2015

Hip-hop Rhymes Reiterate Phonological Typology

Jonah Katz
West Virginia University, katzlinguist@gmail.com

Hip-hop rhymes reiterate phonological typology

Jonah Katz,
Dept. of World Languages, Literatures, & Linguistics
West Virginia University
Chitwood Hall
PO Box 6298
Morgantown, WV 26506
katzlinguist@gmail.com
(617) 448-3598

Acknowledgments: Many thanks to Sverre Stausland Johnsen and two anonymous reviewers for detailed comments on drafts of this paper. The author also wishes to thank Adam Albright, Edward Flemming, and Donca Steriade for helpful discussion. Previous versions of this project have benefited from the comments of audiences at the LSA Annual Meeting and the Linguistics Departments at UMass Amherst, MIT, and Stanford.

Keywords

rhyme; perception; similarity; contrast

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 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 phonology. We argue that this is because both the rhyme patterns and the 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 are not obviously relevant to English phonology.

Many of the arguments advanced here rely on the idea that the likelihood of any kind of rhyme reflects the perceptual similarity of its rhyming parts. 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. This hypothesis is supported by several previous studies of other genres and/or languages (Steriade 2003 on Romanian poetry; Kawahara 2007, 2009 on Japanese hip-hop rhyme and puns, respectively). Studies of English genres tend to find that rhyme is constrained by some kind of similarity (Zwicky 1976 on rock, Holtman 1996 on hip-hop, Hanson 2003 on Pinsky's verse); it is not clear whether similarity in terms of shared phonological features or similarity in terms of auditory perception is the more relevant notion (note that the two notions will be correlated in the general case). One result of the current study is that rhyme likelihood in American hip-hop reflects perceptual similarity, instead of

or in addition to shared phonological features. This argument largely rests on contextual asymmetries, cases where the same phonological features have different likelihoods of mismatch in different phonological contexts, in ways that track the perceptibility of the relevant contrasts in those contexts. Because English phonology allows these contrasts in all of the contexts considered, shared features or natural classes cannot explain such asymmetries.

The investigation of imperfect rhyme is interesting for several reasons. One 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. Observing segment similarity in a perceptual experiment does not mean that listeners have knowledge about that similarity.

Implicit knowledge about perceptual distinctiveness is important in part because 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). 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 less distinct. If rhyme facts and 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.

Another reason to study implicit perceptual knowledge is that it bears on foundational questions of explanation in phonology. One influential hypothesis holds that the perceptual optimization of phonological phenomena just discussed is an emergent effect of the fact that languages are transmitted from one generation to the next via speech perception (Ohala 1975, Blevins 2004). An attractive property of this view is that principles of grammatical inference, generalization, and analogy can be expressed in extremely simple symbolic terms without the need for complex constraints involving perceptual knowledge. If rhyme demonstrates that such perceptual constraints are independently necessary, however, this argument is largely obviated.

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; 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* (see Stallworthy 1996 for an overview). In English verse poetry and hip-hop, rhyme is a similarity or identity relation that holds between various phonetic strings. In English, monosyllabic rhyme involves every part of a phonetic or phonological string except consonants at the beginning of the syllable (onsets). 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, with stress marked by an accute accent. 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.

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

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

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 (Stallworthy 1996, Hanson 2003), rock music (Zwicky 1976), and particularly in hip-hop (Horn 2010), 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 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 intervocalic position. The third correspondent pair mismatches for both consonant features and the presence/absence of a consonant. The fourth pair mismatches for number of consonants and place of those consonants.

Pair	**Rhyme domain**	**Source**
right-pipe	/ɑɪt/-/ɑɪp/	Nas, *The World is Yours*
super-bazooka	/úpə/-/úkə/	MF Doom, *El Chupa Nibre*
differences-witnesses	/ífɪənsəz/-/ítnəsəz/	MF Doom, *El Chupa Nibre*
fiendin'-screamin'	/índən/-/ímən/	Slick Rick, *Kill Niggaz*

Table 2. Imperfect rhymes from the corpus.

The existence of imperfect rhyme in the genre under discussion is important: given the hypothesis that rhyme likelihood correlates with similarity, it means that rhyme data can help reveal a rapper's knowledge of similarity. If only perfect rhymes were allowed, the only conclusion we could draw is that rappers consider segments most similar to themselves.

One problem that immediately arises in the presence of imperfect rhyme, however, is how to tell what rhymes with what. Relying on listener intuition is inappropriate here, because data from the corpus are 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 (Walser 1995, Pihel 1996, Alim 2003, Adams 2009, Horn 2010). The solution described in section 3.2 is to count as a rhyme any correspondent pair that satisfies a very loose definition of rhyme in terms of rhythmic and phonological properties. These criteria undoubtedly introduce some noise, in the form of false positives, into the database. In section 3.2, we offer evidence that this noise does not qualitatively affect the pattern of results.

Another crucial aspect of the rhyme data examined here is that they include 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.

