Published in final edited form as:

Lang Cogn Neurosci. 2020 ; 35(3): 383–392. doi:10.1080/23273798.2019.1665190.

Word Recall is Affected by Surrounding Metrical Context

Amelia E. Kimball,

Laboratoire de Linguistique Formelle, Université Paris Diderot

University of Illinois at Urbana-Champaign

Loretta K. Yiu,

Department of Human Centered Design and Engineering, University of Washington

University of Illinois at Urbana-Champaign

Duane G. Watson

Department of Psychology and Human Development, Vanderbilt University

University of Illinois at Urbana-Champaign

Abstract

It has been claimed that English has a metrical structure, or rhythm, in which stressed and unstressed syllables alternate. In previous research regular, alternating patterns have been shown to facilitate online language comprehension. Expanding these findings to downstream processing would lead to the prediction that metrical regularity enhances memory. Research from the memory literature, however, indicates that regular patterns are less salient and therefore less well remembered, and also that strings of similar sounds are harder to remember. This work suggests that, like lists of words with similar sounds, lists of words with similar metrical patterns are *less* accurately remembered than comparable metrically irregular patterns. The present study tests these conflicting predictions by examining the effects of metrical regularity in a recall task. We find that words are better recalled when they do not match their metrical context, suggesting that a regular metrical structure may not be beneficial in all contexts.

Keywords

word recall; prosody; metrical regularity; stressed syllables; recall

Introduction

In English, metrical structure is formed by patterns of stressed syllables, and regular patterns are formed when stressed and unstressed syllables alternate. For example, metrical structure at the lexical level differentiates the verb *pre-SENT*, i.e. to give someone something, from the noun *PRE-sent*, i.e. a gift. Metrical structure in the form of rhythmic regularity also exists across strings of words, as is sometimes found in English poetry (e.g., if **music be** the

food of love, play **on**), or in conversational speech (**I hate to ski in icy snow**). The goal of this paper is to determine how metrical structure affects memory for words.

Metrical structure at the level of the utterance has been widely investigated in psycholinguistics, and there is a great deal of work suggesting that metrical regularity may benefit language processing. For example, lexical stress plays an important role in speech segmentation (Butterfield & Cutler 1988; Cutler et al. 2018; Cutler & Norris, 1988). We also know that adult listeners are sensitive to metrical regularity and irregularity (Geiser et al. 2008 ; Harding et al, 2019; Henrich et al. 2014; Rothermich et al. 2012) as well as expected and unexpected metrical structures (Bohn et al 2013; Cason & Schön 2012; Dohmas et al., 2013; Kandylaki et al. 2017 ; Rothermich & Kotz 2013; Rothermich et al. 2010, among others).

The Attentional Bounce Hypothesis (Pitt & Samuel, 1990) provides one explanation for this benefit. Pitt & Samuel (1990) argue that metrically regular stretches of speech provide a pattern that the listener uses to predict upcoming stressed syllables. Because stressed syllables are information-heavy (Altman & Carter, 1989), comprehenders can use predictive processing to parse the rapidly unfolding signal, and lessen the noise and ambiguity inherent in the speech stream. This view is supported by evidence that metrical regularity affects online language processing (Pitt & Samuel, 1990; Quené & Port, 2005; Shields, McHugh, & Martin, 1974). For example, if a listener is asked to press a button when given a target speech sound, reaction times in metrically regular speech are faster. This suggests that listeners use the timing of stressed syllables to anticipate upcoming stressed syllables. Similar results have been found in a duration detection task in which lengthening is detected more accurately for stressed syllables and trochaic patterns (Zheng & Pierrehumbert, 2010). Metrical regularity may also facilitate production: speakers are significantly slower and make more errors when producing irregular sequences of trisyllabic nonwords than when producing regular sequences (Tilsen, 2011). These results suggest that regularity offers a processing advantage for low-level processing tasks in both production and perception, and perhaps for perception more generally (see Dynamic Attending Theory, Bauer et al. 2015; Large & Jones, 1999; Jones, 2010; Jones et al., 2002).

The above work suggests that listeners are sensitive to the timing of strong beats, but we also know that listeners are also sensitive to global meter. Sensitivity to metrical regularity has also been shown in online processing as measured by eyetracking in silent reading and in the visual world paradigm. Brown, Salverda, Dilley, & Tanenhaus (2015) found that the pitch and timing of syllables before a target word can disambiguate metrically ambiguous words: listeners' looks to differently stressed cohort competitors (e.g. jury/giraffe) are influenced by the patterns of pitch and timing leading up to the target word. They conclude that listeners create expectations about upcoming prosody, including word stress and rhythm, and interpret incoming acoustic information based on these expectations. Even infants are sensitive to the global rhythm and timing of their native language (e.g. Cutler, McQueen, Norris, & Somejuan, 2001; Nazzi, Bertoncini, & Mehler, 1998; Weber, Hahne, Friedrich, & Friederici, 2004) and are able to differentiate between their native language and other languages with different rhythm (Nazzi & Ramus, 2003; Ramus, 2002). Listeners also use prosody to perceptually group ambiguous syllables (e.g. Dilley & McAuley, 2008).

Evidence also suggests that comprehenders are aware of metrical patterns even in silent reading; words that do not match a predictable metrical pattern in silent reading incur a reading cost (Breen & Clifton, 2011).

