Dynamically Scoring Rhymes with Phonetic Features and Sequence Alignment

Abstract

We present a formalized rhyme function for machine approximation of human rhyme. Words are represented as sequences of phonemic features that facilitate the use of alignment mechanisms to compute different types of phonemic similarities between words. The rhyme function computes a weighted hierarchical combination of these similarities, with the weights determined using an evolutionary approach. We present empirical and qualitative analyses that demonstrate the rhyme function's ability to successfully detect rhyme, and we briefly discuss the model's linguistic basis and its resulting generality.

1 Introduction

Rhyming words is a simple task for humans, but an involved one for machines. A machine may use a human-made corpus of rhymes, but this is a primitive way to approximate human rhyming. A concrete knowledge base is static; it is subject to human error and requires human labor to adapt to changes in language. If the problem of rhyme could be instead represented as a function R in which two words w_1 and w_2 are considered and a numeric value is returned, we might better approach it. Our goal is to decompose the concept of rhyme, and use its constituents to build a rhyme function that closely mimics rhyming in humans across all languages.

An important use case for rhyme is in creativity, a valued form of intelligence in humans. The device of rhyme looms large in songwriting and poetry as a means for creative expression. In natural language processing (NLP) contexts and text-based computationally creative (CC) systems, rhyme score constraints are necessary to automatically approximate human rhyme technique.

Given two words w_1 and w_2, our algorithm R outputs a score ranging from 0 to 1 as a representation of the fitness of the word pair as a rhyme. Perfect rhymes receive a score of 1. Partial rhymes receive some lower score. Given a score 0.3 and another 0.8, we assume the word pair corresponding to the latter score to be the better rhyme.

The concept of rhyme functions is not new. In 2009, Hirjee et al. presented a rhyme function based on a phoneme scoring matrix of likelihoods [Hirjee and Brown, 2010]. More recently, Hinton et al. of the Wall Street Journal built an algorithm that scored rhymes from the musical *Hamilton*, focusing on vowel phonemes, stress, and consonants following vowels (codas) [Hinton and Eastwood, 2015]. These are intelligent approaches because they require the decomposition of words into their sounds, phonemes, and they find patterns in which phonemes are commonly paired with each other.

Our rhyme function takes these concepts a few steps farther by

- defining general rhyme in relation to the stress tail (all syllables beginning from the nucleus of the greatest stress)
- considering all parts of a syllable (including the onset)
- using dynamic alignment to allow for words with multi-phonemic consonant sequences,
- decomposing phonemes into their basic parts, known as phonetic features. and
- drawing data from a large, human-annotated general rhyme database (rather than a specialty data set).

To date, no other rhyme function with these attributes exists.

By taking apart the phoneme and asking "what makes an English sound a human sound?", we hope to better understand rhyme and thus better approximate it. Since this rhyme function uses phonetic features rather than individual phonemes for scoring, it may be easily extended to any other space with new phonemes, such as other languages. This is due to that fact that while phonemes differ across languages, the International Phonetic Alphabet is constant throughout. Additionally, this type of rhyme function gives better insight into why certain phonemes make better rhymes than others.

This rhyme function expands upon the one we presented at ICCC 2017 [**?**] by employing likelihood scoring matrices, onset and coda alignment, discretizing vowel features, including the additional features of rounding, tensing, and stress, and using a genetic algorithm to optimize weights.

2 Methods

Broadly speaking, rhyme is the repetition of similar sounds across multiple words. While we acknowledge that there are many types of rhyme that may be formalized differently, we submit this description as a generalized definition of rhyme in order to automate the assessment of rhymes.

2.1 Rhyme Definition

We define a *rhyme* as phonemic similarity between the stress tails of two or more words. We define *stress tail* as the nucleus and coda of a word's greatest and first stress, followed by all its remaining syllables. This is based on the intuition that the greatest stress in a word is also the syllable with which phonetic similarity begins to matter for rhyme. For example, $station$ and $creation$ rhyme; though $station$ has 2 syllables and $creation$ has 3, the primary stress in $creation$ is in its second syllable. Furthermore, we define 0 as the lowest possible rhyme score and 1 as the highest rhyme score, reserved for perfect rhymes.

A word is made from a sequence of syllables. A syllable is made of an optional onset ω, nucleus ν, and an optional coda κ. The *nucleus* is the central vowel phoneme. The *onset* is the consonant phoneme(s) preceding the nucleus. The *coda* is the consonant phoneme(s) following the nucleus. Both the onset and/or coda may be empty.

