Experimental Evidence for the Rhythm Rule in English

Sam Beames

I. Introduction

In English, rhythm of speech is usually thought of as the relationship of stressed and unstressed syllables to each other and where they fall in a word, as well as how they change within connected speech. For example, the primary stress on a word in isolation may not always be in the same place as the primary stress on that same word in the context of a sentence.

(1) Example of Stress Retraction (Hayes 1984):

a. *Mississíppi* /mɪsɪˈsɪpi/

b. *Mìssissippi législature* /ˌmɪsɪsɪpi ˈlɛʤəsleɪʧəɹ/

The term "rhythm rule" describes a phenomenon wherein the stress of certain words depends on their environment. In the example above, the word *Mississippi* is usually stressed on the penultimate syllable (mɪsɪˈsɪpi), but when followed by a word with first syllable stress, like *legislature*, stress retracts to the first syllable (ˌmɪsɪsɪpi ˈlɛʤəsleɪʧəɹ). However, this only happens in certain environments. The word *thirteen*, for example, is usually stressed on the second syllable (θɚˈtin), but in the phrase *thirteen trumpets* stress shifts to the first (ˈθɚtin ˈtɹʌmpəts). In other phrases, however, this stress shift is less likely (e.g. θɚˈtin trɑmˈboʊnz S). Some further examples follow in (2).

(2) Examples of rhythm rule (Hayes 1984):

a. *Cornéll → Còrnell hóckey* /kɔɹˈnɛl/ → /ˌkɔɹnɛl ˈhɑki/

b. *analýtic → àналytic thóught* /ænəˈlɪtɪk/ → /ˌænəlɪtɪk ˈθɔt/

c. *Alabáma → Àlabama rélatives* /æləˈbæmə/ → /ˌæləbæmə ˈɹɛlətɪvz/

d. *Oklahóma → Òklahoma cóngressman* /oʊkləˈhoʊmə/ → /ˌoʊkləhoʊmə ˈkɑŋgɹəsmən/

This project had two goals: to show that the rhythm rule is productive, and to investigate what the factors are that make it more or less likely to apply. In particular, I will explore the hypothesis that the English rhythm rule is sensitive to the total phonetic duration between two stresses. To give an example, the idea is that stress is likely to shift in "thìrteen trúmpets" because, in the alternative "thirtèen trúmpets", there would not be enough time in milliseconds between the two stresses. Hayes (1984) hypothesizes that phonetic duration may play a role in the English rhythm rule, the hypothesis has not previously been substantiated.

According to Hayes' hypothesis, the less time in milliseconds between the two stresses, the more likely the rhythm rule is to occur. To test this hypothesis, I did a production study with 29 native speakers of American English. Several native speakers of American English that participated in this study did not rhythm-rule on words like *thirteen, fourteen, fifteen,* or *sixteen,* and always stressed the first syllable independent of the context, which raised the question of whether or not the rhythm rule is

psychologically real. The goal of the experiment was to determine whether or not the rhythm rule is productive and, if so, if phonetic duration plays a role in how likely it is. The results indicate that the rhythm rule is psychologically real, and that phonetic duration may play a subtle role.

II. Background

For the rhythm rule to occur, certain conditions must be met. The syllable to which the stress is shifted must have some degree of stress already. For example, the word *diacritic* has secondary stress on the first syllable and primary stress on the third syllable (ˌdaɪəˈkɹɪtək), so when stress shifts in connected speech, like in the phrase diacritic markings (ˈdaɪəkɹɪtək ˈmɑɹkɪŋz), it must shift to the first syllable because that is the only syllable that had some degree of stress previously.