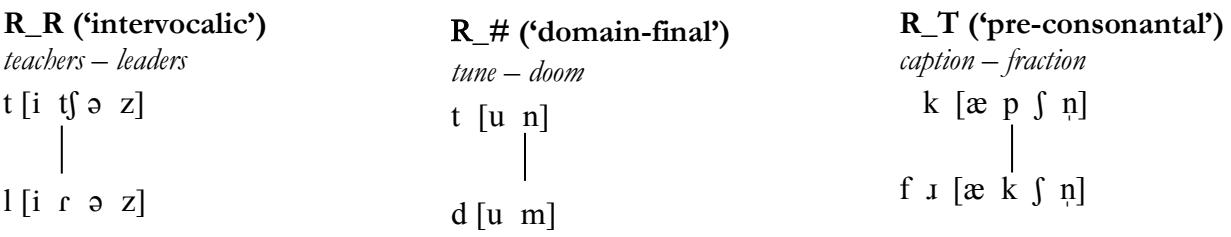

Figure 1. Imperfect rhyme in three contexts (left to right): in between two vowels, glides, or approximants; following a vowel and preceding a rhyme-domain boundary; and following a vowel and preceding a non-sonorant consonant. Brackets indicate rhyme domains. Vertical lines indicate imperfect rhyme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

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. The contrastiveness of the three features is also comparable, because English allows them to contrast in (almost) all of these contexts. Differences in rhyme likelihood across contexts thus cannot be explained by phonological features alone. In contrast, the phonetic correlates of these features, and hence their perceptual distinctiveness, do differ across contexts, as explored in section 2.4.

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 present, and may be important in signaling rhymes. Adams (2009) gives an accessible and interesting overview of hip-hop's rhythmic properties across several historical stages. The sequence of musical beats is organized in stronger and weaker levels, like linguistic stress (see Lerdahl & Jackendoff 1983, Palmer & Krumhansl 1990 for discussion 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 conventions to those of song (although they may be violated more often 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 some 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 do, 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

[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 Sommers (ed.) 2011.

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.

	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.

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 (see, for instance, Adams' (2009) transcription of a 1984 Kurtis Blow recording). 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 (Adams 2009). Walser (1995), for instance, gives a detailed analysis of Chuck D's 1988 performance

in *Fight the Power*: this 'transitional' piece includes extensive mismatch between linguistic units and musical metrical ones, as well as certain types of non-line-final rhyme; Walser offers some observations on how this marks a departure from earlier pieces. Pihel's (1996) partial transcription of a Big L piece from the early 1990s suggests that non-line final rhymes are present, but that linguistic constituents are broadly constrained to align with musical metrical landmarks.

In the corpus considered here, musical rhythmic units have some tendency to align with linguistic constituents, but mismatch between the two types of constituent is fairly frequent as well. Although rhymes generally do occur at some more or less predictable rhythmic interval, they are not constrained to appear only in this position (Adams 2009, Horn 2010). 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.

(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).

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 researchers 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 not very distinct, 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 used formal statistical modeling to argue 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. Kawahara's studies thus overlap with the current one regarding the role of perceptual knowledge in verbal art, but Kawahara's work is not explicitly concerned with perceptually-driven phonological asymmetries.

Stausland 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. This could be interpreted as more evidence

[2] Although there are websites that explicitly discuss rhyming practice with the goal of educating aspiring hip-hop artists (e.g. http://z13.invisionfree.com/Rhyme_Schemes/ar/t93.htm), they arose long after the careers of the artists examined here had begun.

that perceptually similar rhymes are more common, given the additional assumption that better rhymes occur more often (an assumption that seems well-supported by the research mentioned above and will be confirmed again in this study).

The current study, then, attempts to replicate and extend several findings from this previous literature. We use regression models to characterize rhyme frequency 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 displays frequent and often phonetically-distant imperfect rhymes, allows us to examine a wide variety of features. 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 *pu<u>b</u>* and *pu<u>p</u>*, we say that that feature contrasts; when the feature cannot distinguish between lexical items, as in *la<u>p</u>se* and hypothetical **la<u>b</u>se*, 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).

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. 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).

[3] Some dialects of both languages have a marginal second nasal /m/ appearing only in loanwords.

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 obstruents 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 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).

(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-tiːŋ-tiːŋ/	→	natiːntiːŋ 'He cut (it) through'	* natiːŋtiːŋ
/fan-fan/	→	faɱfan '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.

[4] In some dialects the word-final nasal is velar and the coronal variant is absent.

The examples of place neutralization given here all involve nasals, which are especially prone to such neutralization cross-linguistically. Oral stops also undergo place neutralization in some languages, although less frequently than nasals (Jun 1995). Before stops, this can result in the presence of geminates and absence of heterorganic stop clusters, as in Italian (Bertinetto & Loparcaro 2005); or in debuccalization, as in Arbore (Harris 1990) and Tiriyó (Parker 2001). Stops debuccalize to [h] in both word-final and pre-consonantal position in Slavey (Rice 1989). As is the case with nasals, we are not aware of a language that allows major place contrasts for oral stops before non-sonorant consonants but neutralizes those contrasts word finally.

Voicing and place thus display similar contextual profiles, summarized in table 3.

(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.

Several researchers have proposed that the nature of these typological implications follows from the perceptual properties of segments in the contexts under discussion (Ohala 1990a, Jun 1995, Steriade 1999, 2001). More generally, neutralization is influenced by speech perception: contrasts tend to neutralize in positions where they are less distinct (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 (see Wright 2004 for an overview). 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), closure duration, and amplitude of the burst; these properties are often obscured by the closure of a following non-sonorant consonant. To the extent that non-stop consonants are coarticulated with a following consonant, their inherent place cues should also be weakened in this context relative to one without a following consonant. Place contrasts for nasals are likely to be more difficult to perceive than those for oral stops, because nasalization in adjacent vowels will interfere with the perception of formant transitions that cue place (Jun 1995).

