Timing is Stressful:

Do Listeners Combine Meaning and Rhythm to Predict Speech?

By

Molly H. McLeod

* * * * * * * * *

Submitted in partial fulfillment

of the requirements for

Honors in the Department of Psychology

UNION COLLEGE

June, 2022

Abstract

MCLEOD, MOLLY Timing is Stressful: Do Listeners Combine Meaning and Rhythm to Predict Speech? Department of Psychology, June 2022.

ADVISOR: Chad Rogers

English and other languages such as German are stress-timed languages: the timing of the speech is determined by stressed and unstressed syllables, providing structure for sentences. While natural speech is not generally metrically regular, like in Shakespearean poetry, it still conveys timing cues through stress. Prior research has found that metric regularity enhances the processing of words (Rothermich et al, 2012), potentially because it attunes listeners' attention to the predictability of stressed, and therefore important, syllables. Other work (e.g., Rogers, 2017) has suggested that predictability in the form of semantic associations (e.g., hearing "barn" facilitates understanding of "hay") is a driving force for speech understanding, so much so that people falsely "hear" words predicted by semantic context (e.g., hearing "barn" leads to hearing "hay", even if "pay" was presented).

In the current study, we aimed to examine how stress patterns and semantic associations may interact in listeners' understanding of speech, as they both provide bases for predictions on the part of the listener. We measured speech understanding by masking the final word of a sentence in noise, then asking participants to identify what that word was (e.g., Jake visits the park to walk his DOG). We manipulated each sentence's rhythmic predictability (whether the sentence was in natural speech, with a rhythm emphasized, or with a drum beat matching rhythm preceding the rhythmic speech) and semantic predictability (whether the last word made sense with the sentence, e.g. Jake visits the park to walk his dog/log). There was also a baseline condition for each of the rhythmic conditions wherein the sentence predictability was low. The results indicated that the beat prime improved processing of the rhythmic speech in conditions where expectancy effects played a role (semantically congruent and incongruent) but had a negligible impact in the baseline condition.

Keywords: semantic priming, rhythmic priming, speech-in-noise

Acknowledgements

Foremost I would like to thank my thesis supervisor, Dr. Chad Rogers, for his instrumental role in this project, walking me through so many steps and encouraging my progress. I've learned so much from him, and I've loved the chance to get excited about science! I would also like to thank the rest of the SPAN lab team for their help: Ummi Coats for piloting the study, Abe Chait for assisting with audio file segmentation, Isa Albuja for explaining Gorilla and Prolific a number of times, and to Avanti Khare, Hannah Soliman, and Travis Rathwell for their support in the lab.

Thank you also to the participants of the study.

I would like to acknowledge the Union College Undergraduate Research Program, which provided the funding that made this study possible.

I would also like to thank my friends and family; thank you so much for supporting me, and for listening when I incoherently went on about rhythm patterns and got very excited about the IPA chart.

Table of Contents

Introduction

Timing provides us with important cues in all parts of life, but in language processing certain dimensions are incredibly relevant. In metrical poetry like in Shakespeare or Dickinson, the timing of the sentences is regular and follows a pattern, in English commonly divided into iambs (unstressed, **stressed**: "Shall **I** com**pare** thee **to** a **sum**mer's **day**?"(Shakespeare, 1609). In everyday speech, people don't tend to speak in meter, but there are still underlying stress patterns. English is a stress-timed language, meaning that speakers tend to adjust the speed of their speech so that stressed syllables are roughly equidistant temporally from each other (e.g. Nespor et al., 2011). Stress can also provide important grammatical and semantic information, such as in individual words (i.e. **Re**cord is a noun, but shift the stress to the second syllable and re**cord** is a verb). Another type of information that listeners get from language is semantic context. Semantic congruence refers to when the meanings of words make sense together, and semantic incongruence when they do not match in meaning. In this experiment, we aim to examine the effects that rhythmic and semantic context provide to the listener, i.e. if they work together or are separate processes.

Semantic context can provide a number of cues to help people better process language. It is a much-studied domain in the field of language processing, with the general conclusion that semantic priming facilitates word recognition. One of the seminal studies in the field, Meyer & Schvaneveldt (1971), used a lexical decision task measuring how quick participants were to judge visual two strings as either words or nonwords. They found that people were the fastest at declaring strings to be words if the two were semantically associated (e.g. HOUSE and

BUILDING). The implication, then, is that the semantic cue provides information which helps people make a judgment call about the second word on the basis of meaning.