There are also suggestions that phonetic information has an effect on stress with regards to the rhythm rule. The first real mention of this possibility was in the appendix of Hayes 1984. In one of Hayes's examples, he notes that the phrases *Korbel tequila* and *Korbel champagne* are not equally likely to undergo the rhythm rule. He states that a pronunciation like [ˈkɔɹbəl ʃæmˈpeɪn] (shifted from [kɔɹˈbəl ʃæmˈpeɪn]) is far less likely than a pronunciation like [ˈkɔɹbəl təˈkilə] (shifted from [kɔɹˈbəl təˈkilə]). If the rhythm rule acted entirely to eliminate sequences of stressed syllables, this difference would not be predicted: the unshifted versions of both [kɔɹˈbəl ʃæmˈpeɪn] and [kɔɹˈbəl təˈkilə] have one syllable between the two stresses. Hayes' hypothesis is that, because the /ʃ/ in *champagne* is inherently longer than the /t/ in *tequila*, a speaker of American

English is more likely to apply the rhythm rule to *Korbel champagne* because the distance in milliseconds between the two stresses in [kɔɻˈbəl ʃæmˈpeɪn] would be smaller than that in [kɔɻˈbəl təˈkilə]. A major limitation of this work, however, is that Hayes' hypothesis is based only on his intuition and other hints in the literature, and has not yet been explored within this particular empirical domain in a more rigorous way.

Some recent work on other aspects of English stress support the more general idea that stress is at least in part sensitive to phonetic information. Stanton (2019) argues that the factors that govern whether or not "-at-" bears stress in words ending in "-ative" must make reference to phonetic duration. The empirical generalization (verified through a dictionary study) is that speakers of American English are more likely to stress "-at-" when it is preceded by a sequence of multiple consonants. For example, most speakers prefer to stress "-at-" in *legislative* (ˈlɛʤəˌsleɪtɪv), but not in *speculative* (ˈspɛkjələtɪv). This difference cannot be explained by referring to syllables: both *legislative* and *speculative* are four syllables long, with stress on the first. Stanton's idea is that the likelihood of stressing "-at-" depends on the phonetic duration of the consonant(s) preceding "-ative": speakers are more likely to stress "-at-" in *legislative* than in *speculative* because /sl/ is longer than /l/. Through two different studies of speaker preference for nonce word stimuli ending in "-ative," Stanton shows that speakers are sensitive to these small differences in phonetic duration and that their preferences align with the hypothesis. This phenomenon is a strong piece of evidence in favor of the more general hypothesis that the distribution of stress in English depends in part on phonetic information.

III. Methods

For this experiment, my aim was to record native American English speakers from a variety of backgrounds reading a list of sentences designed to force a stress shift. Each participant read a list of 72 sentences. In each sentence, the penultimate word was a two-syllable word with final stress when spoken in isolation, but depending on the composition of the final word, stress sometimes retracted.

To construct the sentences, I chose four possible “first words”, nine one-syllable “second words”, and nine multi-syllabic “second words” that did not have stress on the first syllable.

word1 (orthography / IPA)	word2 (orthography / IPA)
bamboo / bæmˈbu *concrete* / kɑnˈkɹit *fifteen* / fɪfˈtin *complete* / kəmˈplit	*likes* / ˈlaɪks *bikes* / ˈbaɪks *strikes* / ˈstɹaɪks *aches* / ˈeɪks *flakes* / ˈfleɪks *wakes* / ˈweɪks *cakes* / ˈkeɪks *rakes* / ˈɹeɪks *fakes* / ˈfeɪks
	mistakes / mɪsˈteɪks *divides* / dɪˈvaɪdz *requests* / ɹɪˈkwɛsts *withdrawals* / wɪθˈdɹɔəlz *routines* / ɹuˈtinz *machines* / məˈʃinz *guitars* / gɪˈtɑɹz *collections* / kəˈlɛkʃənz *selections* / səˈlɛkʃənz

My options for the first word were *bamboo*, *concrete*, *fifteen*, and *complete*. The word “bamboo” in isolation is always pronounced with stress on the second syllable, but stress shift is actually quite common before one-syllable words with shorter onsets. For example, speakers are much more likely to say [ˈbæmbu ˈbaɪks] than [bæmˈbu ˈbaɪks] for the phrase *bamboo bikes*. The words *concrete* and *fifteen* can be pronounced with stress on either syllable in isolation, depending on the speaker, and there is even variation within a single speaker. I included these words where stress is acceptable on either syllable because they are very likely to be affected by rhythm. I also included the word *complete* as a sort of control word, because stress is always on the final syllable in isolation and should always be on the final syllable in connected speech as well. This is because the first syllable of *complete* has no measure of stress on the first syllable, so stress retraction would not be expected in any context.