Based on these theoretical considerations, we expect various contextual asymmetries in perceptibility, mirroring the typological facts. Many but not all such asymmetries have been tested and confirmed in the perceptual literature. First, we have indirect tests of the asymmetries involving pre-vocalic and non-prevocalic consonants: listeners attend to both place (Fujimura *et al.* 1978, Ohala 1990a) and voicing (Raphael 1981) cues in a following vowel more closely than those in a preceding vowel, as indicated by their categorization of cross-spliced stimuli. For instance, in Raphael's (1981) study, a following vowel cuing voicelessness perceptually 'outweighs' a preceding VC sequence cuing voicing, resulting in more 'voiceless' responses than voiced.

Slightly more direct (though less controlled) tests of these asymmetries come from confusability studies. Wang and Bilger (1973), for instance, find that both voicing and place are less successfully transmitted in final position than they are in initial pre-vocalic position, based on English speakers' consonant identification in noise. Based on analysis of the raw confusion matrices from a similar experiment presented by Woods *et al.* (2010), pairwise similarity measures (Luce 1963, Shepard 1972) are higher for final than for initial pairs of stops and nasals mismatching for place. Similarity is higher for final than initial pairs of stops mismatching for voice in this data, but evidence is mixed for fricatives. Similar results obtain for place contrasts in Dutch, based on analysis of the raw confusion data presented by Pols (1983). Note that tests of the initial vs. final asymmetry for voicing contrasts using English utterances are overly conservative, because the final voicing contrast in this language correlates with a massive difference in preceding vowel duration, the magnitude of which is unusually large from a cross-linguistic perspective (Chen 1970, Mack 1982).

Fewer studies have examined (or can be adopted to examine) asymmetries between final and preconsonantal consonants. Kawahara & Garvey (2014) provide evidence from an identification in noise paradigm that English place contrasts for nasals and stops are more confusable in pre-stop than word-final position. Comparing the two experiments reported in Kochetov & So (2007), one with final stops at the ends of isolated words and the other with a following consonant-initial context, suggests the same conclusion. For voicing, there are not many languages that allow morpheme-internal voicing contrasts before obstruents *and* are widely spoken in places where many phoneticians work. The closest materials we can get in English involve word junctures (e.g. *a peg shorter* vs. *a peck shorter*); these figure in Raphael's (1981) study, although it is not specifically concerned with the relative perceptibility of contrasts across contexts. The results nonetheless show

that speakers are more likely to misidentify *peg* as *peck* when followed by [ʃ] than when presented in isolation; the results also show that replacing the following context (the word *shorter* in this case) with a token of the same string which had originally followed a voiceless stop makes no difference to identification, confirming that subjects are getting little or no information about voicing from the following context in such cases.

The perceptual literature also contains much support for the hypothesis that place distinctions are more perceptible in oral stops than they are in nasals. Kawahara & Garvey (2014) test this directly for English speakers with both similarity judgments and identification in noise: nasal place is less perceptible than oral-stop place in both domain-final and preconsonantal position. Similarity measures derived from raw confusion matrices in English (Woods *et al.* 2010) and Dutch (Pols 1983) show that nasals differing in place are more confusable than stops differing in place in both initial prevocalic and domain-final position. The same is true of the comparison between prevocalic voiced oral stops and nasals in the confusion matrices presented by Miller & Nicely (1955), but not for initial (presumably aspirated) voiceless stops; this may be due to the fact that the stimuli were masked with noise, which could have a disproportionate effect on place cues that are themselves largely contained in (aspiration) noise.

There is thus broad agreement amongst phonologists that contexts with less perceptible voicing and place contrasts tend to be the same contexts with fewer available contrasts across languages. It is not the case, however, that all of the researchers mentioned above agree on *how* or *why* perception and phonology are linked in this way. We can broadly distinguish two kinds of views on the subject: Ohala (1975 *et seq.*) and Blevins (2004) argue that the reason phonology reflects perception is that speakers acquire their language through perception, and more distinct contrasts are more likely to be

retained in the iterative process of language acquisition across generations. On this view, speech perception affects phonology during the course of learning but the mental grammar that is eventually learned need not make any reference to perception at all. Flemming (1995), Jun (1995), and Steriade (1999), on the other hand, argue that phonology is optimized for speech perception because the mental grammar includes constraints on sound correspondences and/or contrasts that specifically reference perceptual properties. On this view, language learners possess fine-grained implicit knowledge about speech perception and use it to constrain their search for a grammar. This debate is relevant to the rhyme data analyzed here because the two approaches entail rather different views about perceptual knowledge, to be discussed in more detail in section 5.

With regard to the current study, the hypothesis that rhyme likelihood reflects perceptual similarity makes several predictions given the perceptual asymmetries described above. One prediction 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. A second prediction is that place mismatches for nasals should be more likely than for oral stops across all positions, mirroring the typological facts. These are the effects that will be tested in the rhyme database 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.

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

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.

All rhymes were transcribed by the author 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 with regard to the sequence of rapped syllables according to empirical music theory (Lerdahl & Jackendoff 1983, Deliege 1987) as occurring at an inter-onset interval (defined in terms of musical beats) that is longer than the surrounding ones. In a sequence of syllables ('musical events') *e1e2e3e4*,

for instance, a group boundary occurs between *e2* and *e3* if and only if the amount of time (measured in musical beats rather than ms here) between the onset of *e2* and *e3* is greater than the amounts of time in between both the onsets of *e1* and *e2* and the onsets of *e3* and *e4*. This criterion essentially declares that a musical event longer than the surrounding musical events is group-final. Stressed vowels were defined disjunctively as: (a) 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 (b) those bearing a pitch accent. The (b) clause is mainly to deal with the fact that [i] and [o] appear in both stressed and unstressed syllables.