Other studies, including Rogers (2017) have repeated variations of this task using an auditory method, measuring participants' response rates identifying words in noise that were primed by either a semantically unrelated word (e.g. JAW-PASS), semantically related word (e.g. ROW-BOAT), or a word semantically related to a phonological neighbor of the word (e.g. ROW-GOAT). Participants generally did very well in the congruent condition (ROW-BOAT), and less well in the baseline condition (JAW-PASS); these findings indicate that the semantic association in the congruent case facilitated correct hearing. The results also found, however, that participants were more likely to mishear in the misleading conditions (i.e. reporting having heard BOAT when they were presented with GOAT). This phenomenon is called false hearing. These findings indicate that not only does semantic context appear to increase intelligibility, but in some cases the influence of semantic context is so strong that it may lead to false perceptions. Age has been found to be a significant predictor of false hearing, too; the change in abilities over time indicates that older adults are more dependent on semantic context for listening. This may indicate an increased reliance on heuristics, i.e. patterns of interpreting speech that they have developed over time, relying more heavily on top-down listening (Rogers, Jacoby, & Sommers, 2012).

Semantic cues provide important information about *what* will occur: for example, a listener hearing "He mailed the letter without a…" might be ready to hear "stamp." In addition, though, the specific timing, the *when*, also plays an important role in speech perception. One specific application of rhythm is musical rhythm. Haegens and Golumbic's (2018) analysis of musical speech priming reveals that rhythm in its temporal sense is connected to language processing, and that training in this area can improve speech processing performance, including for those who have deficits in their language processing. Recent studies have looked into specific effects of musical rhythm on speech perception; Cason et al. (2015) found that rhythmic priming (i.e. a match between the rhythmic structures of a priming sentence and a target sentence) facilitated faster phoneme detection in the target sentence. The implications of this analysis are that a greater understanding of how rhythmic structures improve performance may be able to facilitate strategies for helping people to better comprehend language.

Cason et al. (2015) measured their outcomes in terms of phoneme distinction, which allowed them to determine the speed of processing given rhythmic cues. Another important aspect to look into, though, is how the semantic information in a given sentence is integrated into this process.Rothermich et al. (2012) examined this connection between semantic and timing cues by measuring N400 amplitude, an ERP response which has been shown to be associated with semantic ease of processing. In this study, they manipulated metrical congruence by creating sentences that were either metrically regular (bisyllabic trochaic, to match common German speech patterns) or metrically irregular. They manipulated semantic congruence by having endings to the sentences which either matched the meaning of the sentence (e.g., in English translation, "*Norbert picked last Tuesday Gina's roses and carnations*") or mismatched ("*Stefan picked last Tuesday Maren's pipes and cables*"). They found that participants had a lowered N400 component for semantically irregular words in the metrically regular condition, indicating that the metrical regularity of the sentence aided lexical processing.

In the current study, we aim to examine the relationship between semantic integration and rhythmic priming. The Rothermich et al. (2012) study found that metrical regularity facilitated semantic integration, but would the same be true for more "natural" speech that is not based on a metrical pattern? In the current study, we will examine non-metrical speech to determine if priming participants with a sentence's rhythmic pattern beforehand will enhance expectancy effects and result in better distinction of words in noise. We expect to find a benefit of semantic priming (as in Meyer & Schvaneveldt, 1971; Rogers, 2017; etc.). We hypothesize that listeners will perform best in the semantically congruent case, and worse in the baseline and incongruent cases. Additionally, given that meter provides a basis for phoneme facilitation in the Rothermich et al. study (2012), and that priming listeners with the rhythm of a sentence facilitated phoneme detection (Cason et al., 2015), we would hypothesize that the beat prime condition would help the most in that congruent case by providing additional information, making the semantically congruent rhythmic speech + beat condition the most successful for participants. Under an attentional model, we would expect that a greater basis for expectation would result in better performance in the incongruent condition, also following Rothermich et al.'s (2012) findings that metric regularity facilitated lexico-semantic integration.

Methods

All stimuli, data, and pre-registrations are publicly available for viewing on the Open Science Foundation website:
https://osf.io/8bns4/?view_only=73ff7413186349fcab257f888a5ea3e4

Participants

The participants in this study were drawn from the survey site Prolific.co. All participants were native monolingual English speakers from the United States. Participants were paid $2.50 per quarter hour for their time.