The work reviewed here suggests that listeners are sensitive to metrical patterns and therefore maintain a representation of metrical stress. Furthermore, it shows an advantage for a regular pattern and a cost for irregular patterns. However, this observed advantage for regular meter has largely focused on online processing, that is, to perception and comprehension as speech is unfolding. It is unclear whether the advantage for regular meter extends to downstream comprehension and memory. In fact, two areas of literature suggest that this same regular pattern, while helpful in online perception, may be disadvantageous when it comes to recalling recently heard material.

First, auditory and visual stimuli that are surprising, emphasized, or dissimilar to neighboring context attract attention (Barnes & Jones 2000; Duncan & Humphreys, 1989; Fairhall et. al. 2001 ; Large & Jones 1999; McAuley & Jones 2003) and enhance memory formation, learning (Ranganath & Rainer, 2003), and associative learning (Schultz and Dickinson, 2000, Fletcher et al., 2001).

There is a long record of consistent findings of an isolation effect for stimuli that differ from their neighbors (Hunt, 1995). Stimuli that stand out from their surroundings drive increases in attention and memory across different cognitive domains, including visual attention (Itti & Baldi, 2005; 2009). In language experiments, it has been shown that changes in behavior for stimuli that do not match their context is not limited to attention, but instead, affects processing as a whole, as evidenced by increased performance in comprehension tasks in which new information is accented, and therefore salient (Birch & Clifton, 1995; J. K. Bock & Mazzella, 1983). As Winkler and colleagues (2009) put it “deviance detection is an essential function of the human sensory systems, because deviance, defined as a violation of some detected regularity, [...] represents new information for the organism.” These results suggest that irregular deviations from a fixed pattern may play an important part in language comprehension.

Second, work on verbal working memory, language production, and language comprehension suggests that sets of words that share linguistic features such as similar phonemes are more difficult to process. Metrically regular strings of words by definition have alternating stress. Because conversational English has a bias toward initial strong syllables (Cutler & Carter 1987) and toward trochaic stress in disyllabic words (Clopper 2002), these regular strings may contain words that have a similar metrical structure. It is possible that this shared structure leads to greater processing difficulty. This has been extensively studied in sets of words that share phonemes (see Acheson & MacDonald, 2009; Conrad, 1964; Conrad, Freeman, & Hull, 1965). For example, a list of phonologically similar words like “cab, can, cab, cad” are more difficult to recall than a list of words that share fewer phonemes such as “pit, few, cow, pen.” (Baddeley, 1966)

This disadvantage for words that are linguistically similar is also present in linguistic tasks, such as language production. Sentences that contain words with overlapping initial

phonemes are more difficult to produce than sentences that do not have overlapping phonemes (Shattuck-Hufnagel, 1987). Speakers tend to avoid producing words that overlap phonologically with a previously uttered word (T. F. Jaeger, Furth, & Hilliard, 2012), and phonological overlap between two words in a discourse can lead to slowdowns in speech (O'Seaghdha & Marin, 2000; Sevald & Dell, 1994; Yiu & Watson, 2015). Both recall difficulty and production difficulty seemed to be linked to phonological similarity, and it has been argued that the psychological processes that are measured by verbal working memory tasks are really tapping into language processing mechanisms (Acheson & MacDonald, 2009; MacDonald & Christiansen, 2002). Similar effects have been found in language comprehension. Processing syntactically complex sentences is more difficult if words in the sentence overlap phonologically (Acheson & MacDonald, 2011). In short, these results suggest that a disadvantage in memory for overlapping representations correlates with broader deficits in production, perception, and comprehension. Based on this work, we should expect that strings of words with similar metrical structure will be less well remembered than strings of words that vary in their metrical structure.

The goal of the current study is to test the effect of metrical patterns of stressed and unstressed syllables on memory. There are two possible predictions based on the literature cited above. If memory for metrical structure is similar to online processing of metrical structure, we would expect words that share the same metrical pattern as their neighbor to be recalled more accurately. If, on the other hand, metrical overlap and segmental overlap have similar effects on memory, or irregular meter draws attention, we would expect words that match the metrical structure of neighboring words to be recalled less accurately. In Experiment 1, we test participants' recall of lists of words that are either perfectly metrically regular, or were metrically regular except for one mismatching word.

Experiment 1

Methods

Participants—Sixty participants recruited from Amazon Mechanical Turk performed the task online and were paid \$5 for participating. All were self-reported native English speakers with normal hearing.

Materials and Design—We constructed two lists of 52 items each containing eight bisyllabic nouns for participants to recall. For nouns that could also be interpreted as verbs, we selected nouns in which the noun frequency was higher than verb frequency using the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). The experiment was a two by two design, comparing four conditions that differed in pattern (strong-weak vs. weak-strong) and regularity (regular vs. irregular). Half of the items had a strong-weak (SW) stress pattern (i.e., stress on the first syllable of every word), and the other half had a weak-strong (WS) pattern (i.e., stress on the second syllable of every word). To clarify, we are not investigating whether one word pattern (SW vs. WS) will be more accurately recalled. Instead, we investigate whether switching from one pattern to another has a cost in memory. In designing an experiment to test this, it is important to take into account word frequency, because it is well known that high-frequency words are recalled with better accuracy than

low frequency words in free recall (Hall 1954; Mandler 1980, though list design can change or remove this effect, see Bousfield & Cohen, 1955; Shepard, 1967; Underwood, Ekstrand, & Keppel, 1965). Because SW words are higher in both type and token frequency in American English, we expect that SW words will be better recalled.

