

MCFlow: A Digital Corpus of Rap Transcriptions

NATHANIEL CONDIT-SCHULTZ
Ohio State University

ABSTRACT: This paper describes a new digital corpus of rap transcriptions known as the Musical Corpus of Flow (MCFlow). MCFlow currently contains transcriptions of verses from 124 popular rap songs, performed by 86 different rappers, containing a total of 374 verses, and consisting of 5,803 measures of music. MCFlow transcriptions contain rhythmic information, encoded in musical durations, as well as prosodic information, syntactic information, and phonetic information, including the identification of rhymes. In the second part of the paper, preliminary analyses of the corpus are presented, describing the “norms” of several important features of rap deliveries. These features include speed, rhyme density, metric position of stressed syllables, metric position of rhymes, phrase length, and the metric position of phrases. Several historical trends are identified, including an increase in rhyme density and phrase variability between 1980 and 2000. In each analysis, variance between different performers is compared to variance between songs. It is found that there is generally more variability between songs than between performers.

Submitted 2015 September 15; accepted 2015 December 30.

KEYWORDS: *rap, corpus, rhythm, rhyme, phrasing, historical trends*

INTRODUCTION

RAP is a unique form of vocal artistry which straddles the boundary between song and poetry. As a song form, rap is unusual for its deemphasis of pitch structures in favor of rhythmic and poetic structures. As a poetic form, rap is unusual for its rhythmic articulation in relation to an independent metric hierarchy, provided by musical accompaniment. Due to its setting in a metric hierarchy, rhythm in rap can be understood (at least in part) in musical terms: metric positions, syncopation, rhythmic motives, etc.. However, other elements of rap are not typically considered musical, but are rather poetic. Still, poetry and music share many common features. Notably, parallelism and repetition are important features of both poetry and music. Phrasing, as regards basic grouping as well as formal structure, is also important in both music and poetry. The musical aspects of rap—the rhythm, phrasing, and parallelism—are referred to as flow. The artists who perform rap are known as *emcees*.

History

The ancestry of rap can be traced to various sources in African, Caribbean, and African-American traditions, including West African *griots*, Jamaican *toasting*, *signifying*, the *dozens*, jazz poetry, and patter song (Keyes 2002, 17-121). Rap itself emerged in the 1970s, with its first commercial recordings appearing at the end of that decade. Rap gradually expanded its presence in commercial music throughout the 1980s, rapidly accelerated its growth through the 1990s, and reached a commercial peak in 2004 with nearly a third of Billboard's top 100 singles (Figure 1). In this period rap achieved a commercial and social status comparable to the other major genres of western popular music (e.g. rock, pop, country). Rap's short and well-documented history offers a unique opportunity to observe the development of a musical genre from its genesis to the present. We can study how rap flow has evolved since its roots in the 1980s, through the Gangsta 1990s, and up to the present day. In particular, the stylistic shift between *old-school* and *new-school* rap, typically placed in the mid to late 1980s, deserves special attention (Adams 2009).



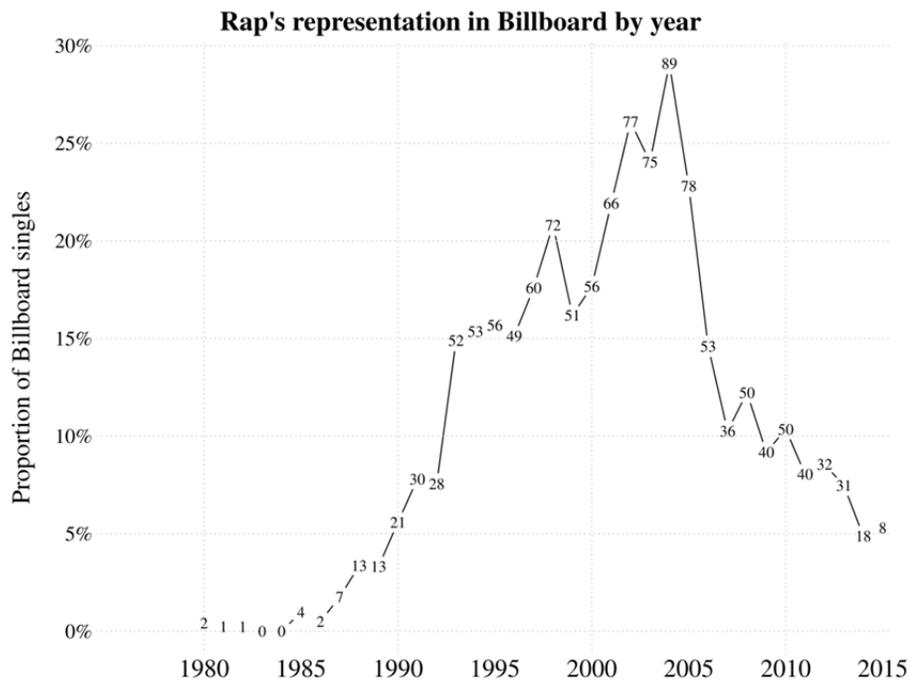


Fig. 1. Rap's representation on Billboard between 1980–2014. The y-axis represents the proportion of Billboard singles while the numbers on the chart represent the actual number of singles.

Literature

Scholarship concerning rap has predominantly explored rap's social and cultural dimensions, with work published in fields such as sociology, social psychology, education, and African-American studies. Much of this research has explored the impact and/or role of rap in urban communities. Generally, much more attention is given to the content of rap lyrics than any aspect of rap's sonic musicality. Even within music, most work on rap has been musicological or ethnomusicological in focus. This musicological work also tends to focus on rap culture, but several scholars have given some consideration to rap's sonic artistry: Musicologist Robert Walser (1995) includes a handful of illustrative transcriptions of rap. Music analyst Adam Krims' book (2001) was the first work to include a generalized theory concerning rap flow, though it is limited in scope mainly to stylistic classification. Krims identifies three primary flow styles. The most substantive contribution to musical rap analysis has been the work of Kyle Adams (2008; 2009). Adams' 2009 paper presents detailed musical theoretic close readings of three raps, focusing his analyses on the placement of rhyming syllables, the placement of accented syllables, the degree of correspondence between syntactic units and measures, and the number of syllables per beat. Adams contrasts different styles of rap flow, particularly the difference between the type of flow that is extremely regular in its placement of rhyme and its relationship to the meter, and the type of flow which misaligns or conflicts with the meter to various degrees. This contrast roughly corresponds to the difference between Krims' sung and effusive style categories. Adams also gives examples of rhythmic motives in several raps (2008).

Rap has also received some research attention in linguistics, mostly regarding rhyme. H. Samy Alim (2003) gives a close analysis of rhyme in works released by the emcee Pharoahe Monche. Linguist Jonah Katz (2008; 2010) has presented more detailed linguistic analysis of rhyme in rap, including work which applies theories from music cognition. Finally, in computer science, Hirjee and Brown (2010) conducted a corpus analysis of the texts from several thousand rap songs by twenty-five popular artists. Hirjee and Brown noted a marked increase in the complexity of rhyme patterns found in rap between 1980

and 2000. They also found that different emcees could be distinguished with fairly high accuracy based on their usage of rhyme.

The existing literature has significant weaknesses. Foremost, most of this research has focused on very small samples, case studies of one work or a handful of works. Sampling strategies have been problematic as well, as pieces have generally been picked by authors with little explanation. It is thus difficult to say how much observations made by these authors generalize to other rap works. These projects have not considered the date of rap works as well, essentially regarding approaching flow as a static phenomenon. Only Hirjee and Brown's (2010) and, to a lesser extent Katz' (2008), work are based on significant samples of music, with only Hirjee and Brown specifically looking at historical trends. However, these linguistic analyses do not include any rhythmic information, approaching rap songs essentially as written texts. What's more, all these researchers divide rap into poetry-like lines—an approach made almost inevitable by a text-based approach. As an oral art form, lines are not explicitly articulated in rap. Though it is certainly the case that phrases in rap are often very much akin to poetic lines, it is important not to superimpose the concept of poetic lines onto rap phrasing.

Current Project

This paper describes a research project intended to develop a systematic understanding of rap musicality. This paper is divided into two sections. The first part of the paper describes a new corpus of rap transcriptions created by the author. This corpus, known as the Musical Corpus of Flow, is freely available to be shared with other scholars at www.rapsience.net. By facilitating the analysis of a large, randomly sampled, body of rap, using both including poetic and musical information, MCFlow is intended to overcome the limitations of previous research. MCFlow currently contains transcriptions of rapped verses from 124 popular hip-hop songs. The second part of the paper contains preliminary exploratory analyses conducted using the corpus. These preliminary analyses focus on three goals of a cumulative character, each building off the last.

- To describe the “norms” of rap flow.
- To determine how flow varies between artists and songs.
- To describe changes in flow over time.

Not all aspects of rap are analyzed. The current project follows Adams' (2009) lead in its emphasis on what he calls “metrical techniques.” Analysis is focused on three basic categories alluded to in the introductory paragraph: rhythm, phrasing, and parallelism. The phenomena of rhyme is also analyzed as its usage in rap is closely related to the three analysis categories: Rhyme is itself a form of phonemic parallelism as well as a crucial device for structuring phrases and creating rhythmic emphasis.

THE MUSICAL CORPUS OF FLOW

The following section presents a brief overview of the contents and encoding scheme of the Musical Corpus of Flow. Complete details of the transcription process and the rationale for encoding and sampling decisions can be found in the author's dissertation (Condit-Schultz, 2016). However, this is an ongoing project, so for the most up to date information regarding the current version of corpus, please visit www.rapsience.net; the dataset itself is also available for download at this site.

Sampling

The first research goal, establishing the norms of rap, calls for a representative sample of the population of rap flow. However, as is often the case in arts research, what constitutes the “population” of a musical genre is not clear: all raps ever performed? Ever recorded? Or all “great” raps? To circumvent this thorny issue, I formulate the population of interest as the experience listeners have listening to rap: targeting rap consumption rather than rap production. I can conveniently operationalize rap consumption in terms of sales of commercially recorded rap music, using the Billboard Top 100. The Billboard chart by its nature is intended to represent the most widely listened to music, thus forming a reasonable basis for a representative

sample of rap consumption. Billboard sales do not reflect the biases or opinions of a single individual but rather the collective choices of millions of consumers. By targeting consumption I defer the judgment of quality to the population of rap listeners. The appearance of a song on the Billboard chart, as well as the position on the chart, indicates a certain level of artistic success, which we can hope correlates with “greatness.” This use of the Billboard chart of course has potential pitfalls: In particular the Billboard Top 100 only represents songs released as singles, which may differ systematically from album tracks.

