AUTOMATIC DETECTION OF INTERNAL AND IMPERFECT RHYMES IN RAP LYRICS

Hussein Hirjee **Daniel G. Brown**
University of Waterloo
Cheriton School of Computer Science
{hahirjee, browndg}@uwaterloo.ca

ABSTRACT

Imperfect and internal rhymes are two important features in rap music often ignored in the music information retrieval community. We develop a method of scoring potential rhymes using a probabilistic model based on phoneme frequencies in rap lyrics. We use this scoring scheme to automatically identify internal and line-final rhymes in song lyrics and demonstrate the performance of this method compared to rules-based models. Higher level rhyme features are produced and used to compare rhyming styles in song lyrics from different genres, and for different rap artists.

1. INTRODUCTION

Song lyrics have received relatively little attention in music information retrieval, but can provide data about song style or content that is missing from raw audio files or user-input tags. Recent work focusing on lyrics [1–3] involves using lyric text to extract song topic, theme, or mood information; the pattern and sound of the words themselves is usually ignored.

These sound features are central to rap music, providing information about vocal delivery and rhyme scheme. This data can be characteristic of different rappers, as MCs often boast of the uniqueness and superiority of their rhyming style. Lyric rhymes have previously been studied as an aid in characterizing different musical genres [4], but this prior work ignores two stylistic features of rap lyrics: imperfect rhymes, where syllable end sounds are similar but not identical, and internal rhyme, which occurs in the middle of lines.

To study these features, we have developed a system for automatic detection of rap music rhymes. We train a probabilistic scoring model of rhymes using a corpus of rap lyrics known to be rhyming, using ideas derived from bioinformatics. We then use this model to find and categorize various rhymes in different song lyrics, and assess the model's success. Finally, we calculate high-level statistical rhyme scheme features to attempt to quantitatively model and compare rhyming styles between artists and genres. Our work allows the automated study of new features in rap music, and may be extensible to other genres of song lyrics or for poetry analysis.

2. BACKGROUND

Hip hop music is characterized by lyrics with intermittent rhymes being rhythmically chanted (rapped) to an accompanying beat. In "Old School" rap (dating from the late 1970s to mid 1980s), lyrics typically followed a simple pattern and contained a single rhyme falling on the fourth beat of each bar [5]. Contemporary rap features more varied delivery and many complex rhyme stylistic elements that are often overlooked. Key among these are rhymes that are imperfect, extended, or internal. Holtman [6] provides a good overview of the abundance of imperfect rhyme in rap lyrics. A normal rhyme involves two syllables that share the same nucleus (vowel) and coda (ending consonants). Two syllables form an imperfect rhyme if one of these two parts does not correspond exactly. However, these types of rhymes are not just composed of vowels and consonants being paired randomly; there is a constraint to the amount of dissimilarity in these rhymes, determined by the shared articulatory features of matching phonemes.

In Holtman's hierarchy, the most similar consonants are nasals, fricatives, and plosives differing only in place of articulation, as in the line-ending /m/ and /n/ phonemes in:

> Entertain and tear you out of your **frame**
> Leave you in a puddle of blood, then let it **rain**. [7]

(Rhyming syllables in quoted lyrics are displayed with the same font style.) Less similar consonant pairs include those with the same place of articulation, but differing in voice or continuancy, such as the /k/ and /g/ pair in:

> Bring a bullet-proof vest, nothin' to **ricochet**
> Ready to aim at the brain, now what the **trigger say**? [7]

Vowels are most similar when differing only in height or "length" (advanced tongue root), such as the penultimate vowels in:

> I'm the alpha, with no **omega**
> Beginning without the, end **so play the**. [7]

Holtman's work is largely taxonomic and describes known rhymes, rather than discovering them. Hence, we used a statistical model of phonetic similarity based on frequencies in actual rap lyrics. However, the patterns we automatically discovered largely validate her taxonomy.

