, N. (2019). "Beatbox sounds recognition using a speech-dedicated HMM-GMM based system," in *Models and Analisys of Vocal Emission for Biomedical Applications.*

Feld, S., and Fox, A. A. (1994). Music and language. *Annu. Rev. Anthropol.* 23, 25–53. doi: 10.1146/annurev.an.23.100194.000325

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *J. Acoust. Soc. Am.* 84, 115–123. doi: 10.1121/1.396977

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *J. Phon.* 8, 113–133. doi: 10.1016/S0095-4470(19)31446-9

Fukuda, M., Kimura, K., Blaylock, R., and Lee, S. (2022). "Scope of beatrhyming: segments or words," in *Proceedings of the AJL 6,* (Singapore: Asian Junior Linguists), 59–63.

Goldsmith, J. A. (1976). *Autosegmental Phonology* (Dissertation). Cambridge, MA: Massachusetts Institute of Technology.

Greenwald, J. (2002). Hip-hop drumming: the rhyme may define, but the groove makes you move. *Black Music Res. J.* 22, 259–271. doi: 10.2307/1519959

Guinn, D., and Nazarov, A. (2018). "Evidence for features and phonotactics in beatboxing vocal percussion," in *15th Old World Conference on Phonology* (London).

Halle, J., and Lerdahl, F. (1993). A generative textsetting model. *Curr. Musicol.* 55, 3–23.

Hayes, B. (2009). "Textsetting as constraint conflict," in *Towards a Typology of Poetic Forms: From language to metrics and beyond*, eds. J. L. Aroui, and A. Arleo. Amsterdam: John Benjamins Publishing Company, 43–62.

Hayes, B., and Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *Linguist. Rev.* 13, 3–4. doi: 10.1515/tlir.1996.13.3-4.243

Hayes, B., and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguist. Inqu.* 39, 379–440. doi: 10.1162/ling.2008.39.3.379

Himonides, E., Moors, T., Maraschin, D., and Radio, M. (2018). Is there potential for using beatboxing in supporting laryngectomees? Findings from a public engagement project. *MET.* 2018, 165–168.

Icht, M. (2018). Introducing the Beatalk technique: Using beatbox sounds and rhythms to improve speech characteristics of adults with intellectual disability: using beatbox sounds and rhythms to improve speech. *Int. J. Lang. Commun. Disord.* 54, 401–416. doi: 10.1111/1460-6984.12445

Icht, M. (2021). Improving speech characteristics of young adults with congenital dysarthria: An exploratory study comparing articulation training and the Beatalk method. *J. Commun. Disord.* 93: 106147. doi: 10.1016/j.jcomdis.2021.106147

Icht, M., and Carl, M. (2022). Points of view: Positive effects of the Beatalk technique on speech characteristics of young adults with intellectual disability. *Int. J. Dev. Disabil.* 1–5. doi: 10.1080/20473869.2022.2065449

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108, 1252–1263. doi: 10.1121/1.1288413

Katz, J. (2022). Musical grouping as prosodic implementation. *Linguist. Philos.* 46, 1–30. doi: 10.1007/s10988-022-09365-y

Kiparsky, P. (2006). "A modular metrics for folk verse," in *Formal Approaches to Poetry: Recent Developments in Metrics*, eds B. E. Dresher and N. Friedberg (Walter de Gruyter), 7–49.

Kirby, J. P. (2011). Vietnamese (Hanoi Vietnamese). *J. Int. Phon. Assoc.* 41, 381–392. doi: 10.1017/S0025100311000181

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130397. doi: 10.1098/rstb.2013.0397

Lederer, K. (2005/2006). *The Phonetics of Beatboxing*. Available online at: https://www.humanbeatbox.com/articles/the-phonetics-of-beatboxing-abstract/ (accessed April 27, 2022).

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

List, G. (1963). The boundaries of speech and song. *Ethnomusicology* 7, 1–16. doi: 10.2307/924141

Llorens, M. (2022). *Signs of Skill in the Co-Speech Ticcing of Adult Tourette's Ticcers Dissertation*. Los Angeles (CA): University of Southern California.

Martin, M., and Mullady, K. (n.d.). *Education. Lightship Beatbox*. Available online at: https://www.lightshipbeatbox.com/education (accessed June 6, 2022).

McPherson, L., and Ryan, K. M. (2018). Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94, 119–156. doi: 10.1353/lan.2018.0003

Moors, T., Silva, S., Maraschin, D., Young, D., and Quinn, V., J., et al. (2020). Using beatboxing for creative rehabilitation after laryngectomy: experiences from a public engagement project. *Front. Psychol.* 10, 2854. doi: 10.3389/fpsyg.2019.02854

Palmer, C., and Kelly, M. H. (1992). Linguistic Prosody and Musical Meter in Song. *J. Mem. Lang.* 31, 525–542. doi: 10.1016/0749-596X(92)90027-U

Paroni, A. (2022). *Human Beatboxing: Pushing the Boundaries of Human Voice Production.* (Dissertation). Grenoble: Université Grenoble Alpes. Available online at: https://theses.hal.science/tel-03992210 (accessed September 13, 2023).

