Phonotactic Probability Predicts Perception of Rhyme?

Robert Kimelman and Nat Condit-Schultz

1 Introduction

Rhyme is a ubiquitous feature in musical lyrics, playing a particularly important role in hip hop. Though folk definitions of rhyme are pervasive—being introduced to children at a young age—there has been little coherent scientific theory or empirical research published, especially in a musical context. The function rhyme serves in musical and poetic artistry is not fully understood. Rhyme has often been noted as a memory aid in poetic traditions [cite something], but evidence of its effectiveness as a (pneumonic?) mnemonic is ambiguous [cite]. Rhyme is often used in predictable "schemes", thus eliciting obvious anticipatory qualia. However, it is not clear why "rhymes" specifically are suited to serving these roles. What's more, rhyme is known to elicit a distinct phenomenological experience even when unpredictable—a unexpected rhyme often "pops out" when you hear it. The question remains: *why* is rhyme so powerful?

In this project, we propose a novel theory regarding the perception of rhyme, based in statistical learning. We propose that the perceptual quality of rhyme is provoked by improbable phonemic coincidences: During normal speech, the specific speech *sounds* that occur are essentially random, as their mapping to semantic meaning is arbitrary. For example,... Thus, we hypothesize that the salience of rhymes can be directly predicted from the phonostatistics of the language.

We conducted a behavior experiment to test our theory. We focus on the form of rhyme known as *assonance* (vowel rhyme) since the prevalence and sequence of vowels is simple to model statistically. Participants listened to random pseudo-English stimuli, with the vowel content manipulated to create more or less probable utterances.

1.1 Literature Review

There is substantial evidence for the validity of statistical learning (Huron, 2006);

Cite: people are sensitivity to phonological details of native language. —We can recognize pseudo words that are "English-like" or not. Phonostatistics play a major role in language acquisition, and the perception of syllable boundaries.

There have been few studies that have explored how phonological information structure could be applied to rhyme perception (Wagner & McCurdy, 2010).

2 Methodology

2.1 A Statistical Model of Rhyme

Obviously, listeners do not consciously model the probabilities of utterances. Rather, we presume that some (unconscious) mental processes respond to and anticipate linguistic linguistic utterances in a manner which can be approximated by a statistical analysis. How exactly that analysis takes place is not obvious. By focusing on vowels alone (assonance) we simplify the task. Ignoring, meaning we can think of the vowels that appear in spoken utterances as random draws from a a languages' vowel inventory—though some vowels are overall more frequent than others. If we look at the distribution of vowels in corpora of spoken English, we can see that this model looks plausible: Take the XXX corpus of spoken English, which includes vowel pronunciation and sentence/phrase boundaries. The overall distribution of vowels in the corpus is shown in Table XXX. If we take this distribution as representing a multinomial distribution, we can compute a probablity density for any particular sample of vowels. If we take each of the xxx utterances in the corpus and calculate the associated probability (p value), we see that their distribution closely matches XXX the theory: $p \leq .05$ occur approximately 5% of the time; $p \geq .90$ occur approximately 10% of the time; etc. XXX

Given a stream of vowels in a spoken utterance, it is not clear exactly what constitutes a "sample" from the hypothetical multi-nomial distribution. How large of a time window is considered? If a speaker uses 95 "As" over the course of 2 minutes (approximately 2x the expected value) would this be salient? Though spoken, conversational, English rarely forms neat orderly "samples," artistic language *is* frequently divided into relatively discrete phrases or lines. Unfortunately, "relatively" is still fairly complex—identifying phrases in music is a subjective endeavor [cite].

We will model the phonotactic probabilities in the following way:

1. Given the lack of available data for variables 1 & 2, we determine the probability of a phrase using only 3. Also, it is known that the probability of one phoneme occurring is dependent on the previous phoneme, but this seems to only be true for words. Since we are looking specifically at the vowel sounds across different words, we can assume independence of each vowel sound. Note that if we included consonants in our model, the probabilities would be even lower! Given the probability distribution in (Vitevitch, 2004) (phonotactic probability calculator), we will determine the likelihood of a phrase as follows:

Let n = the number of vowel phonemes in a particular phrase where X is a discrete random variable with X = vowel phoneme type. Then the probability that a phrase with n vowel phonemes, each with a particular type X, will occur is:

$$P(X = x) = \prod_{i=1}^{n} p(x_i),$$

where $p(x_i)$ is the probability that $X = x_i$, where i denotes the position of phoneme x.