Rap music often features triplet or longer rhymes with unstressed syllables following the initial stressed pair, which may span multiple words (mosaic rhymes). Longer rhymes can also include more than one pair of stressed syllables:

> Maybe my sense of **húmor gets ínto you**
> But girl, they can make a per**fúme from the scént of you**.
> [8]

(Here the accents mark the syllables with primary stress.) Finally, contemporary rap music features dazzlingly complex internal rhyme. Alim [9] analyzes Pharoahe Monch's 1999 album Internal Affairs [10] as a case study, and identifies chain rhymes, compound rhymes, and bridge rhymes. Chain rhymes are consecutive words or phrases in which each rhymes with the previous, as in:

> New York **City gritty** com**mittee pity** the fool that
> Act **shitty** in the midst of the calm the **witty**, [10]

where "city", "gritty", "committee", and "pity" participate in a chain. Compound rhymes are formed when two pairs of line internal rhymes overlap within a single line. A good example of this is given in "Official":

> Yo, I stick around like hockey, now what the **puck**
> Cooler than **fuck**, ma*neuver* like Van*couver* Ca**nucks**,
> [10]

where "maneuver" and "Vancouver" are found between "fuck" and "Canucks." Bridge rhymes are internal rhymes spanning two lines:

> **How I made it** you **salivated** over my **calibrated**
> RAPS that **validated** my ghetto credi*bility*
> *Still I be* PACKin a*gilities* unseen
> Forreal-a my killin a*bilities* unclean fa*cilities*. [10]

Here, we call pairs in which both members are internal (such as "agilities" / "abilities") bridge rhymes, and those where the first word or phrase is line-final (such as "calibrated" / "validated"), link rhymes.

3. FINDING RHYMES AUTOMATICALLY: A PROBABILISTIC MODEL

We modeled our rhyme detection program after local alignment protein homology detection algorithms using BLOSUM (BLOcks of amino acid SUbstition Matrix) [11]. In this framework, pairs of proteins are modeled as sequences of symbols generated either randomly or based on shared ancestry (homology). Pairs of matched amino acids receive a log-odds score in the BLOSUM matrix M: a positive score indicates the pair more likely co-occurs due to homology, and a negative score indicates the pair is more likely to co-occur due to chance. Scores are in log-odds: $\mathrm{M}[i, j] = \log_2(\Pr[i, j|\mathrm{H}]/\Pr[i, j|\mathrm{R}])$, where H is a model of related proteins (obtained by counting the frequency with which we see symbols i and j matched to each other in proteins known to be homologous) and R is the frequency of the symbols i and j in random proteins (obtained from frequency counts over all proteins). If a pair of protein sequences contains regions in which the amino acids align to give high scores, the pair is considered to be homologous.

In our work, song lyrics are transcribed into sets of sequences of syllables, with each sequence corresponding to a line of text. Similar to Kawahara's [12] treatment of consonants in Japanese rap lyrics, probabilistic methods are used to calculate similarity scores for any given pair of syllables. Phonemes which match with each other in rhyming phrases more often than expected by chance receive positive scores, while those which match less often than expected receive negative scores. Regions with syllables that, when matched to each other, have total score surpassing a threshold are identified as rhymes.

4. RHYMING SYLLABLES

To generate models of rhyming and randomly co-occurring syllables in rap lyrics, we needed a data set of known rhymes. Our training corpus includes the lyrics of 31 influential albums from the "Golden Age" of rap (1984-1994), chosen because they received the highest rating from The Source, the top-selling US rap music magazine of the time, plus nine additional albums by influential artists from the time period (Run-DMC, LL Cool J, The Beastie Boys, Public Enemy, Eric B. and Rakim). We downloaded lyrics from the Web and manually corrected them to fix typos and ensure that pairs of consecutive lines ended with matching rhymes, yielding 27,956 lines of lyrics (13,978 rhymed pairs), approximately 700 lines per album.

We first transcribe plaintext lyrics into sequences of phonemes using a wrapper we built around the Carnegie Mellon University (CMU) Pronouncing Dictionary [13], which gives phonemes and stress markings for words in North American English. We augmented the dictionary with slang terms and common elements of hip-hop vernacular (e.g., the "-in" ending in "runnin' ", or the "-a" ending in "brotha" or "killa"), and reduced the stress assigned to common one-syllable words of minor significance in rhyme ("a", "I", etc.). To handle words not found in the augmented dictionary, we added the Naval Research Laboratory's text-to-phoneme rules [14].