The “second words” I chose were broken up into two categories: target words, or one-syllable words of varying onset lengths, and filler words, which were multisyllabic and did not have a stress on the first syllable. I had nine of each type of word, and every possible combination with the first words yielded 36 target sentences and 36 filler sentences, for a total of 72 sentences for each participant. The target second words could be further divided up into four categories: words without an onset (*aches*), words with an obstruent onset (*bikes, cakes, fakes*), words with a sonorant onset (*likes, wakes, rakes*), and words with a cluster onset (*flakes, strikes*). I tried to find words with similar vowels, so each second content word has either the diphthong /aɪ/ or /eɪ/. The filler words were designed to balance the experiment, in order to break up the rhythm of

the 36 content sentences. This was done to prevent participants from learning over the course of the experiment that most of the first words should have first syllable stress, in an attempt to produce more variation in responses. Every possible combination of first and second word was generated, and I created a sentence frame that had some lexical meaning. Some examples are “when you least expect it, bamboo strikes” and “whenever I leave New York, I have concrete withdrawals”.

The words of interest were always in phrase-final position to control for phrasal intonation. In total, I ran three pilots and 29 participants whose data I included in this study. These participants were recruited by reaching out to other NYU students over email and through posts on social media.

To record the participants, I used a Shure SM35 head-mounted microphone and a Marantz PMD661 MKIII recording device. Participants were recorded in a soundproof booth to reduce background noise. I then went back and listened to each recording in Praat, a speech analysis software used in phonetics, and scored each token with either a 1 or 2, which reflected whether the stress of the first word fell on the first or second syllable.

I then took these data and created a spreadsheet that denoted several things in addition to the participant number, each of the two words, and where stress falls on the first word.

partici-pant	word1	word1.rr.exp	word2	word2.rr.exp	word2.seg	word2.segtype	word2.segtype.broad	stress
1064	bamboo	yes	flakes	yes	fl	CC	O.CC	2
1083	complete	no	strikes	yes	str	CC	O.CC	2
1069	fifteen	yes	requests	no				2

Word1.rr.exp notes whether or not the rhythm rule is expected based on the identity of the first word, and word2.rr.exp does the same based on the second word. I also categorized each second word in the content group based on the identity of the onset in one column: as either beginning with a vowel, obstruent, sonorant, or cluster in another; and finally, as belonging to either a vowel and sonorant group or a obstruent and consonant cluster group. These several layers of distinction were created in order to look at different variables within the data to see which variables have the largest effect on stress placement. Relevant variables include the number of syllables in the second word, the identity of the onset of the second word (and whether it is an obstruent, sonorant, cluster, or null), and whether it is possible to ever stress the first syllable of the first word. I used R, a program for statistical computing, as well as regression analysis, to isolate the variables of interest and to see their effect on stress placement. If Hayes' view of the rhythm rule is correct, I expected to find that the rhythm rule is more frequent when it is expected based on the first word (more frequent for the words *bamboo*, *concrete*, and *fifteen*, but not for *complete*), and when the rhythm rule is expected based on the second word (it would not be expected for the filler words). If phonetic distance between the stresses matters, I would also expect to see the effect of

the onset type of the second word, so the stress would retract in the first word when the second word is one syllable and has a shorter onset.

IV. Results

Overall, the results of this study show two important findings: that the rhythm rule is, in fact, psychologically real, and that durational properties of individual segments have a small effect. The trends present in my data are consistent with what we currently know and hypothesize about the rhythm rule.

Figure a shows the effect of the first word on stress. The dark part shows how often each word gets first syllable stress, and the light part shows how often each word gets second syllable stress. This graph only takes into account the data where the rhythm rule is expected given the second word, so only the cases where the second word has one syllable.