Billboard data was collected via the site bullfrogspound.com. I found this data to contain some errors in terms of genre indications: many songs had no genre indication while others were clearly mislabeled (e.g. many R&B acts were labeled as Rap). Through a combination of automated and manual error correction I corrected these errors as best possible. Some errors certainly remain but I believe any remaining errors are not egregious. The corrected data set includes a total of 1,314 rap songs that peaked on the Billboard Top 100 chart between January 1980 and March 2015.

Even limited to the Billboard data, there are multiple ways to operationalize charting success. After some exploration of the data set I decided simply to rate songs by their peak position, with the number of weeks they spent in the Top 40 as a tie-breaker. By this measure, the top five songs from each year were selected, creating a “Yearly Top-5” sample target. Songs were sampled by the year that they peaked, which was not always the year the song was released or entered the chart. If you consider Figure 1, a problem with this sampling strategy is evident. Only 43 (out of 1,314) rap singles appeared on the chart before 1990. Since a historical comparison is one of the goals of this study—especially regarding old school rap from before 1990—I decided to add the 43 songs which charted before 1990 to the sample. This created a new “Top-5 + old school” sample target, containing a total of 213 targeted songs.

The “Top-5 + old school” sample contains works by 88 artists. However, most artists (48) appear only once in the sample. In order to make meaningful comparisons between the styles of different artists, I estimated that at least five songs from each artist would be required. Fortunately, fourteen artists already appeared five or more times in the sample. To better fulfill the second research goal, I decided to target more artists for extra sampling. Returning to the complete Billboard data set, I ranked artists by their number of singles on the chart, with time in the Top 40 again used as a tie-breaker. Based on this ranking the top thirty artists were selected. Thirteen of the fourteen artists who appeared five or more times in the “Top-5” appear in this “Top-30” list. As a secondary sampling goal, the top five songs from each of the additional 17 artists were targeted. In addition, two artists who first achieved commercial success before 1990 were targeted: the Beastie Boys and Run D.M.C. The resulting target of thirty-three successful artists are, from most successful to least: Jay-Z, Eminem, Lil Wayne, Kanye West, T.I., Nelly, Ludacris, Pitbull, 50 Cent, LL Cool J, Snoop Dogg, 2pac, Busta Rhymes, Bow Wow, The Black Eyed Peas, OutKast, Rick Ross, Fabolous, The Notorious B.I.G., Fugees, Wiz Khalifa, Nas, Puff Daddy, Ja Rule, Will Smith, Missy Elliott, Fat Joe, Young Jeezy, DMX, M.C. Hammer, Twista, The Beastie Boys, and Run D.M.C.. The list of all songs which have been targeted for sampling can be accessed at www.rapscience.net.

CURRENT SAMPLE

Though the total sampling target has not yet been reached, sufficient data has already been gathered to begin some basic analysis. The list of songs that have currently been transcribed are listed on www.rapscience.net. The corpus includes 124 songs by 47 artists, containing a total of 374 verses, and consisting of 5,803 measures of music. These measures contain 51,474 words (62,466 syllables). Some of the 47 artists are actually groups of two or more emcees; for instance, OutKast consists of the emcees Big Boi and Andre 3000. What’s more, guest emcees are often featured on songs (17 out of 124 songs). As a result, the sample includes flow from a total of 86 emcees (80 males, 6 females).

The top plot of Figure 2 shows the proportion of measures in the corpus rapped by each emcee. Dashed pie wedges indicate the thirty-three highly successful rappers, who make up 59% of the measures in the corpus. The lower plot in Figure 2 shows the proportion of measures in the corpus by year.

Official studio-recorded single versions of songs were accessed via YouTube in order to make transcriptions. Due to rap’s frequent use of offensive lyrics, censored—“clean”—versions of many songs exist. In most cases, “clean” versions of songs simply censor offensive words by muting or “bleeping” them—presumably, alterations created by a censor, not the artist. In other cases, artists write and record clean versions of songs using alternate words, which themselves can be quite creative. The Billboard data available to me does not tell us which version of songs (“clean” or “dirty”) are listened to more. Thus, I

was forced to make a principled decision, not a data-driven decision. My approach is to use the “dirty” versions of recordings, because they represent the uncensored artistic intent of the artist.

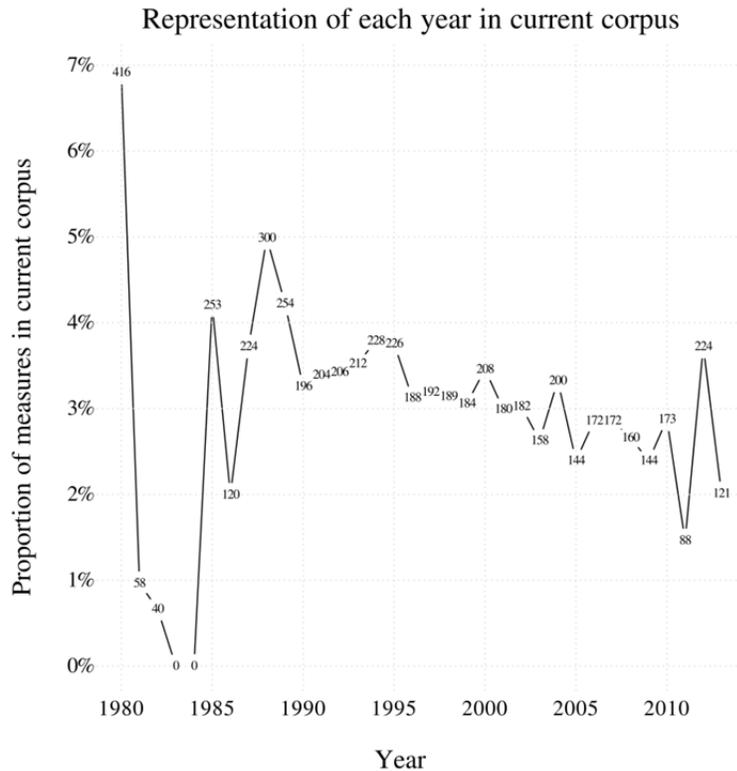
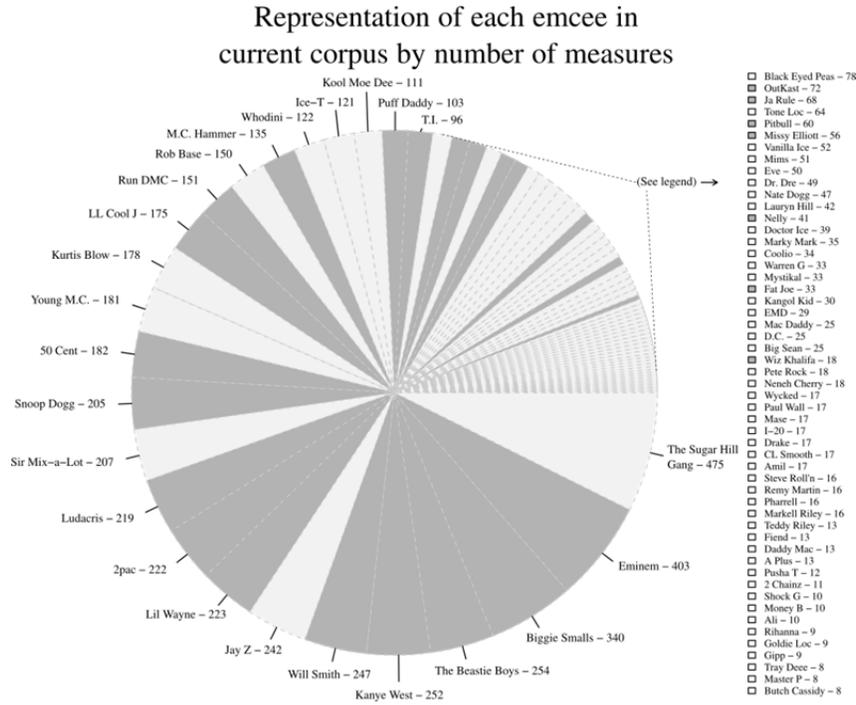


Fig. 2. Representation of emcees and years in current corpus. Greyed pie wedges, and greyed boxes in the legend, indicate artists in the "Top-30" sample target. In both figures, numbers represent measures of music.

Coding

Any transcription of music is necessarily a simplified model of the true performance. Only a limited amount of information can be encoded, and choosing what to encode is inevitably a subjective endeavor. The analytical goals of the current project shaped the choices made during transcription: focusing on rhythm, phrasing, and parallelism, including rhyme.

Transcriptions are coded in ****kern** format, consistent with the Humdrum Toolkit. Individual syllables are encoded as the rhythmic unit, each syllable occupying one line. Seven data spines are currently included in each transcription:

- ****recipx** (rhythm)
- ****stress** (Syllable stress)
- ****tone** (Intonational accents/breaks)
- ****break** (Prosodic boundaries)
- ****rhyme**
- ****ipa** (Pronunciation)
- ****lyrics**

PROSODY

Prosody is a term which refers to the sonic features of entire speech utterances, as opposed to individual syllables or phonemes. The principle dimensions of prosody are rhythm, stress, intonation, and timbre. Prosody is important in speech for several reasons, many of which carry over into rap. For one, rhythm, stress, intonation, and timbre are used to give syllables different levels of prominence. Variations in prominence influence the meaning of utterances, as particular words or clauses are emphasized or deemphasized. In rap, variations in prominence are also important rhythmically, as they help articulate different *rhythmic layers*, a term I borrow from Adams (2009). Figure 3 illustrates various rhythmic layers in an excerpt from the corpus. Prosody is also very important for segmentation; in particular, rhythmic breaks and pitch contours serve as important boundary cues in speech and in rap. These boundaries create prosodic units, which in rap play a role similar to that of musical phrases, written sentences, and/or poetic lines. Prosodic units can be nested within each other, just as melodies can contain sub phrases within larger phrases.