5. SCORING POTENTIAL RHYMES

To generate a log-odds scoring matrix for rhyming syllables, we need models for random syllables and for rhymes. For any pair of syllables i and j, the random model, $\Pr[i, j|\mathrm{Random}]$, gives the likelihood of i and j being matched together by chance while the rhyme model, $\Pr[i, j|\mathrm{Rhyme}]$, gives the likelihood of i and j being paired in a true rhyme. As in BLOSUM [11], the log-odds score is calculated as $\ln(\Pr[i, j|\mathrm{Rhyme}]/\Pr[i, j|\mathrm{Random}])$. To avoid overfitting, we reduce each syllable to its vowel (nucleus), end consonants (coda), and stress—the relevant features for determining rhyme. We approximate the coda by taking the first half (rounded up) of the consonants between adjacent pairs of vowels. Both models are trained using the occurrence frequencies of phonemes in the training data.

In the random model, the likelihood of vowel a matching with vowel b is calculated by taking the product of the frequencies of a and b. The likelihoods for consonants and varying stress are calculated in the same manner. For the rhyming model, the likelihood of vowels a and b being matched is calculated by taking the number of times a and b are seen matching in known rhymes, and dividing by the total number of matched vowel pairs in known rhymes. Then the log-odds score for the vowels is calculated as $\text{vowelScore}(a, b) = \ln(\Pr[a, b|\text{Rhyme}] / \Pr[a, b|\text{Random}])$.

The likelihood for consonants is more complicated since we must also consider unmatched consonants when aligning syllable codas of differing size. We use an iterated approach to solve these problems. In the first pass over the training data, we produce initial vowel and consonant scoring matrices by calculating the statistics above. We consider rhymes in paired lines to be all syllables following the final primary-stressed syllable, after Holtman [6]. In the second pass, we identify the start of rhymes by moving backwards from the end of the line while initial scores for stressed syllables are positive. We perform global alignment [15] on matched codas to determine frequencies for consonants pairing with other consonants, and being unmatched at the start or end of the coda. This distinction is useful since some consonants (such as /l/ and /r/) are more likely to be unmatched at the beginning of clusters, and others (often coronals, such as /d/ and /z/) are more likely to be unmatched at the ends of clusters. A simple example of this is found in the repeated occurrences of "alarmed" rhyming with "bomb" in Public Enemy's "Louder Than A Bomb." [16]

Using these frequency statistics, we produce the rhyming model and log-odds scores for consonants and stress in the same way as for vowels. Finally, we normalize the consonant score by dividing by the length of the coda to avoid the problem of syllables with long codas having the consonant score dominate. Intuitively, "win" and "gin" rhyme as well as "splints" and "mints." Since all the constituent scores are log-odds, they can be added together to form a combined probabilistic log score. The final score for two given syllables is the sum of the vowel score, normalized consonant score, and stress score.

Tables 1 and 2 show the pairwise scoring matrices. The symbols "_*" and "*_" indicate scores for unmatched consonants at the beginning and end of codas, respectively. High scores for pairs like (/m/,/n/) and (/k/,/p/) largely validate Holtman's hierarchy [6].

6. RHYME DETECTION ALGORITHM

With our probabilistic scoring method for matched syllables in place, we need a procedure to identify internal and end rhymes. Our technique is a variant on local alignment [15]; for each syllable, we identify its closest preceding rhyming syllable, and longest preceding rhyming phrase within the current and previous lines. For example,

