Hip-hop rhymes mirror phonological typology

Jonah Katz, Dept. of Linguistics, UC Berkeley
1223 Dwinelle Hall
UC Berkeley
Berkeley, CA 94720
katzlinguist@gmail.com
(617) 448-3598

Abstract

A database of rhymes from African-American Vernacular English hip-hop shows that rappers possess detailed implicit knowledge about speech sounds that is largely irrelevant to the sound pattern of their particular language, but has been argued to play a role in constraining possible sound systems across all human languages. Rhyme data is relevant to phonological typology because many sound patterns involve the neutralization of contrasts in contexts where they are less perceptible, and rhyme involves implicit judgments of perceptual similarity between sounds. The general finding is that sounds are more likely to mismatch in a rhyme in those contexts where the contrast between the sounds is more likely to neutralize cross-linguistically, even though the contrasts studied here do not neutralize in these contexts in English. The results contribute to the evidence for detailed implicit phonetic knowledge and bear on the issue of synchronic explanation in linguistic theory.

1 Introduction

The question of what a speaker knows about the sounds of her language, the relationship of that knowledge to extrinsic properties of the human communication system, and the role of both domains in constraining the range of possible human languages are central topics in phonetic and phonological theory. This paper is part of a growing literature that uses rhyme and verbal games to examine speakers' phonetic knowledge. A database of rhymes from a corpus of African American English (AAE) hip-hop is shown to reflect certain typological generalizations about possible and impossible human languages. By hypothesis, this is because both the rhyme patterns and these specific typological facts are partly determined by perceptual properties of the strings under investigation. The result is especially interesting because some of the perceptual factors investigated here, involving voicing and major place, are not obviously relevant to English phonology.

The arguments advanced here rely on the idea that rhyme goodness or likelihood reflects perceptual similarity (Steriade 2003, Kawahara 2007, Johnsen 2011). While many rhymes are perfect, in the sense that their rhyming parts are phonologically identical, some rhymes feature parts that mismatch in one or more ways. We refer to these as *imperfect rhymes*. Not all types of mismatch are equally likely, and this is by hypothesis related to the perceptibility of those mismatches. One important property of imperfect rhyme data is that it reflects not only perceptual similarity, but a rhymer's implicit knowledge of similarity. Rhymers do not confuse imperfect rhymes for perfect ones and use them by mistake. Instead, they tolerate imperfect rhymes in proportion to how perceptually similar the rhyming pairs are. This means that, unlike most laboratory experiments on speech perception, we can study similarity independently from errors and confusion. This allows us to examine the implicit knowledge of speakers in ways that, for instance, identification and discrimination experiments do not. If rhymers consistently allow rhymes mismatching for some feature more often in some context, it is evidence that they implicitly know that contrasts for the relevant feature are less perceptible in that context.

One reason to study implicit knowledge about perceptual distinctiveness is that it has been argued to play a role in the typology of certain phonological processes and contrasts (e.g. Flemming 1995, Steriade 1999, Côté 2004). In particular, some phonological contrasts are easier to perceive in certain contexts than others. Some, but by no means all, phonological processes appear to reflect these perceptual factors, being systematically more likely to neutralize contrasts in positions where they are perceptually indistinct. If rhyme facts and some phonological processes both reflect implicit knowledge of perceptual similarity, then we expect perceptually-driven phonological phenomena to be reflected in rhyme data. The results presented in section 4 suggest that this prediction is correct.

The current study makes several empirical and theoretical contributions. The results show that rappers' verbal behavior reflects fine-grained perceptual distinctions that go above and beyond what is required to learn their grammars, and thus contribute to the study of phonetic knowledge. This is the first study that uses formal statistical hypothesis-testing to show that rhyme data parallel phonological typology and that these effects generalize across individual artists. Finally, the results here bear on a central issue in current phonological theory: the balance between transmission-related factors and factors related to the mental grammar in explaining the constrained relationship between phonetics and phonology.

The paper is organized as follows: section 2 contains background about rhyme, hip-hop, and positional neutralization; section 3 describes the construction and analysis of a database of hip-hop rhymes; section 4 examines how various featural mismatches pattern in the corpus and compares their patterning to the phonological typology discussed in section 2; section 5 discusses the findings and their implications for phonological theory.

2 Background

2.1 Rhyme

All of the data discussed here involve the notion of *rhyme*. In English verse poetry and hip-hop, rhyme is a similarity or identity relation that holds between various phonetic strings. In most English verbal traditions, monosyllabic rhyme involves every part of a phonetic or phonological string except consonants at the beginning of the syllable (onset consonants). This constituent, which also plays a role in phonology (Selkirk 1982, Harris 1983, Steriade 1988), will be referred to here as *rime*, in order to distinguish it from the rhyme relation itself. It is defined as the string of segments beginning at the nucleus of a syllable and extending to the end. For instance, the rime of the English word *dogs* is/ɔgz/.

In rhymes that extend over more than one syllable, all unstressed syllables following the initial one also participate in the rhyme relation. We refer to the entire string involved in a rhyme as the *rhyme domain*: the rime of a stressed syllable and the entirety of zero or more succeeding unstressed syllables (after Holtman 1996). When two strings stand in a rhyme relation, it is their rhyme domains that correspond. We say that those domains form a *correspondent pair* consisting of two *correspondents*. This is somewhat similar to the OT concept of output-output correspondence (Benua 1997; see Holtman 1996 and Horn 2010 for applications to rhyme). Rhyme correspondence is illustrated in table 1. The first and third correspondent pairs rhyme: the rhyme domains of the two correspondents are identical. The second pair features different stressed vowels in the two correspondents, and does not rhyme. The fourth pair features the same stressed vowel in the two correspondents, but all other segments in the two rhyme domains are not the same; this pair constitutes at most a defective or marginal rhyme.

[TABLE 1 HERE]

The rhyming pairs in table 1 are *perfect* rhymes: the rhyme domains of the two correspondents are identical. These examples thus make it appear that there is an all-or-nothing criterion for rhyme: if the two domains mismatch in any way, then the correspondent pair is not a rhyme. This is not, in fact, true: in English verse poetry, and particularly in hip-hop, we also observe imperfect rhymes. These are rhymes whose correspondent domains mismatch in one or more ways, but still somehow ‘count’ as a rhyme, in the sense of being perceived as a rhyme or of being allowed to occupy metrical positions that are constrained to rhyme.

Some examples of imperfect rhyme from the corpus are illustrated in table 2. The correspondent pair in the first rhyme mismatches for consonant place. The second pair displays a similar mismatch in between two vowels. The third correspondent pair mismatches for both consonant features and the presence of a consonant. The fourth pair mismatches for number of consonants and place of those consonants.

[TABLE 2 HERE]

The existence of imperfect rhyme in the genre under discussion is important: given the hypotheses that rhyme is a similarity relation and that better rhymes are used more frequently, it means that rhyme data reveals something about the similarity of various linguistic objects. If only perfect rhymes were allowed, the only conclusion we could draw is that segments (or tokens thereof) are most similar to themselves (or other tokens of the same categories).

One problem that immediately arises in the presence of imperfect rhyme, however, is how to tell what rhymes with what. Relying on listener intuitions is inappropriate here, because the data from the corpus is used to support arguments about perceptual similarity, and listener intuitions presumably have their source in exactly this domain. It would be circular to claim that *x* corresponds with *y* infrequently because *x* and *y* are perceptually very distinct, if the basis for counting rhymes in the first place is perceived similarity. In more rhythmically rigid genres, this problem is avoided by defining the rhyme position rhythmically and then counting everything that occurs in that rhythmic position as a rhyme. In contemporary hip-hop, however, rhythmically predictable rhymes are accompanied by a large number of rhymes in other, less predictable rhythmic positions (Alim 2003, Horn 2010). The solution described in section 3.2 is to count as a rhyme any correspondent pair that satisfies a very liberal definition of rhyme in terms of rhythmic and phonological/phonetic properties. These criteria undoubtedly introduce some noise, in the form of false positives, into the database. Noise, however, can be overcome through statistical modeling, unlike circularity.

Another crucial aspect of the rhyme data examined here is that they includes consonants in correspondence in a number of different segmental contexts. The first rhyme in table 2, for instance, involves two correspondent consonants mismatching for place features in post-vocalic, domain-final position. The second rhyme involves a similar featural mismatch in intervocalic position. The third rhyme, if we use segmental alignment as a guide, involves /t/ and /f/ corresponding in V_C position. The analysis of contextual differences in rhyme likelihood features prominently in the analysis here. The three contexts just mentioned are illustrated with minimal pairs in figure 1. Here and in what follows, *R* stands for an approximant, glide, or vowel; *T* for an obstruent; # for a rhyme-domain boundary.