Materials

For the stimuli in this task, sentences were taken from the completion norm study conducted by Peelle et al. (2020). The 90 baseline sentences, intended to reflect a situation where the last word would not be predictable, were among the lowest entropy sentences out of the 3085 collected in the Peelle study. The last words for these sentences were selected from among the responses given by participants of the Peelle study, under the conditions that they were semantically sensible and one-syllable (e.g. "The man trembled at the thought of a… fight"). The 180 experimental sentences were chosen from among those with the highest response percentages, i.e. the most participants answered with the same word (e.g. "On sunny days, Jake visits the park to walk his… dog"). For each of these sentences, the word which the Peelle et al. participants identified as completing the sentence served as the semantically congruent condition. In addition, another word was generated for each sentence to fit semantic incongruity while retaining phonetic similarity (e.g. "On sunny days, Jake visits the park to walk his… log"). Each of these incongruent words was either a minimal pair or had only 2-3 phonetic changes (i.e. in manner, place of articulation, voicing).

Each sentence was also recorded in a number of rhythmic conditions. For each sentence, a rhythmic beat was recorded following the stress patterns of each sentence using MIDI technology in the program Garageband. The beats were recorded using the Bongos Performance sounds from the Garageband library. The sentences were first recorded by a native English speaker in natural speech, and then again in rhythmic speech in time with the beat recordings. All recordings were made in the program Audacity using an Audio Technica 2035 microphone. The sentences and words were leveled to -25dB and the words given a 5ms cosine volume ramp in the beginning using Adobe Audition. The final words for each sentence were recorded separately in the same conditions and were masked by speech-shaped noise. The speech shaped noise was generated according to the frequencies of the recorded stimuli in PRAAT, and overlaid onto the words using Adobe Audition.

Table 1 displays the nine different conditions that the participants were presented with. Each condition had 30 associated trials. Trials were counterbalanced using a Latin square, such that each participant received all of the sentences across the study, but only ever heard each sentence once. There were a total of six different versions of the study, counterbalancing the conditions that a participant might receive.

Table 1: Experimental Conditions

	Natural Speech	Rhythmic Speech	Rhythmic Speech With Beat Prime
Semantically Congruent	On sunny days, Jake visits the park to walk his dog.	On **sun**ny **days**, Jake **vi**sits the **park** to **walk** his **dog**.	◡ — ◡ — ◡ — ◡ ◡ — ◡ — ◡ — On **sun**ny **days**, Jake **vi**sits the **park** to **walk** his **dog**.
Semantically Incongruent	On sunny days, Jake visits the park to walk his log.	On **sun**ny **days**, Jake **vi**sits the **park** to **walk** his **dog**.	◡ — ◡ — ◡ — ◡ ◡ — ◡ — ◡ — On **sun**ny **days**, Jake **vi**sits the **park** to **walk** his **log**.
Baseline	The man trembled at the thought of a fight.	The **man trem**bled at the **thought** of a **fight.**	◡ — — ◡ ◡ ◡ — ◡◡ — The **man trem**bled at the **thought** of a **fight.**

N.b. — or **bold** indicate a stressed syllable, ◡ or normal text indicate an unstressed syllable

Procedure

The experiment was administered to participants using the online data collection tool Gorilla.sc. Participants were instructed to set up their listening environment to provide the best audio conditions, wifi quality, and minimal distractions. The basic structure of the task remained the same across all of the conditions: after listening to the stimulus, participants were asked to identify the final word of each sentence and type it out. The participants, after the task had been explained to them, were initially presented with a sequence of nine practice trials so they could get used to the task. After, they entered the main part of the experiment in which they were presented with 270 trials, of all nine conditions, in a random order. In the natural speech and rhythmic speech conditions, the spoken sentences were the only stimulus. For the rhythmic speech with beat prime conditions, participants were additionally primed with the associated beat before they heard the sentence. After completing the experimental trials, participants answered questions regarding their listening environment and completed a demographic questionnaire. Participants were then thanked for their participation and debriefed on the nature of the experiment.

Results

We report all effects found to be significant ($p < .05$) that are not otherwise involved in a higher order interaction. Post-hoc F-tests applying a Bonferroni correction for Type I error rates were used to determine the source of the interaction. If Mauchly's test for sphericity was found to be significant we applied the Greenhouse-Geyser correction for F values and degrees of freedom.