Let A be the first event (first sequence of phonemes occurring) and B be the second event (second sequence of phonemes occurring).

In this model of independence, the probability that the same vowel sound occurs again is just $P(A)$. Thus, the probability of assonance is

$$P(B = Rhyme \mid A) = P(A) = P(first\ sequence\ of\ vowel\ phonemes\ occurs).$$

The probability of a non-rhyme, however, would be:

$$P(B = Non - Rhyme \mid A) = 1 - P(A) = P(any\ other\ sequence\ of\ phonemes\ occurs)$$

since the probability of any other phoneme(s) occurring is the probability that assonance does not occur, i.e. the probability of its complement.

There is substantial evidence for the validity of statistical learning (Huron, 2006); to the best of our knowledge, there have been few studies that have explored how phonological information structure could be applied to rhyme perception (Wagner & McCurdy, 2010), but there have not been any that have tested the following theory: We propose that the perceptual qualia of rhyme is caused by the perceived improbability of utterances repeating phonemic content by chance.

3 Method

Our experimental stimuli consist of meaningless pseudo-English "utterances," spoken by an artificial speech generator. Each utterance was nine-syllables long. Three variables are manipulated: 1) length of assonant syllable sequences 2) position of assonant sequences in phrases and 3) assonant vowel phonemes. Multiple studies have confirmed a processing difference when subjects are presented with non-words as compared to real words (Goldstein, 2016). We aim to eliminate the effect of connotation in order to control for the effect of rhyme. We predict an inverse relationship between the phonostatistic likelihood of vowel sequences and the likelihood of a rhyme being perceived (Wagner & McCurdy, 2010).

Given that we are controlling for other aspects of classical poems (i.e., rhythm and meter), rhyme would be the most salient characteristic to discern when determining if the stimuli are poems (Tizhoosh & Dara, 2006; Tsur, 1996).

We focus on the form of rhyme known as *assonance* (vowel rhyme) since the prevalence of vowels and sequences of vowels is simple to model statistically. Our experimental stimuli consist of meaningless pseudo-English "utterances," spoken by an artificial speech generator.

Many studies (Wagner & McCurdy, 2010; Macmillan, 2002) have used explicit measures of rhyme, asking participants directly if they perceived a rhyme. However, the traditional definition of the "perfect rhyme" serves as explicit confound to participants' judgement of rhyme; thus, we utilize an implicit task: Participants were informed that the stimuli were either prose or poetry that had been transcribed, stripped of meaning, and then artificially generated; their task was to guess which utterances were originally poems.

We recruited participants (n = 24) through convenience sampling, and obtained a mix of different demographics (age 18 - 65, both African American English and American English dialects).

3.1 Generation of Stimuli

English pseudo-words were generated from the ARC Nonword Database (Rastle et al. 2002). Phonotactic probability was determined through the web-based Phonotactic Probability Calculator made by Vitevitch & Luce (2004). We mapped the vowel permutations to random words containing each specific vowel for our stimuli containing assonance. We filled the rest of the remaining slots with random words from the database. Assonance stimuli were generated such that almost half of all possible permutations of one, two, and three vowels were used in instances of assonance out of a total of eight possible vowel sounds (IPA Phoneme Representations): [ɑ, æ, ɐ, e, e ɪ, ɪ, iy, and u]. An additional highly unlikely vowel sound, er, was used in a handful of one vowel rhymes. The values for each of these permutations are:

$$\begin{aligned} {}^{8}P_1 &= \tfrac{8!}{(8-1)!} &&= 8 \\ +\ {}^{8}P_2 &= \tfrac{8!}{(8-2)!} &&= 56 \\ +\ {}^{8}P_3 &= \tfrac{8!}{(8-3)!} &&= 336 \\ &= 400\, total\ permutations \end{aligned}$$

Once these vowel permutations were determined, we mapped the corresponding vowel phonemes onto the non-word database such that words were generated in these permutations in lines of 9 words of varying rhyme type (internal, slant, and end). An example of a three vowel internal rhyme stimulus is pictured below:

fliz	plQk	Tr{b	spuvd	strQndZ	dwVp	skwi	SQld	z{nT

which, in plain English, translates to:

fleez pluhk trahb spooved strahnj dwup skwee suhld zahnt

Some real-world examples of assonance are detailed below:

One Vowel Internal Rhyme

“You only get one shot, do not miss your chance to blow.”