[FIGURE 1 HERE]

The examination of these contexts allows us to explore parallels with implicational universals in phonology, some of which are stated over such contexts. It also offers a test of the hypothesis that rhyme likelihood involves perceptual similarity rather than being wholly determined by phonological features. The nature of the phonological features examined in this study, such as [voice] and [continuant], are fundamentally the same in the three contexts. This is plausibly part of what we mean when we call them *phonological* features. Differences in rhyme likelihood across contexts thus cannot be explained by phonological features. In contrast, the phonetic correlates of these features, and hence their perceptual distinctiveness, do differ across contexts.

2.2 Hip-hop

Hip-hop is a verbal art form that arose in African-American communities in 1970s New York.[1] It involves setting words to an isochronous musical beat, much like the lyrics of a song, but generally without musical pitch; linguistic pitch is, of course, present, and may be important in signaling

[1] Any statement more specific than this about the origins of hip-hop would be controversial. For more on the social and cultural history of hip-hop, see Summers (ed.) 2011.

rhymes. The sequence of isochronous musical beats is organized into stronger and weaker levels, much like linguistic representations of stress (see Lerdahl & Jackendoff 1983, Palmer & Krumhansl 1990 for discussions of musical metrical structure). In song, there are a number of principles constraining the association of linguistic syllables to musical beats, explored at length by Halle & Lerdahl (1993), Hayes & Kaun (1996), Hayes & MacEachern (1996), and Hayes (2009); hip-hop follows broadly similar rules to the genres discussed in these works (although the rules or constraints are weaker and more likely to be violated in hip-hop). For instance, syllables receiving word-level stress tend to be mapped to strong beats in the musical meter. Horn (2010) gives a detailed account of aspects of hip-hop textsetting and how they differ from previously-studied genres.

The textsetting properties of hip-hop are not a central concern here, although they are certainly interesting and worthy of further investigation. We will, however, highlight one aspect of textsetting that affects our study: rhyme alignment. As described in section 2.1, the availability of imperfect rhyme makes it difficult to determine which domains stand in rhyme correspondence. In (rhythmically) simpler genres such as those examined by Hayes & MacEachern (1996), as well as in early hip-hop, rhymes are located at and (more or less) only at the right edge of constituents referred to as *lines*. Lines in turn are defined by their tendency to be aligned with linguistic constituents such as phrases or sentences and the fact that they occupy a particular number of beats in the metrical structure. All of these properties are visible in figure 2.

It is simple to extract rhymes from this example: they occur at and only at the right edges of consecutive lines. The line can be identified as occupying a particular number of musical beats, indicated here with metrical 'X' marks underneath the lyrics. These marks stand in for two particular levels of periodicity in the song, at roughly 60 and 120 beats per minute, which would probably be notated as half and quarter notes, respectively. Lines predictably span four beats at the half-note level. Each line also corresponds to a large linguistic constituent, roughly a clause (or, in prosodic terms, an intonational phrase). The rhymes here are perfect: there are no mismatches between the rhyme domains of *chicken-lickin'* and *funny-money*. This is typical of this period in hip-hop. If the material in the corpus discussed here displayed predictable lines and rhyme positions like this example, we could say with some degree of certainty which strings stand in rhyme correspondence.

Later hip-hop, however, has considerably more freedom on all of the dimensions mentioned above. Although musical rhythmic units tend to align with linguistic constituents, *enjambment* (mismatch between rhythmic and linguistic constituency) is fairly frequent. Although rhymes generally do occur at some more or less predictable rhythmic interval, they are not constrained to appear only in this position. And although perfect rhyme is frequent, imperfect rhyme is the rule: a previous version of the corpus with rhymes coded by listener intuition showed that about 65% of the perceived rhymes mismatched for one or more features/segments. Several types of non-line-final rhyme are illustrated in Figure 3.

[FIGURE 3 HERE]

In (3a), *dreams-jeans* occurs at the right edge of a line, but another rhyme (*niggaz-figure*) occurs internal to the second line, underlined here. In (3b), the rightmost syllable of each line rhymes (*jam-gram*), but there is also a series of multiple 'stacked' rhymes preceding this: *slow-pro* and *poundin' a – down wit' the*. In (3c), a series of two- or three-syllable rhyme correspondents follow each other in quick succession (note that these rhymes involve a vowel merger not present in general American English). This last example also illustrates the rhythmic complexity of later hip-hop: although the song is generally in a duple meter, which is reinforced by the preceding context and the instrumental background, the stress contour and rhyme alignment in this section instead reinforce a periodicity of three beats (at the level immediately below the notated Xs). Coupled with the absence of pause or (musically) long syllables anywhere in the local context, this makes it difficult to even define a line level. For all of these reasons, the current study essentially gives up on trying to find a principled, structure-based way to locate rhymes in the musical surface and instead adopts an overly-inclusive string-based heuristic, described in section 3.2.

2.3 Previous literature on rhyme and similarity

Rhyme and verbal wordplay and their relationship to phonetic and phonological similarity are the topic of a growing literature. A number of studies have found that some types of imperfect rhyme in English are more common than others. Within the generative linguistics tradition, Zwicky (1976) was one of the first authors to examine imperfect rhyme in terms of phonological features. In a corpus of rock lyrics, he finds that place, voicing, and continuancy are most likely to mismatch. Holtman (1996) finds that in English verse poetry and hip-hop, single-feature mismatches are most common, especially place mismatches. Hanson (2003), in a study of English slant rhyme (where only final consonants, and not vowels, correspond) in the work of Robert Pinsky, and Horn (2010), in a study of hip-hop artist Snoop Dogg (known at various times as Snoop Doggy Dogg, Snoop Lion), replicate some of these findings.

None of these papers include statistical tests of the differences in prevalence between various types of rhyme. Additionally, none of these studies distinguish between bias and similarity in analyzing the relative likelihood of various mismatches. For instance, if /t/ and /k/ or /m/ and /n/ frequently correspond in final position in English, it may be because place contrasts in this context are perceptually indistinct, or it may be because /t/, /k/, /m/, and /n/ are all very frequent segments in final position in English. We can't conclude anything about similarity until we examine rhyme frequency data that has been corrected for segmental bias (i.e., frequency).

Several later studies attempt to correct for bias through the use of contextually-conditioned probabilities, observed over expected ratios, or frequency-balanced experimental stimuli. Steriade (2003) argues that Romanian poets make use of imperfect rhyme in ways that reflect perceptual similarity, and *not* phonological features, lexical knowledge, or knowledge of Romanian alternations. She further argues that these perceptual asymmetries are the same ones implicated in phonological typology. For instance, voicing mismatches are more common after nasal consonants and domain-finally than they are intervocalically; this corresponds to the cross-linguistic fact that voicing contrasts are frequently neutralized in post-nasal position (e.g. Arusa, Japanese; see Hayes & Stivers 2000 for an overview) and domain-final position (e.g. Russian, Totontepec Mixe, see Steriade 1999 for an overview) without being neutralized in pre-vocalic position. Steriade's study thus has substantial overlap with the current one; the main differences here are the inclusion of a greater variety of feature mismatches and contexts, the use of formal statistical modeling and hypothesis-testing, and the incorporation of between-subjects variance with the goal of generalizing beyond the few artists under consideration. The current study also differs from Steriade's in examining a form that is not *learned*, in the bisyllabic sense of being the topic of scholarly literature and conventions that are explicitly taught to aspiring artists. This is not to suggest that learned genres are less valuable as objects of study, simply that the current study expands the empirical domain of rhyme-phonology parallels to a different kind of genre.[2]

Several other papers have argued that rhyme likelihood is best explained with reference to perceptual similarity rather than phonological features. Kawahara (2007, 2009) argues that Japanese hip-hop rhymes and imperfect puns reflect phonetic similarity, and cannot be explained by phonological factors alone. He finds no evidence that phonological alternations mediate similarity judgments. For instance, /h/ and /ɸ/, although they alternate in Japanese, are no more likely to correspond with each other than other comparable pairs of obstruents. He also shows that voicing for sonorants, although it is inert in Japanese phonology, nonetheless affects pun likelihood. Johnsen (2011) demonstrates that American English speakers' explicit judgments of rhyme 'goodness' across domain-final consonantal mismatches are better predicted by perceptual confusability data than by phonological feature metrics.