Paroni, A., Bernardoni, N. H., Loevenbruck, H., Gerber, S., Baraduc, P., and Savariaux, C. (2022). "Etude comparative acoustique et articulatoire de la plosion entre parole et beatbox," in *34e Journées d'Études sur la Parole (JEP2022) - Parole, Geste, Musique: des unités à leur organisation.*

Paroni, A., Henrich Bernardoni, N., Savariaux, C., Lœvenbruck, H., Calabrese, P., Pellegrini, T., et al. (2021a). Vocal drum sounds in human beatboxing: An acoustic and articulatory exploration using electromagnetic articulography. *J. Acoust. Soc. Am.* 149, 191–206. doi: 10.1121/10.0002921

Paroni, A., Lœvenbruck, H., Baraduc, P., Savariaux, C., Calabrese, P., and Bernardoni, N. H. (2021b). "Humming Beatboxing: The Vocal Orchestra Within," in *MAVEBA 2021 - 12th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications*. Florence: Universita Degli Studi Firenze.

Parrell, B., Goldstein, L., Lee, S., and Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *J. Phonet.* 42, 1–11. doi: 10.1016/j.wocn.2013.11.002

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.

Patil, N., Greer, T., Blaylock, R., and Narayanan, S. S. (2017). Comparison of basic beatboxing articulations between expert and novice artists using real-time magnetic resonance imaging. *Interspeech* 2017, 2277–2281. doi: 10.21437/Interspeech.2017-1190

Pike, K. L. (1946). Phonemic pitch in Maya. *Int. J. Am. Linguist.* 12, 82–88. doi: 10.1086/463894

Pillot-Loiseau, C., Garrigues, L., Demolin, D., Fux, T., Amelot, A., and Crevier-Buchman, L. (2020). Le human beatbox entre musique et parole: Quelques indices acoustiques et physiologiques. *Volume!* 16, 125–143. doi: 10.4000/volume.8121

Prince, A., and Smolensky, P. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Hoboken, NJ: John Wiley and Sons.

Proctor, M., Bresch, E., Byrd, D., Nayak, K., and Narayanan, S. (2013). Paralinguistic mechanisms of production in human "beatboxing": a real-time magnetic resonance imaging study. *J. Acoust. Soc. Am.* 133, 1043–1054. doi: 10.1121/1.477 3865

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *J. Mathem. Music* 5, 35–53. doi: 10.1080/17459737.2011.573676

Rohrmeier, M., and Pearce, M. (2018). "Musical Syntax I: Theoretical Perspectives," in *Springer Handbook of Systematic Musicology,* ed. R. Bader. Cham: Springer, 473–486.

Saltzman, E. L., and Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* 1, 333–382. doi: 10.1207/s15326969eco0104_2

Schellenberg, M. (2012). Does language determine music in tone languages? *Ethnomusicology* 56, 266–278. doi: 10.5406/ethnomusicology.56.2.0266

Schellenberg, M., and Gick, B. (2020). Microtonal variation in sung cantonese. *Phonetica* 77, 83–106. doi: 10.1159/000493755

Shih, S. S., and Inkelas, S. (2014). A subsegmental correspondence approach to contour tone (dis)harmony patterns. *Proc. Annual Meet. Phonol.* 1, 1. doi: 10.3765/amp.v1i1.22

Shih, S. S., and Inkelas, S. (2018). Autosegmental aims in surface-optimizing phonology. *Linguist. Inquiry* 50, 137–196. doi: 10.1162/ling_a_00304

Shih, S. S., and Zuraw, K. (2017). Phonological conditions on variable adjective and noun word order in Tagalog. *Language* 93, e317–e352. doi: 10.1353/lan.2017.0075

Smith, C. M. (2018). *Harmony in Gestural Phonology* (Dissertation). Los Angeles (CA): University of Southern California.

Smolensky, P., Goldrick, M., and Mathis, D. (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cogn. Sci.* 38, 1102–1138. doi: 10.1111/cogs. 12047

Steedman, M. (1984). A generative grammar for jazz chord sequences. *Music Percept.* 2, 52–77. doi: 10.2307/40285282

Steedman, M. (1996). "The blues and the abstract truth: music and mental models," in *Mental Models in Cognitive Science,* eds. A. Garnham and J. Oakhill. Mahwah: Erlbaum, 305–318.

Steriade, D. (1993). "Closure, release, and nasal contours," in *Phonetics and Phonology Volume 5, Nasals, Nasalization, and the Velum,* eds. M. K. Huffman, and R. A. Krakow, (San Diego: Academic Press), 401–470.

Steriade, D. (1994). "Complex Onsets as Single Segments: The Mazateco Pattern," in *Perspectives in Phonology*, eds. Cole, J. and Kisseberth, C (Stanford: CSLI Publications), 203–291.

Stowell, D. (2003). *The Beatbox Alphabet.* Available online at: http://www.mcld.co.uk/beatboxalphabet/ (accessed April 27, 2022).

Stowell, D., and Plumbley, M. D. (2008). "Characteristics of the beatboxing vocal style," in *Technical report C4DM-TR-08–01, 1–4.* London: Centre for Digital Music, Department of Electrical Engineering, Queen Mary, University of London.

Stowell, D., and Plumbley, M. D. (2010). Delayed decision-making in real-time beatbox percussion classification. *J. New Music Res.* 39, 203–213. doi: 10.1080/09298215.2010.512979

Taylor, A., and Taylor, J. (1806). *Rhymes for the Nursery*.

Tyte, G., and Splinter, M. (2002/2014). *Standard Beatbox Notation (SBN).* Available online at: https://www.humanbeatbox.com/articles/standard-beatbox-notation-sbn/ (accessed December 8, 2019).