- Eminem, Lose Yourself

Three Vowel Internal Rhyme

“I’m a Young Money milli-on-aire, tougher than Nigeri-an - hair.”

- Lil Wayne, A Milli

Two Vowel Internal Rhyme and One Vowel End Rhyme

“Sleep with, keep the ep a secret, why not? (Uh)

Why blow up my spot ‘cause we both got hot?”

- The Notorious B.I.G., Big Poppa

By reading/hearing these examples, we can see how the more vowels that are repeated, the more noticeable they are, which makes sense given our probabilistic model.

We covered a large portion of the total vowel permutations with distribution of rhyme type taking first priority, while keeping the number of stimuli in each block constant and including more rhymes we predict to be more salient (namely, end and slant rhymes) in each block; the layout of each block with permutations considered is as follows:

Block 1: Block 1 stimuli consisted of 24 non-rhyming lines for a total of 20 stimuli (4 had two lines) and 14 two vowel end rhymes, 7 three vowel end rhymes, 22 three vowel internal rhymes, 8 one vowel end rhymes, 25 three vowel slant rhymes, and 4 one vowel slant rhymes (grand total permutations = 80 (one vowel rhymes included 4 vowels + additional); total stimuli = 100; total end rhymes = 29, total slant rhymes = 29, total internal rhymes = 22).

Block 2: Block 2 stimuli included 26 non-rhyming lines for a total of 20 stimuli (6 had two lines) and 4 one vowel slant rhymes, 14 three vowel internal rhymes, 8 one vowel internal rhymes, 25 three vowel slant rhymes, and 29 three vowel end rhymes (grand total permutations = 160; total stimuli = 100; total end rhymes = 29, total slant rhymes = 29, total internal rhymes = 22).

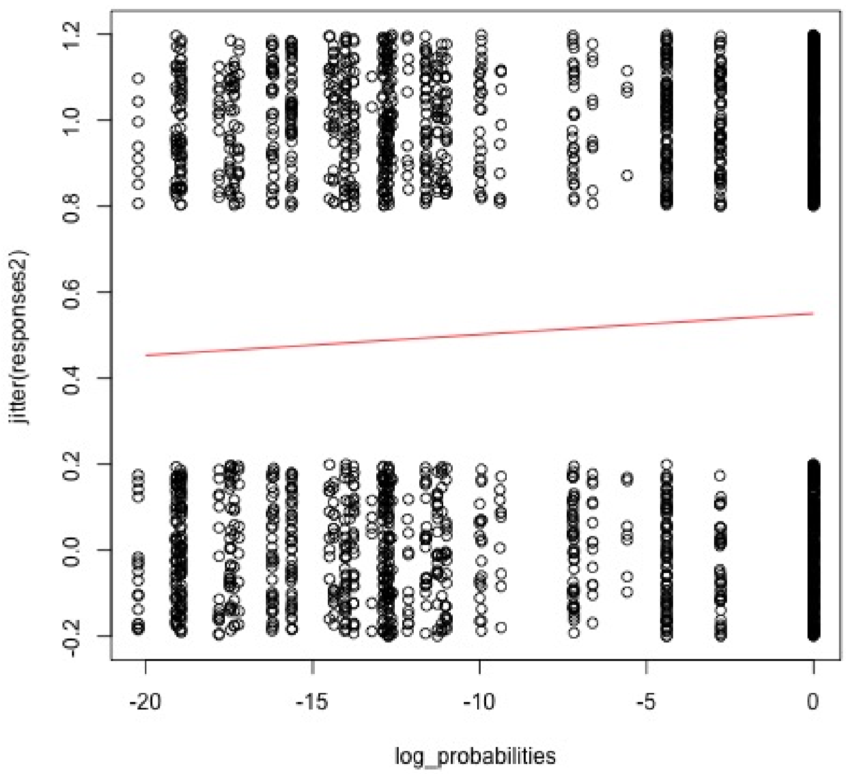
Therefore, we covered 40% of all possible vowel permutations in our rhymes. An additional vowel ”er” was used in the one vowel rhymes. Future studies may want to include the other 60% of permutations to ensure validity. (include the exact permutations we had, or potentially mention that we would be open to sharing them if asked for it?)