The current study attempts to replicate and extend several findings from the previous literature while improving the methodology used in the study of rhyme likelihood. In particular, regression modeling allows us to characterize rhyme frequencies in a framework with well-understood quantitative properties that can be used to test the statistical significance of various asymmetries in the corpus. The input to the regression model consists of a distance metric, derived from Luce's (1963) Biased Choice Model, that corrects for bias. The use of mixed-effects models allows us to generalize across multiple rhymers while still taking the variation between rhymers into account. The use of hip-hop, which allows frequent and often phonetically-distant imperfect rhyme

[2] Although there are websites that explicitly discuss rhyming practice with the goal of educating aspiring hip-hop artists, they arose long after the careers of the artists examined here had begun.

correspondences, allows us to examine a wide variety of features with a smaller corpus than would be necessary for more rigid genres. And the phonotactics of English allow us to examine consonantal mismatches in a wider variety of contexts than, e.g., Japanese, where nearly all consonantal correspondences occur in intervocalic position.

2.4 Contrast and positional neutralization

This study focuses on mismatches between rhyme correspondents for major place and voicing, two features that have been particularly well studied from both a phonetic and phonological standpoint (e.g. Fuimura *et al.* 1978, Jun 1995, Lisker & Abramson 1964, Steriade 1999). Linguistic contrasts involving both of these features frequently *neutralize* in one or more contexts. If a feature is capable of distinguishing between lexical items in a given position, like [voice] in *pub* and *pup*, we say that that feature contrasts; when the feature cannot distinguish between lexical items, as in *lapse* and hypothetical **labse*, we say that it is neutralized. The term *neutralization* thus covers *assimilatory* cases like *lapse*, where the voicing of /p/ is predictable from the following consonant, and *non-assimilatory* cases, such as final devoicing (Stampe 1973), where the voicing of a segment is predictable from its position in a phonetic or phonological string.

The hypothesis explored here is that, because the cross-linguistic distribution of phonological contrasts and the distribution of featural mismatch in rhyme are both influenced by perceptual properties, the likelihood of rhymes mismatching for these features should mirror their cross-linguistic distribution. This entails that if a phonological neutralization process is affected by perceptual asymmetries, then that process should find a parallel in the domain of rhyme. We do not claim that all phonological neutralization processes are driven by perceptual asymmetries; it is entirely possible that perceptually-grounded neutralization is present in grammars alongside neutralizations that pertain to articulatory efficiency, abstract markedness, paradigmatic morphological effects, or any number of other linguistic factors. In this section we summarize how voicing and place contrasts, which are plausibly affected by positional perceptual differences, pattern typologically. More extensive reviews for voicing are given by Lombardi (1991) and Steriade (1999); for major place, Steriade (2001) and Jun (2011). We also briefly examine processes that neutralize contrasts such as nasality, approximancy, and continuancy.

Both major place and voicing contrasts are least likely to be neutralized before a vowel or sonorant consonant. Every language that neutralizes one of these contrasts in pre-vocalic position also neutralizes it in all other positions (Steriade 1999, 2001, Jun 2011). These are languages that have only one phonemic nasal, like Mohawk (Mithun 1996) and Tlingit (Maddieson *et al.* 2001)[3]; and languages with no (obstruent) voicing contrasts, like Yukulta (Keen 1983) and Canela-Krahô (Popjes & Popjes 1986).

Both major place and voicing contrasts are less likely to neutralize in word- (or phrase-) final position than before a non-sonorant consonant. Every language that neutralizes one of these contrasts domain-finally also neutralizes it before obstruents, but some languages which neutralize one of these contrasts before obstruent consonants do not do so domain-finally (Steriade 1999, Jun 2011). This is illustrated in figure 4a for nasal place in Spanish, which neutralizes both finally and pre-consonantally;[4] Selayarese (Mithun & Basri 1986) and Greek (Arvaniti 1999) pattern similarly. Figure 4b illustrates neutralization of nasal place pre-consonantally but not domain-finally in Diola

[3] Some dialects of both languages have a marginal second nasal /m/ appearing only in loanwords.
[4] In some dialects the word-final nasal is velar and the coronal variant is absent.

Fogny (Sapir 1965); Ponapean (Ito 1986) and Malayalam (Jun 1995) pattern similarly. For voicing, domain-final and pre-obstruent neutralization occurs in Russian (Padgett 2002) and Lithuanian (Kenstowicz 1972); pre-obstruent but not domain-final neutralization occurs in French (Dell 1995) and Hungarian (Lombardi 1991).

[FIGURE 4 HERE]

[TABLE 3 HERE]

Voicing and place thus display similar contextual profiles, summarized in table 3. Several researchers have proposed that the nature of these typological implications follows from the perceptual properties of segments in the contexts under discussion (Ohala 1990, Jun 1995, Steriade 1999, 2001). More generally, neutralization is influenced by speech perception: contrasts tend to neutralize in positions where they are perceptually indistinct (Liljencrants & Lindblom 1972, Ohala 1983, Flemming 1995, Steriade 1999, Blevins 2004). Both voicing and place contrasts are cued in part by properties of adjacent sonorant segments: for place, this is primarily formant transitions; for voicing, F0 and F1 of adjacent sounds, duration of a preceding sound, and VOT in a following sound. The more flanking sonorant sounds (R_R compared to the other two contexts), the more cues. For stops, place and voicing are also cued in part by spectral properties (for place), duration, and amplitude of the burst; these properties are often obscured by the closure of a following non-sonorant consonant.

If these perceptual asymmetries also affect the likelihood of feature mismatch in rhyme, the most straightforward hypothesis is that rhymes mismatching for voicing and major place should be more likely in contexts with more gray cells in their columns in table 3: least likely in between sonorants, more likely domain-finally, and most likely before non-sonorant consonants.

The current study also examines nasals and liquids. The features that distinguish these segments from one another and from obstruents, [son], [nas], [cont], and [approx], are unlike voicing and place in that they rarely display perceptually-driven positional neutralization. We briefly summarize some phonological alternations involving these features here.

Nasality frequently spreads non-locally (Hansson 2001, Rose & Walker 2004), but nasal harmony does not affect consonants in any particular phonological position. Nasal stops become oral adjacent to nasal stops in several Bantu languages including Punu and Lingala (Hyman 2003). Although the motivation for this process is not entirely clear, it may be related to the difficulty of perceiving nasal place contrasts coupled with a general prohibition on geminates in these languages. In any case, it is not clear that it is related to the perceptibility of the nasal contrast *per se*.

Medial stops (especially in between vowels) frequently become approximants or fricatives in the course of spirantization (see Kaplan 2010 for an overview), but this virtually never neutralizes existing phonological contrasts (Gurevich 2003). Continuants frequently harden to stops following nasals, and this can neutralize contrasts (Steriade 1993, Padgett 1994). There is no reason, however to believe this particular form of positional neutralization is related to perceptual properties of the continuancy or approximancy contrast. The authors mentioned above see it as primarily an articulatory phenomenon, and we concur. Some languages display neutralizing glide hardening in pre-consonantal (Bergüner Romansh, Kaisse 1992) or post-consonantal (Cypriot Greek, Kaisse 1992) position, or neutralizing stop gliding in word-final position (Lama, Gurevich 2003). As the brief description just given suggests, these neutralizations occur in distinct and non-overlapping contexts, and there does not appear to be any implicational relationship between those contexts.

As such, the only straightforward prediction from phonological typology about obstruents and sonorants is that mismatches between the two classes should be generally less common than mismatches within either one. This is because, while place and voice neutralization show clear contextual implicational asymmetries across languages, which are plausibly driven by perceptibility, continuancy neutralizations (including obstruent-liquid and obstruent-glide) either show no clear contextual implicational asymmetries or show asymmetries related to production rather than perception. The prediction about relative mismatch frequency is already confirmed in the studies mentioned in section 2.3, and will be tested again in what follows.

3 Methods

3.1 Materials

The songs included in the corpus were recorded from 1993 to 2007. There is no particular thematic or generic unity to the corpus; it includes a mix of 'conscious' (e.g. Talib Kweli), 'hardcore' (e.g. Big Pun), and commercial (e.g. Jay-Z) hip-hop, for instance. The artists represented in the corpus were all selected in part because they impressionistically have a high proportion of inner and multi rhymes as illustrated in figure 3. These rhymes are harder to objectively locate than line final ones, but investigation of a pilot version of this corpus also suggests that these rhymes are more likely to be imperfect than the line-final ones. As such, they are a valuable source of data: the more imperfect rhymes in the corpus, the easier it is to statistically test hypotheses about the relative likelihood of various mismatches.