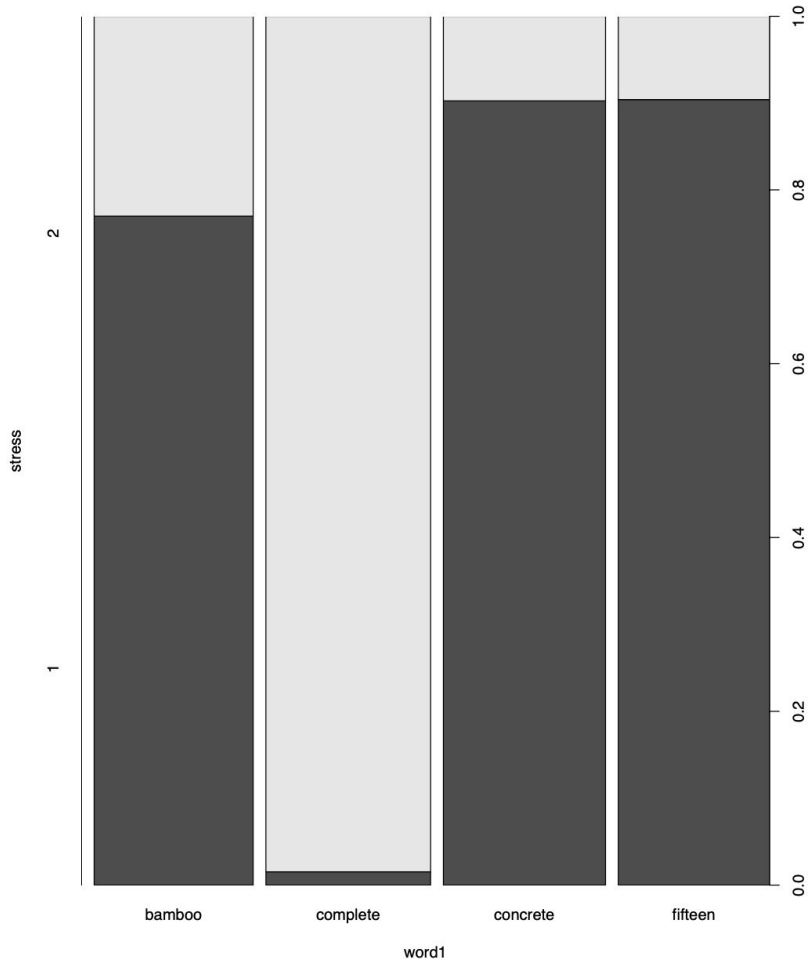

Fig. a

Unsurprisingly, we see the word “complete” behaving very differently from the other three words, as it almost never has stress on the first syllable. This is in line with my hypothesis, that the rhythm rule only occurs when it is expected based on the identity of the first word.

Figure b shows the effect of the second word on the stress of the first word. I have excluded cases where the first word is “complete”, because that would cause the

rhythm rule to not be expected no matter the identity of the second word. “No” refers to second words where the rhythm rule is not expected, which are the multisyllabic filler words, and “yes” refers to second words where the rhythm rule is expected, which are the one syllable target words.