	AA	AE	AH	AO	AW	AY	EH	ER	EY	IH	IY	OW	OY	UH	UW
AA	2.3	-3.3	-0.8	1.6	-1.7	-2.7	-7.2	-0.6	-3.9	-4.8	-3.9	-1.0	-1.7	-3.3	-3.9
AE		2.1	-1.5	-6.6	-1.9	-3.3	-1.5	-3.4	-1.8	-2.0	-4.3	-4.6	-4.5	-3.7	-6.7
AH			2.2	-1.2	-1.4	-1.4	-0.6	-0.2	-1.7	-0.3	-3.0	-1.0	-0.6	-0.9	-1.5
AO				3.1	-1.0	-3.8	-6.5	-1.1	-3.9	-4.2	-6.3	-0.3	-0.4	1.1	-3.3
AW					3.8	-0.3	-6.0	-4.2	-5.7	-6.0	-5.7	-2.0	-2.9	-4.5	-1.4
AY						2.5	-4.2	-1.1	-7.0	-1.8	-3.2	-4.3	-1.1	-5.7	-6.4
EH							1.9	-1.2	-1.5	0.2	-2.1	-7.0	-4.5	-6.1	-4.3
ER								3.9	-5.6	-1.5	-5.5	-1.6	-2.7	-1.3	-2.6
EY									2.5	-3.4	-2.7	-4.4	-4.3	-5.8	-6.5
IH										2.0	-0.9	-7.1	0.2	-2.2	-3.7
IY											2.4	-4.4	-4.2	-5.8	-6.4
OW												2.8	-4.0	-2.5	-1.5
OY													4.9	0.1	-3.7
UH														2.6	-0.5
UW															3.1

Table 1. Scoring Matrix for Vowels

given the line

> *Unob**tain**able* to the **brain** it's *unex**plain**able* what the verse'll do [10]

from Pharoahe Monch's "Right Here," the middle "ain" syllables all rhyme, while the whole of "unexplainable" also rhymes with "unobtainable."

For every pair of consecutive lines in a set of lyrics, we first construct a two-dimensional matrix of the score for every pair of syllables. Entries in this matrix (corresponding to pairs of syllables in the lines) are selected as "anchors" if they have score above a threshold and contain a stressed syllable or are line-final. From these anchor positions, rhymes are extended forward, ensuring that the length-normalized score is above a syllable threshold. In addition to the iterative extension, a "jump"-type extension is also allowed, in which one or two syllables can be skipped over if the following syllable pair is an anchor type with score above a higher threshold. This was included since longer polysyllabic mosaic rhymes often contain one or two syllables that do not rhyme in the midst of three or four that do. A good example of this can be found in Fabolous' "Can't Deny It":

> I **keep spittin', them clips copped on those calicos**
> **Keep shittin', with ziplocks of that Cali 'dro** [8]

where the two lines rhyme in their entirety, with the exception of "them"/"with" and "those"/"that."

We filtered the set of rhymes to remove one-syllable rhymes including unstressed syllables, as these tended to be noise. After a set of rhymes was identified, we removed duplicates and consolidated consecutive and overlapping rhymes together.

7. VALIDATING THE METHOD

Our first test verifies that our probabilistic score for syllable rhyming is better at identifying perfect and imperfect rhymes than rules-based phonetic similarity measures. We did a 10-fold cross validation where we chose 36 albums from the training data, trained a rhyme model for those albums, and used it to score the known rhyming lines from the other four albums (true positives) as well as randomly selected lines from those four albums (presumed to be true