All of the artists examined here were born or raised in New York City. This was done to keep regional dialectal variation to a minimum, making it easier to generalize across the artists in the corpus. Of course, this introduces a limitation on the interpretation of any results: we don't know if these results will generalize to a wider variety of regional accents. While AAE has a (gradient and variable) tendency not to reflect geographically-based variants in local white dialects, it still undoubtedly displays some level of regional variation (See Labov 2010, ch. 16, for an overview). In any case, the hypotheses investigated here involve the *existence* of a kind of phonetic knowledge; showing that one dialect reflects that knowledge is therefore sufficient for our purposes.

Seven artists are examined here: Slick Rick, Nas, MF Doom, Talib Kweli, Big Pun, Mos Def, and Jay-Z. All of them were born and raised in New York, except for Slick Rick and MF Doom, who were born in the UK and moved to New York as children. All of them rap in AAE, broadly construed (Green 2002); they display such features as (near-)merger of *raw-roar* ('non-rhotic'), *pin-pen*, *cycle-psycho* (vocoid /l/), *pride-prod* (/ɑɪ/ merges with /ɑ/ before voiced consonants and word boundaries), and invariant pronunciation of the inflectional morpheme –*ing* with a coronal nasal. These features are reflected in transcriptions and used in rhyme-domain segmentation as described below. For each artist, enough songs were transcribed to extract around 500 consonantal correspondences (before the data filtering described in the next section); the number of songs thus varied between artists. Table 4 contains more details about the database.

[TABLE 4 HERE]

All rhymes were transcribed by hand in a broad phonetic transcription based on recorded performances; those rhymes were identified using the criteria described in the next section. An alternative would have been to script the transcription process using a pronouncing dictionary. This process, however, ignores prosodically-influenced factors (such as vowel reduction) and a fair bit of allophonic variation, as well as failing to transcribe non-standard lexical items, which are frequent in this genre. When automatic transcription was used, virtually all of the materials needed to be re-transcribed by hand; it was therefore abandoned.

3.2 Data collection

The criteria used for rhyme are as follows: if some rhyme domain has the same number of syllables and the same stressed vowel as another rhyme domain that appears within 16 beats at the most salient metrical level (generally around 60-120 beats per minute), the two domains are counted as a rhyme. The 'most salient level' here corresponds to the music-theoretic notion *tactus*, which is generally defined as the most natural periodicity for listeners to tap or clap along with a piece of music (Lerdahl & Jackendoff 1983). There is some evidence that this most-salient level is independently motivated by accentual patterns and fine-grained timing regularities in music performance (Temperley 2001), although we have not investigated those factors here.

The rhyme domain, recall, is the string of segments beginning at a stressed vowel and extending to the end of the *rhythmic group* or the next stressed syllable. A rhythmic group boundary was defined according to empirical music theory (Lerdahl & Jackendoff 1983, Deliege 1987) as occurring at an inter-onset interval that is longer than the surrounding ones. Stressed vowels were defined as those having qualities other than the unstressed English vowels [ə], [i], [o] and occupying prominent metrical positions according to empirical music theory (Lerdahl & Jackendoff 1983), or those bearing a pitch accent.

For instance, the phrase *chíllin' in Kentúcky Fríed Chícken* from figure 2, if each syllable spans one metrical beat and each content word bears a pitch accent and/or full vowel on its stressed syllable, is segmented and transcribed (assuming a more or less typical phonetic implementation for this dialect) as [(tʃ)ɪlənəŋkən][(t)ʌki][(fɹ)ɑd][(tʃ)ɪkən]. The second half of the next line, *eatin' food and finger lickin'*, is [itən][(f)udən][(f)ɪŋgə(ɹ)][(l)ɪkən]. This illustrates several important points about transcription.

The treatment of syllabic consonants and schwa is quite difficult to resolve. The current study, however, ends up omitting unstressed-syllable rime data for independent reasons, so these issues do not need to be resolved here.

The final /ɹ/ in *finger* is notated inside parentheses because it was treated as present just in case there was another consonant that could correspond with it. Although there is clearly no phonetic approximant in this position, it's much harder to tell if phonetic reflexes of this segment survive in the form of an offglide or vowel coloring (especially in stressed syllables); as such, we simply left all 'deleted' rhotics and voicoid liquids in the transcription as optional. This decision does not directly affect obstruent voicing and within-manner place contrasts. It could, however, affect data for features like [son] and [approx] that we use as a kind of control in what follows. If these consonants are really absent, it means we will systematically introduce noise into the analysis of /ɹ/ that is not present for other segments. One type of evidence that stressed-syllable coda /ɹ/ is present in some form is its likelihood of corresponding with itself (or various other coda consonants) relative to corresponding with nothing (open syllables). In stressed, domain-final position, which is unambiguously an /ɹ/-'deletion' context, /ɹ/ corresponds with itself 40 times, with other coda consonants 17 times, and with an open syllable only 9 times. In 6 of the 9 deletion-like correspondences, the open syllable in question contains /ɔ/ (e.g. *raw* corresponding with *more*), which is often realized with a schwa-like offglide somewhat similar to a vocoid /ɹ/ in this dialect. All of these facts suggest that /ɹ/ is present in these syllables in whatever representation drives rhyme correspondence.

Finally, realization of /t/ and /d/ as an apical tap in intervocalic non-pre-stress positions is not reflected in this transcription; instead, we notated cases where /t/ or /d/ is *not* tapped in a context where it could or generally would be, and used the coding of phonological feature mismatch in the statistical model to capture this variation.

The example above also illustrates the fact that the rhyme criteria used here induce some false positives: because *finger* has the same syllable count and stressed vowel as *chicken* and *lickin'*, it would be characterized as rhyming with those words. Most listeners would probably not characterize this as a rhyme. The rhyme database contains many such probable false positives; they should have the aggregate effect of adding essentially random consonant correspondences, that is, noise. The statistical model of rhyme similarity will thus need to overcome that noise in order to coherently characterize the data; the results presented in section 4 suggest that it succeeded in doing so. A pilot version of this project using a more varied (but smaller) corpus and intuition-based coding of rhymes produced essentially the same results as this iteration, but effects were more robust with smaller amounts of data. As such, the current procedure can be seen as relatively conservative.

The rhyme correspondent pairs extracted from the corpus were decomposed into individual segmental correspondences. Only unambiguous correspondences, where the same number of consonants occur in the same context in each rhyme correspondent, were included in the database, because when unequal numbers of consonants occur in the two correspondents there is no theory-neutral way of deciding which ones are in correspondence. For instance, pairs like [aska] – [apta] would be treated as containing two correspondences, [s] – [p] and [k] – [t]; pairs like [aska] – [ata] would not have any correspondences included in the database, because it is ambiguous whether [t] in the second string corresponds with [s] or [k] in the first and it is difficult to characterize what the context of the correspondence is because it differs for the consonants in the two strings. After this type of exclusion was applied, there were 3,442 unambiguous segmental correspondent pairs across all segments and contexts.

The three contexts reported on here are those shown in table 3: R__R, where both corresponding segments are flanked by vowels, glides, or liquids; R__#, where both correspondents follow a vowel, glide, or liquid and precede a rhyme-domain boundary; and R__T, where both correspondents follow a vowel, glide, or liquid and precede a stop, fricative, or nasal. Because data was quite sparse for R__R and R__T contexts in unstressed positions, the statistical model is limited to consonants following stressed vowels.

The R__T context was further winnowed down to exclude certain positions with obligatory or near-obligatory place assimilation in English. The reasoning is that if a segment α appears in a context where it is subject to place assimilation, then it is impossible for α to mismatch for place with a corresponding segment β unless β either appears in a different context or differs from α in more features than just place. In other words, these contexts differ from the other ones considered in this study in not allowing minimal place mismatches. For instance, in the context of tautomorphemic /V_k/, it is impossible for nasals to minimally mismatch for place, because only the velar nasal appears here; for a second correspondent to mismatch the place of [ŋ], the context consonant (/k/ here) would also need to differ between the two strings, or the correspondent segments would need to differ for more features than just place (e.g. [ŋ] – [s]). For nasals, assimilation contexts were defined as occurring before stops and all fricatives except the inflectional morpheme /-z/, and /ð/, which is unambiguously the start of a distinct morpheme in such sequences. Assimilation contexts for obstruents were somewhat more complicated: some place contrasts for these segments are *de facto* neutralized by the impossibility of adjacent identical obstruents within an English word, e.g. /æpt/ is a word of English but */ætt/ is impossible. Such contexts were defined separately for each set of obstruents (the contexts differ according to the voicing and continuancy of the first consonant in the sequence). Correspondent pairs were excluded from the corpus if both segments appeared in assimilation contexts meeting these definitions.