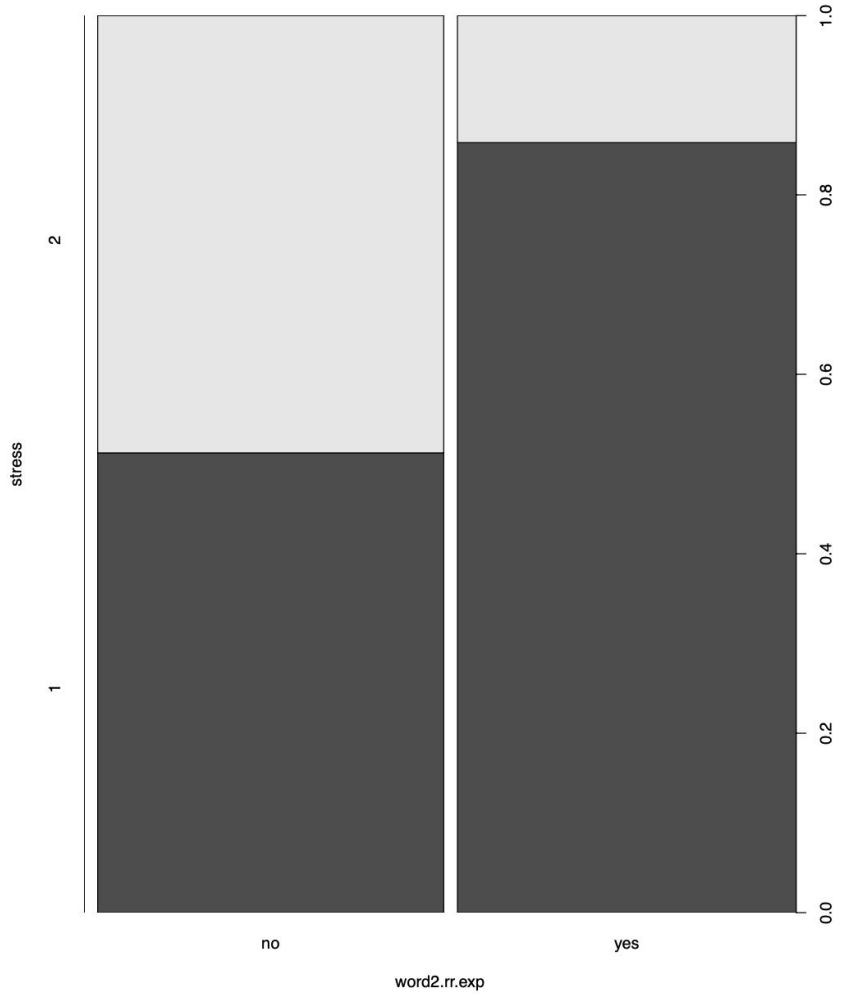

Fig. b

As predicted, we see the two syllable filler words tend to cause the first word to have stress on the second syllable at a much higher rate than the content words. Put differently, the rhythm rule applies less frequently in contexts where it is not expected.

This is another important piece of evidence that supports the psychological reality of the rhythm rule.

I used R (R Core Team, 2013) to run a model that looked at whether the identity of the first word and the identity of the second word contributes to the likelihood of the rhythm rule.

Predictor	Coefficient	z-value	Significant?
(Intercept)	6.83	-	-
Word1: RR expected	-6.88	-7.27	yes ($p < .001$)
Word2: RR expected	-2.18	-9.47	yes ($p < .001$)

Fig. c

Figure c shows that both the identity of the first word and the identity of the second word have quite a bit of influence on whether or not the rhythm rule occurs: as predicted, 2nd syllable stress (in the first word) is less likely to occur when the rhythm rule is expected. Word1 has a lower coefficient than Word2, which suggests that Word1 is a better predictor of when the rhythm rule will occur. However, because both the z-values of Word1 and Word2 are lower than -2, they are both great predictors of whether or not the rhythm rule will occur.

In figure d, I grouped each onset type into one of four categories: consonant clusters (CC), obstruents (O), sonorants (R) , and vowels (V).