	B	CH	D	DH	F	G	JH	K	L	M	N	NG	P	R	S	SH	T	TH	V	Z	ZH	_*	*_
B	4.3	-4.8	1.1	0.4	-5.5	1.9	1.9	-6.9	-0.3	-0.5	-1.6	-5.5	0.1	-0.9	-1.6	-4.6	-1.0	-4.3	2.3	0.3	-2.5	-0.6	-1.5
CH		4.2	-1.6	-4.9	-0.3	0.3	0.4	1.5	-6.8	-6.6	-2.8	-5.5	1.1	-6.7	0.3	0.6	0.9	1.4	-6.1	-2.0	-2.5	-6.0	-2.6
D			2.3	-7.0	-7.6	0.1	0.2	-3.1	-1.7	-2.2	-2.2	-3.0	-1.8	-0.9	-9.0	-2.1	0.2	0.0	-0.2	0.0	-4.6	-0.2	1.2
DH				3.5	-5.6	-5.1	-4.2	-0.4	-0.2	-2.0	-7.5	-5.6	-6.2	-1.4	-7.0	-4.8	-0.3	1.3	2.8	1.1	-2.6	-6.0	-3.4
F					3.4	-1.2	-4.9	-0.3	-1.5	-1.3	-3.5	-1.6	1.1	-2.7	1.1	1.2	-0.9	4.0	0.6	-7.3	-3.2	-1.4	-2.9
G						4.2	1.9	0.0	-0.2	-1.0	-1.9	-5.7	-0.6	-0.8	-2.5	-4.9	-1.1	-4.5	0.3	-0.3	-2.7	-0.9	-2.8
JH							5.2	-6.3	-1.5	0.1	-0.5	-4.8	-0.2	-0.3	-0.6	0.6	-1.1	-3.6	1.4	1.0	4.1	-5.3	0.5
K								2.6	-2.9	-2.1	-2.6	-1.3	1.7	-2.1	-0.7	-0.6	0.9	0.5	-1.8	-3.1	-4.7	-1.0	-1.8
L									2.8	-1.8	-1.8	-2.8	-8.1	-0.5	-2.9	-6.6	-2.9	-6.3	-1.3	-1.6	-4.5	0.4	-1.0
M										2.7	1.8	0.7	-3.2	-1.2	-2.9	-1.1	-2.5	0.4	-0.6	-3.7	-4.2	-0.8	-1.7
N											2.2	1.2	-2.5	-1.0	-2.3	-0.7	-1.5	-0.6	-1.5	-2.1	-5.1	-0.4	-2.3
NG												4.1	-6.8	-2.7	-2.3	-5.3	-3.5	-5.0	-2.1	-2.0	-3.2	0.2	-3.9
P													3.3	-2.0	-1.1	-0.7	1.1	0.9	-0.6	-7.9	-3.8	-0.7	-0.8
R														2.8	-2.3	-0.8	-1.2	-6.1	-2.1	-2.2	-4.3	1.7	-0.7
S															2.6	2.4	-1.0	1.0	-2.4	0.5	0.0	0.6	0.6
SH																5.2	-0.6	-4.1	-1.3	-0.2	3.6	-5.8	-7.7
T																	1.7	1.6	-0.9	-9.2	-5.2	0.0	0.7
TH																		4.4	0.5	-6.1	-2.0	-5.4	-0.6
V																			2.9	-0.4	1.6	-1.2	-1.7
Z																				2.6	3.0	-1.3	1.1
ZH																					6.8	-3.7	-5.6

Table 2. Scoring Matrix for Consonants

negatives). We developed implementations of the minimal mismatch of articulatory features and Kondrak alignment [17] metrics to compare the performance of these scoring measures, which are based on the physical process of the human voice. We show receiver operator characteristic (ROC) curves comparing the true positive rate to false positive rate when varying the score threshold for each of the three methods in Figure 1. The probabilistic method significantly outperforms both simpler rules-based methods.

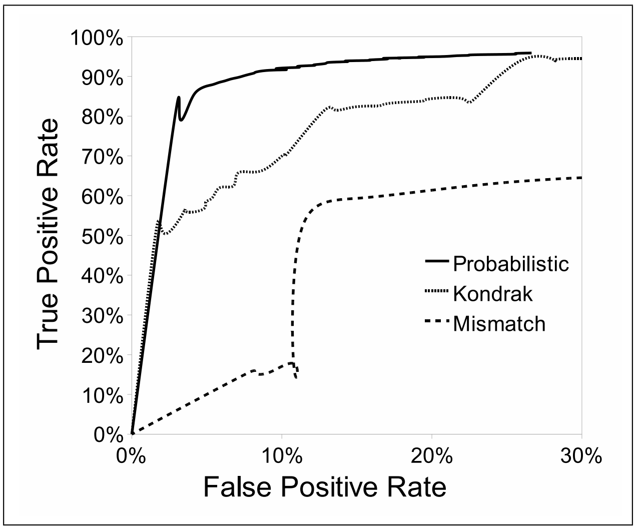

Figure 1. ROC curves for the three different scoring methods, comparing percentage of actual rhymes found by algorithm on the y-axis with percentage of unrelated syllables detected as rhyming on the x-axis