Some of the consonants that were included in the database appear in positions of *voicing* assimilation in English (e.g. obstruents in the context /V_s/). Following the same logic applied above for place assimilation, these segments should be less likely to mismatch for voicing because the following segments would also need to mismatch in this case. Because excluding this data would result in a near total lack of obstruents in pre-consonantal position in the database, they were kept in the analysis. Because the prediction is that voicing should be more likely to mismatch in pre-consonantal position than other contexts, this confound should go against the experimental hypothesis. This means that the database as currently constituted will result in an overly conservative test of contextual hypotheses.

Segmental correspondence data were examined for the segments /p, t, k, b, d, ɡ, f, s, v, z, m, n, ŋ, l, ɹ/. These segments constitute the closest thing that English allows to an exhaustive crossing of phonological features such as [voice], [continuant], [sonorant], [nasal], [approximant], and major place; they are also reasonably frequent in most contexts in the corpus. There were 1,755 unambiguous correspondences for these segments included in the database.

3.3 Data analysis

Correspondences involving the segments mentioned above were analyzed as stimulus-response pairs, with the first segment treated as stimulus and the second treated as response. Pairwise distance measures for each pair of segments in each context for each artist were computed using the *d* measure from Luce's (1963) Biased Choice Model (BCM). This measure, which estimates the perceptual distance between any two segments based on their confusability, is defined as follows: for any pairwise contingency table of correspondences between segments α and β, the BCM measure *d* is the sum of the negative log odds of α appearing given segment β and the negative log odds of β given α. This measure thus characterizes each segment as being distance 0 from itself, with distances between different segments calculated relative to this baseline. The *d* measure distinguishes between bias and similarity, where bias in the corpus will be essentially equivalent to segmental frequency. This property is crucial in analyzing correspondence data; it ensures that a pair is not judged as similar simply because its component segments are frequent.

The BCM in general and the *d* measure in particular are generally construed as characterizing perceptual distance between various categories based on a subject's likelihood of labeling an instance of one category as a different category in an identification task. This is subtly different from rhyme, which cannot be straightforwardly characterized as an identification task. As such, I refer to the *d* measures used here as measures of *rhyme distance* rather than perceptual distance. Under the hypothesis, supported by Steriade (2003), Kawahara (2007), and Johnsen (2011), that more perceptually similar segments are more likely to correspond in rhymes, the BCM is also applicable to rhyme data.

The BCM distance measure is particularly useful for this study because it takes as an input categorical data, such as responses to a stimulus or the occurrence of rhyming pairs, and produces as output a continuum of values on an interval scale, construed as a kind of distance. This interval data can be subjected to a variety of parametric statistical tests, unlike the categorical data that we began with. So, for instance, we can ask if the difference in rhyme distance for voicing mismatch between domain-final and pre-consonantal position is significantly different from the difference in rhyme distance for place mismatch between the two contexts. In the current study, we submit the distance values derived from categorical data to a regression model, described below, which allows us to estimate the effect of various types of mismatch and context on rhyme distance or likelihood.

BCM distance data for non-matching segments (recall, all matching segments are set to have distance 0) were subjected to linear mixed-effects regression analysis using the lme4 package in R (Bates 2007). This model estimates the effect of mismatch for various features, changes in context, and combinations of feature mismatches and contexts on rhyme distance. A positive effect of some parameter indicates that that parameter increases rhyme distance or, equivalently, decreases rhyme likelihood. The mixed-effects property allows us to explicitly model between-artist variance while attempting to generalize across individual artists to the larger population. That larger population could be construed in various ways; the most conservative characterization would be something like '20-35 year-old male African American professional rappers who spent most of their childhoods in New York City'. Parameters of the model that characterize variance between artists are treated as *random effects*, variables whose levels (individual rapper identities) are sampled from a larger population. The linguistic properties of interest here are treated as *fixed effects*, variables whose levels are systematically controlled and examined in the study. For more background on mixed models and their uses in linguistics, see Baayen *et al*. 2008.

The fixed effects in the model were phonological featural mismatch, context, and mismatch x context interactions. Artist identity was a random effect. The features used are [voice], [continuant], [sonorant], [nasal], [approximant], and major place. Featural mismatches were dummy-coded with place mismatch set as the baseline and all other mismatches compared to place. Contexts were dummy coded with domain-final set as the baseline. Dummy coding is a way of comparing two or more categorical predictor variables to each other. In this particular model, the intercept term corresponds to the rhyme distance of place mismatches in domain-final position, and all other combinations of feature mismatches and contexts are assigned some unique set of values for contextual and featural variables.

This model is set up to ask a series of questions about differences between features, differences between the same feature in various contexts, and differences between differences across contexts and features. Specifically, how much more or less likely is major place to mismatch than to not mismatch in domain-final position? How does segmental context affect this likelihood for major place? Do other features differ from major place in the way that their likelihood of mismatching varies across contexts? These questions are encoded, respectively, by the effect of major place, by the effects of context, and by the interactions between other features and context.

The interaction terms here are thus somewhat complicated, but this is inevitable given the kinds of questions we are trying to ask in this study. The issue examined here is not just whether major place and voicing show a contextual profile that matches their typological patterning, but whether they differ from each other or from other features that don't show the same typological patterning. The other features included here can thus be seen as a control to assure that contextual differences we find for voicing and place are not just artifacts of some general property of rhyme data. Major place is set as the baseline because the sources discussed in section 2.3 all report that this is the most frequent type of mismatch in the English genres they examine. Domain-final position is set as the baseline because we predicted on the basis of typology that it should be intermediate in rhyme distance (for voicing and place) between pre-consonantal position (lower distance) and intervocalic position (higher distance); setting domain-final position as the baseline allows us to test both steps in this hypothesized scale.

The significance of fixed effects was assessed with Markov chain Monte Carlo (MCMC) sampling, using the pvals.fnc function in lme4. This function, to simplify slightly, generates hypothetical sets of parameters over and over again, then compares these parameters to the actual ones the model has fitted to the data in order to assess the probability of obtaining such extreme parameters by chance. Baayen *et al.* (2008) describe the procedure in more detail. In what follows, effects are reported with the coefficient β, which gives the slope of the regression line fitted to the data for that variable; the standard error, which is the model's estimated standard deviation for the size of β; the *t* statistic, which measures the size of β in standard error units; and a *p* value derived from MCMC sampling.

4 Results

Rhyme distance parameters for obstruent voicing mismatch and place mismatch within manner are shown in figure 5. Both types of mismatch show a decline from left to right: this corresponds to increasing rhyme likelihood (decreasing rhyme distance) in contexts where neutralization is more common, as predicted. For instance, the left panel shows that obstruents mismatching for voicing have the greatest rhyme distance in intervocalic position (left box), intermediate rhyme distance in domain-final position (middle box), and smallest rhyme distance in pre-consonantal position (right box).

[FIGURE 5 HERE]

The statistical model of rhyme distance parameters estimates the independent contribution of each feature to rhyme distance in each context; results are shown in table 5. The term *independent* here relates to the fact that some segments mismatch for more than one feature, and some featural mismatches characterize more than one pair of segments. The model attempts to generalize across all of these pairs and features. For instance, the likelihood of /d/ and /ɹ/ corresponding is modeled as the sum of the likelihood associated with each feature in which they mismatch: [sonorant], [continuant], and [approximant].

[TABLE 5 HERE]

Place mismatches are significantly less distant in pre-consonantal position than in domain-final position and more distant in intervocalic position than in domain-final position. Obstruent voicing mismatches show no significant differences from place in contextual changes, although they are more distant in general.

The other features of the consonants examined here, voicing, continuancy, nasality, sonority, and approximancy, are all signficantly more distant domain-finally than place is. While sonorancy and continuancy do not significantly differ from place in their contextual profiles, nasality and approximancy do. Specifically, mismatches for both features decrease in rhyme distance from domain-final to intervocalic position relative to the change for place. Nasality mismatches also decrease in distance in pre-consonantal position relative to place.