Metrically regular and irregular items were formed by manipulating the stress pattern of the word in position three in the item (henceforth the target word). In half of the items, the stress pattern of the target word matched that of its surrounding context, creating a regular item (examples 1 and 2, table 1), and in the other half, it conflicted with the stress pattern of the other words, creating an irregular item (examples 3 and 4, table 1). Target words were always in position three in the item to control for serial order effects. The experiment was a within-subjects design. All participants heard all 52 items, with 13 items from each of the four conditions; the order of items was randomised for each participant.

To ensure that one pattern of target words was not better recalled due to frequency and imageability, we matched every SW target to a WS target that was equivalent in frequency and imageability (Schock, Cortese, & Khanna, 2012). Target words were reversed with their yoked versions to create a second set of 52 items. In other words, target words that were embedded in regular conditions in the first list of items were also targets in the irregular conditions in the second set. This allowed each of the 52 target words to be heard in both regular and irregular patterns. Participants were randomly assigned to one of the lists.

Each item was recorded as 44.1 kHz waveform audio files in a sound attenuated booth by the first author, a native speaker of English. For Experiment 1, each item was recorded in one take, using a metronome to control timing. This method was chosen in order to preserve any acoustic cues to metrical pattern in the distal prosody of the utterance (Dilley & McAuley, 2008; Heffner, Dilley, McAuley, & Pitt, 2013) rather than splicing words in from different recordings which would remove these cues (we address the potential complications introduced by natural recordings below in Experiment 2). Still, it was important to establish that the target words were not pronounced in an exaggerated way so as to draw attention to them above and beyond effects of metrical structure. To rule out experimenter bias in the productions of the recordings, an acoustic analysis was conducted comparing target words in irregular and regular conditions. Paired t-tests revealed no differences between the target words across the regular and irregular conditions on mean duration ($t(51)=-0.58, p=0.57$), mean f_0 ($t(51)=-1.43, p=0.16$), and mean intensity ($t(51)=0.20, p=0.85$) (see Table 2).

Procedure—Participants heard separate recordings of each of the 52 items. The order of the items was randomised. They were only allowed to play each item once. After they heard the audio clip, they were presented with a textbox to report all the words they recalled from the item in any order.

Results

Two coders scored the recalled items for accuracy. Spelling errors were accepted as correct if there was no doubt as to the intended word (e.g., tigre for tiger), but not if they could be interpreted as a word with an alternate stress pattern (e.g., desert for dessert). Participants

correctly reported at least one word in 99% of the items, indicating that they could hear the recordings and were engaged in the task.

To ensure that the current result was not due to a quirk of the task in which listeners consciously identified irregular targets and strategically set out to identify them, thirty subjects on Amazon Mechanical Turk listened to the recordings used in the first study and were told to check any of the words that “stood out” or to check “none” if no words stood out to them. We did not specifically ask about metrical pattern. At the end of the task, subjects were asked why they selected the words that they did. Only two out of thirty mentioned word stress. Each word was given a score from 0 to 30 based on the number of subjects who marked it as standing out.

Recall accuracy was analyzed in a mixed effects logistic regression using the lme4 software package in R (Bates, Mächler, Bolker, & Walker, 2015). The model included a factor for metrical regularity (irregular vs. regular) pattern of surrounding context (SW vs. WS), and target status (target vs. other). In addition, we included as a predictor the number of times a word was checked as “standing out” in the normalization task described above. The random effects structure included predictors for each subject, as well as for each word, following (Barr et al. 2013). The maximal model that achieved convergence included random intercepts only.¹

The data are presented in Figure 1. This graph shows general trends in the data across participants by presenting the proportion of words recalled in each list position, grouped by regularity and word type. Two patterns are immediately apparent, and are predicted from the literature: First, the U-shaped curve shows a classic primacy and recency effect (Murdock, 1962). Second, for words that are not in the target position, SW words were more accurately recalled than WS words. This difference is likely driven by the fact that frequency of non-target words were not controlled, and SW words are more frequent in English than WS words, and frequent words are better recalled than non-frequent words (Hall 1954; Mandler 1980).

These two effects serve as manipulation checks: they suggest that listeners are sensitive to metrical patterns and that we see standard recall effects associate with ordered lists. However, they shed no light on our central question: does a switch in metrical pattern lead to a cost in recall?

Results showed a small but significant effect of regularity. Any word in an irregular list was slightly more likely to be remembered, with an increase in odds of .01 or approximately 1% . This is above and beyond any effects of memorability or frequency of individual words, which are encompassed by the random effect structure of the model.

The key result for our hypothesis is the presence of an interaction of regularity and target status: words that had the opposite metrical pattern as their surroundings (that is, words that were in target position in an irregular list) were much more likely to be remembered (an

¹For details on the random effects and the model selection process, see model code available at the following url: https://osf.io/svx5m/?view_only=53168e7af030483a82cb47fb0cf1ba74

increase in odds of .22 or about 18%) , compared to the same word surrounded by words of the same metrical pattern. For details of all effects, see the regression table in the online supplemental data.