3.2 Interface

We used jsPsych to program our experiment and obtained permission to conduct our study from the Georgia Institute of Technology IRB Review Board.

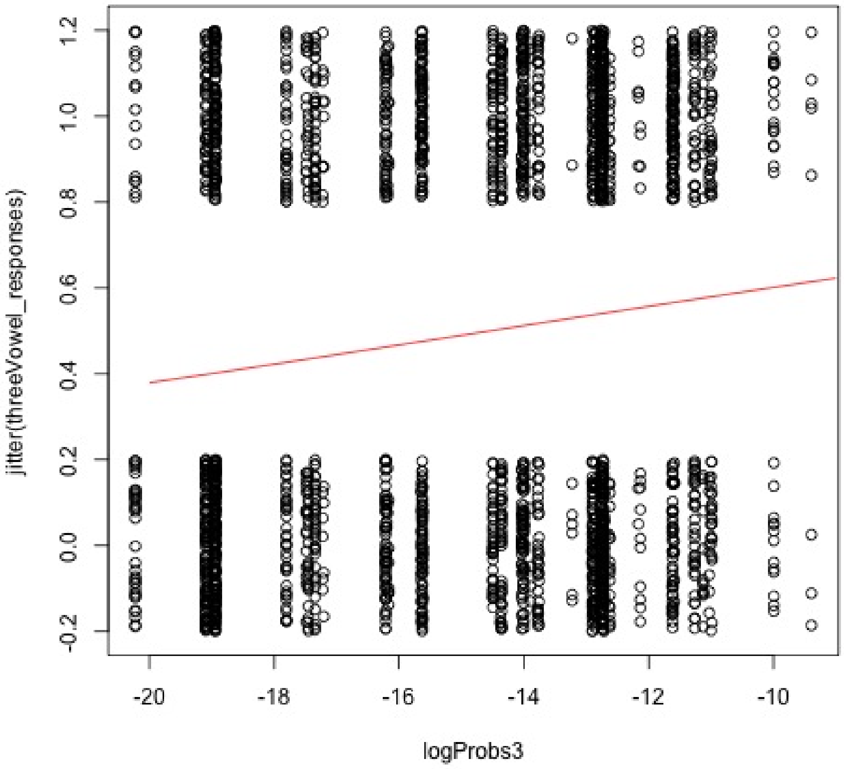
4 Results

We tested our theory by plotting log probabilities against response within all instances of assonance and running a logistic regression model. We found that the probability of a sequence of vowels occurring is a significant predictor of response ($p < 0.05$). The relationship was also in the predicted direction, with lower probabilities corresponding to more poetry responses. The scatter plot and logistic regression graph is shown below.



To further validate our theory, we ran a mixed effects model with number of vowels as a random intercept, this time including non-rhymes having probability 1. The log probabilities still significantly predicted rhyme perception ($p < 0.05$), and there was not a significant difference in the level of variance that each number of vowels accounted for in the perception of rhyme ($p > 0.05$).

We also observed how only instances of assonance with three vowels predicts rhyme perception and we see the same significant relationship again (we did not consider one and two vowel assonance because of the relative small sample of them used in our experiment due to the nature of the permutation calculations). This validates the theory further since the number of vowels was not varied and all probabilities were within a tighter range. (plot on next page)



5 Conclusion

Based on our results, there is evidence in support of the hypothesis, i.e. that there is an inverse relationship between the likelihood of a particular sequence of speech sounds and the likelihood of rhyme being perceived. Artists and poets can use this evidence to their advantage by incorporating more noticeable rhymes in particular places in their work to fit their artistic themes and messages. In future work, we hope to have more control over the speech characteristics of our stimuli by hiring a voice actor to speak each utterance individually. This will also serve as further validation of the theory if we obtain similar results.