Analysis of the hip-hop rhyme database thus confirms our principal hypothesis: mismatches for voicing and place in rhyme parallel patterns of phonological neutralization for these features across languages. Voicing and place are most likely to mismatch in these rhymes before obstruents and nasals, less likely at the ends of rhyme domains, and least likely in between vowels or sonorant consonants.

Other features parallel the contextual profile of place to various degrees. Continuancy and sonority appear to be roughly the same as place in this regard, while nasality and approximancy show different contextual patterns. We concluded in section 2.4 that there is no particularly strong generalization in phonology about perceptually-driven positional neutralization for these features. They all, however, plausibly neutralize less often than voicing and place; corresponding to this asymmetry, segments mismatching for these features are less likely to correspond across the board. This generally holds true even of the individual features distinguishing obstruents and sonorants, but it holds *a fortiori* at the level of the segment: a segment that differs from an obstruent in sonorancy necessarily also differs from the obstruent in either nasality or approximancy (and possibly more features), meaning that the rhyme distance for such pairs is estimated as the sum of the distances associated with individual features.

5 Discussion

These findings bear on specific questions that arise in the literature on rhyme discussed in section 2.4. First, certain features are more likely to mismatch than others. This means that the rhyme data are not consistent with a model where rhyme likelihood is determined only by phonological features. Second, the likelihood of mismatch varies by context. Because the features examined here are by hypothesis the same (in phonological terms) in all contexts, this suggests that phonetics plays a role in determining rhyme likelihood. For at least voicing and major place, the data is consistent with the idea that featural mismatch is more likely in contexts where the feature is less perceptible.

The study also bears on a larger question within linguistic theory: why typological patterns in phonology display parallels to asymmetries in speech perception. As we saw in section 2.4, for instance, languages that neutralize nasal place contrasts word-finally also neutralize them before oral stops, but the converse is not true. Corresponding to this implicational asymmetry, nasal place contrasts are difficult to discriminate before stops (Ohala 1990a, Hura *et al.* 1993). Given that the typology of certain phonological processes and contrasts reflects asymmetries in speech perception, the question of how and why this parallelism holds immediately arises. At least two explanations have been offered, and we briefly summarize them here.

Nasal place contrasts may frequently neutralize before stops because language learners are more likely to misperceive place of articulation in this context and subsequently learn word-forms different from the ones intended by the speaker (Ohala 1990a), or they may neutralize because language learners organize their phonological grammars to allow unfaithful mappings for nasal place in contexts where nasal place contrasts are less perceptually distinct (Jun 1995, Steriade 2001). More generally, contrasts may neutralize in indistinct contexts because they are more likely to be miscategorized during the process of transmission from one generation to the next (Ohala 1975, Blevins 2004, Garrett & Johnson 2013), or they may neutralize because individual speakers' grammars optimize for the perceptual distinctiveness of contrasts (Flemming 1995, Steriade 1999, Hayes, Kirchner, & Steriade (eds.) 2004, Kawahara 2006, Zuraw 2007). Note that these explanations are not mutually exclusive; both factors may well influence typology.

One of the principal objections to phonetically-based phonology (Anderson 1981) and to synchronic phonological theory more generally (Ohala 1990b, Blevins 2004) is that it is needlessly complex: because languages are acquired by individuals through speech perception, those languages will inevitably reflect asymmetries in confusability, whether or not speakers' grammars are optimized to exploit such asymmetries. Positing such specialized knowledge should thus be avoided on grounds of parsimony. The current study provides evidence that the argument from parsimony, while it may be valid, is essentially irrelevant: speakers behave in ways that reflect subtle differences in perceptibility, and thus must be able to mentally represent this information at some level. Rhyming pairs mismatch more often for voicing and place in precisely the contexts where those features are more likely to neutralize cross-linguistically (but *not* in English). The simplest explanation for this parallelism is that rhyme and phonology share a common cognitive core: speakers' knowledge of perceptual asymmetries in speech. Note that typical speech perception studies do not provide this kind of evidence. Identification and discrimination tasks investigate 'errors' in speech transmission, e.g. cases where two distinct sounds are heard as identical or cases where one sound is miscategorized as another. These studies do not show and are not meant to show that speakers know (explicitly or implicitly) anything about the distinctiveness of linguistic contrasts.

Of course, the current study should be interpreted with some caution. It does not provide direct evidence for phonetically based grammatical constraints, e.g. speakers using phonetic optimization as a means of constraining the language acquisition process. What the study does show is that parsimony alone does not favor either approach over the other, and that evidence bearing on the question will need to come from other domains. For instance, the existence of idiosyncratic phonological phenomena with no obvious phonetic motivation would tend to favor a role for diachronic explanation (Bach & Harms 1972, Blevins 2004), while directionality in certain phonological changes and repair strategies favors a role for synchronic optimization (Hura *et al.* 1993, Steriade 2001, Kiparsky 2006).

A second caveat is that the artists examined here constitute a self-selected subject group, namely people who have become famous for their skill at composing aesthetically pleasing hip-hop. Although it is far from obvious that aesthetic appreciation of hip-hop is linked to phonetically similar rhymes (it may just as well be more aesthetically pleasing to hear surprising, dissimilar rhymes), this is a possibility. Some of the perceptual subtleties therefore may be characteristic of extraordinary individuals but fail to generalize to the population of African-American English speakers, or of speakers in general. Several other strands of research, however, provide converging arguments from phonology and phonetics that 'normal' individuals possess a wealth of implicit knowledge about speech perception, speech production, and the links between the two (Kingston & Diehl 1994, Berent *et al.* 2007, Zuraw 2007). There is as yet no reason to believe that hip-hop artists are special in this particular regard.

Tables

Pair	*Rhyme domain*	*Rhyme?*
beat-seat	/it/-/it/	Y
beat-suit	/it/-/ut/	N
barrier-carrier	/æɹiɹ̩/-/æɹiɹ̩/	Y
barrier-fatuous	/æɹiɹ̩/-/ætʃuəs/	N

Table 1. Illustration of rhyme domains and rhyme correspondence in English.

Pair	**Rhyme domain**	**Source**
right-pipe	/ɑɪt/-/ɑɪp/	Nas, *The World is Yours*
super-bazooka	/upə/-/ukə/	MF Doom, *El Chupa Nibre*
differences-witnesses	/ɪfɹənsəz/-/ɪtnəsəz/	MF Doom, *El Chupa Nibre*
fiendin'-screamin'	/indən/-/imən/	Slick Rick, *Kill Niggaz*

Table 2. Imperfect rhymes from the corpus.

(a)

Place	**R__R**	**R__#**	**R__T**
Mohawk	No contrast	No contrast	No contrast
Spanish	Contrast	No contrast	No contrast
Diola Fogny	Contrast	Contrast	No contrast

(b)

Voicing	**R__R**	**R__#**	**R__T**
Yukulta	No contrast	No contrast	No contrast
Russian	Contrast	No contrast	No contrast
Hungarian	Contrast	Contrast	No contrast

Table 3. The typology of major place neutralization (a) and voicing neutralization (b). The leftmost column contains an example of each pattern. Neutralization in any cell of the table asymmetrically entails neutralization in all cells to the right.

Artist	Songs	Recording date	Lines	Correspondences
Slick Rick	I Run This	1999	400	511
	Why, Why, Why			
	Kill Niggaz			
	Street Talkin'			
	Trapped in Me			
Nas	New York State of Mind	1993	513	531
	One Love			
	The World is Yours			
MF Doom				
(Danger Doom)	El Chupa Nibre	2005	543	591
	Sofa King			
	Basket Case			
	Mince Meat			
Talib Kweli	Goin' Hard	2004	418	482
	Broken Glass			
	I Try			
	Listen	2007		
Big Pun	The Dream Shatterer	1998	443	511
	Beware			
	Glamour Life			
Mos Def	Mathematics	1999	589	591
	Miss Fat Booty			
	Hip Hop			
Jay-Z	What More Can I Say?	2003	624	562
	Justify My Thug			
	Change Clothes			
	Moment of Clarity			

Table 4. Information about the corpus. 'Lines' refers to the total number of rhyme-final domains transcribed in the corpus for each artist; some lines contain more than one rhyme domain. 'Correspondences' refers to the total number of consonantal correspondences extracted for each artist, before the filtering process described in section 3.2.