Figure 1 displays the accuracy rates of participants in each of the nine conditions. As shown in the graph, the different conditions yielded different rates. Participants were worse in the incongruent trials than the other conditions, but the specifics of the interactions reveal a more complicated pattern.

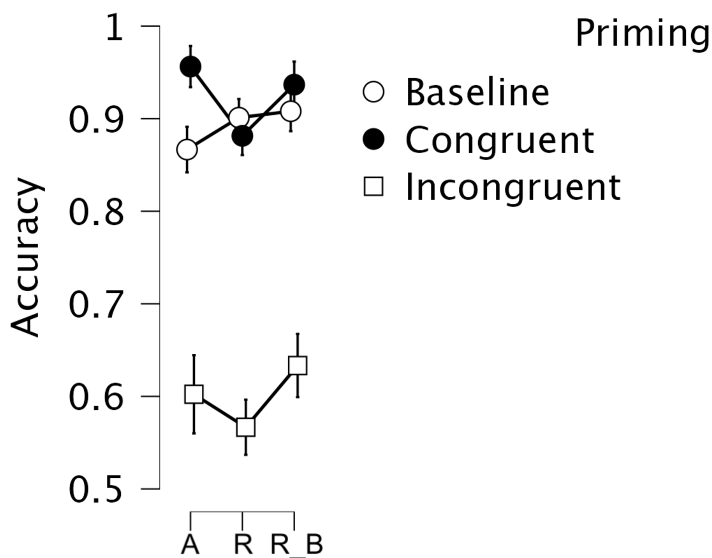

Fig. 1 Accuracy, based on hit rates, for each condition. A = Arhythmic speech, R = Rhythmic speech, and R_B = Rhythmic speech primed with a beat.

To confirm statistical reliability of these results, we ran a 3 (Semantic Congruence: Congruent, Incongruent, Baseline) X 3 (Rhythmic Priming: Arhythmic, Rhythmic, Beat + Rhythmic) mixed-model repeated measures ANOVA on correct identification rates across all participants. The ANOVA revealed significant main effects of both Semantic Priming, $F(1.20, 33.46) = 214.65$, $p < .001$, $\eta_p^2 = .89$, and Rhythmic Priming, $F(1.96, 54.81) = 15.69$, $p <.001$, $\eta_p^2 = .36$. There was also a significant interaction between Semantic Priming x Rhythmic Priming ($F[2.95, 82.53] = 8.88$, $p < .001$, $\eta_p^2 = .24$).

To follow up on the interaction, we performed a series of post-hoc tests to determine where the interaction occurred. In the baseline trials, performance did not not significantly differ between rhythmic conditions ($M_{Arhythmic} = .87$, $M_{Rhythmic} = .90$, $M_{Rhythmic\_Beat} = .90$, all $p > .05$). In congruent and incongruent conditions, however, results were significantly different. For congruent trials, participants performed better when the rhythmic speech was preceded by a beat ($M = .94$) than when it was not ($M = .88$, $t = 4.03$, $p < .01$). This was also true in incongruent trials ($M_{Rhythmic\_Beat} = .63$, $M_{Rhythmic} = .57$, $t = 4.86$, $p < .001$).

The difference between the Arhythmic and Rhythmic conditions for each semantic category was only significant in the congruent case, with participants performing better in the Arhythmic condition ($M = .96$) than in the Rhythmic ($M = .88$, $t = 5.46$, $p < .001$). There were no significant differences between the Arhythmic and Rhythmic + Beat in any of the semantic conditions.

Discussion

The results of the current experiment showed that the effects of the rhythmic priming differed based on the type of semantic priming, indicating an interactive relationship between rhythmic and semantic processing. As predicted by the prior research (e.g. Rogers, 2017), participants were significantly worse in the incongruent condition than they were in the congruent condition, indicating that participants may have been expecting the congruent word and so misheard the incongruent as congruent. The baseline conditions not differing significantly from one another seems to be an indication that the added information in the other rhythmic conditions did not facilitate or harm perception. Perhaps this was because in the baseline conditions, participants did not have an expectancy effect for the final word; the sentence was ambiguous enough that participants were not waiting for one specific utterance, and so the expectancy effect was not present.

Also in line with expectations from prior research (Cason et al. 2015), the participants were better at identifying the word correctly when they were primed with the sentence of the beat than with just the rhythmic speech. This pattern could be an indication that the expectancy effects allowed listeners to attend better to the final syllable of the sentence where they knew to expect the masked word.