Next, we considered false positives and negatives for detected end rhymes, using the score threshold of 1.5 (meaning matched syllables are at least $e^{1.5}$ times more likely to rhyme than expected by chance). Out of 1000 pairs of unrelated random lines from our training data, 79 syllables were marked as parts of end rhymes ("false positives") by our procedure. Of these, 22 were in fact true rhymes, with scores higher than 3.0. 30 were near-rhymes; that is, that they could be found (though less frequently) as line final rhymes in actual lyrics. Usually scoring above 2.0, they included matches such as "stiff"/"fit", "pen"/"thing", and "cling"/"smothering", with more than one articulatory difference or different stress. 14 matched end syllables (often suffixes), typically with high scores (greater than 3.0). Examples such as "weak**er**"/"drumm**er**" and "tapp**in'**"/"posit**ion**", may have exact matches, but are not relevant rhymes due to their lack of stress. The remaining 13 moderately high scoring (between 1.5 and 2.5) pairs featured either high consonant scores (like "bust"/"test") or high vowel scores due to matching rare vowel sounds ("box"/"wrong").

From a set of 1000 matched pairs of lines, we used the iterative method (moving backwards from the end of the line while scores for stressed syllables are positive) to see which true rhymes would be missed. Pairs with all such matches scoring less than 1.5 were marked and treated as false negatives. Out of 132 such syllables, the largest group (48) were moderately low scoring (between -1.0 and 1.5) pairs participating in polysyllabic and mosaic rhymes. A good example of this is "battery"/"battle me" in Eric B. and Rakim's "No Omega" [7]; many of these were flanked by high scoring pairs, and would be included in rhymes using the jump extension described in the above section. 35 were very low scoring pairs (less than 0.0) which were either caused by words having been transcribed improperly or the lack of a true rhyme in the lyrics. 22 were caused by the rhyme start being extended too far back and starting with a low positive scoring pair. Again, this would not cause problems in our actual detection algorithm since, in that case, rhymes are extended forward from stressed anchors. 17 were caused by differences between the actual pronunciation and the dictionary's pronunciation ("poems" treated as one syllable, or "battles" specifically being pronounced to rhyme with "shadows"). Finally, 10 were caused by deliberate mismatch in syllable stress.

The probabilistic model is quite good at finding both perfect and imperfect rhymes. Quite few syllable pairs (less than 15 in the 1000 line pairs) scored highly without being perceivably rhyming, and most low scoring "true" rhyme pairs take part in complex mosaic and polysyllabic rhymes.

Finally, we used our model on a set of manually annotated rap lyrics, to measure the ability of the program to find both internal and line-final rhymes. We used five songs of varying style: the Beastie Boys' "Intergalactic", a Grammy-winning song in the old-school style; Pharoahe Monch's "The Truth" (featuring Common and Talib Kweli) and "Right Here", which were annotated by Alim [9] and feature high rhyme density and a complicated scheme; Jay-Z and Eminem's "Renegade", which features very high rhyme density; and Fabolous' "Trade It All (Part 2)", a song specifically mentioned by Alim for its prevalence of long (five or six syllable) rhymes. We show the ROC curves for this test set in Figure 2; the best overall performance is for specificity and sensitivity just above 60%. Most "false positive" are rhymes that were not annotated due to lack of rhythmic importance or accidental omission. False negatives included several where the performer created a rhyme from words that do not appear to rhyme as text, and some longer rhymes that were cut off prematurely due to too many non-rhyming syllables within them and lower scoring syllable pairs surrounding them. Finally, some rhymes were missed due to intervening rhymes being found between the rhyming parts, particularly when the threshold for rhymes is set low. This is especially evident in the ROC curves at lower cut-off thresholds, where true positive rates peak around 80% and begin to decline as the threshold is lowered.