Distance for mismatch in	compared to	β	Std. Err.	t	p(MCMC)	Sig.
place in R_#	0 (match)	4.88	0.58	8.37	<0.001	*
[voice] in R_#	place in R_#	1.54	0.43	3.60	<0.001	*
[cont] in R_#	place in R_#	0.75	0.30	2.47	0.012	*
[nas] in R_#	place in R_#	2.40	0.40	5.92	<0.001	*
[son] in R_#	place in R_#	1.17	0.41	2.85	0.004	*
[app] in R_#	place in R_#	3.87	0.42	9.32	<0.001	*
place in R_T	place in R_#	-1.97	0.49	-4.00	<0.001	*
[voice] in R_T vs. R_#	place in R_T vs. R_#	-0.77	0.61	-1.27	0.216	
[cont] in R_T vs. R_#	place in R_T vs. R_#	-0.67	0.43	-1.57	0.113	
[nas] in R_T vs. R_#	place in R_T vs. R_#	-3.16	0.57	-5.52	<0.001	*
[son] in R_T vs. R_#	place in R_T vs. R_#	-0.36	0.58	-0.63	0.524	
[app] in R_T vs. R_#	place in R_T vs. R_#	-0.36	0.59	-0.62	0.542	
place in R_R	place in R_#	1.01	0.47	2.12	0.030	*
[voice] in R_R vs. R_#	place in R_R vs. R_#	-0.70	0.58	-1.21	0.228	
[cont] in R_R vs. R_#	place in R_R vs. R_#	-0.65	0.42	-1.56	0.119	
[nas] in R_R vs. R_#	place in R_R vs. R_#	-2.48	0.56	-4.39	<0.001	*
[son] in R_R vs. R_#	place in R_R vs. R_#	-0.25	0.57	-0.43	0.670	
[app] in R_R vs. R_#	place in R_R vs. R_#	-1.78	0.58	-3.06	0.002	*

Table 5. Statistical model of rhyme distance, including featural mismatch terms, context terms, and feature x context interactions. Columns show the effect coefficient β, the estimated standard error of β, the *t* statistic associated with the effect, and a *p*-value derived from MCMC sampling.

FIGURES

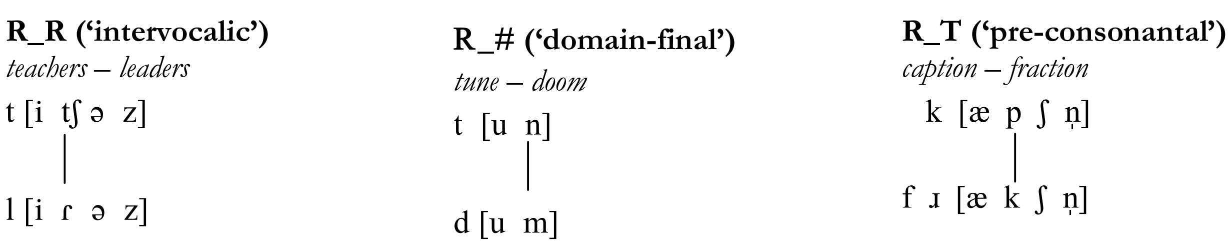

Figure 1. Imperfect rhyme in three contexts (left to right): in between two vowels, glides, or approximants; following a vowel and preceding a rhyme-domain boundary; and following a vowel and preceding a non-sonorant consonant. Brackets indicate rhyme domains. Vertical lines indicate imperfect rhyme correspondence with featural mismatch.

	One	day when	I was	chillin'	in Ken	tucky	Fried	Chicken
	Just	mindin'	my business		eatin'	food and	finger	lickin'
	This	dude	walked	in	lookin'	strange and	kind of	funny
	Went	up	to the	front	with a	menu	and his	money
		X		**X**		**X**		**X**
Meter:		**X**	**X**	**X**	**X**	**X**	**X**	**X**

Figure 2. Metrical and orthographic representation of an 'old-school' hip-hop song, Run DMC's *You Be Illin'*, 1986. Two salient levels of metrical pulse are shown.

(a)

	small	city	girl	with	big	city	dreams
	niggaz	try to	figure	how to	get up	in them	jeans
	X		**X**		**X**		**X**
Meter:	**X**	**X**	**X**	**X**	**X**	**X**	**X**

(b)

			mu-	sic	pound-	in' a	slow	jam
	with a	girl	knowin'	and	down	wit' the	pro-	gram
	X		**X**		**X**		**X**	
Meter:	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

(c)

	higher	and hot-	ter than	lava	this scho-	lar ad-	visor	is smart
	as Mac-	Gyver	to put	honor	in-side	the heart	of a	liar
	X		**X**		**X**		**X**	
Meter:	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Figure 3. Non-final rhymes from the corpus. (a) Inners, from Talib Kweli, *Broken Glass* (2004). (b) Multis, from Slick Rick, *Why, Why, Why* (1999). (c) Chains, from Big Pun, *The Dream Shatterer* (1998).

(a)

Coronal	Labial	Velar
tanto ‘so much’	* tampoko	* blanko
* tamto	tampoko ‘neither’	* blamko
* taŋto	* taŋpoko	blaŋko ‘white’
tan ‘so’	* tam	* taŋ

(b)

/ni-gam-gam/	→	nigaŋgam ‘I judge’	* nigamgam
/na-ti:ŋ-ti:ŋ/	→	nati:nti:ŋ ‘He cut (it) through’	* nati:ŋti:ŋ
/fan-fan/	→	faŋfan ‘lots’	* fanfan

Figure 4. Illustrations of the implicational universal governing nasal place neutralization. (a) Spanish neutralizes nasal place contrasts before stops (first three rows) and domain-finally (last row). (b) Diola Fogny neutralizes nasal place contrasts before stops, but not domain-finally. No attested language neutralizes nasal place contrasts domain-finally but licenses them before stops. Asterisks indicate impossible phonological forms.

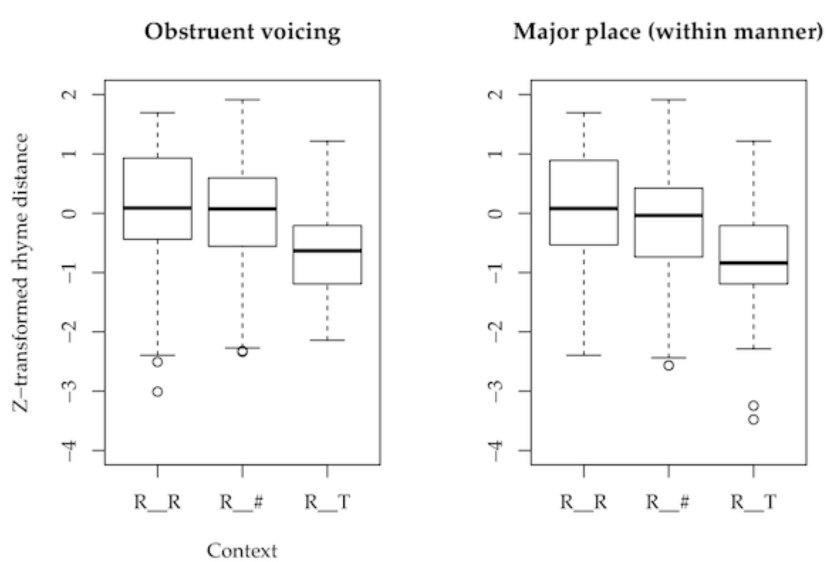

Figure 5. Rhyme distance associated with obstruent voicing mismatch (left panel) and major place mismatch within the classes of obstruents and nasals (right panel) in three contexts. Vertical axis

shows BCM *d* measures subjected to a by-subject Z transform for comparison across subjects. Dark line indicates median, boxes indicate inter-quartile range, whiskers indicate range up to 1.5 times inter-quartile range, open circles indicate potential outliers.

References

Alim, S. (2003). On some serious next millenium rap ishhh: Pharaoh Monch, hip hop poetics, and the internal rhymes of Internal Affairs. *Journal of English Linguistics, 31,* 60-84.

Anderson, A. (1981). Why phonology isn't "natural." *Linguistic Inquiry, 12,* 493-539.

Arvaniti, A. (1999). Standard Modern Greek. *Journal of the International Phonetic Association, 29,* 167-172.

Bach, E., & Harms, R. (1972). How do languages get crazy rules? In R. Stockwell & R. Macaulay (eds.), *Linguistic change and generative theory* (pp. 1-21). Bloomington: Indiana University Press.