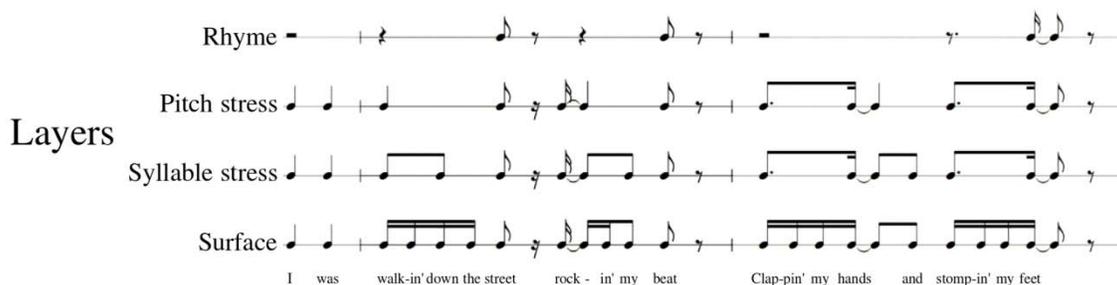


Fig. 3. An illustration of four rhythmic layers in Kool Moe Dee's "Go See the Doctor."

In MCFLOW transcriptions, four data spines encode prosodic information: The ****recipx** spine, ****stress** spine, ****tone** spine, and ****break** spine. These four prosodic spines represent only a very limited amount of the complexities of prosody in rap. Most notably, timbre is completely ignored. Two of the prosodic spines in MCFLOW (****tone** and ****break**) are modeled on the ToBI intonational transcription system (Beckman et al. 2005). Full implementation of the ToBI scheme is not possible because isolated recordings of rap vocals are not available for all songs. As a result, the prosodic information encoded in MCFLOW is less precise than would be found in a standard ToBI analysis. What's more, the musicality of rap gives it certain prosodic features which cannot be captured by ToBI.

RHYTHM

A key thesis of the current project is that rap is made musical, as opposed to poetic, by its rhythm. Thus, the current project conceptualizes rhythm from a musical standpoint, not a linguistic or poetic one. There are of course many subtleties of speech rhythm which are not well served by this musical perspective. The precise point in time at which a given syllable is experienced as occurring—known as its perceptual-center—is highly variable between syllables, contexts, and listeners. As a result, the rhythms of spoken syllables are somewhat vague and imprecise, and a certain amount of random timing variability is found in any spoken, or rapped, utterance. What’s more, like related musical styles (Funk, R&B, Jazz) rap deliveries contain systematic rhythmic nuances which are difficult to describe or notate. Particularly common is a tendency for rappers to lag “behind the beat.” All these factors introduce a good deal of rhythmic variety and nuance in rap delivery which may or may not be important to listeners, and which is in any case difficult to represent. In the MCFLOW this detailed rhythmic nuance is ignored. Rhythms are transcribed at the discretion of the author, quantized to the nearest 16th, triplet-16th, or 32nd note as appropriate. Passages which lag “behind the beat” are transcribed on the beat. The subjective quantization of the corpus in its current form means that it is not suitable for analysis of rhythmic nuance or fine rhythmic details in rap.

Rhythm is encoded as metrical durations in ****kern** format, which is based upon traditional western rhythmic notation—the Humdrum ****recip** spine designation is used to represent ****kern** rhythmic information without pitch information. Rests were indicated by a /R/ in place of a syllable. Consistent with common practice in music notation, only very explicit silences are encoded as rests. The tempo in beats per minute (BPM) was also recorded for each verse of each song.

STRESS

The term stress refers to the relative prominence of syllables created by contrasts in loudness, vowel articulation, and vowel length. Stress is used to differentiate multi-syllable words (lexical stress) and to clarify grammatical and semantic meaning (prosodic stress). From a rhythmic perspective, unstressed syllables are relatively unimportant—they often act like musical “ghost notes” in rap, part of the lowest rhythmic layer. For instance, the passage shown in Figure 4 is heard mainly as a series of dotted-eighth notes, delineated by the stressed syllables, not a stream of sixteenth notes. At a minimum English makes use of two stress levels, though speakers and listeners likely differentiate more levels. The current corpus is limited to two levels of stress: stressed or unstressed, coded as 1 and 0 respectively. Pitch accents can combine with syllable stress to create greater syllable prominences. The contribution of pitch to syllable prominence is encoded in the ****tone** spine.

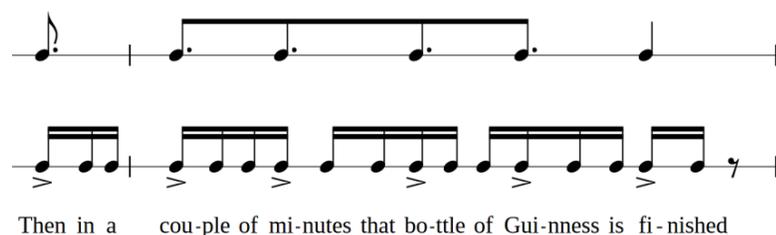


Fig. 4. Stressed syllables forming a syncopated rhythmic layer in Eminem's “Drug Ballad.”

PITCH

Pitch intonation has several functions in rap, some that parallel its usage in normal speech and others that are distinctly musical. Three pitch intonation features are included in the current corpus. Firstly, pitch intonation contributes to syllable prominence via pitch accents. Secondly, certain pitch intonation patterns contribute to the marking of phrasing boundaries. Finally, pitch intonation can also create musical parallelism, especially when in combination with rhythm and rhyme. For instance, rappers often deliver multi-syllable rhymes with the same distinct pitch contour, like a melodic motif. These three types of pitch information are encoded in the ****tone** spine.

Table 1. **tone spine key.

Symbol	Meaning
+	Local pitch peak
–	Local pitch nadir
-	Local average pitch
/	Pitch glide up
\	Pitch glide down
^	Overall increase in pitch register
v	Overall decrease in pitch register

PROSODIC BOUNDARIES

The final prosodic spine is the **break spine. The **break spine parallels the ToBI Break-index tier (see Table 2), encoding the perceived disjuncture between consecutive syllables—the primary marker of prosodic phrase boundaries in MCFLOW. Disjunctures between syllables can be created by a variety of prosodic features, including rhythm, stress, and pitch. Notably, the presence of a rest is often a cue to a break. Since, the **tone spine (or ToBI tone-tier) records pitch patterns which indicate prosodic boundaries, the **break spine is somewhat redundant with the **tone spine. However, due to the musical nature of rap, the redundancy/accord between the **break and **tone spines is not as clear cut as that between the corresponding ToBI tiers. MCFLOW, like ToBI, does not contain annotations regarding lower level prosodic boundaries, such as prosodic feet.

Table 2. **break spine key.

Symbol	Meaning	Explanation
0	Weak syllable boundary	
1	Normal syllable boundary	
2	Sub-phrase boundary	Intonation boundary without rhythmic break, or vice versa
3	Phrase boundary	ToBI intermediate phrase
4	Hyper-phrase boundary	ToBI full intonation phrase

LYRICS

Lyrics for sampled songs were accessed via the Internet and encoded in the **lyrics spine. Many lyrics had minor errors which were corrected based on the judgment of the author. Since the transcriptions are syllable based, multi-syllable words are split across records using a /-/ symbol before and/or after each syllable. In addition, grammatical and semantic boundaries are indicated in the **lyrics spine using traditional punctuation. Thus, information regarding syntactic units is encoded in the **lyrics spine, as opposed to prosodic units in the **tone and **break spines. In rap, syntactic and prosodic units do not always align, creating effects similar to *enjambment* in poetry (Adams 2009). The independent encoding of syntactic and prosodic units allows MCFLOW to be a resource for studying enjambment and similar effects. Since rap is not normal speech, nor prose, the usage of punctuation in the **lyrics spine is not exactly standard (Table 3).

Table 3. Punctuation in **lyrics spine.

Symbol	Meaning
.	End of sentence
;	Conjunction between independent sentences
:	After prefix-like dependent clause
,	Before affix-like dependent clause

PRONUNCIATION

Lyrics were translated to IPA (the International Phonetic Alphabet) and encoded in the **ipa spine. Transcriptions are encoded in Unicode format text files, allowing for the direct encoding of IPA symbols. The translator makes use of a modified version of the MRC Psycholinguistic Database (at the University of Western Australia) and CMU dictionaries. Entries in the dictionary were altered to represent American pronunciations. Other perceived errors in the dictionary were also fixed. Finally, a large number of words and terms that were encountered in rap lyrics were added to the dictionary, as each rap was translated. In many cases, rappers pronounce words in idiosyncratic manners. For example, the word “and” was pronounced many different ways by various rappers in various contexts, including /ænd/, /æm/, /In/, and /n/. I carefully examine and correct each translation, so that it represents as closely as possible the actual pronunciation in the recording.

RHYME

Rhyme is a perceptual phenomenon which is evoked by phonemic parallelism. Rhyme seems to draw attention to a syllable or word, adding to its perceived prominence, and creating a sense of connection between rhymed words. I will refer to a sequence of parallel phonemes as the rhymes' motive. For instance, the motive of the rhyme /think/pink/ is /,I,Nk/. The phonemes in a rhyme motive are not necessarily adjacent, but may be interspersed with other non-motivic (non-rhyming) phonemes. For instance, the motive of the rhyme /rake/rock/ is /r, ,k/. I use the word parallelism as opposed to repetition, since similar sounding phonemes may evoke rhyme. For example, /cat/tap/ where /t/ is substituted for /p/.

Though couplets are common in rap, rhyme motives frequently repeat more than once. I call the total set of all appearances of a particular rhyme motive a rhyme chain. Each repetition of a motive forms a link in the chain. For example, in the line

To all the ladies in the place with style and grace, allow my to lace these lyrical douches in your bushes—from “Big Poppa,” by the Notorious B.I.G..

the words /place/grace/lace/ form a rhyme chain with three links. In long rhyme chains the full motive may not appear each time. Thus, rhyme motives are abstractions which may have variable realizations. In most cases, syllables are the smallest unit capable of rhyming, and thus serve as a logical “unit” for labeling rhyme. However, rhyme motives may spread across multiple syllables, forming multi-syllabic rhymes.

In the **rhyme spine of each MCFlow transcription, each rhymed syllable is annotated with a unique letter, matching it to the syllable(s) with which it rhymes. Uppercase letters are used for stressed syllables, lowercase letters for unstressed syllables. Multi-syllable rhymes are encoded by indicated with parenthesis grouping individual rhyme tokens. Table 4 gives several examples of MCFlow rhyme encodings.

Table 4. Examples of MCFlow rhyme encodings.