2.2 Phonetic Features

Phonemes, the constituents of syllables, can be further broken down into phonetic features. These features are define what phonemes are and are universal to all human languages. Some are quantifiable as continuous variables, but are more commonly expressed as equivalence classes.

Vowel Features

In a departure from our previous work with rhyme, we chose to more closely follow linguistic standards by discretizing the values of all vowel features. The 5 vowel features we use are:

- *height* (h) – refers to the height of the tongue when a vowel phoneme is formed. Its three discrete equivalence classes are high, mid, and low. It is also known as the first formant.
- *frontness* (f) – refers to the distance of the tongue from the back of the mouth when a vowel phoneme is formed. Its three discrete equivalence classes are front, central, and back. It is also known as the second formant.
- *rounding* (r) – refers to whether the lips make a round shape when a vowel phoneme is formed, and may be represented as a Boolean value.
- *tensing* (t) – refers to whether the mouth's width is narrowed when a vowel phoneme is formed, and may also be represented as a Boolean value for tense and not tense (lax).
- *stress* (s) – refers to the emphasis placed on a particular vowel phoneme. Its three discrete equivalence classes are primary, secondary, and none.

The relationship between the first four of these features with regards to frontness and height can be observed in Figure 2.

Consonant Features

Three features create what we know as consonant phonemes:

1. *manner of articulation* (m) – refers to the configuration and interaction of the tongue, lips, and palate when forming a consonant phoneme. Its seven discrete categories are affricate, aspirate, fricative, liquid, nasal, semivowel, and stop.

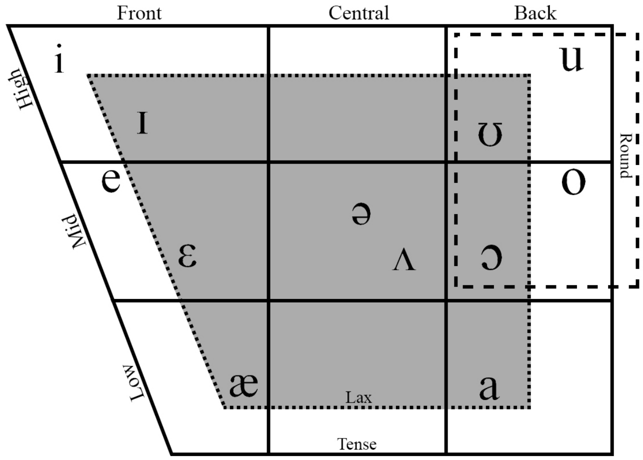
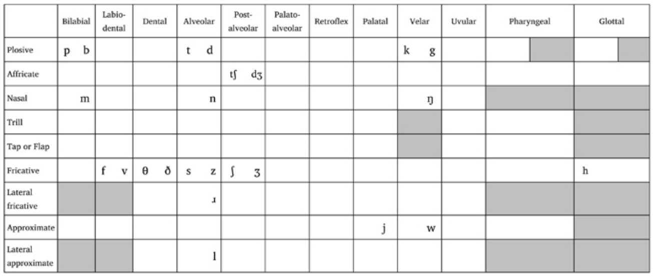


Figure 1: The standard IPA English Vowel Chart [Association, 1999]. Here we see the 12 common English vowel phonemes and their 4 features of height (the vertical axis), frontness (the horizontal axis), rounding, and tensing.



	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palato-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d					k g		▨	▨
Affricate					tʃ dʒ							
Nasal	m			n					ŋ		▨	▨
Trill									▨			▨
Tap or Flap									▨			▨
Fricative		f v	θ ð	s z	ʃ ʒ							h
Lateral fricative	▨	▨		ɹ							▨	▨
Approximate								j	w			▨
Lateral approximate	▨	▨		l							▨	▨

Figure 2: The standard IPA English Consonant Chart [Association, 1999]. Here we see the 24 common English consonant phonemes and their features of manner of articulation (the vertical axis), place of articulation (the horizontal axis), and voicing (voiceless on left, voiced on right).

2. *place of articulation* (p) – refers to the point of contact where an obstruction occurs in the vocal tract to produce a consonant phoneme. Its seven discrete categories are bilabial, labial, interdental, alveolar, palatal, velar, and glottal.
3. *voicing* (v) – refers to whether vocal chords are used to pronounce a phoneme, and may be represented as a Boolean value.