Bates, D. (2007). lme4: An R package for fitting and analyzing linear, nonlinear and generalized linear mixed models. Software application.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104,* 591-630.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns* Cambridge, UK: Cambridge University Press.

Côté, M. (2004). Syntagmatic distinctness in consonant deletion. *Phonology, 21,* 1-41.

Deliege, I. (1987). Grouping conditions in listening to music: an approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception, 4,* 325-360.

Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua, 95*, 5-26.

Flemming, E. (1995). *Auditory representations in phonology*. PhD dissertation, UCLA.

Fujimura, O., M. Macchi, & L. Streeter. (1978). Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech* **21**, 337-346.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. Yu (ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97). Oxford: Oxford University Press.

Green, L. (2002). *African-American English: A Linguistic Introduction*. New York: Cambridge University Press.

Gurevich, N. (2003). *Functional constraints on phonetically conditined sound changes*. PhD dissertation, University of Illinois at Urbana-Champaign.

Halle, J., & Lerdahl, F. (1993). A Generative Text-Setting Model. *Current Musicology, 55,* 3-26.

Hanson, K. (2003). Formal variation in the rhymes of Robert Pinsky's The Inferno of Dante. *Language and Literature, 12,* 309-337.

Hansson, G. (2001). *Theoretical and typological issues in consonant harmony*. PhD dissertation, UC Berkeley.

Harris, James W. (1983). Syllable structure and stress in Spanish: a nonlinear analysis. Cambridge, Mass.: MIT Press.

Hayes, B. (2009). Textsetting as constraint conflict. In J. Aroui & A. Arleo (eds.), *Towards a Typology of Poetic Forms* (pp. 43-61). Amsterdam: John Benjamins.

Hayes, B. & Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *The Linguistic Review, 13,* 243-303.

Hayes, B., Kirchner, R. & Steriade, D. (eds.). (2004). *Phonetically Based Phonology*. Cambridge, UK: Cambridge University Press.

Hayes, B. & MacEachern, M. (1996). Are there lines in folk poetry? In C. Hsu (ed.), *UCLA Working Papers in Phonology 1* (pp. 125-142). Los Angeles: UCLA Linguistics Department.

Hayes, B., & Stivers, T. (2000). Postnasal voicing. Ms., UCLA.

Holtman, A. (1996). *A Generative Theory of Rhyme*. PhD dissertation, Utrech Institute of Linguistics OTS.

Horn, E. (2010). *Poetic Organization and Poetic License in the lyrics of Hank Williams, Sr. and Snoop Dogg.* PhD Dissertation, University of Texis at Austin.

Hura, S., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, *35*, 59-72.

Hyman, L. (2003). Segmental phonology. In D. Nurse & G. Philippson (eds), *The Bantu languages* (pp. 42-58). London: Routledge/Curzon.

Ito, J. (1986). *Syllable Theory in Prosodic Phonology*. PhD dissertation, UMass Amherst.

Johnsen, S. (2011). Rhyme acceptability determined by perceived similarity. Paper presented at the 29th West Coast Conference on Formal Linguistics, Tucson, April 22, 2011, University of Arizona.

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: an Optimality-Theoretic approach*. PhD dissertation, UCLA.

Jun, J. (2011). Positional effects in consonant clusters. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (eds.), *The Blackwell Companion to Phonology* (pp. 1103-1123). Malden, Mass.: Wiley-Blackwell.

Kaisse, E. (1992). Can [Consonantal] Spread? *Language, 68,* 313-332.

Kaplan, A. (2010). *Phonology Shaped by Phonetics: The case of intervocalic lenition*. Ph.D. dissertation, UC Santa Cruz.

Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale: the case of [+voice] in Japanese. *Language, 82,* 536-574.

Kawahara, S. (2007). Half-rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics, 16,* 113-144.

Kawahara, S. (2009). The role of psychoacoustic similarity in Japanese puns: a corpus study. *Journal of Linguistics, 45*, 111-138.

Keen, S. (1983). Yukulta. In R.M.W. Dixon & B.J. Blake (eds.), *Handbook of Australian Languages 3* (pp. 191-304). Amsterdam: John Benjamins.

Kenstowicz, M. (1972). Lithuanian phonology. *Studies in the Linguistic Sciences, 2,* 1-85.

Kingston, J. & Diehl, R. (1994). Phonetic knowledge. *Language, 70*, 419-454.

Kiparsky, P. (2006). Amphichronic linguistics vs. Evolutionary Phonology. *Theoretical Linguistics, 32,* 217-236.

Labov, W. (2010). *Principles of Linguistic Change, Volume III: Cognitive and Cultural Factors*. Chichester, UK: Wiley & Sons.

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

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48,* 839-862.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word, 20,* 384-422.

Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization.* PhD Dissertation, UMass Amherst.

Luce, R. (1963). Detection and recognition. In Luce, Bush, & Galanter (eds.), *Handbook of Mathematical Psychology* (pp. 103-189). New York: Wiley & Sons.

Maddieson, I., Smith, C., & Bessell, N. (2001). Aspects of the Phonetics of Tlingit. *Anthropological Linguistics, 43,* 135-176.

Mithun, M. (1996). The Mohawk Language. In J. Maurais (ed.), *Quebec's Aboriginal Languages: History, Planning, and Development* (pp. 159-173). Clevedon, UK: Multilingual Matters.

Mithun, M., & Basri, H. (1986). The Phonology of Selayarese. *Oceanic Linguistics, 25,* 210-254.

Ohala, J. (1975). Phonetic explanations for nasal sound patterns. In C. Ferguson, L. Hyman, & J. Ohala (eds.), *Nasalfest: Papers from a symposium on nasals and nasalization* (pp. 289-316). Stanford: Language Universals Project.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (ed.), *The production of speech* (pp. 189 – 216). New York: Springer-Verlag.

Ohala, J. (1990a). The phonetics and phonology of aspects of assimilation. In Kingston & Beckman (eds.), *Papers in Laboratory Phonology I*. Cambridge, UK: Cambridge University Press.

Ohala, J. (1990b). There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics, 18,* 153-171.

Padgett, Jaye. (1994). Stricture and nasal place assimilation. *NLLT, 12*, 465-513.

Padgett, J. (2002). Russian Voicing Assimilation, Final Devoicing, and the Problem of [v]. Ms., UC Santa Cruz.

Padgett, J. (2003). Contrast and post-velar fronting in Russian. *NLLT, 21*, 39-87.

Palmer, C., & Krumhansl, C. (1990). Mental representation for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 728-741.

Popjes, J., & Popjes, J. (1986). Canela-Krahô. In D. Derbyshire & G. Pullum (eds.), *Handbook of Amazonian Languages I* (pp. 128-199). Berlin: Mouton de Gruyter.

Rose, S., & Walker, R. (2004). A typology of consonant agreement at a distance. *Language, 80,* 475-531.

Sapir, J. (1965). *A Grammar of Diola-Fogny.* Cambridge, UK: Cambridge University Press.

Selkirk, E. (1982). The syllable. In H. van der Hulst & N. Smith (eds.), *The structure of phonological representations* (pp. 337–383). Dordrecht: Foris

Stampe, D. (1973). *A Dissertation on Natural Phonology*. PhD dissertation, University of Chicago.

Steriade, D. (1988). Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology*, *5*, 73–155

Steriade, Donca. (1993). Closure, release, and nasal contours. In M. Huffman & R. Krakow (eds.), *Nasals, Nazalization, and the Velum* (pp. 401-470). San Diego: Academic Press.

Steriade, D. (1999). Phonetics in phonology: The case of laryngeal neutralization. In M. Gordon (ed.), *UCLA Working Papers in Linguistics 2* (pp. 25-146). Los Angeles: UCLA Linguistics Department.

Steriade, D. (2001). Directional assymetries in place assimilation: a perceptual account. In Hume & Johnson (eds.), *The Role of Speech Perception in Phonology*. New, York: Academic Press.

Steriade, D. (2003). Knowledge of similarity and narrow lexical override. In Nowak, Yoquelet, & Mortensen (eds.), *Proceedings of BLS 29* (pp. 582-598). Berkeley, CA: Berkeley Linguistics Society.

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language, 83,* 277-316.

Zwicky, A. (1976). Well, this rock and roll has got to stop. Junior's head is hard as a rock. *Chicago Linguistics Society, 12*, 676-697.