English	rap tunes / Sassoon	west / Ness / chest / vest / sex	women / linen / diamonds in em' / finest women	deny me / need me / try me / believe me
IPA	ræp-tunz / sæ-sun	w3st/n3st/tʃ3st/v3st/s3 ks	wi-mIn / lI-nIn / dæl-mInz In-Im / fæl-nIs wI-mIn	nal mi / nid mi / tral mi / liv mi
Encoding	(A a) / (A a)	B / B / B / B / B	(E e) / (E e) / (F f E e) / (F f E e)	(E F) / (F F) / (E F) / (F F)

In actual artistic contexts, identifying when a rhyme occurs can be somewhat subjective. This is especially the case with singleton rhyme motives, such as a single vowel or consonant. Since there are a limited number of vowels, utterances will frequently repeat vowels without evoking the qualia of rhyme. Only if a vowel is used often enough, or regularly enough, does it start to sound like a rhyme. For instance, the phrase /he threw the apple back to the girl/ is unlikely to be perceived as containing any rhymes while the phrase /he threw the apple back at the cad/, with four repetitions of the /æ/, might evoke a rhyme, especially if the /æ/ vowels were delivered with similar inflection.

The phonemic parallelism of rhyme is nearly always accompanied by prosodic parallelism. In

addition, rhymes are usually placed in utterances/phrases at parallel points, creating a structural parallelism. In some cases, non-phonemic parallelism can stand in for rhyme in rap—a word which clearly doesn't rhyme is delivered as if it does. Rhyme is easiest to identify, and the least debatable, when reinforced by prosodic or structural parallelism. The correlation between rhyme, phrasing, and prosodic parallelism potentially creates confounds in the transcription process. The presence of an obvious rhyme may lead the transcriber to note a phrase boundary where prosody does not suggest one. Likewise, the presence of a clear prosodic boundary might lead a transcriber to annotate a rhyme where they might otherwise not.

The first link in any rhyme chain represents a special case. If we hear a word such as /maze/ we don't know that it is part of a rhyme chain until we hear the word /blaze/. Thus, it is unclear whether it is appropriate to treat /maze/ as part of the rhyme chain. However, regular structural parallelisms often create clear expectations of when rhymes will occur—for example, at the end of each phrase. As listeners, we actively anticipate the appearance of the next rhyme. When a syllable is heard which we expect to be rhymed, I refer to this as a marked syllable. When a marked syllable is “resolved” by the appearance of the next link, I refer to this as the rhyme being clinched. Rhymes placed in clear structures which evoke expectations, marking and clinching, are very common, and form one of the most artistically powerful uses of rhyme. In this type of context, it becomes possible for the first link in a rhyme chain to be heard as marked, and thus to participate in the rhyme chain. However, many rappers also use rhymes in unpredictable patterns where the first link in the chain is unmarked, and thus only retrospective, and where expectations are imprecise. In these cases, each new link in a chain is only retrospectively heard in its connection. In most analyses, my approach is to treat first links as rhymes—future research is needed to properly explain this complex topic.

ANALYSES

In the remainder of this paper, I will present some simple preliminary analyses of MCFlow. These analyses are not exhaustive, but are rather a demonstration of how rap flow can be studied using the MCFlow.

My first analytical goal is to establish descriptions of the norms of rap as regards rhythm, phrasing, and parallelism. I set out to quantify features relating to each of these categories, creating a basic “taxonomy” of flow features. To construct such a taxonomy, zeroth-order counts (meaning counting and characterizing objects in isolation, without taking into account context) of a variety of measures were taken. Pursuant to my other research goals, I also consider how variance in these measures is nested between artists, songs, and over time.

Surface features

I first can consider “surface” rhythm features, without reference to phrasing or meter. The two surface features considered here are speed/tempo, and rhyme frequency.

SPEED

To characterize the overall pace of rap, we need to consider the rate at which syllables appear in flow. Syllable durations, and rest durations, are not relevant to this measure, so all rest durations were appended to their preceding syllable durations, creating inter-onset-interval (IOI) for all syllables. Figure 5 (left panel) presents plots of the inverse of IOI, syllables per second. The grey background histogram represents all 62,466 syllables in the corpus, by syllable rate. As can be seen, this distribution is bimodal, with peaks at approximately three and six syllables per second. These peaks appear because most rap primarily alternates between two principle rhythmic values—usually eighth-notes and sixteenth-notes, but in other cases quarters vs eighths, or sixteenths vs thirty-seconds, depending on tempo. To be precise, within the rap verses in the corpus, 85% of durations in a given rap verse alternate between two primary durations related by the ratio of two to one. However, in this case these zeroth-order raw counts do not give a clear impression of the overall pace of rap delivery. The red overlay in the left panel of Figure 5 shows a histogram of the actual number of syllables to fill each second in the corpus—with dashed lines below indicating the middle 50% and 95% quantiles. This red overlay gives a more realistic impression of the typical pace of rap deliveries, approximately 4.5 syllables per second. As can be seen, it is extremely rare for rappers to rap fewer than two syllables per second while, at the other extreme, we see some rare cases

of rappers rapping as fast as ten or more syllables per second.

We can now ask, how much rap speed varies between songs and emcees, and across time? To approach this question a mixed-effects multiple regression model was created. Since the emcees and songs in the corpus are random samples from populations of songs and emcees, these factors are treated as random effects. The year on the other hand was included as a fixed effect—1996 was subtracted from the years to center the model at that year. Table 5 presents the results of this analysis— χ^2 values were calculated using likelihood-ratio tests.

Table 5. Results of mixed-effects regression model predicting rap syllable rate.

Fixed effects:					
Effect	Estimate	Std. error	χ^2	<i>p</i>	
Intercept (year 1996)	4.52	0.075			
Year	0.003	0.008	0.15	.7	
Random effects:					
Effect	Variance	Std. deviation	χ^2	<i>p</i>	
Song	0.334	0.578	7280	<.0001	
Emcee	0.169	0.412	676	<.0001	
Residual	1.465	1.210			

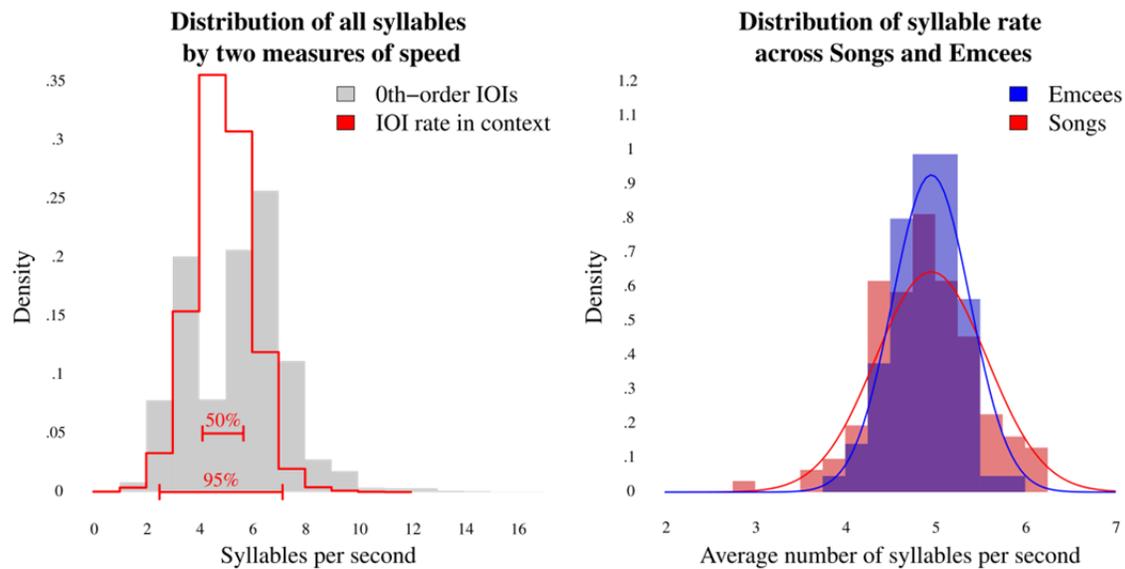


Fig 5. Left: histogram of syllable rate in corpus. Right: distribution of random intercepts in model.

There is no evidence in the data that the average speed of rap has changed over time—this is evident in the statistical results, as well as graphical depictions of the data. However, the two random effects in the model are each highly significant, consistent with the idea that speed in rap varies both between emcees and between songs. The most interesting result here is that the model suggests that the variance between songs is greater than the variance between emcees. This is illustrated in the right panel of Figure 5: a histogram of the models' random-effect intercepts, where the histogram of emcees' intercepts is noticeably narrower than the song histogram. This gives us some sense of how speed in rap varies: for instance, two emcees rapping on the same song will vary in speed less, while two songs rapped by the same emcee will vary more.

The lack of variation in rap speed across time becomes especially interesting in light of a clear change in rap's tempo across time (Figure 6). Each point in Figure 6 represents the tempo of a song in the

corpus. The variety of shapes and colors are coded by artist. As can be seen, the typical tempo of rap beats has gradually decreased over the three decades of rap history. It seems that rappers have tended to use slower and slower tempos over time, yet have continued to rap at roughly the same average speed. Though the overall trend is clear in the plot, the distribution of tempos remains skewed towards faster tempos. This skew violates the assumptions of simple linear regression, leading to poor model fit. This skew is most evident in artists like Pitbull and M.C. Hammer who use far faster tempos than the bulk of their contemporaries. Using a Bayesian approach, a multi-level linear regression model was estimated using a Markov Chain Monte Carlo (MCMC) method. Random intercepts were encoded for each artist. The Bayesian approach is more flexible, allowing for a “robust” formulation of the model, specially using a skewed t-distribution to model variation in artists' intercepts. To protect against type-1 errors, a weak but skeptical prior distribution was specified on the slope of the regression line, specifically a normal distribution centered at a slope of zero, with a standard deviation of one. A random sample of one hundred plausible regression slopes from the MCMC chain are plotted in Figure 6. As can be seen, despite the skeptical prior distribution, the most plausible regression slopes are around -0.8 BPM per year. The 99% credible interval for the regression slope is between -0.31 and -1.19.