8. EXPERIMENTS

We used our procedure to examine a variety of features about the rhymes in several sets of lyrics. We computed the number of syllables per line, the number of rhymes per line, the number of rhymes per syllable, average end rhyme scores, and proportion of rhymes having two, three, four, or more syllables. We also counted all of the complex rhyming features (bridge, link, chain and internal rhymes) per line.

We hypothesized that these features would show differences between genres of popular music, and calculated them for four sets of data: the top 10 songs from Billboard Magazine's 2008 year-end Hot Rap Singles chart; the top 20 songs from the 2008 year-end Hot Modern Rock Songs chart; the first 400 lines of Milton's "Paradise Lost" [18], as a similar-sized sample of non-rhyming verse; and the top 10 songs from the 1998 year-end Hot Rap Singles chart. To compare the verses most of all, the song lyrics were modified to remove intro/outro text, repeated lines, and additional choruses. Our results are in Table 3. High end rhyme scores are indicative of song lyrics in general (relative to unrhymed verse); rap has higher rhyme density, internal rhyme, link rhymes, and bridge rhymes. Interestingly, blank verse and rock lyrics have similar amounts of rhyming per line, but rock lyrics have more rhymes per syllable. The data from 1998 and 2008 rap songs suggest that in their rhyming pattern, there has not been much shift in style.

	Rap '08	Rap '98	Rock	Blank
Number of Lines	476	613	502	400
Number of Syllables	4646	6492	4053	4146
Syllables per Line	9.76	10.59	8.07	10.37
Number of Rhymes	794	1118	476	393
Rhymes per Line	1.67	1.82	0.95	0.98
Rhymes per Syllable	0.17	0.17	0.12	0.09
Rhyme Density	0.28	0.27	0.19	0.12
Average End Score	5.28	5.21	4.36	2.49
per Syllable	3.75	3.67	4.01	2.28
Doubles per Rhyme	0.23	0.29	0.15	0.18
Triples per Rhyme	0.08	0.06	0.04	0.03
Quads per Rhyme	0.02	0.03	0.05	0.00
Longs per Rhyme	0.03	0.02	0.04	0.01
Internals per Line	0.62	0.60	0.27	0.28
Links per Line	0.20	0.28	0.13	0.16
Bridges per Line	0.43	0.48	0.28	0.40
Chaining per Line	0.32	0.18	0.15	0.07

Table 3. Rhyme Features for Different Genres

We also hypothesized that features of individual rappers might also be informative, so we produced these statistics for albums by nine famous MCs from a diverse range of styles and eras: Run-DMC, Rakim, Notorious B.I.G., 2Pac, Jay-Z, Fabolous, Eminem, 50 Cent, and Lil' Wayne. Features were calculated for segments of at least 40 lines to produce means and standard deviations of the statistics for each album. The results indicate that many of these features can be characteristic of different artists' styles. For example, Run-DMC's (1984) old-school style has lower rhyme density and less internal rhyme with an average of 1.5 rhymes per line and only 6% of rhymes being longer than 2 syllables; while Rakim (1987), known for his more complex style, is detected as using more internal rhymes (0.63 per line to Run-DMC's 0.48) and more rhymes longer than 2 syllables (12%). Rival rappers Notorious B.I.G. (1994) and Tupac Shakur (1995) display fairly similar style characteristics: 28% of their rhymes are 2 syllables long, 6% are three syllables, and 3% are longer. However, Biggie's lines are significantly shorter in length, with, on average, 10.8 syllables to 2Pac's 11.6.

Artists from the early 2000s like Jay-Z (2001), Eminem (2000), and especially Fabolous (2001) favour longer rhymes, with 15%, 17%, and 30% respectively of their rhymes being longer than 2 syllables. They also have the most rhyme density overall, with 2.2, 2.3, and 1.9 rhymes per line respectively. Jay-Z and Eminem tend to use more internal rhyme as well, having 0.8 internal rhymes per line–about 25% higher than the average among other MCs. Although he portrays a "thug" persona, 50 Cent (2003) uses the most syllables per line (12.1), while Lil' Wayne (2008) has the fewest (10.2). However, he manages high rhyme density (0.3 rhymed syllables for each syllable used) with relatively few (only 1.8) rhymes per line. In general, we find that automatic rhyme detection can yield characteristic data about performers and genres.