2.3 Scoring

Our rhyme scorer works by

1. extracting the stress tails s_1 and s_2 from two words w_1 and w_2,
2. aligning the stress tails' syllables,
3. aligning the onset ω, nucleus ν, and coda κ of each syllable,

4. aligning the consonant phonemes in multiphonemic onsets and codas, and
5. scoring each aligned phoneme pair.

The rhyme score for two words w_1 and w_2 is defined as

$$R(w_1, w_2) = \frac{\sum S(s_{1_i}, s_{2_i})}{n_s} \tag{1}$$

where $R : W^2 \rightarrow [0 \dots 1]$, W is the set of all words, s_1 and s_2 are stress tails of equal length of words w_1 and w_2 respectively, and n_s is the number of stress tail syllables in s_1 and s_2. The syllable score for two syllables σ_1 and σ_2 is defined as

$$\Sigma(s_1, s_2) = w_\omega A(\omega_1, \omega_2) + w_\nu R_v(\nu_1, \nu_2) + w_\kappa A(\kappa_1, \kappa_2) \tag{2}$$

where $\Sigma : {\Sigma'}^2 \rightarrow [0 \dots 1]$, Σ' is the set of all syllables, A represents a greedy consonant sequence alignment using R_c to score individual consonant pairs. This alignment follows the principles of Needleman-Wunsch alignment [Gotoh, 1982].

$$A(x, y) = max(a) \tag{3}$$

where x and y are consonant phoneme sequences and a is the set of all possible alignments between x and y. Each pair of vowels is scored by

$$\begin{aligned} R_v(v_1, v_2) = {} & \alpha_h M_h(h_1, h_2) + \alpha_f M_f(f_1, f_2) \\ & + \alpha_r M_r(r_1, r_2) + \alpha_t M_t(M_1, M_2) \\ & + \alpha_s M_s(s_1, s_2). \end{aligned} \tag{4}$$

where $R_v : V^2 \rightarrow [0 \dots 1]$, V is the set of all vowel phonemes, v_1 and v_2 are two vowel phonemes, and tables M are scoring matrices. Individual consonant pairs are scored by

$$R_c(c_1, c_2) = \alpha_m M_m(m_1, m_2) + \alpha_p M_p(p_1, p_2) + \alpha_v M_v(v_1, v_2). \tag{5}$$

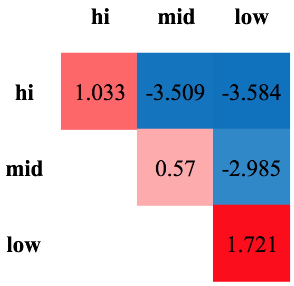
where $R_c : C^2 \rightarrow [0 \dots 1]$, C is the set of all consonant phonemes, c_1 and c_2 are two consonant phonemes, and tables M are scoring matrices. These scores are used by the dynamic programming function A to determine the highest-scoring consonant alignment.

To obtain our final likelihood scoring tables M, we performed the following:

1. created likelihood scoring tables for all phonetic features of vowels,
2. created likelihood scoring tables for all phonetic features of consonants, excluding words with onsets or codas with more than one phoneme, and
3. used our monophonemic consonant likelihood scoring tables to greedily align consonant sequences (onsets and codas) and create multiphonemic consonant likelihood scoring tables.

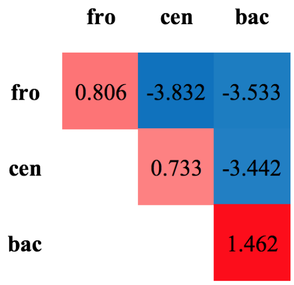
Each cell of a likelihood table is given by

$$\frac{P_r}{P_p} \tag{6}$$



	hi	mid	low
hi	1.033	-3.509	-3.584
mid		0.57	-2.985
low			1.721

Figure 3: Rhyme scoring correlation matrix for height M_h. The feature categories in order are high, middle, and low.



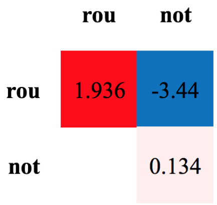
	fro	cen	bac
fro	0.806	-3.832	-3.533
cen		0.733	-3.442
bac			1.462

Figure 4: Rhyme scoring correlation matrix for frontness M_f. The feature categories in order are front, central, and back.

where P_r is the probability that 2 phonetic features are paired in a rhyme and P_p is the probability that 2 phonetic features are paired in random word pairings.