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 transcribed (assuming a more or less typical phonetic implementation for this dialect) as [(tʃ)ɪlənəŋkən][(t)ʌki][(fɹ)ɑd][(tʃ)ɪkən]. The brackets here represent rhythmic groups. 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, especially in cases where the syllabic consonant is a vocoid or is absent in this dialect. The current study omits unstressed-syllable rime data for independent reasons, so these issues do not need to be resolved here. Note that realizations 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 random consonant correspondences, that is, noise. We discuss this at length in section 3.3. Note that there is no objective way to reject *finger* as a rhyme correspondent without appealing to some notion of (ostensibly perceptual) similarity; if we wish to argue from the database that perception affects rhyme likelihood, any such exclusion criterion would result in circularity. For this reason, we use inclusion criteria based only on syllable count and vowel quality, with no reference to consonantal properties: the procedure is virtually guaranteed to result in noisier data than intuition-based coding, but this is unavoidable given the hypotheses that are being tested.

The rhyme correspondent pairs extracted from the corpus were decomposed into individual segmental correspondences, which included both perfect and imperfect 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 ('intervocalic'), where both corresponding segments are flanked by vowels, glides, or liquids; R__# ('domain-final'), where both correspondents follow a vowel, glide, or liquid and precede a rhyme-domain boundary; and R__T ('pre-consonantal'), where both correspondents follow a vowel, glide, or liquid and precede a stop, fricative, or nasal. Because data was quite sparse for intervocalic and pre-consonantal contexts in unstressed positions, the statistical model is limited to consonants following stressed vowels.

The pre-consonantal 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. Contexts where identical sequences would be an issue were defined separately for each set of obstruents: the contexts differ according to the voicing and continuancy of the first consonant in the sequence, so for instance we would exclude voiceless stops that occur before other voiceless stops, voiced stops occurring before voiced stops, etc. 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 minimal mismatch is impossible here. Excluding this data would result in a near total lack of obstruents in pre-consonantal position in the database, so they were kept in the analysis. We predicted that voicing should be more likely to mismatch in pre-consonantal position than other contexts. If the impossibility of minimal voicing mismatch in this context affects rhyme likelihood, the effect should go against the experimental hypothesis. This means that the database will result in a conservative test of contextual hypotheses.

Segmental correspondence data were examined for the segments /p, t, k, b, d, g, f, s, v, z/ with regard to voicing and place, and separately for /p, t, k, b, d, g, m, n, ŋ/ with regard to the relative likelihood of place mismatch in obstruents and nasals. These segments constitute the closest thing that English allows to an exhaustive crossing of the phonological features [voice], [continuant], [nasal], and major place; they are also reasonably frequent in most contexts in the corpus. There were 1,270 unambiguous correspondences for these segments included in the database: a breakdown by context and feature mismatch is shown in table 5.

Context	Total correspondences	Place mismatches	Voicing mismatches
R_R	457	143	79
R_#	629	230	106
R_T	184	71	46

Table 5. Correspondence counts, place and voicing mismatches by context.

3.3 False positives and noise

As outlined in section 3.2, the rhyme algorithm used here introduces some false positives into the data. Here we briefly investigate the nature of the noise introduced by false positive rhymes. We take advantage of an earlier version of this project that used a different corpus, with rhymes coded by listener intuition (Katz 2010). Two songs are included in both that earlier corpus and the present one, and we compare the coding of those two songs (referred to as 'mutual songs' in what follows) in the two corpora.

Note that *any* method of coding this corpus for rhymes is problematic. Using listener intuitions, even if validated across multiple listeners, would be circular given that we are using the corpus as evidence for propositions pertaining to perceptual similarity. It would also plausibly introduce false positives and miss true positives, if the transmission of rhyme from composer to listener is less than perfect. Using the method of the research discussed in section 2.3, which generally only counts as rhymes units that occur in certain structural positions (e.g. line-final), would result in a large proportion of missed true positives if applied to this corpus, due to the frequency of non-line-final rhyme. Using automatic detection algorithms will result in both misses and false alarms, given that we don't have a perfect understanding of where rhyme is licensed in this genre. The relative risk of misses and false alarms will depend on the stringency of the criteria used for the algorithm: the more stringent the definition of 'rhyme', the greater the ratio of misses to false alarms. We have chosen the algorithmic approach here, and used an extremely liberal definition of 'rhyme', because out of the possibilities just described, this is the only one that is guaranteed to result in more false alarms than misses. Not having enough data to fit a model would be catastrophic for this project, whereas having false positives in the data may be overcome, as we attempt to show directly.