Figure 2. Rhyme Detection Syllable ROC Curves for Different Songs. The y-axis indicates the percentage of true rhymes identified by the algorithm, while the x-axis shows the percentage of automatically identified rhymes not considered to be true rhymes.

9. CONCLUSION

Using a probabilistic scoring model, we were able to identify both perfect and imperfect rhymes with a higher level of accuracy then simpler rules-based methods. The heuristic rhyme detection methods achieved moderate success at finding both internal and line-final rhymes in song lyrics. More importantly, statistical features of these rhymes did correspond to real world characterizations of rhyme style, and many of these features are quite consistent within individual artists' lyrics and varied between different artists. This leads to the possibility that automatically calculated rhyme statistics can be used to make meaningful categorizations and recommendations based on rhyme style.

10. REFERENCES

[1] B. Wei, C. Zhang, M. Ogihara: "Keyword Generation for Lyrics," *Proceedings of the International Conference on Music Information Retrieval*, 2007.

[2] F. Kleedorfer, P. Knees, T. Pohle: "Oh Oh Oh Whoah! Towards Automatic Topic Detection in Song Lyrics," *Proceedings of the International Conference on Music Information Retrieval*, pp. 287–292, 2008.

[3] H. Fujihara, M. Goto, J. Ogata: "Hyperlinking Lyrics: A Method for Creating Hyperlinks Between Phrases in Song Lyrics," *Proceedings of the International Conference on Music Information Retrieval*, pp. 281–286, 2008.

[4] R. Mayer, R. Neumayer, A. Rauber: "Rhyme and Style Features for Musical Genre Classification by Song Lyrics," *Proceedings of the International Conference on Music Information Retrieval*, pp. 337–342, 2008.

[5] Adam Bradley: *Book of Rhymes: The Poetics of Hip Hop*, Basic Civitas Books, 2009.

[6] Astrid Holtman: *A Generative Theory of Rhyme: An Optimality Approach*, PhD dissertation, Utrecht Institute of Linguistics, 1996.

[7] Eric B. and Rakim: *Let the Rhythm Hit 'Em*, MCA Records, 1990.

[8] Fabolous: *Ghetto Fabolous*, Elektra Records, 2001.

[9] H. Samy Alim: "On Some Serious Next Millennium Rap Ishhh: Pharoahe Monch, Hip Hop Poetics, and the Internal Rhymes of Internal Affairs," *Journal of English Linguistics*, Vol. 31, No. 1, pp. 60–84, 2003.

[10] Pharoahe Monch: *Internal Affairs*, Rawkus Records, 1999.

[11] S. Henikoff and J.G. Henikoff "Amino Acid Substitution Matrices from Protein Blocks" *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 89 No. 22 pp. 10915-10919, 1992.

[12] Shigeto Kawahara: "Half rhymes in Japanese rap lyrics and knowledge of similarity," *Journal of East Asian Linguistics*, Vol. 16, No. 2, pp. 113–144, 2007.

[13] Kevin Lenzo: *The CMU Pronouncing Dictionary*, `http://www.speech.cs.cmu.edu/cgi-bin/cmudict,2007`.

[14] H.S. Elovitz, R.W. Johnson, A. McHugh, J.E. Shore "Automatic translation of English text to phonetics by means of letter-to-sound rules," *Interim Report Naval Research Lab*. Washington, DC., 1976 `http://www.speech.cs.cmu.edu/comp.speech/Section5/Synth/text.phoneme.3.html`.

[15] R. Durbin, S. Eddy, A. Krogh, G. Mitchison *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, 1999.

[16] Public Enemy: *It Takes a Nation of Millions to Hold Us Back*, Def Jam Recordings, 1988.

[17] Grzegorz Kondrak: "A New Algorithm for the Alignment of Phonetic Sequences," *Proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 288–295, 2000.

[18] John Milton: *Paradise Lost*, Samuel Simmons, 1667.