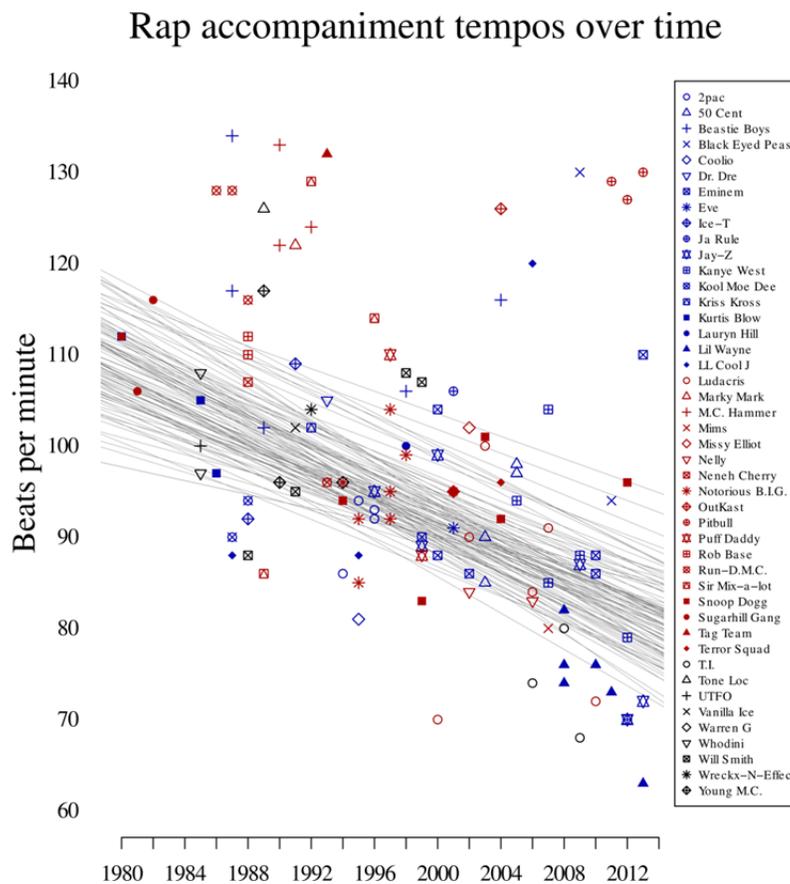


Fig 6. Changes in rap accompaniment tempos over time. Shape and color of points indicate artist. Grey lines represent random sample of plausible regression lines from Bayesian analysis.

RHYME

The second surface feature considered here is the frequency of rhyme. Rhymes are typically focused around stressed syllables, so only stressed syllables were considered in this analysis. Overall, approximately one quarter of stressed syllables in MCFLOW are rhymed. For a more rigorous analysis, a mixed-effect logistic regression model was created, predicting which stressed syllables are rhymed. As in the previous model, song and emcee were included as random effects in the model, while year was included as a fixed effect. As before, song and emcee do appear to be significant sources of variation, and once again there seems to be slightly more variation between songs than between emcees (Figure 7 left panel). Unlike rap speed however, there does appear to be a change in the density of rhyme over time (Figure 7 right panel). However, though the linear model achieves reasonably good fit, the trend in rhyme density does not appear to be linear. Rather, there seems to be an increase in rhyme density until 2002, followed by a reduction in the use of rhyme. This pattern is consistent with results found by Hirjee and Brown (2010). The dramatic increase in rhyme usage in the early 1990s may be associated with the transition between old-school and new-school rap.

Table 6. Results of mixed-effects logistic regression model predicting rhymes.

Fixed effects:

Effect		Estimate (log odds)	Std. error	χ^2	<i>p</i>
Intercept	(year 1996)	-0.923 (.28)	0.043		
Year		0.021 (1.02 odds)	0.004	22	<.0001

Random effects:

Effect	Variance (log odds)	Std. deviation	χ^2	<i>p</i>
Song	0.075	0.275	256	<.0001
Emcee	0.061	0.247	50	<.0001

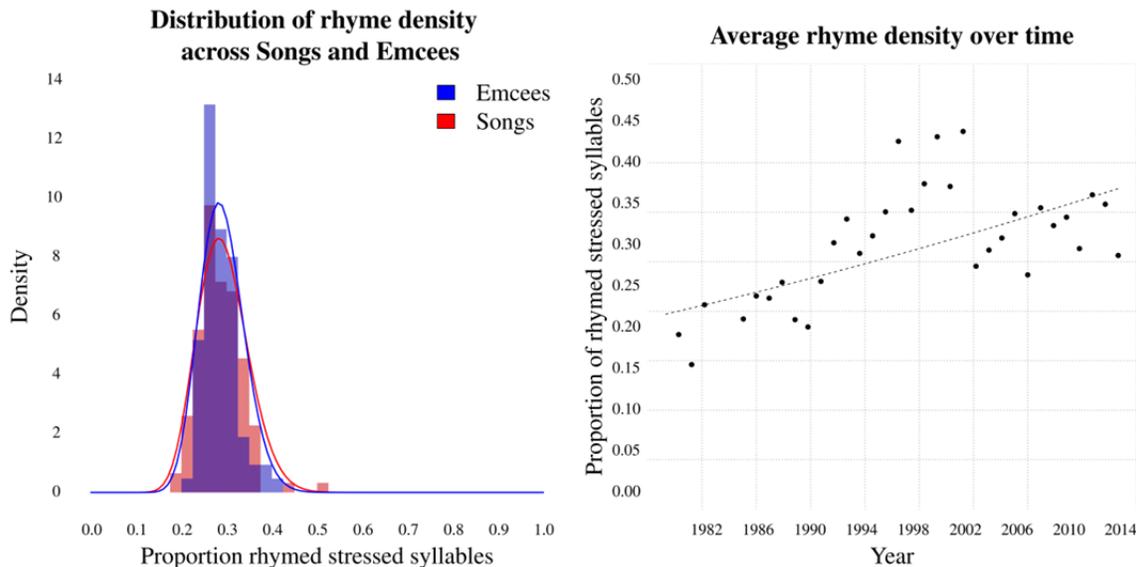


Fig. 7. Left: Distribution of random intercepts in rhyme prediction model. Right: Rhyme density in rap over time.

Another feature of rhyme worthy of consideration is the length of rhyme chains—how many times do rappers reuse a rhyme motive? Figure 8 shows the length of rhyme chains in the corpus over time. As can be seen in the left panel of Figure 8 there is a significant increase in the average length of rhyme chains

over time. In this case, a large shift towards longer rhyme chains seems to occur slightly later (about 1998), and seems more sustained, than the overall increase in rhyme density we've already observed. However, there is a strong correlation between average rhyme length, and overall rhyme density ($r = .77$). The right panel of Figure 8 shows the 50%, 95%, and 100% quantiles of rhyme chain lengths for each year. Thick bars near the bottom of the graph enclose the 50% quantile (between 0 and the median), while the next higher, smaller, bar marks the 95% quantile. The thin line running to the top represents the longest rhyme chain in that year. As can be seen, in nearly all years more than half of rhyme chains are couplets—only in 2010 does the 50% quantile reach as high as four. In all years, the 95% quantile never reaches higher than sixteen links. In short, couplets continue to dominate rap flow until the present, yet the usage of longer rhyme chains became much more popular around 1998. As before, estimates from a linear mixed-effects regression model suggest that variation between songs is a much more important factor than variance between emcees (Table 7).

Table 7. Results of mixed-effects regression model predicting rhyme chain length.

Fixed effects:

Effect	Estimate	Std. error	χ^2	<i>p</i>
Intercept (year 1996)	3.212	0.137		
Year	0.061	0.014	17	<.0001

Random effects:

Effect	Variance	Std. deviation	χ^2	<i>p</i>
Song	1.141	1.068	141	<.0001
Emcee	0.334	0.586	16	<.0001
Residual	6.144	2.479		

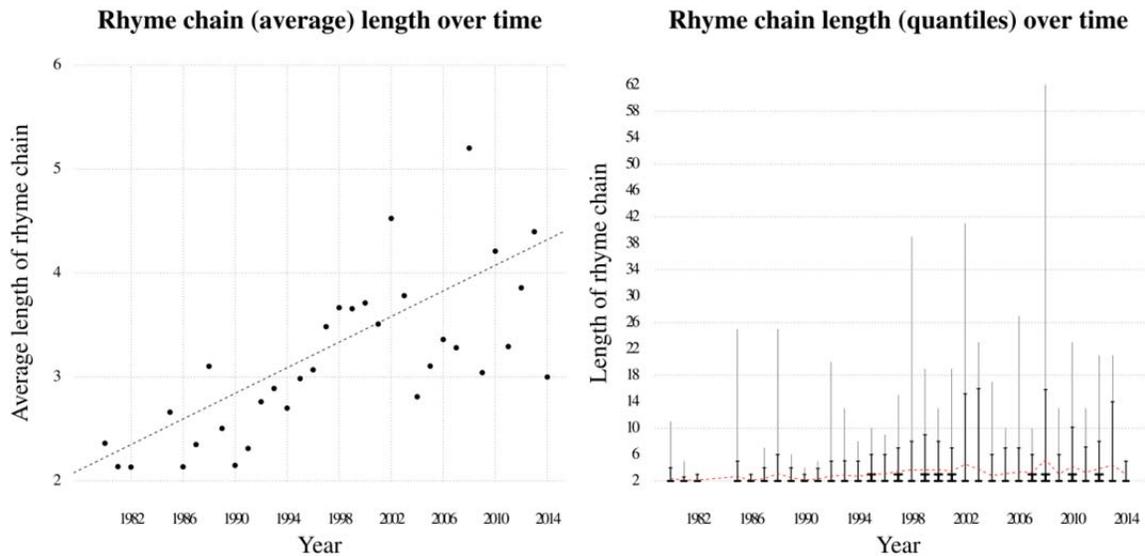


Fig. 8. Left: Average length of rhyme chains over time. Right: Distribution of rhyme chain lengths over time.

Metric Features

I next consider how rhythms are embedded in meter. The meter-based measures considered in this paper include syncopation density, the distribution of syllables across beats, and the distribution of rhymes across beats.