This effect can be clearly seen when examining average subject performance. Mean accuracy scores were calculated for each participant for each word position. The distribution of those mean accuracy scores is displayed below in the boxplots in Figure 2. Boxplots for each word position show the median, first and third quartiles of the scores.² Again, it is clear that in position 3, the target position, we see an increase in words recalled in the irregular condition compared to the regular condition.

Discussion

Overall, Experiment 1 showed that a word that differs metrically from its surrounding context is more likely to be recalled. This is consistent with previous studies examining effects of phonological similarity on recall: words that were metrically dissimilar to their neighbors were better recalled. In addition, words from lists with a mismatching metrical target were better recalled overall than words from regular lists.

In Experiment 1, the same target word was recorded separately in regular and irregular items in order to preserve any acoustic details that comprehenders use in predicting metrical pattern. However, this strategy leaves open the possibility that uncontrolled acoustic differences between target words across conditions led to differences in recall rates. We test this in Experiment 2 by creating stimuli through splicing, thereby eliminating the possibility of experimenter bias or acoustic cues biasing participant responses. In Experiment 2, we test whether the irregularity effect still obtains when the *same recording* of a target word is embedded in different metrical contexts.

Experiment 2

Participants

Sixty participants who had not participated in Experiment 1 were recruited from Amazon Mechanical Turk. They performed the task online and were paid \$5 for participating. All were self-reported native English speakers with normal hearing.

Materials and Design

To create stimuli, each word was spliced out of recordings from the regular conditions from Experiment 1. Each target word was placed in position five of a nine-word item, surrounded by four words on each side. The left and right context either matched or mismatched the metrical pattern of the target. The left and right context were taken from the last four words of a regular recording. All words were taken from recordings made in the same recording session. In Experiment 1, sounds were recorded continuously and therefore between the words was ambient noise present in the sound attenuating booth. To mimic this with spliced

²Notches are $(1.58 * \text{the interquartile range}) / (\text{sqrt}(n))$. This gives a roughly 95% confidence interval for comparing medians. if the notches of two boxes do not overlap, this suggests that the medians are significantly different.

stimuli, in Experiment 2 and 3 words were spliced together with a 500ms ISI of ambient noise from the recording booth.

In order to keep metrical patterns intact, the flanking context of four words consisted of words that were originally recorded as a sequence. The right context, left context, and target words were all taken from recordings in the regular condition in Experiment 1, and were then recombined using splicing to create both regular and irregular conditions for Experiments 2 and 3. The target words were taken from items that differed from those from which the surrounding contexts were extracted, eliminating the possibility that acoustic cues from the context would yield systematic differences across conditions in recalling the target. To control the inter-stimulus interval (ISI) between words, which varied in the natural recordings, these items were spliced apart and then spliced back together with a 500ms ISI. All words were normalised to 65 dB using Praat (Boersma & Weenink, 2017) so that words recorded in separate sessions would nonetheless sound similarly loud. Each word was repeated twice throughout Experiment 2. Participants were exposed to an equal number of conditions, which were counterbalanced in a Latin square design.

Procedure

Participants heard separate recordings of each of the 52 items. The order of the items was randomised. They were only allowed to play each item once. After they heard the audio clip, they were presented with a textbox to report all the words they recalled from the item in any order. Due to browser incompatibility problems, two recordings did not play for a majority of the participants. Data from these two recordings were discarded for all participants, and only data from the remaining 50 items were analyzed.

Results

Responses were coded as correct or incorrect by the first author. As in Experiment 1, spelling errors were accepted as correct if there was no doubt as to the intended word (e.g., *chaufer* for *chauffeur*), but not if they could be interpreted as a word with an alternate stress pattern (e.g., *moral* for *morale*). Participants correctly reported at least one word in 98% of the items, indicating that they could hear the recordings and were engaged in the task.

Recall accuracy was analyzed in a mixed effects logistic regression using the *lme4* software package in R (Bates et al., 2015). The model included a factor for metrical regularity (irregular vs. regular) pattern of surrounding context (SW vs. WS), and target status (target vs. other). The random effects structure included intercepts for each subject, as well as for each word. The maximal model that achieved convergence included random intercepts only.

Results (see figure 3 & 4 below) mirrored those in Experiment 1. As in Experiment 1 and in previous literature, there was an evident show a primacy and recency effect, as well as an advantage for the more frequent SW words, indicating that our manipulations worked as expected.

The statistical model showed small but significant effect of regularity: Any word in a regular list was remembered with slightly lower accuracy, (a decrease in odds of .01, or

approximately 1%,). This is above and beyond any effects of memorability or frequency of individual words, which are encompassed by the random intercepts of the model. Again, the key result for our hypothesis is the presence of an interaction of regularity and target status: words that had the opposite metrical pattern as their surroundings (that is, words that were in target position in an irregular list) were much more likely to be remembered (an increase in odds of about .25 or just over 20%), compared to the same word surrounded by words of the same metrical pattern.