Out of the 222 rhymes coded from the mutual songs in the current corpus, 61 (27%) do not appear in the older corpus. This figure overestimates the proportion of noise in the data in two ways. First, a non-trivial portion of the 'false positives' here are actually plausible rhymes that we missed during our intuition-based transcription; in these cases, the inclusive algorithm may do *better* than our intuition (this illustrates one of the pitfalls of using intuition to code such a corpus). The second reason this figure overestimates the amount of noise in the data has to do with the treatment of different-length strings of consonants. If two strings were marked as rhyming but contained different numbers of consonants in the relevant positions, those consonants were excluded from the

final analysis for reasons explained in section 3.2. If 'real' rhyme domains are constrained to be similar, then they should match for number of segments more often than randomly selected strings. In the sample here, 29% of rhymes transcribed in both corpora ('hits') mismatch for number of consonants, while 44% of false positives do. This means that the false positives would be less likely to generate segmental correspondences that would appear in the final data. Incorporating this factor, the upper bound for the proportion of false positive rhymes in the corpus is estimated at 23%. Note that this is a high estimate, because it assumes that anything not included in the intuition-based corpus is 'noise'; in reality, a non-trivial portion of these data are probably 'signal'.

If around 20% of the data here is actually noise, one might wonder whether that noise biases the conclusions in one way or the other. We have two circumstantial pieces of evidence that it does not. First, we analyzed all of the (small amount of) false-positive segmental correspondences from the mutual songs. With one exception, this small sample had place and voicing mismatches roughly evenly distributed across the three contexts investigated here: 35-40% of segment pairs failed to match for voicing; 60-65% failed to match for major place, across all contexts. The exception is that this sample appeared to undergenerate place mismatches in domain-final position; only 35% of pairs mismatched for place here.

The second piece of evidence that false positives are not unduly influencing the results here comes from comparing the results here to the earlier study mentioned above. Those results that can be compared across the two studies are largely the same. The earlier study investigated voicing of coronal consonants, for instance, and found that voicing mismatches are more likely in pre-consonantal position than they are in word-final or intervocalic position; the current study reports

similar results. We thus tentatively conclude that our results are unlikely to be an artifact of unevenly distributed noise in the data.

3.4 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 and their independent frequencies, 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. It 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.

BCM distance data for non-matching segments (recall, matching segments are used to define distance 0) were subjected to linear mixed-effects regression analysis using version 1.1-7 of the lme4 package in R (Bates *et al.* 2014). 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 the parameter increases rhyme distance. 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 models were phonological featural mismatch, context, and mismatch x context interactions. Artist identity was a random effect. The features used for the first model, examining obstruents, are [voice], [continuant], and major place; the second model examined nasals and (voiced and voiceless) oral stops. In the first model, featural mismatches were dummy-coded with place mismatch set as the baseline and all other mismatches compared to place. For the second model, difference-coded (scalar) fixed effects compared nasal place mismatch to oral-stop place mismatch, and compared oral-stop place mismatch to all other kinds of mismatch. Contexts were

difference coded along the scale intervocalic < domain-final < pre-consonantal. Dummy coding is a way of comparing two or more categorical predictor variables to each other. In the first model, for instance, the intercept term corresponds to the rhyme distance of place mismatches in intervocalic position, and all other combinations of feature mismatches and contexts are assigned some unique set of values for contextual and featural variables. Difference coding is a way to test the significance of steps along a hypothesized scale.

The first, more complex 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. For instance, how much more or less likely is major place to mismatch ($d > 0$) than to not mismatch ($d = 0$) in intervocalic 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 in this regard. 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.

The significance of fixed effects was assessed by examining the *t*-statistics returned by the model. Values larger than 2 or smaller than -2 indicate a probability of type-I error smaller than 5%. Values close to the significance criterion were double-checked by a likelihood-ratio test. The significance of random effects was assessed using a likelihood-ratio test. We follow Baayen *et al.* (2008) in testing all fixed effects in the models for by-subjects slopes, but only retaining significant by-subjects terms in the final model.

4 Results

4.1 Obstruent voicing and place by context

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).

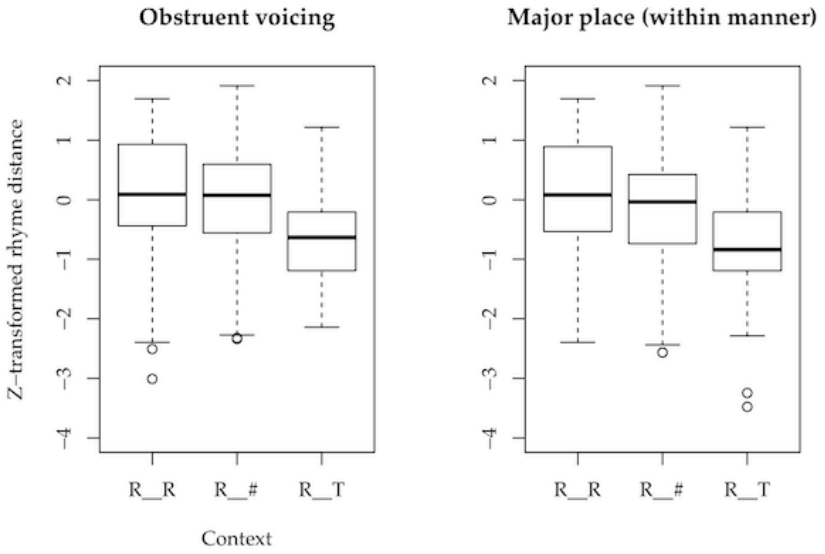

Figure 5. Rhyme distance associated with obstruent voicing mismatch (left panel) and major place mismatch within the classes of obstruents and nasals (right panel) in three contexts. Vertical axis shows BCM *d* measures subjected to a by-subject Z transform for comparison across subjects. Dark line indicates median, boxes indicate inter-quartile range, whiskers indicate range up to 1.5 times inter-quartile range, open circles indicate potential outliers.