SYNCOPATION

I define syncopation as a rhythmic attack on a metric subdivision which is not followed by an attack on the subsequent stronger subdivision. From a musical perspective, it is more appropriate to consider syncopation of the stressed-syllable rhythmic layer, ignoring unstressed syllables. This is because it is possible for a passage to be un-syncopated on the surface (all syllables), but have syncopations between stressed syllables (Figure 4). However, results of syncopation analyses on layer were highly correlated, so the results presented here concerning stressed syllables are similar to what is found if all syllables are analyzed. Figure 9 and Table 8 report results concerning the syncopation analysis. As can be seen, there is an overall trend for less syncopation in the corpus over time, though this trend is not statistically significant. The extremely high syncopation scores from the period 1980–1981 come from only three songs by the same artist, The Sugar Hill Gang, and these outliers may be driving the apparent trend. This analysis does not yet consider the metric level of the syncopation (8th-note, 16th-note etc.), nor the tempo, so it should be considered highly preliminary. Again, variance between songs is greater than variance between emcees.

Table 8. Results of mixed-effects logistic regression model predicting syncopations.

Fixed effects:					
Effect		Estimate (log odds)	Std. error	χ^2	<i>p</i>
Intercept	(year 1996)	-1.805 (.14 probability)	0.093		
Year		-0.018 (0.98 odds)	0.009	3.8	.0513
Random effects:					
Effect		Variance (log odds)	Std. deviation	χ^2	<i>p</i>
Song		0.553	0.743	1060	<.0001
Emcee		0.198	0.445	110	<.0001

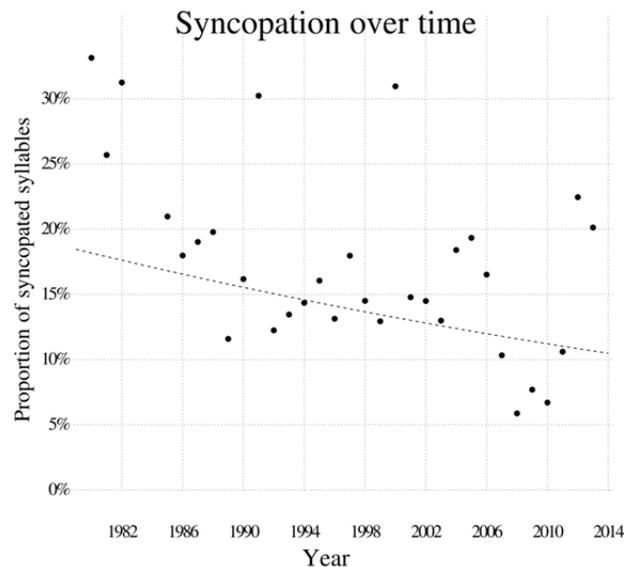


Fig. 9. Syncopation density over time. Regression slope shown here is *not* significant.

METRIC PLACEMENT OF RHYTHMIC LAYERS

I next consider the placement of syllables in the meter. The left panel of Figure 10 shows the frequency of metric positions being occupied by either stressed or unstressed syllables—metric positions between 16th-notes are rounded to the previous 16th-note. The tendency for stressed syllables to be placed on relatively strong beats is clear. Note also, the relatively few syllables landing in beat four, where phrase boundaries are likely to occur. The right panel of Figure 10 shows the placement of stressed rhymed syllables. Rhymed syllables, a yet higher rhythmic layer, are clearly even more likely to land on strong beats. However, rhymes are much more likely to occur later in the measure, especially around beat four, where phrases often end. This concentration of rhymes around beat four, is a pattern previously noted by Kyle Adams (2009).

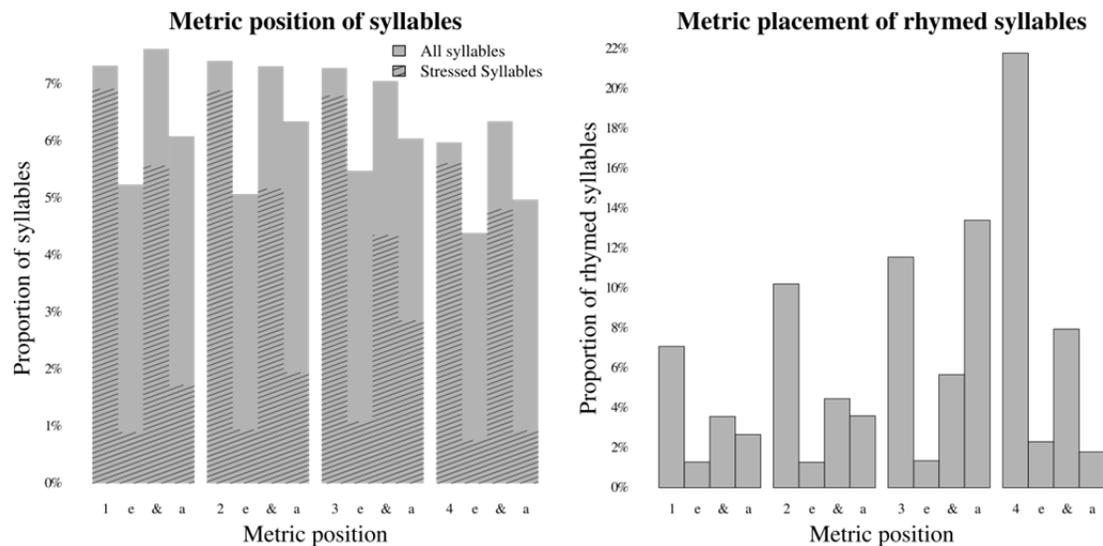


Fig. 10. Distributions of three rhythmic layers across metric positions.

How might we quantify variability in the metric position of rhythmic layers? One approach would be to calculate the number of distinct metric positions used, on average, within a small window. For instance, how many distinct metric positions have rhymes on them in an eight measure window. This approach, however, does not take into account the relative proportion of rhymes landing on each metric position. A useful mathematical construct which has fruitfully been applied to similar musical questions is entropy (Margulis, 2008). Though the literal information-theoretic meaning does not apply here, entropy serves as a single useful calculation which captures the variability of choices in a set of discrete categories. Both the number of unique metric positions used and their distribution contribute to the entropy value. Figure 11 shows the entropy of metric positions of rhymed syllables for each emcee. Comparing the metric distribution plots for Jay-Z and the Beastie Boys gives a good impression of the difference between high and low entropy values, regarding rhyme positions (Figure 12). As can be seen, the Beastie Boys place about half of their rhymes on beat four, or the sixteenth before beat four, with most of the rest of their rhymes on each of the strong beats. In contrast, Jay-Z places far more of his rhymes on various weak subdivisions.

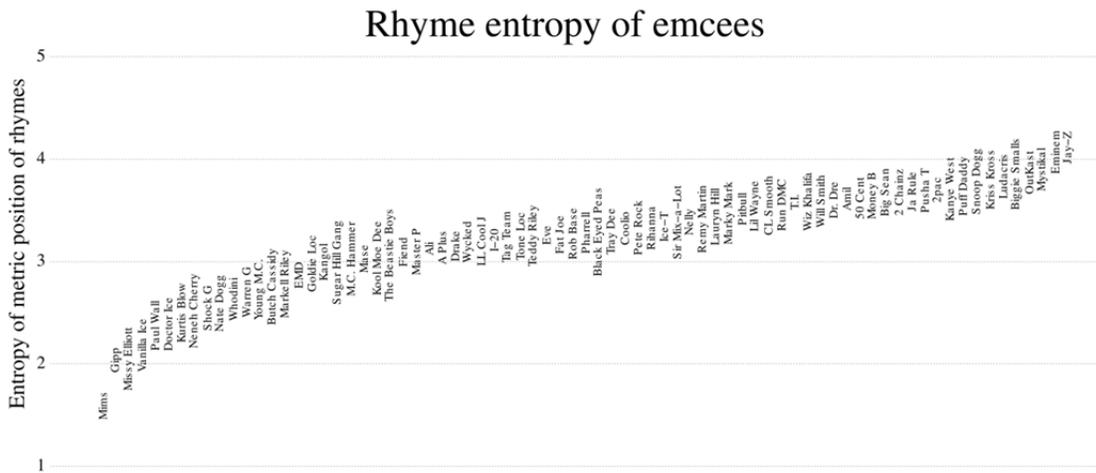


Fig. 11 Distribution of emcees by entropy of metric position of rhymes.

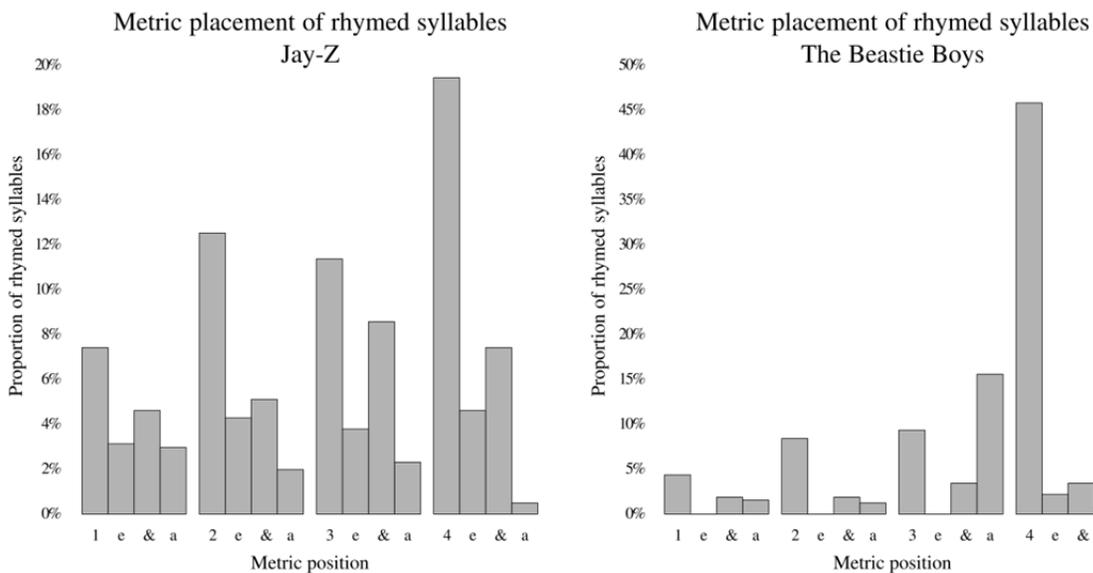


Fig.12. Comparison of distribution of rhymes in meter between Jay-Z (high entropy) and the Beastie Boys (low entropy).

PHRASES

We next consider how rhythms are embedded within phrases, and how phrases relate to meter. The preliminary analyses related to phrasing presented here are phrase length in syllables, phrase length in beats, and the metric position of phrases. In the current paper, phrases are simply defined by the appearance of either a /3/ or a /4/ in the prosodic **break spine. This approach is very simplistic, doing little justice to the complexity of phrasing in rap flow. MCFlow contains a wealth of other prosodic, and syntactic, information which could potentially enable much more complex, nuanced, analysis of rap phrasing.