Discussion

Experiment 1 demonstrates that a word that has a metrical structure that mismatches its surrounding context is more likely to be recalled. Experiment 2 replicates this effect in sentences that were controlled acoustically. Experiments 1 and 2 used the same metrical pattern of irregularity: one isolated word, surrounded by words of the opposite stress pattern. The effects appear to be driven chiefly by this aberrant target word.

The data suggest that words that do not match their metrical context are more likely to be recalled. This is consistent with both the literature on the isolation effect, as well theories proposing that disruptions in regularity draw listeners' attention and lead to deeper processing (Calhoun, 2007; Cutler & Darwin, 2017; Cutler & Foss, 1977; Kakouros & Räsänen, 2015).

General Discussion

Previous work has shown an advantage for metrical regularity, particularly in tasks measuring online processing. It has been argued that these advantages are a direct result of regularity, which enables comprehenders to entrain to a pattern and use predictive processing for upcoming speech. However, previous work on memory also shows that words that stand out from their neighbors are better attended to and remembered, and that strings of phonologically similar words are poorly remembered. We tested memory for lists of words that were either metrically regular or had one word that metrically mismatched its neighbors. We found that words that mismatched were better recalled. We believe that this result is important for two reasons.

First, while metrical regularity has been shown to advantage processing, irregularity has been found to have an effect on comprehension in speech prosody. Breaks in intonation patterns are perceived as prominent (Calhoun, 2007; Cutler & Darwin, 2017; Cutler & Foss, 1977), and metrical structure specifically has been identified as useful for predicting upcoming prosody in work by Shields and colleagues (1974). In addition, it has been shown that abrupt changes in pitch enhance recall (Frankish, 1995). Further direct evidence of the connection between (un)predictability and prominence is found in the work of Kakouros and Räsänen (2015) who find that perception of stress is correlated with the unpredictability of certain acoustic correlates of prosody. Kakouros and Räsänen take this as evidence that stress perception is connected to unpredictability and auditory attention. Our results are in line with this literature, and suggest further that words that break prosodic regularity, in this case metrical stress, are processed more deeply by listeners, resulting in improved recall.

Another reason why it is crucial to consider metrical *irregularity* in models of speech comprehension is that irregularity may be the norm in everyday spontaneous speech. While many formal linguistic theories of metrical structure argue that speech has a regular, hierarchical metrical structure (Hammond, 1999; Hayes, 1995; Nespor & Vogel, 1989; Selkirk, 1984), empirical measurements do not consistently support this argument. First of all, acoustic measurements have repeatedly shown that stressed syllables are not isochronous (i.e. evenly spaced in time) (Dauer, 1983; Grabe & Low, 2002; Lehiste, 1977). The acoustic signal itself is, at best, only “quasi-regular” (Peelle & Davis, 2012). Some results indicate that speech is perceived as being regular, even if it is acoustically irregular (Beckman, 1992; Dilley, Wallace, & Heffner, 2012; Laver, 1994), though listeners may not agree on the location or patterns of word stress (Vogel, Bunnell, & Hoskins, 1995). Corpus studies which attempt to measure metrical regularity similarly show mixed results. Analysis of stressed syllable locations in two corpora annotated for prosody show that regularity is present in both, but it is much more common in the formal speech of radio announcers in the Boston University Radio News Corpus (Ostendorf, Price, & Shattuck-Hufnagel, 1995) than in casual, telephone conversations from 45 minutes of the Switchboard Transcription Project (Greenberg, Hollenback, & Ellis, 1996; Temperley, 2008). If regularity provides wide-ranging benefits to communication, why has it been so difficult to detect by language scientists? One answer to this question is that there may be a benefit to metrically unexpected linguistic information. If occasional irregularity provides a memory or processing benefit, widespread irregularity in language is less surprising.

A key difference between this study and most previous work is the timescale on which responses are measured. Previous work in the framework of Dynamic Attending Theory and the Attentional Bounce Hypothesis shows that stimuli are better perceived or more accurately judged when they are presented in time with an entrained rhythm (e.g., Zheng & Pierrehumbert 2010). This effect holds both for stimuli that are similar to their neighbors, and oddballs that are presented either in sync or out of sync with a temporal pattern (McAuley & Fromboluti 2014). However, these studies do not test the effects of pattern regularity on later memory as our study does (although see Dilley & McAuley, 2008 for an example of using recognition memory to explore effects of metrical structure on lexical processing). These results motivate further investigation into whether processing advantages in the moment are directly linked to later downstream comprehension and memory.

Of course, there are also limitations to the current study. We used lists of words as stimuli in order to precisely control metrical contexts, but memory for lists of words likely differs from memory for conversational speech in important ways. More work is needed to see whether the current results generalize to natural speech. Future work might also consider the role of musical training or musicality on this type of task, as it has been shown that musicians may process speech rhythm differently than non-musicians (Harding et al., 2019; Marie, Magne & Besson 2011). Finally, this study concerns alternating metrical patterns, but does not investigate the spacing of syllables in time, i.e. isochrony or perceptual isochrony. Future work that includes uncontrolled samples should take both into account, as even perfectly regular alternating stress patterns could sound irregular if their timing was highly uneven.