The first statistical model estimates the independent contribution of various features to rhyme distance between obstruents; results are shown in table 6. 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 /s/ and /b/ corresponding is modeled as the sum of the likelihood associated with each feature in which they mismatch: [voice], [continuant], and place.

Effect no.	Distance for mismatch in	compared to	β	S.Err.	t	sig.
1	**place in R_R**	0 (match)	6.20	0.96	6.48	*
2	**[voice] in R_R**	place in R_R	0.62	0.42	1.47	
3	**[cont] in R_R**	place in R_R	-0.11	0.42	-0.26	
4	**place in R_#**	place in R_R	-1.71	0.80	-2.15	*
5	**[voice] in R_# vs. R_R**	place in R_# vs. R_R	1.04	0.63	1.66	
6	**[cont] in R_# vs. R_R**	place in R_# vs. R_R	1.43	0.62	2.28	*
7	**place in R_T**	place in R_#	-1.14	0.62	-1.83	?
8	**[voice] in R_T vs. R_#**	place in R_T vs. R_#	-1.26	0.66	-1.93	?
9	**[cont] in R_T vs. R_#**	place in R_T vs. R_#	-1.37	0.65	-2.10	*

Table 6. 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, and the *t* statistic associated with the effect.

The rhyme distance of place mismatches is significantly smaller in domain-final position (R_#) than in intervocalic position (R_R, effect 4); there is a trend for distance to be smaller in pre-consonantal position (R_T) than word-final, which is close to significance (effect 7). A one-tailed *t*-test on this value returns a 3.4% probability of type I error, but a likelihood ratio test gives a larger value: $\chi^2 = 3.36$ on 1 df; $p = 0.067$.

The rhyme distance of voicing mismatches is not significantly different from place in intervocalic position (effect 2), and this (lack of) difference does not interact with domain-final position (effect 5). The difference between domain-final and pre-consonantal positions, however, is a fair bit larger for voicing than for major place (effect 8). A likelihood ratio test suggests this is marginally significant: $\chi^2 = 3.73$ on 1 df; $p = 0.054$.

We made no particular predictions about how continuancy would pattern in the corpus, but it was included in this model because it is a feature that varies amongst the segments examined here and should thus be accounted for. Its rhyme distance is not significantly different from place in intervocalic position (effect 3), but this (lack of a) difference does change in the other two contexts: distance of continuancy mismatches is somewhat larger than that of place in domain-final position (effect 6) and somewhat smaller before consonants (effect 9).

One by-subject random slope was found to significantly improve model fit. The amount by which featural mismatch in general was more likely to occur in word-final position than in intervocalic position differed in magnitude (though not in its direction) for different subjects: $\chi^2 = 16.8$ on 2 df; $p < 0.001$.

4.2 Place in oral stops vs. nasal stops

Rhyme distance parameters for place mismatch in oral stops and nasals are shown in figure 6. The data display increased rhyme likelihood (decreased rhyme distance) for nasals, which are more likely to neutralize for place cross-linguistically.

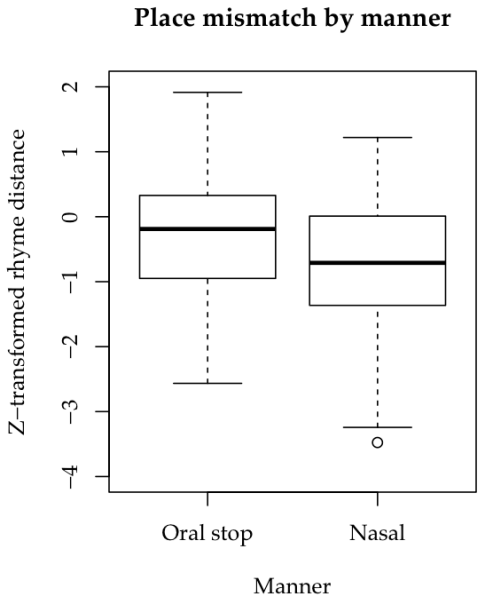

Figure 6. Z-transformed rhyme distance associated with place mismatch for oral stops (left) and nasals (right), pooled across contexts.

The second statistical model estimates the relative contribution to rhyme distance of place mismatch in oral stops and place mismatch in nasal stops; results are shown in table 7.

Effect no.	Distance for mismatch in	compared to	β	S.Err.	t	sig.
1	**R_R position**	0 (match)	6.83	0.93	7.38	*
2	**place for oral stops**	other mismatches	0.25	0.33	0.75	
3	**place for nasals**	place for oral stops	-2.51	0.53	-4.69	*
4	**R_# position**	R_R position	-0.09	0.61	-0.15	
5	**R_T position**	R_# position	-3.47	0.36	-9.70	*

Table 7. Statistical model of rhyme distance for place mismatches amongst oral and nasal stops.

Rhyme distance of place mismatches is significantly smaller for nasals than for stops (effect 3). The difference between oral and nasal stops was tested for interactions by position: none were significant, and these interactions were not retained in the final model.

There is no significant difference between the rhyme distance of place mismatches for oral stops and the other mismatches contained in the data set (voicing, nasality, and place in nasal-oral pairs; effect 2). Average rhyme distance (across all mismatches) in domain-final position is not significantly different from that in intervocalic position (effect 4). Average rhyme distance in pre-consonantal position is significantly less than in domain-final position (effect 5).