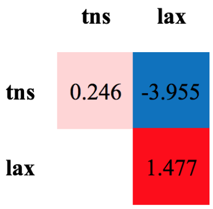
Since syllables and therefore all nuclei are aligned, no sequence alignment beyond syllable alignment for vowels is necessary. Figures 3 through 7 show the resulting scoring tables.

In English syllables, many consonants stand alone and thus are easily paired and scored. But unlike vowel phonemes, consonants can also be found in contiguous sequences and therefore must be aligned before scoring. This makes derivation of the consonant score significantly more involved. In addition to the discrete categories of the three consonant features, we include gaps in order to cover the case when a phoneme is paired with nothing. We distinguish between three types of gaps: beginning gap (G1), middle gap (G2), and end gap (G3). Figures 8 through 10 show the resulting scoring tables.



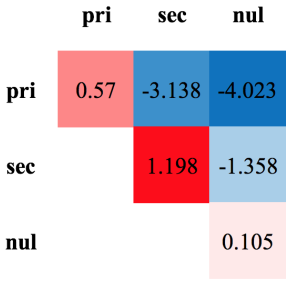
	rou	not
rou	1.936	-3.44
not		0.134

Figure 5: Rhyme scoring correlation matrix for rounding M_r. The feature categories in order are rounded and unrounded.



	tns	lax
tns	0.246	-3.955
lax		1.477

Figure 6: Rhyme scoring correlation matrix for tensing M_t. The feature categories in order are tense and lax.



	pri	sec	nul
pri	0.57	-3.138	-4.023
sec		1.198	-1.358
nul			0.105

Figure 7: Rhyme scoring correlation matrix for stress M_s. The feature categories in order are primary stress, secondary stress, and unstressed.

Upon viewing the likelihoods in these scoring tables, the strong positive diagonals are apparent. This makes sense, because similar phonemes are rhymed more frequently. And the irregularity throughout the tables proves that people rhyme not only identical phonemes, but also phonemes of similar makeup. The most common pairing between different phoneme features is labiodental and interdental.

Also worth noting is that some feature categories are rhymed with themselves more than others. For example, voiceless consonants are more likely to be rhymed with each other than voiced consonants, and unstressed vowels have a very low likelihood of being found in a rhyme.

In consonant sequence alignments, gaps are uncommon. For the majority of features, the most frequent gap type used in rhyme is middle gaps.

2.4 Data

We use the CMU Pronouncing Dictionary [Kominek and Black, 2004] to assign phonemes and stresses to words. This resource uses the English phonetic transcription code ARPAbet, which has a symbol for 15 vowel phonemes and 24 consonant phonemes. We used only words with a single pronunciation. We use a custom syllabifier using the 14 phonotactic rules of English [Harley, 2006]. We use data from RhymeZone.com [Datamuse, 2017] to construct our likelihood tables, and for our genetic fitness function.

2.5 Genetic Optimization

After developing likelihood scoring tables for all 8 phoneme features, we used a genetic algorithm to optimize weights in our rhyme function. Each genetic individual has a weight for the 5 vowel features, the 3 consonant features, and the 3 syllable parts, for a total of 11 evolutionary dimensions: frontness w_f, height w_h, rounding w_r, tensing w_t, stress w_s, manner of articulation w_m, place of articulation w_p, voicing w_v, onset w_ω, nucleus w_ν, and coda w_κ.

Our fitness function returns the mean squared error, defined as

$$\frac{1}{n}\sum_{i=0}^{n}(R - R_d)^2 \tag{7}$$

where R is our algorithm's rhyme score and R_d is the normalized score from the data source (Rhyme Zone). Our genetic algorithm produced populations of 100 individuals from the 20 individuals of the greatest fitness (lowest error) of the past generation. Figure 11 shows the evolutionary process over 300 generations.

We found many individuals of high fitness with diverse differences in weights. Our most fit genetic individual has these normalized weights:

Vowel features

- $w_f = .355$
- $w_h = .921$
- $w_r = .979$
- $w_t = .053$
- $w_s = .398$

Consonant features

- $w_m = .933$
- $w_p = 0.0$
- $w_v = 1.0$

Syllable components

- $w_\omega = .013$
- $w_\nu = .355$
- $w_\kappa = .014$

While these three tiers of weights influence one another and thus cannot be directly compared, these results suggest a few things:

- in vowels, height and tensing stand out as important rhyming features, while tensing is practically meaningless.
- in consonants, place of articulation has no effect on rhyme quality.
- the nucleus of a syllable is by far its most important component, and the importance of the coda is about equal to that of the onset.