In the incongruent condition, the misleading semantic context leads the listener to expect a different word as the final utterance, causing poorer performance. We expected that in the rhythmic with beat condition, the incongruency effects might be somewhat mitigated given that the listener was attending to the syllable where the word was going to occur. We found that the beat prime did facilitate comprehension in the incongruent case, but interestingly it also did in the congruent context.

In the congruent context, participants seemed to struggle more with the rhythmic speech than they did with natural speech. Perhaps this difference was due to the unexpected nature of the sing-songy speech. In the sing-songy rhythmic speech, utterances take longer; perhaps this manipulation of the duration of the sentence disrupted the participants' focus or ability to comprehend the speech. In any case, this struggle with the rhythmic speech seemed to be mitigated by the presence of the beat in the rhythmic plus beat condition, indicating that the presence of the prime helped enough for participants to better understand the rhythmic speech. Perhaps the prime clued the participants in to the modified duration of the sentence so that they were prepared.

These findings have an interesting implication on the idea of using music to facilitate language processing. While participants struggled with the sing-songy language with no beat prime, songs generally have an established beat which would functionally provide the expectancy effects that the beat prime functioned as in this study. In the congruent condition, participants generally did very well in the arhythmic and rhythm with beat prime conditions. Perhaps this was somewhat of a ceiling effect in these conditions, and having a more challenging listening environment would reveal more of a difference between those conditions. If, in general, the rhythm + beat conditions facilitated speech perception in some participants, this may be helpful for communication in noisy or distracting environments. If the beat is providing important contextual information which would help the listener make sense of speech in noise, then perhaps this beat method could be used to facilitate understanding in these environments.

While the results of the study indicate that participants performed better in the rhythmic speech condition primed with the beat than without, it is also important to note that in this condition the participants experienced a greater duration of time before the sentence. Perhaps this "rest" time was responsible for the increase in hits. A follow-up study could investigate the potential effects of this issue, by including a duration of silence of similar length to the beat files played before the sentence file. This would determine whether it was the duration of time responsible for the increase in performance, or whether it was the prime stimuli contents that improved performance.

Conclusion

This study found that, in line with prior research, there were helpful effects of congruent priming in both a rhythmic and semantic context. Furthermore, the results indicate that the rhythmic priming effects differ based on the semantic nature of the sentence, facilitating sentences where expectancy effects would be present but having a negligible effect on sentences where the listener could not predict the final word. These results could be further investigated by looking into the effects of a misleading rhythmic prime, to examine whether the misleading prime would hinder the performance of listeners across semantic conditions.

References

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception. *Acta Psychologica, 155*, 43-50. https://doi.org/10.1016/j.actpsy.2014.12.002

Failes, E., Sommers, M.S. & Jacoby, L.L. Blurring past and present: Using false memory to better understand false hearing in young and older adults. Mem Cogn 48, 1403–1416 (2020). https://doi.org/10.3758/s13421-020-01068-8

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews, 86*, 150-165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234. https://doi.org/10.1037/h0031564

Nespor, M., Shukla, M. and Mehler, J. (2011). Stress-Timed vs. Syllable-Timed Languages. In The Blackwell Companion to Phonology (eds M. Oostendorp, C.J. Ewen, E. Hume and K. Rice). https://doi.org/10.1002/9781444335262.wbctp0048

Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behavior Research Methods, 52*(4), 1795-1799. 10.3758/s13428-020-01351-1

Rogers, C.S. (2017). Semantic priming, not repetition priming, is to blame for false hearing. *Psychonomic Bulletin & Review, 24*, 1194-1204.

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging, 27*(1), 33–45. https://doi.org/10.1037/a0026231

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia, 50*(2), 232-244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Shakespeare, W. (1609). Sonnet 18. In I. Overdrive (Ed.), Shakespeare's Sonnets (2021). Duke Classics.

Appendix

Table 2: Mean proportion correct in each condition, SD indicated standard deviation from the mean, N indicates the number of participants in each condition

Priming	Rhythm	Mean	SD	N
Baseline	Arhythmic	0.867	0.092	29
	Rhythmic	0.901	0.079	29
	Rhythmic_Beat	0.908	0.092	29
Congruent	Arhythmic	0.956	0.050	29
	Rhythmic	0.881	0.065	29
	Rhythmic_Beat	0.937	0.057	29
Incongruent	Arhythmic	0.602	0.182	29
	Rhythmic	0.567	0.135	29
	Rhythmic_Beat	0.633	0.163	29