One by-subject random slope was found to significantly improve model fit. The magnitude and direction of the difference between rhyme distances in the intervocalic and domain-final contexts differed for different subjects: $\chi^2 = 12.0$ on 2 df; $p < 0.01$.

5 Discussion

The findings from the rhyme database broadly support the hypothesis put forth in section 1 that rhyme mismatch for a feature is more likely in contexts where that feature is more likely to phonologically neutralize cross-linguistically. Major place is more likely to mismatch domain finally than intervocalically, and voicing does not differ significantly from place in this regard. Both features are more likely to mismatch before non-approximant consonants than they are domain finally, though this difference is statistically marginal for place. And nasals are more likely to neutralize for place than oral stops. All of these results parallel the perceptual and typological facts discussed in section 2.4.

The statistical significance of the comparison between place mismatch in domain-final and pre-consonantal positions deserves some comment. Unlike general linear models, the mixed effects

models used here are not associated with universally-accepted procedures for determining the probability of type I error. As the hypothesis regarding this particular effect is directional ('place will mismatch *more often* in pre-consonantal than domain-final position'), one way to estimate this probability is by performing a one-tailed test on the *t*-statistic associated with the effect's coefficient β and the estimated standard error for this coefficient. This test returns an estimate of $p = 0.034$ in this case. A reviewer, however, notes that this method may be anticonservative for these models, and suggests a likelihood-ratio test instead. This test measures the improvement in model fit due to a particular effect, and is non-directional; in this case, it returns an estimate of $p = 0.067$. It is not clear which estimate is closer to the truth, and unfortunately the two fall on opposite sides of the traditional (and arbitrary) significance criterion of 5%. What we can say for certain is that the positional asymmetry for obstruent place goes in the predicted direction, is associated with a *p*-value somewhere in the 3-7% range, and is not as large as the one for voicing (as indicated by the near-significant interaction between feature and context reported in section 4.1). This seems broadly consistent with the experimental hypothesis.

As to why this effect is less robust than others discussed here, two post-hoc explanations spring to mind. First, there is somewhat less data in pre-consonantal position than the other two, due to consonant clusters being less common than singletons. It may be the case that parameter estimates are just less accurate in this context, although they do not appear to be more variable based on the estimated standard errors in the statistical model. Another possibility has to do with the existence of pseudo-neutralizing patterns in English phonology. In particular, we noted in section 3.2 that the impossibility of identical obstruent sequences entails *de facto* partial place neutralization in obstruent clusters. This means that changing the place of a pre-obstruent obstruent will sometimes be

unavailable as a minimal mismatch in rhymes. Given that English makes such minimal mismatches impossible, the 'leftover' possible correspondences may not be quite as frequent as expected on the basis of other contexts and features, where such considerations do not come into play.

The findings also bear on specific questions that arise in the literature on rhyme discussed in section 2.4. First, features are more likely to mismatch in some contexts than others. This means that the data are not consistent with a model where rhyme likelihood is determined only by the number of matching phonological features. There are more sophisticated ways of quantifying feature-based similarity, involving feature-weighting, shared natural classes, or contrastivity (e.g. Frisch *et al.* 2004). These data, however, show that the likelihood of mismatch for a given feature can vary by context. Because the features examined here are by hypothesis the same (in contrastive, phonological terms) in all contexts, this suggests that something above and beyond phonological features plays a role in determining rhyme likelihood. For 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 domain-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). We call these two explanations the *confusion* approach and the *optimization* approach, respectively. Note that they are not mutually exclusive; both factors may well influence typology.

One of the principal objections to optimization approaches (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.

A proponent of the confusion approach might counter that the argument from parsimony still holds despite a speaker's knowledge of perceptual similarity.[5] The reasoning is that describing the mental grammar as making use of perceptual optimization is still more complicated than describing it as not using such optimization, even if speakers clearly possess the means for optimizing. We contend that the truth of this statement depends on what a theory uses *instead* of perceptually-optimized constraints. The alternative is not 'nothing': any theory needs to explain how generalizations and alternations exist in the mental grammar and how they come to be there. In the confusion approach, whatever such knowledge is posited (analogy, purely symbolic rules, etc.) will be quite different from the constraints that govern rhyme. As such, we believe that the choice is between describing phonology and rhyme as using fairly similar principles as opposed to entirely different kinds of principles. Put in a different way, the confusion approach views confusability and perceptual 'optimization' in phonology as being due to the same principles, while perceptual optimization in rhyme is due to a different set of principles. The optimization approach views perceptual optimization in phonology and rhyme as subject to the same type of constraints, while confusability itself provides the basis for these constraints but is not identical to them. Put in this way, it seems clear that parsimony cannot favor the confusion approach and may in fact favor optimization.

We attempted to use rhyme data here to draw conclusions about a speaker's phonetic knowledge. 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 are extremely useful for determining the facts of how similar or dissimilar various sounds are in various contexts. But they are not meant to show that speakers *know* (explicitly or implicitly) anything about the distinctiveness of linguistic contrasts, in the sense that they actively

[5] Many thanks to Sverre Stausland Johnsen for raising the issues discussed in this paragraph.

make use of fine-grained distinctions in this domain. To be clear, this is not to dismiss traditional speech perception studies; they have built a tremendously important body of knowledge that bears on foundational issues in phonetics and phonology. The suggestion is rather that the rhyme data analyzed here allow us to draw different kinds of conclusions about perceptual properties.