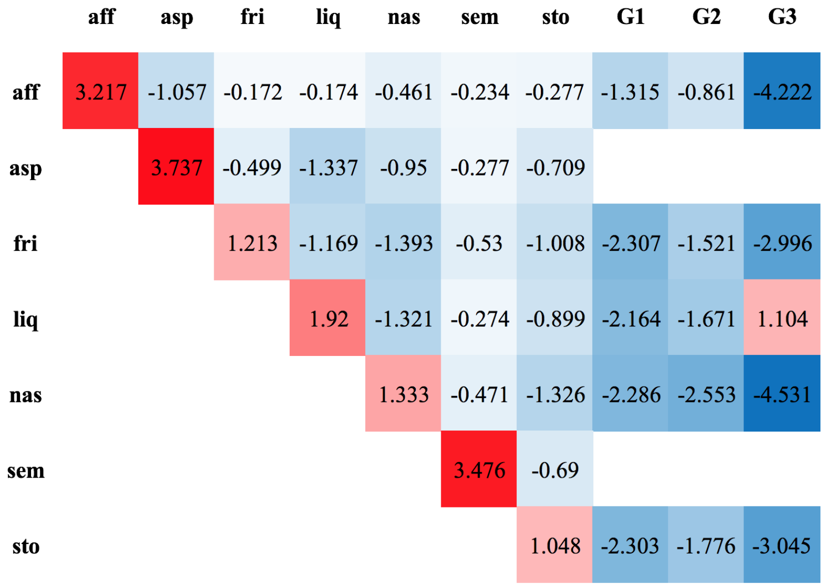
While the latter observation is somewhat intuitive and mirrored in many rhyme functions, the other two observations are more novel and interesting.

3 Results

With likelihood scoring tables and genetically-optimized weights, the rhyme function is ready to score word pairs. Figure 12 gives an example of rhyme function output.

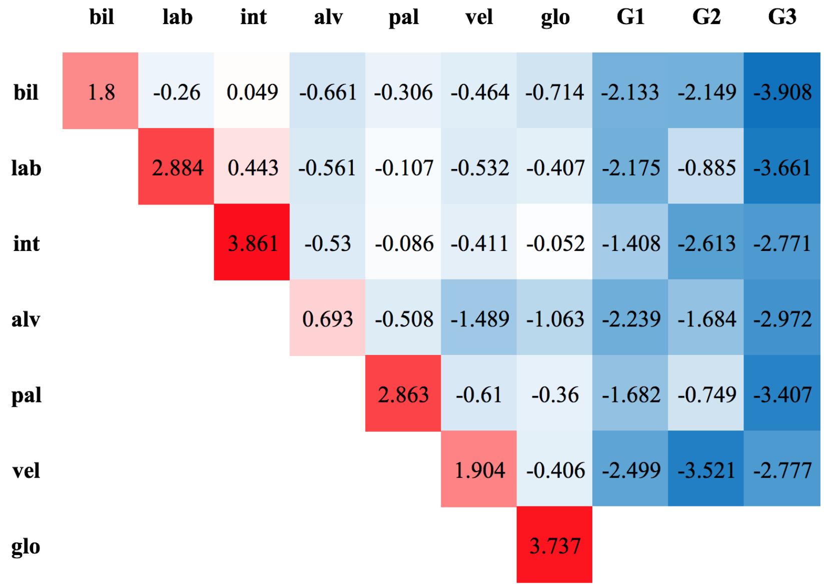
Additionally, this rhyme function can be used to find rhymes for challenging words. For example, the word *keyboard* pairs with the 10 following words with a score of .97 or greater:

In this paper, we present a new rhyme function based on likelihoods that carries the novel characteristics of the stress tail, including all three syllable parts, allowing for multi-phonemic consonant sequences, and decomposing phonemes into phonetic features. We further improved our results via genetic weight optimization.



	aff	asp	fri	liq	nas	sem	sto	G1	G2	G3
aff	3.217	-1.057	-0.172	-0.174	-0.461	-0.234	-0.277	-1.315	-0.861	-4.222
asp		3.737	-0.499	-1.337	-0.95	-0.277	-0.709			
fri			1.213	-1.169	-1.393	-0.53	-1.008	-2.307	-1.521	-2.996
liq				1.92	-1.321	-0.274	-0.899	-2.164	-1.671	1.104
nas					1.333	-0.471	-1.326	-2.286	-2.553	-4.531
sem						3.476	-0.69			
sto							1.048	-2.303	-1.776	-3.045