To conclude, we find that words that have a different metrical pattern than their surrounding context are better recalled than words whose metrical patterns match their context. Based on these results, we argue that listeners are attuned to metrical structure, and that while regular metrical patterns lead to entrainment and better processing in the moment, breaks in those same patterns play a beneficial role in language retention and memory.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This project was supported by Grant Number R01DC008774 from the National Institute on Deafness and Other Communication Disorders and a grant from the James S. McDonnell Foundation. Special thanks to Daniel Escobar, Sara Geoghegan, and Minqi Wang for data coding and running subjects. Portions of this research were presented at Experimental and Theoretical Approaches to Prosody 3, Urbana, IL, May 2015.

References

- Acheson DJ, & MacDonald MC (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, 135(1), 50–68. 10.1037/a0014411 [PubMed: 19210053]
- Acheson DJ, & MacDonald MC (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language*, 65(2), 193–207. [PubMed: 21743771]
- Altman G, & Carter D (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech & Language*, 3(3), 265–275. 10.1016/0885-2308(89)90022-3
- Baayen RH, Piepenbrock R, & Gulikers L (1995). CELEX2 LDC96L14 Web Download. Philadelphia: Linguistic Data Consortium.
- Baddeley AD (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, 18(4), 362–365. 10.1080/14640746608400055 [PubMed: 5956080]
- Barnes R, & Jones MR (2000). Expectancy, attention, and time. *Cognitive psychology*, 41(3), 254–311 [PubMed: 11032658]
- Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. 10.1016/j.jml.2012.11.001
- Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. 10.18637/jss.v067.i01
- Bauer A-KR, Jaeger M, Thorne JD, Bendixen A, & Debener S (2015). The auditory dynamic attending theory revisited. *Brain Research*, 1626(c), 198–210. 10.1016/j.brainres.2015.04.032 [PubMed: 25934332]
- Beckman ME (1992). Evidence for Speech Rhythms Across Languages In *Speech Perception, Production, and Linguistic Structure* (pp. 457–463). IOS Press.
- Birch S, & Clifton C (1995). Focus, Accent, and Argument Structure: Effects on Language Comprehension. *Language and Speech*, 38, 365–391. [PubMed: 8816086]
- Bock JK, & Mazzella JR (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, 11(1), 64–76. [PubMed: 6855561]
- Boersma P, & Weenink D (2017). Praat: Doing Phonetics by Computer. Available form: URL: <http://www.fon.hum.uva.nl/praat>.

- Bohn K, Knaus J, Wiese R, & Domahs U (2013). The influence of rhythmic (ir) regularities on speech processing: evidence from an ERP study on German phrases. *Neuropsychologia*, 51(4), 760–771. [PubMed: 23333869]
- Bousfield WA, & Cohen BH (1955). The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. *The Journal of General Psychology*, 52(1), 83–95.
- Breen M, Dilley LC, McAuley JD, & Sanders LD (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, cognition and neuroscience*, 29(9), 1132–1146.
- Brown M, Salverda AP, Dilley LC, & Tanenhaus MK (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2), 306–323. 10.1037/a0038689 [PubMed: 25621583]
- Butterfield S, & Cutler A (1988). Segmentation errors by human listeners: Evidence for a prosodic segmentation strategy. In *Seventh Symposium of the Federation of Acoustic Societies of Europe* (pp. 827–833). Institute of Acoustics.
- Calhoun S (2007). Information structure and the prosodic structure of English : a probabilistic relationship. University of Edinburgh.
- Cason N, & Schön D (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, 50(11), 2652–2658. [PubMed: 22828660]
- Clopper CG (2002) Frequency of stress patterns in English: A computational analysis. *Indiana University Linguistics Club Working Papers* 2, 1–9
- Conrad R (1964). Acoustic Confusions in Immediate Memory. *British Journal of Psychology*, 55(1), 75–84. 10.1111/j.2044-8295.1964.tb00899.x
- Conrad R, Freeman PR, & Hull AJ (1965). Acoustic factors versus language factors in short-term memory. *Psychonomic Science*, 3(1–12), 57–58. 10.3758/BF03343017
- Cutler A, & Carter DM (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3), 133–142.
- Cutler A, & Darwin C (2017). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics*, 29(3), 217–224.
- Cutler A, & Foss DJ (1977). On the role of sentence stress in sentence processing, 20(1), 1–10.
- Cutler A, McQueen JM, Norris D, & Somejuan A (2001). The roll of the silly ball. In *Language, brain, and cognitive development Essays in honor of Jacques Mehler* (pp. 181–194).
- Cutler A, Junge C, Spokes T, & Kidd E (2018). Phonological acquisition: Stress-based segmentation in English. Paper at Laboratory Phonology (LabPhon16) Lisbon, Portugal.
- Cutler A, & Norris D (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance*, 14(1), 113.
- Dauer RM (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62.
- Dilley LC, & McAuley JD (2008). Distal prosodic context affects word segmentation and lexical processing, 59(3), 294–311.
- Dilley L, Wallace J, & Heffner C (2012). Perceptual isochrony and fluency in speech by normal talkers under varying task demands. In Niebuhr O & Pfitzinger H (Eds.), *Prosodies Context, Function, and Communication* (pp. 237–258). Berlin.
- Domahs U, Klein E, Huber W, & Domahs F (2013). Good, bad and ugly word stress–fMRI evidence for foot structure driven processing of prosodic violations. *Brain and language*, 125(3), 272–282. [PubMed: 23587493]
- Duncan J, & Humphreys GW (1989). Visual search and stimulus similarity. *Psychological review*, 96(3), 433. [PubMed: 2756067]
- Fairhall AL, Lewen GD, Bialek W, & van Steveninck RRDR (2001). Efficiency and ambiguity in an adaptive neural code. *Nature*, 412(6849), 787. [PubMed: 11518957]
- Fletcher PC, Anderson JM, Shanks DR, Honey R, Carpenter TA, Donovan T, ... & Bullmore ET (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature neuroscience*, 4(10), 1043. [PubMed: 11559855]