Of course, this 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 clearly 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 rules 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 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 in the data 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.

References

Adams, K. (2009). On the metrical techniques of flow in rap music. *Music Theory Online* **15**(4).

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.

Baayen, R.H., D. Davidson & D. Bates. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390-412.

Bach, E., & R. Harms. (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., M. Maechler, B. Bolker, S. Walker, R. Christensen, H. Singmann, & B. Dai. (2014). lme4. v. 1.1-7: Linear mixed-effects models using Eigen and S4. Software application.

Benua, L. (1997). *Transderivational identity: phonological relations between words.* PhD dissertation, UMass Amherst.

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.

Bertinetto, P.M. & M. Loporcaro. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* **35**(2), 131-151.

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

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* **22**, 129-159.

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.

Frisch, S. , J. Pierrehumbert, & M. Broe. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* **22**, 179-228.

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.

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.

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

Harris, John. (1990). Segmental Complexity and Phonological Government. *Phonology* **7**. 255-300.

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.

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.

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

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.

Katz, J. (2010). Phonetic similarity in an English hip-hop corpus. Presentation at the LSA Annual Meeting, Baltimore, January.

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.

Kawahara, S. & K. Garvey. (2014). Nasal place assimilation and the perceptibility of place contrasts. *Open Linguistics* **1**, 17-36.

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.

Kochetov, A. & C. So. (2007). Place assimilation and phonetic grounding: a crosslinguistic study. *Phonology* **24**, 397-432.

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.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: monolingual vs. bilingual production. *Journal of the Acoustical Society of America* **71**, 173-178.

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

Miller, G. & P. Nicely. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* **27**(2), 338-352.

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

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

Parker, Steve. (2001). On the Phonemic Status of [h] in Tiriyó. *IJAL* **67**. 105-118.

Pihel, E. (1996). A furified freestyle: Homer and hip-hop. *Oral Tradition* **11**(2), 249-269.

Pols, L. (1983). Three-mode principal component analysis of confusion matrices, based on the identification of dutch consonants, under various conditions of noise and reverberation. *Speech Communication* **2**, 275-293.

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.

Raphael, L. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica* **38**, 126-147.

Rice, Keren. (1989). *A Grammar of Slave*. Berlin: Mouton de Gruyter.

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

Shepard, R. (1972). Psychological representation of speech sounds. In Davis & Denes (eds.), *Human Communication: A unified view*, 67-113. New York: McGraw-Hill.

Sommers, J. (2011). *Hip-Hop: A cultural odyssey*. Los Angeles: Aria Multimedia Entertainment.

Stallworthy, J. (1996). Versification. In Ferguson, Salter, & Stallworthy (eds.) *The Norton Anthology of Poetry*, 2027-2052. New York: W.W. Norton.

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

Stausland 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.

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

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.

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.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, Mass.: MIT Press.

Walser, R. (1995). Rhythm, rhyme, and rhetoric in the music of Public Enemy. *Ethnomusicology* **39**(2), 193-217.

Wang, M. & R. Bilger. (1973). Consonant confusions in noise: a study of perceptual features. *Journal of the Acoustical Society of America* **54**(5), 1248-1266.

Woods, D., E. Yund, T. Herron & M. Cruadhlaoich. (2010). Consonant identification in CVC syllables in speech-spectrum noise. *Journal of the Acoustical Society of America* **127**(3), 1609-1623.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, D. Steriade, and R. Kirchner (eds.), *Phonetically Based Phonology*, 34-57. Cambridge, UK: Cambridge University Press.

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.

Appendix

Raw correspondence counts by context, pooled across subjects.

R_R	p	t	k	b	d	g	m	n	ŋ	f	s	v	z
p	18	9	15	3	1	1	3	1	0	2	5	3	0
t		33	13	1	30	2	5	13	0	3	2	13	5
k			34	2	2	5	2	2	0	4	2	3	1
b				9	2	2	0	2	0	1	1	4	1
d					22	2	3	3	1	3	1	3	4
g						18	1	0	1	0	2	1	1
m							9	13	0	0	1	1	0
n								37	3	0	0	3	0
ŋ									2	0	1	0	0
f										7	3	3	1
s											10	3	4
v												19	3
z													8

R_R context

R_#	p	t	k	b	d	g	m	n	ŋ	f	s	v	z
p	13	28	16	2	4	4	2	0	0	5	1	2	3
t		78	37	0	18	1	0	8	0	18	10	2	3
k			53	0	5	3	5	5	2	5	7	4	4
b				1	1	0	0	2	0	0	0	1	0
d					24	0	1	2	0	1	2	2	4
g						0	1	1	1	1	1	3	1
m							29	36	2	0	0	2	2
n								40	7	1	8	5	3
ŋ									8	0	0	0	0
f										11	9	2	0
s											24	3	10
v												2	1
z													26

R_# context

R_T	p	t	k	b	d	g	m	n	ŋ	f	s	v	z
p	7	6	6	0	0	0	1	1	2	0	2	1	0
t		9	5	0	3	1	1	3	0	1	3	1	1
k			15	0	1	1	2	1	1	0	9	1	1
b				0	0	0	0	0	0	0	1	0	0
d					5	0	0	1	0	1	4	0	0
g						1	0	0	0	0	1	0	0
m							0	11	0	0	2	0	0
n								9	3	3	7	1	3
ŋ									2	0	0	0	0
f										1	4	1	0
s											31	3	0
v												2	0
z													1

R_T context