Figure 8: Rhyme scoring correlation matrix for manner of articulation M_m. The feature categories in order are affricate, aspirate, fricative, liquid, nasal, semivowel, stop, beginning gap, middle gap, and end gap.



	bil	lab	int	alv	pal	vel	glo	G1	G2	G3
bil	1.8	-0.26	0.049	-0.661	-0.306	-0.464	-0.714	-2.133	-2.149	-3.908
lab		2.884	0.443	-0.561	-0.107	-0.532	-0.407	-2.175	-0.885	-3.661
int			3.861	-0.53	-0.086	-0.411	-0.052	-1.408	-2.613	-2.771
alv				0.693	-0.508	-1.489	-1.063	-2.239	-1.684	-2.972
pal					2.863	-0.61	-0.36	-1.682	-0.749	-3.407
vel						1.904	-0.406	-2.499	-3.521	-2.777
glo							3.737			

Figure 9: Rhyme scoring correlation matrix for place of articulation M_p. The feature categories in order are bilabial, labial, interdental, alveolar, palatal, velar, glottal, beginning gap, middle gap, and end gap.

	v	no	G1	G2	G3
v	0.699	-0.96	-2.183	-1.031	-2.945
no		1.188	-2.408	-2.033	-3.013

Figure 10: Rhyme scoring correlation matrix for voicing M_v. The feature categories in order are voiced, voiceless, beginning gap, middle gap, and end gap.

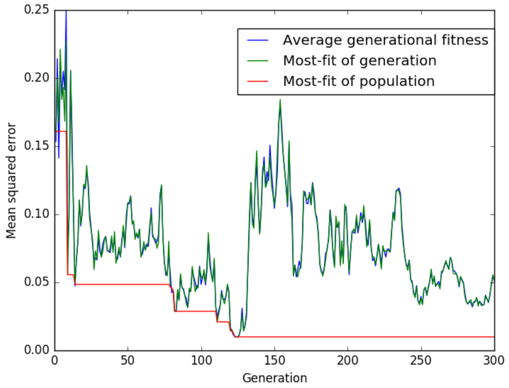


Figure 11: Fitness of algorithm weights over 300 generations of evolutionary training. Note that the best genetic individual has an error of only 0.012 and is reached after 123 generations.

	slant	lies	delight	surprise	eased	kind	blind
slant	1	0.573	0.589	0.573	0.563	0.597	0.597
lies		1	0.96	1	0.843	0.963	0.963
delight			1	0.96	0.824	0.964	0.964
surprise				1	0.844	0.963	0.963
eased					1	0.843	0.843
kind						1	1
blind							1

Figure 12: Rhyme correlation matrix for end-rhymes in Emily Dickson's *Tell all the truth but tell it slant*. Note that all words have a stress tail of syllable length 1. Scores for words with themselves are always 1. Also noteworthy is that scores for $kind$ and $blind$ are identical, since their stress tails are identical.

seewald	freeport	seaport	retort	leadoff	peapod	freefall	seashore	freeform	seaborne
0.982	0.981	0.981	0.981	0.98	0.977	0.976	0.976	0.975	0.973

Figure 13: Single words from the CMU Pronunciation Dictionary that best rhyme with the word $keyboard$ using this rhyme function.

Code for our implementation of this paper can be found on GitHub [**?**]. Instructions for using it can also be found there.

One compelling idea for future work is to optimize weights via a deep neural network. These weights and their overall fitness could then be compared against those of the genetic algorithm. We plan to test this concept in the near future.

References

[Association, 1999] International Phonetic Association. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[Datamuse, 2017] Datamuse. Datamuse api, 2017.

[Gotoh, 1982] Osamu Gotoh. An improved algorithm for matching biological sequences. *Journal of molecular biology*, 162(3):705–708, 1982.

[Harley, 2006] Heidi Harley. *English Words: A Linguistic Introduction*. Blackwell Publishing Ltd., 2006.

[Hinton and Eastwood, 2015] Erik Hinton and Joel Eastwood. Playing with pop culture: Writing an algorithm to analyze and visualize lyrics from the musical "hamilton". 2015.

[Hirjee and Brown, 2010] Hussein Hirjee and Daniel Brown. Using automated rhyme detection to characterize rhyming style in rap music. 2010.

[Kominek and Black, 2004] John Kominek and Alan W Black. The CMU arctic speech databases. In *Fifth ISCA Workshop on Speech Synthesis*, 2004.