- Frankish C (1995). Intonation and Auditory Grouping in Immediate Serial Recall. *Applied Cognitive Psychology*, 9, 5–22.
- Geiser E, Zaehle T, Jancke L, Meyer M, 2008 The neural correlate of speech rhythm as evidenced by metrical speech processing. *J. Cogn. Neurosci* 20, 541–552. [PubMed: 18004944]
- Grabe E, & Low EL (2002). Durational variability in speech and the rhythm class hypothesis. In Gussenhoven C & Warner N (Eds.), *Laboratory Phonology 7* Berlin.
- Greenberg S, Hollenback J, & Ellis D (1996). The Switchboard transcription project. LVCSR Summer Workshop Technical Reports, 72.
- Hall JF (1954). Learning as a function of word-frequency. *The American journal of psychology*, 67(1), 138–140. [PubMed: 13138778]
- Hammond M (1999). *The Phonology of English : A Prosodic Optimality-Theoretic Approach*. Oxford University Press, UK.
- Harding EE, Sammler D, Henry MJ, Large EW, & Kotz SA (2019). Cortical tracking of rhythm in music and speech. *NeuroImage*, 185, 96–101. [PubMed: 30336253]
- Hayes B (1995). *Metrical Stress Theory: Principles and Case Studies*.
- Henrich K, Alter K, Wiese R, & Domahs U (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and language*, 136, 19–30. [PubMed: 25113242]
- Hunt RR (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, 2(1), 105–112. 10.3758/BF03214414 [PubMed: 24203592]
- Itti L, & Baldi P (2005). A Principled Approach to Detecting Surprising Events in Video. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 1–7.
- Itti L, & Baldi P (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306. 10.1016/j.visres.2008.09.007 [PubMed: 18834898]
- Jaeger TF, Furth K, & Hilliard C (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1439–1449. 10.1037/a0027862
- Jones MR (2010). Attending to sound patterns and the role of entrainment In *Attention and Time* (pp. 317–330). Oxford: Oxford University Press.
- Jones MR, Moynihan H, MacKenzie N, & Puente J (2002). Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. *Psychological Science*, 13, 1–7. [PubMed: 11892772]
- Kandylaki KD, Henrich K, Nagels A, Kircher T, Domahs U, Schlesewsky M, ... & Wiese R (2017). Where is the beat? The neural correlates of lexical stress and rhythmical well-formedness in auditory story comprehension. *Journal of Cognitive Neuroscience*, 29(7), 1119–1131. [PubMed: 28294714]
- Kakouros S, & Räsänen O (2015). Perception of Sentence Stress in Speech Correlates With the Temporal Unpredictability of Prosodic Features. *Cognitive Science*, 40(7), 1739–1774. 10.1111/cogs.12306 [PubMed: 26481111]
- Large E, & Jones MR (1999). The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review*, 106(1).
- Laver J (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lehiste I (1977). Isochrony Reconsidered. *Journal of Phonetics*, 5(3), 253–263.
- MacDonald MC, & Christiansen MH (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54. 10.1037/0033-295X.109.1.35 [PubMed: 11863041]
- Mandler G (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, 87(3), 252.
- Marie C, Magne C, & Besson M (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, 23(2), 294–305. [PubMed: 20044890]
- McAuley JD, & Fromboluti EK (2014). Attentional entrainment and perceived event duration. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1658), 20130401. [PubMed: 25385779]