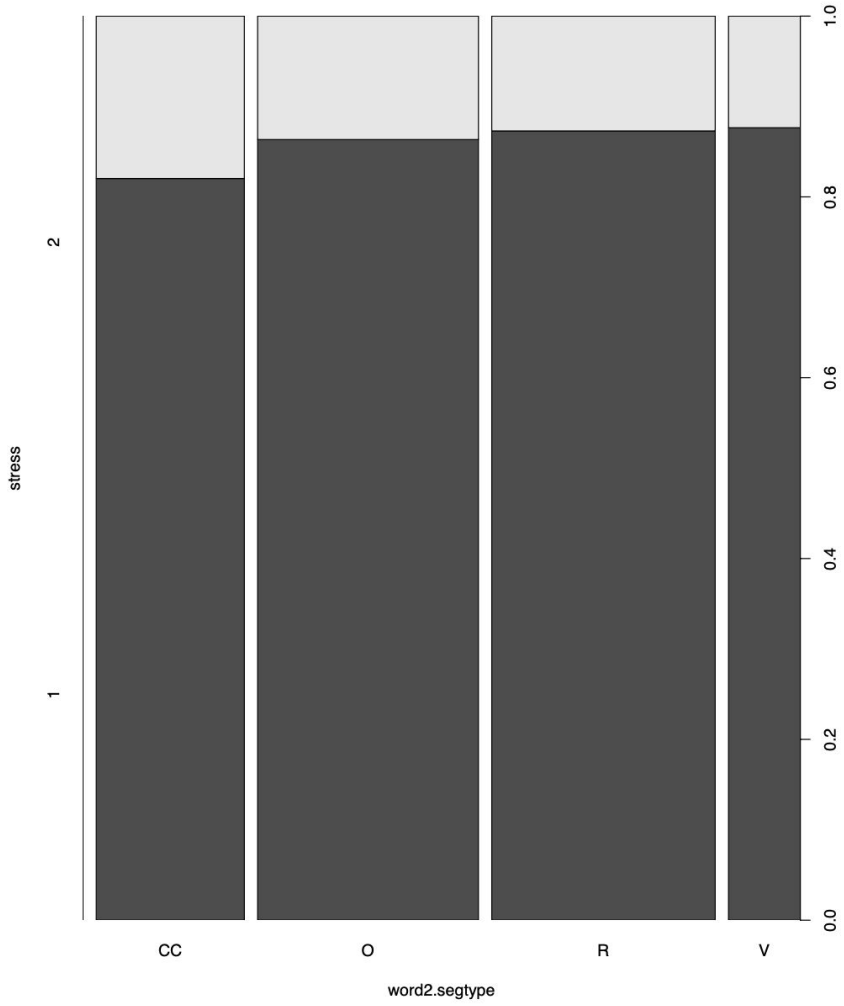

Fig. d

Here, we see that clusters appear to have a slightly higher rate of provoking second-syllable stress.

I fit an additional logistic regression model in R to the subset of the data where the rhythm rule is predicted based on the identity of both the first and second word to see whether or not the segmental profile of word 2 matters in whether or not the rhythm rule applies. The variable corresponding to segmental information was helmert-coded.

Helmert-coding is a technique used to compare levels of a variable with the mean of the subsequent levels of the variable. In this case, I compared the rate in which the rhythm rule is active in vowel-initial second words to the average rate in obstruent-, sonorant-, and cluster-initial second words; the average rhythm rule rate in sonorants with the combined rate in obstruents and clusters; and finally, the average rate in obstruents with the average rate in clusters.

Predictor	Coefficient	z value	Significant?
(Intercept)	-2.17	-	-
O vs. CC	-0.18	-1.28	no (p = .20)
R vs. O, CC	-0.09	-1.12	no (p = .26)
V vs. R, O, CC	-0.07	-0.72	no (p = .47)

Fig. e

In the table above, the model is asking how the identity of the onset of word 2 (broken up into categories of obstruent, sonorant, or vowel) affects the probability of word 1 having second syllable stress. All three categories (obstruent versus cluster, sonorant versus obstruent and cluster, and vowel versus sonorant, obstruent, and cluster) have a negative coefficient. This means that the second member of each comparison, so the consonant cluster, obstruent/cluster, and sonorant/obstruent/cluster groups is less likely to cause second syllable stress in the first word, or in other words more likely to induce the rhythm rule. This small difference is consistent with my hypothesis that the segmental type does have an effect on stress patterns.

The z-value is a measure that shows how far below or above the mean of a specific value is in a given dataset. For a figure to count as significant, this value must be higher than 2 or lower than -2. Though the values in this model are not significant, the data are trending in the right direction. My hypothesis is that this value would become significant if I were to repeat this study with more participants, because these values are nearly double the z-values of when I ran the model with only half of the data included.

V. Implications

The results of this study show us, first and foremost, that the rhythm rule plays an important role for English speakers in determining stress placement. There is, in addition, a suggestion that onset duration is relevant, but more work is necessary to show this convincingly. The rhythm rule as it is currently understood is not the whole story. Stress has historically been discussed only in terms of syllables as the smallest significant unit, but these data show that individual segments also have an effect.

The effect of this durational information on the phonetic level is just shy of significant in its z-value, but this is very likely a reflection of the size of the subject pool rather than the importance of this effect. An increase in the number of subjects caused the z-value increase as well, so it follows that adding more data would only further legitimize these claims.

Though most earlier accounts of stress have not made explicit reference to this, the effect of phonetic duration has always been present in the periphery. Hayes (1984)

mentions the possibility, though his claim was based on intuition rather than empirical evidence. The effect of speaker perception of durational phonetic information may be small, but it is essential in the analysis of stress in American English.

Works cited:

Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15. 33-74.

Stanton, Juliet. 2019. Phonetic lapse in American English *-ative*. To appear in *Glossa*.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.