Figure 13 plots the distribution of phrases by lengths, both in syllables and in beats. As can be seen, the most common phrases in rap are either around five syllables spread over two beats, or around ten syllables spread over four beats. Of more interest than simple phrase length is the placement of phrases within the meter. Figure 14 illustrates the metric position of phrases in the corpus. A single arrow is drawn for each of the 7,927 phrases in the corpus. The left terminus of each arrow represents the metric position of the first syllable in each phrase while the arrowhead at the right terminus indicates the metric position of the last syllable in each phrase. The plotting location of all arrows is randomly jittered a small amount so that a general impression of the density of arrows at each metric position is given. As can be seen, there are two areas where phrases are the most concentrated, representing phrases that begin around beat one and end around beat four, and shorter phrases that begin around beat three and end around beat four. However, there is a huge diversity of other possibilities. The metric position of phrases is an aspect of rap delivery which is completely absent from non-musical poetry, and impossible to measure from text-based analyses, and thus illustrates a novel dimension of analysis that is afforded by MCFLOW.

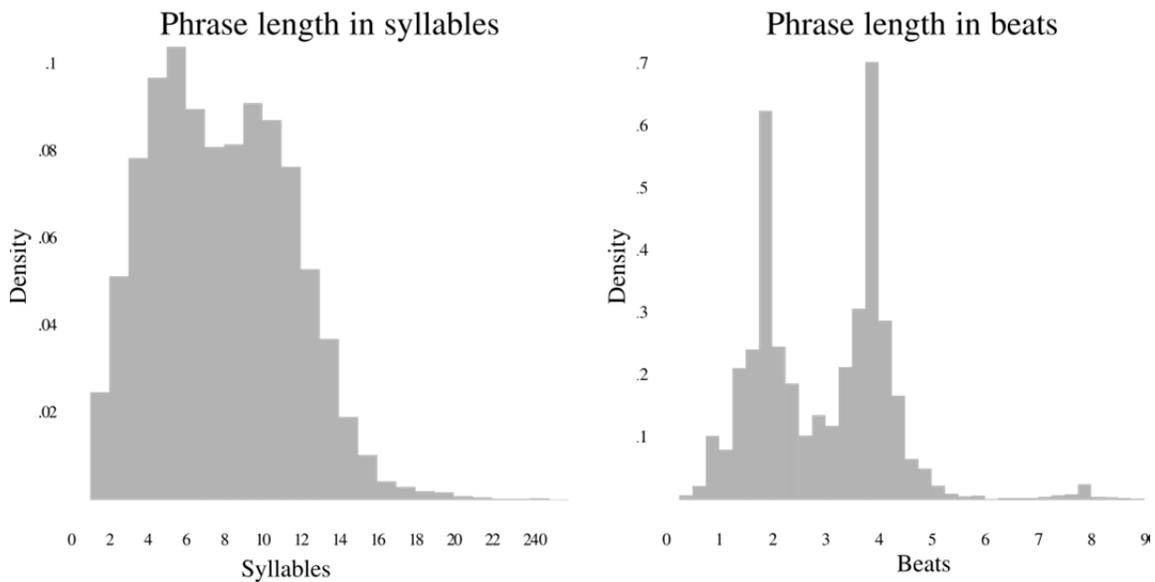


Fig. 13. Distribution of phrases in corpus by length in syllables (left) and in beats (right).

These flow phrasing entropy measures should be taken with a grain of salt. These metric analyses overlook tempo, which certainly interacts with phrasing. At a slow tempo, emcees are more likely to rap two beat phrases, starting around beats one or three, and ending around beats two or four. Conversely, at very fast tempos, emcees are more likely to rap two measure phrases. As a result, some of the observed differences in entropy may be due to variability in the tempos used by a particular emcee.

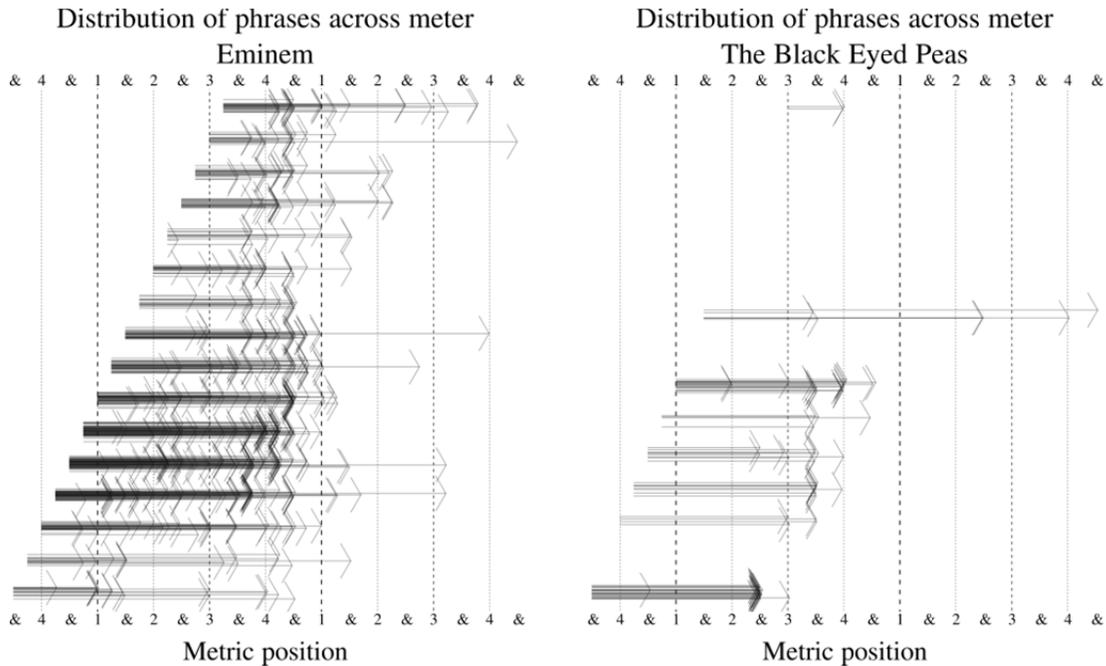


Fig. 16. Comparison of distribution of phrases in meter between Eminem (high entropy) and the Black Eyed Peas (low entropy).

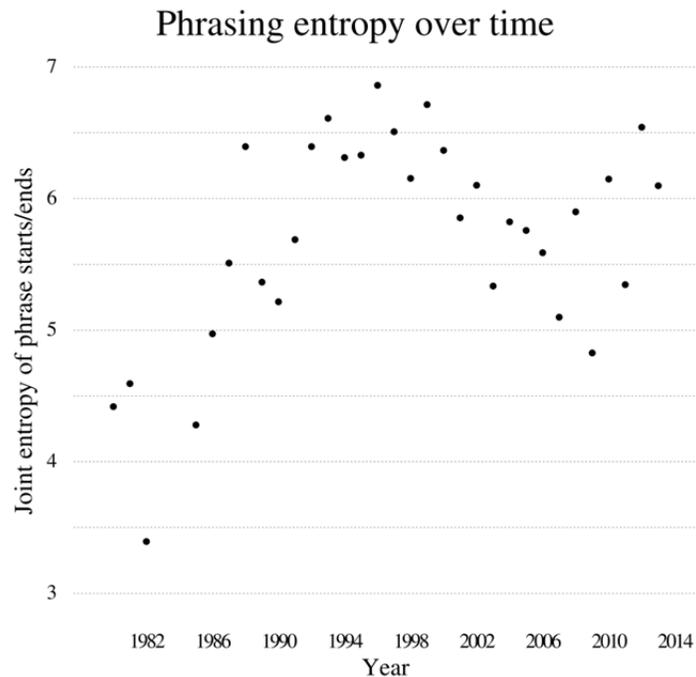


Fig. 17. Joint entropy of metric position of phrase starts and ends by year.

FORMAL STANDARDIZATION

Though most trends in rap seem to be towards greater complexity over the first two decades of rap's history, an interesting counter trend is also evident. Figure 18 plots three features of rap song form over time: number of verses, length of verses in measures, and overall song length. The area of squares over each year represent the proportion of raps/verses with that number of verses/measures. In the top two figures, there is an evident “standardization” of rap form. Since 1995, nearly all rap songs on the Billboard Top 100 have had either two or three verses, with three being much more common. Similarly, around 1995 sixteen-measure verses became dominant. Compare this to the relative variability in number and length of verses before 1995. Thus, in the same period where rap was reaching new heights of commercial success, a standard form of rap developed: three verses of sixteen measures.

The bottom-right panel in Figure 18, plots the total length of each song in the corpus over time. In this case, there seems to be less evidence of significant change over time. Clearly, the majority of successful rap songs have been between three and six minutes in length, a range typical in most pop music genres. This standard length seems to be fully established by the mid-1980s, though clearly the very earliest raps tended to be far longer.