- McAuley JD, & Jones MR (2003). Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1102. [PubMed: 14640833]
- Murdock BB Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482–488. 10.1037/h0045106
- Nazzi T, & Ramus F (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41(1), 233–243. 10.1016/S0167-6393(02)00106-1
- Nazzi T, Bertoncini J, & Mehler J (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756–766. 10.1037/0096-1523.24.3.756 [PubMed: 9627414]
- Nespor M, & Vogel I (1989). On Clashes and Lapses. *Phonology*, 6(1), 69–116.
- O'Seaghdha PG, & Marin JW (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 57–73. 10.1037/0096-1523.26.1.57 [PubMed: 10696605]
- Ostendorf M, Price PJ, & Shattuck-Hufnagel S (1995). The Boston University radio news corpus. Linguistic Data Consortium.
- Peelle JE, & Davis MH (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, 3 10.3389/fpsyg.2012.00320
- Pitt MA, & Samuel AG (1990). The use of rhythm in attending to speech., 16(3), 564–573. 10.1037/0096-1523.16.3.564
- Heffner CC, Dilley LC, McAuley, & Pitt JD, M. A., (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, 28(9), 1275–1302. 10.1080/01690965.2012.672229
- Quené H, & Port RF (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, 62(1), 1–13. 10.1159/000087222 [PubMed: 16116301]
- Ramus F (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, 2, 85–115.
- Ranganath Charan, and Rainer Gregor. “Cognitive neuroscience: Neural mechanisms for detecting and remembering novel events.” *Nature Reviews Neuroscience* 4.3 (2003): 193. [PubMed: 12612632]
- Rothermich K, Schmidt-Kassow M, & Kotz SA (2012). Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia*, 50(2), 232–244. [PubMed: 22178743]
- Rothermich K, Schmidt-Kassow M, Schwartz M, & Kotz SA (2010). Event-related potential responses to metric violations: rules versus meaning. *Neuroreport*, 21(8), 580–584. [PubMed: 20431495]
- Rothermich K, & Kotz SA (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *Neuroimage*, 70, 89–100. [PubMed: 23291188]
- Schock J, Cortese MJ, & Khanna MM (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, 44(2), 374–379. 10.3758/s13428-011-0162-0 [PubMed: 22037976]
- Schultz W, & Dickinson A (2000). Neuronal coding of prediction errors. *Annual review of neuroscience*, 23(1), 473–500.
- Selkirk EO (1984). *Phonology and Syntax: the relation between sound and structure*. Cambridge, MA: The MIT Press.
- Sevald CA, & Dell GS (1994). The sequential cuing effect in speech production. *Cognition*, 53(2), 91–127. [PubMed: 7805353]
- Shattuck-Hufnagel S (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns In Keller E & Gopnik M (Eds.), *Neuropsychology and neurolinguistics. Motor and sensory processes of language* (pp. 17–51). Lawrence Erlbaum Associates, Inc.
- Shepard RN (1967). Recognition memory for words, sentences, and pictures. *Journal of Memory and Language*, 6(1), 156.
- Shields J, McHugh A, & Martin JG (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech, 102(2), 250–255.

- Temperley D (2008). Distributional Stress Regularity: A Corpus Study. *Journal of Psycholinguistic Research*, 38(1), 75–92. 10.1007/s10936-008-9084-0 [PubMed: 18937069]
- Tilsen S (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, 2(1). 10.1515/labphon.2011.006
- Underwood BJ, Ekstrand BR, & Keppel G (1965). An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. *Journal of Memory and Language*, 4(6), 447.
- Vogel I, Bunnell T, & Hoskins S (1995). The Phonology and Phonetics of the Rhythm Rule In Connell B & Arvaniti A (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 111–127). Cambridge University Press 10.1017/cbo9780511554315
- Weber C, Hahne A, Friedrich M, & Friederici AD (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research*, 18(2), 149–161. 10.1016/j.cogbrainres.2003.10.001 [PubMed: 14736574]
- Winkler I, Horváth J, Weisz J, & Trejo LJ (2009b). Deviance detection in congruent audiovisual speech: Evidence for implicit integrated audiovisual memory representations. *Biological Psychology*, 82(3), 281–292. 10.1016/j.biopsycho.2009.08.011 [PubMed: 19733617]
- Yiu LK, & Watson DG (2015). When overlap leads to competition: Effects of phonological encoding on word duration. *Psychonomic Bulletin & Review*, 22(6), 1701–1708. 10.3758/s13423-015-0828-1 [PubMed: 25855202]
- Zheng X, & Pierrehumbert JB (2010). The effects of prosodic prominence and serial position on duration perception. *The Journal of the Acoustical Society of America*, 128(2), 851–859. 10.1121/1.3455796 [PubMed: 20707454]

Figure 1.
Recall accuracy by condition for Experiment 1

Figure 2. Boxplot showing the distribution of mean words recalled across word types. Target words were always in position 3.

Figure 3.
Recall accuracy by condition for Experiment 2

Figure 4. Boxplot showing the distribution of mean words recalled across word types. Target words were always in position 5.

Table 1

Example items from all four conditions

	pattern	regularity	item
1	SW	regular	cupcake, elbow, candy, chicken, pocket, morning, stomach, ocean
2	WS	regular	regime, baboon, guitar, award, debris, offense, revenge, hello
3	SW	irregular	rabbit, mountain, cashier, beetle, student, curtain, apple, engine
4	WS	irregular	tonight, request, pumpkin , allure, degree, brochure, cachet, routine

Note: SW words are in bold

Table 2

Acoustic measures of target words in regular vs. irregular conditions

Regularity	Mean Duration (s)	Mean f0 (Hz)	Mean Intensity (dB)
Regular	0.79 (0.12)	185.45 (38.11)	55.85 (3.73)
Irregular	0.77 (0.12)	176.14 (22.69)	56.01 (3.85)

Table 3

Example items from all four conditions, Experiment 2

pattern	regularity	left context	target	right context
SW-SW-SW	regular	penguin, illness, husband, picture	nickel	chapter, zebra, coffee, dresser
WS-WS-WS	regular	estate, domain, buffoon, croquet	giraffe	descent, prestige, amount, latrine
SW-WS-SW	irregular	virus, bucket, hamster, music	giraffe	photo, beaker, journey, minute
WS- SW -WS	irregular	array, consent, parade, haiku	nickel	critique, ellipse, attempt, chauffeur

Note: SW words are in bold