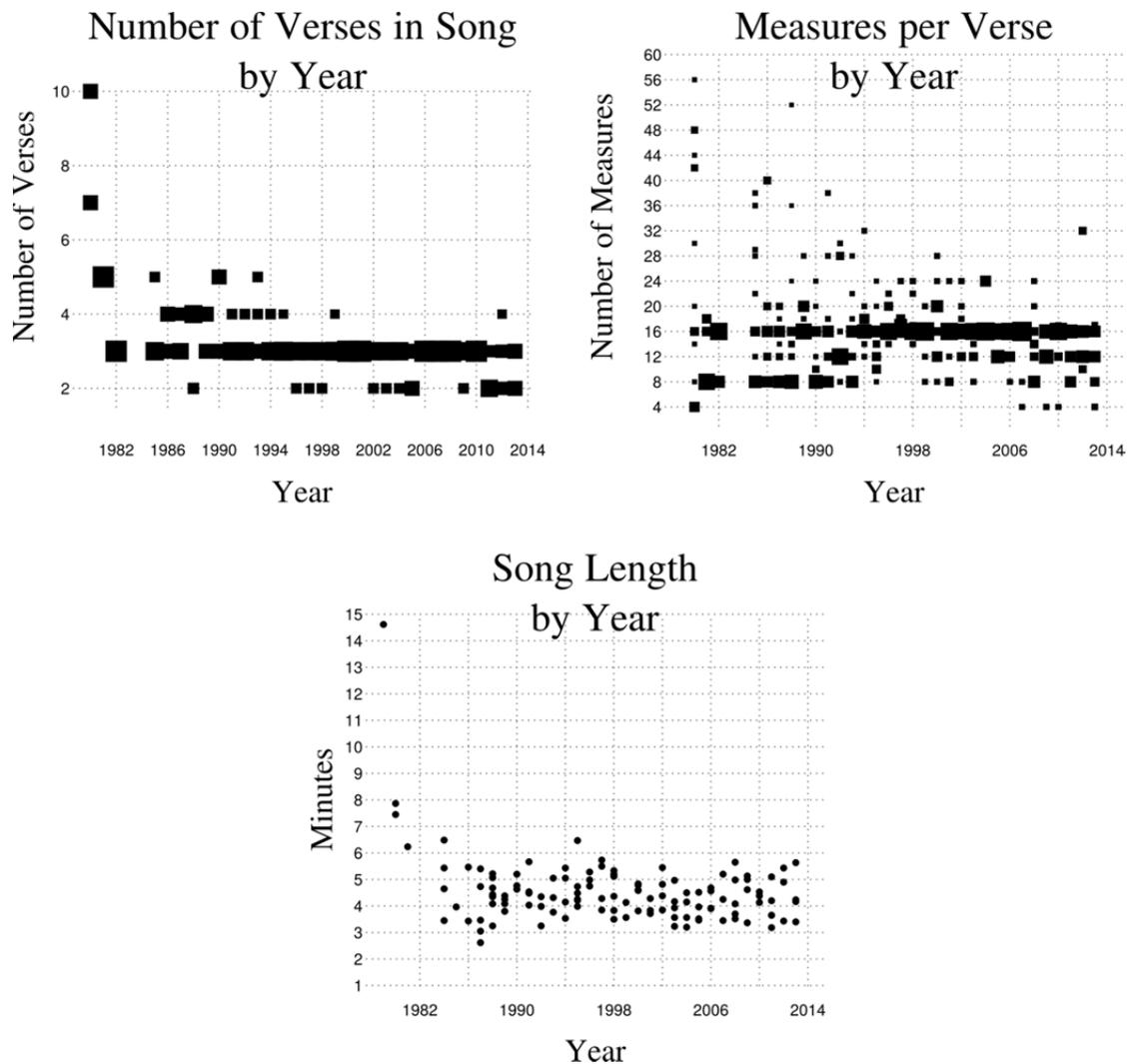


Fig. 18. Formal standardization of rap over time. Top left: number of verses in each song. Top right: number of measures in verses. Bottom: total length of song in seconds.

HIGHER ORDER PROBABILITIES

All the analyses presented here have made use of “zeroth order” counts. Though this approach does give an adequate account of the “norms” of flow, it tells us little about the experience of listening to rap. As in all music, higher order relationships are needed to truly understand the ebb and flow of the experience of the listener. In this paper, I will present only one example, concerning the first-order probabilities of the metric position of rhymes. For instance, if we hear a rhyme land on beat four, where do we expect the next rhyme to occur? Figure 19 illustrates how the probabilities of an antecedent rhyme are altered by different consequent rhymes. The y-axis represents the metric position of the antecedent syllable, which is rhymed by a syllable landing at a metric position represented on the x-axis. Filled circles represent an increase in probability compared to the zeroth-order distribution (right panel of Figure 12), while empty circles represent a decrease in probability. The area of circles represent the magnitude of the increase/decrease. As can be seen, the largest filled circles land on two diagonals. The center diagonal, represents rhyming syllables separated by exactly one measure, while the other two diagonals represent syllables separated by exactly two beats.

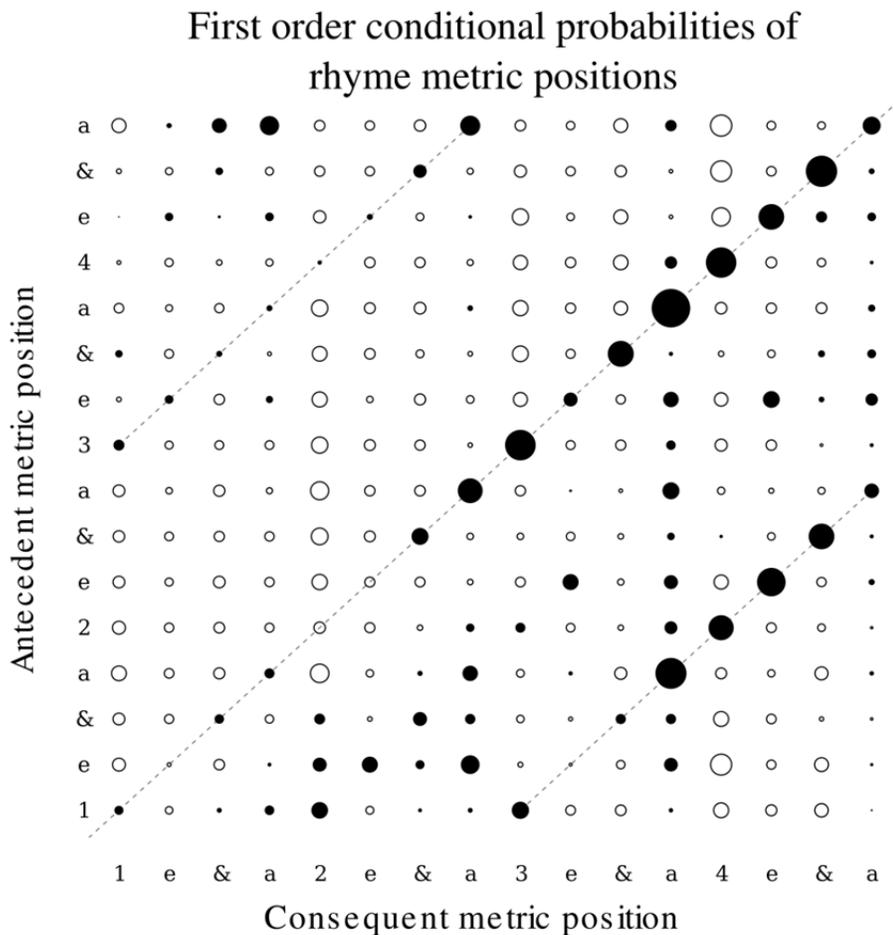


Fig. 19. Metric positions of rhyme, given metric position of previous rhyme. Filled circles indicate increased probability relative to zeroth-order probabilities, open circles indicate decreases.

CONCLUSION

The first part of this paper presented the details of the Musical Corpus of Flow. MCFlow currently consists of 124 transcriptions of the verses of popular rap songs. However, MCFlow is a work in progress and the ultimate sampling target contains over two hundred songs. Each MCFlow contains rhythmic information, prosodic information, syntactic information, phonetic information, lyrics, and rhyme information. Prosodic information includes annotations of prosodic stresses, boundaries, and parallelisms. A more complete description of MCFlow annotations and methodology can be found in my dissertation (Condit-Schultz 2016) as well as online at www.rapsience.net, where the complete dataset is freely available at any time and updates and improvements to the corpus will be posted periodically. The website also includes a Graphical User Interface which produces visualizations of the dataset similar to the figures included in this paper, as well as “flow diagrams” of individual verses.

The second part of this paper presents some simple preliminary analyses of flow, conducted using MCFlow. A number of historical trends were identified, including an increase in rhyme density and rhyme chain length, a decrease in the tempo of rap accompaniments, an increase in the variability of the metric positions of phrases and rhymes, and a standardization of rap songs' formal structure. In most analyses, it was clear that flow features vary significantly between emcees—that each emcee has a personal style of flow. In future work, I hope to expand and refine the taxonomy of flow features presented here, so that I can ultimately create complete profiles of the flow of different emcees. The current analyses suggest that emcees vary their style greatly between songs, and that this is a greater source of variance in flow style than between artists. For instance, an emcee who is known for rapping very rapidly, might nonetheless release a slow ballad where he/she raps slowly. As another example, an emcee who in most songs uses only rhymed couplets, might nonetheless release a song where he/she uses much longer rhyme chains.

Rap is a sophisticated musical/poetic art form, overflowing with creative and beautiful displays of musical ingenuity. The “norms” of flow analyzed here are not only interesting in themselves, but serve as a point of comparison for more detailed analyses of individual musical passages. By affording the detailed analysis of a large, representative sample of rap flow, MCFlow can be a tool for developing a holistic understanding of rap musicality.

REFERENCES

- Adams, K. (2008). Aspects of the Music/Text Relationship in Rap. *Music Theory Online*, 14(2).
- Adams, K. (2009). On the Metrical Techniques of Flow in Rap Music. *Music Theory Online*, 15(5).
- Alim, H. S. (2003). On Some Serious Next Millennium Rap Ishhh: Pharoahe Monch, Hip Hop Poetics, and the Internal Rhymes of Internal Affairs. *Journal of English Linguistics*, 31(1), 60–84.
<https://doi.org/10.1177/0075424202250619>
- Beckman, M. E. & Hirschberg, J. & Shattuck-Hufnagel, S. (2005). The Original ToBI System and the Evolution of the ToBI Framework. In *Prosodic Typology—The Phonology of Intonation and Phrasing* (Ed. Sun-Ah, J.). Oxford, UK: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199249633.003.0002>
- Condit-Schultz, N. (2016). *MCFlow: A Digital Corpus of Rap Flow* (Doctoral dissertation). Retrieved from OhioLINK. (osu1461250949)
- Hirjee, H., & Brown, D. (2010). Using automated rhyme detection to characterize rhyming style in rap music. *Empirical Musicology Review* 5(4), 121–145.
- Katz, J. (2008). Towards a Generative Theory of Hip-Hop. *Music, Language, and the Mind*. Tufts, Medford MA.
- Katz, J. (2010). Phonetic similarity in an English hip-hop corpus. *Linguistic Society of America Annual*

Meeting. Baltimore.

Keyes, C. (2002). *Rap music and street consciousness*. Urbana, IL: University of Illinois Press.

Krims, A. (2001). *Rap Music and the Poetics of Identity*. Cambridge, UK: Cambridge University Press.

Walser, R. (1995). Rhythm, Rhyme, and Rhetoric in the Music of Public Enemy. *Ethnomusicology*, 39(2), 193–217. <https://doi.org/10.2307/924425>