Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Linguistic syncopation: Alignment of musical meterto syntactic structure and its effect on sentence processing

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Hilton, Courtney
Goldwater, Micah

Publication Date

2019

Linguistic syncopation: Alignment of musical meter

to syntactic structure and its effect on sentence processing

Courtney Hilton (courtney.hilton@sydney.edu.au)
School of Psychology, The University of Sydney, Sydney, Australia
Centre for Research on Learning and Innovation, Sydney, Australia
Micah Goldwater (micah.goldwater@sydney.edu.au)
School of Psychology, The University of Sydney, Sydney, Australia

Abstract

Language and music are structured at multiple temporal scales and have been characterized as having meter: a hierarchical and periodic alternation of the prominence of syllables/beats. Meter is thought to emerge from the entrainment of neural oscillators, affording temporal expectations and selective attention. Higher-levels of a metric hierarchy also tend to track syntactic phrase structure, however, it is not clear within the framework of temporal attending why this would be advantageous. Neural oscillations have recently been shown to also track syntactic phrases. We propose that meter aligns to phrase structure so as to make syntactic processing more efficient. In two experiments (both visual and auditory language), we show that certain alignments of meter to syntax influence sentence comprehension and we suggest potential mechanisms for why certain alignments tend to be preferred. Our results underline the rhythmicity of not only low-level perception but also of higher-level cognitive processing of syntactic sequences.

Keywords: Language, time, oscillations, musical meter, syntax, merge

Introduction

Music and spoken language present similar challenges to a listener in that structure at multiple temporal scales must be decoded from a continuous sound signal in real-time. It is, therefore, no surprise that there are some parallels in how this is achieved and, as such, also parallels in musical and linguistic structure that bear the mark of these shared processing means. One such parallel is metrical structure. Meter generally refers to the perceived hierarchical alternation of stress in syllables in speech (Port, 2003), or beats in music (for more detail, see: Lerdahl & Jackendoff, 1983). In this paper, we motivate a view of meter as being something emerging from, on the one hand, the computational problem of extracting a discrete structured representation from a continuous signal, and on the other hand, an algorithmic solution that fits within the implementational oscillatory-constraints of neuro-computation (Rimmele et al, 2018a).

In explaining what meter affords its perceiver, the predominant theory has been that it is a system for *predicting when* and that these temporal expectations then in turn afford the dynamic allocation of attention to expected points in time to optimize processing (Jones, 1976; Pitt & Samuel, 1990). These theories of 'dynamic attending' have been formalized in models using coupled neural oscillators to explain how meter is flexibly entrained to a signal and how the dynamics of hierarchical perceived stress emerge naturally from this mechanism (Large & Jones, 1999; Port, 2003).

More generally, there is an attractive isomorphism between the temporally multi-scaled structure of language and music, and the multi-scaled oscillatory paradigm of neural processing in the brain. The consensus seems to be that the entrainment of one to the other—'tuning the inside to the outside'—is, at least, important if not necessary to both basic perception and perhaps even to deeper analysis and comprehension. As such, the concepts of oscillation and entrainment have become central in recent cognitive and neuroscientific theories of language processing (Giraud & Poeppel, 2012), and in theories of music processing for both rhythm/meter (Large & Kolen, 1994) and tonality (Large et al, 2016).

Specifically for the case of speech, it is proposed that the auditory cortex entrains a cascade of oscillatory sampling windows to the speech envelope: phonemes sampled with gamma oscillations (>30hz), syllables with theta (3-8hz), and intonational phrases with delta (<3hz). And while delta-oscillations are normally observed to follow prosody (Bourguignon et al, 2013) they have recently been shown to track syntactic phrases, even in the absence of prosodic cues (Ding et al, 2016), thus demonstrating top-down linguistic knowledge. Meyer and colleagues (2017) additionally showed that when prosody and syntax are misaligned, delta tracks the syntactic rather than the prosodic phrase. How should this all be interpreted?

One consideration that has been neglected is how meter figures into this: perhaps what delta is really tracking here is meter. This is especially important as the paradigms used to show delta-tracking of syntax employ a frequency-tagging approach where the speech must be presented isochronously and thus may be particularly likely to induce a subjective percept of meter. Indeed, similar paradigms have also been used to show oscillatory tracking of meter where delta too tracks higher-metric levels not present in the acoustic signal (Nozaradan et al, 2011).

While the precise rhythmicity of naturalistic speech is still hotly debated (for example, two contrasting positions: Nolan & Jeon, 2014; Brown Pfordresher, Chow, 2017), and thus the extent to which strict parallels between speech and musical rhythm/meter are valid, it is certainly clear that there many special cases of speech where the parallel with musical meter is clear, such as poetry and song (Lerdahl, 2001). And perhaps more importantly, it seems that rhythm and meter are especially important cues during early language development, in line with the framework of prosodic bootstrapping where infants rely on prosodic information to segment input. In line with this, it has been shown that the perception of meter is present in the first year of infancy and that meter supports the learnability of other structure in a signal such as rhythm and melody (Hannon & Johnson, 2004). More generally, this idea may explain the clear metric structures of nursery rhymes and in children's literature (Breen, 2018; Fitzroy & Breen, 2019).

In recent years, there has also been growing interest in the relationship between musical experience and language abilities (for recent meta-analyses see: Gordon et al, 2015a; LaCroix et al, 2015), with the underlying rationale of some overlap in neural implementation and that music may have certain properties that enable to it to preferentially strengthen these networks (Patel, 2011). Specifically, it seems that the subcomponent of meter is particularly crucial in mediating the transfer of musical abilities to the processing of speech and syntax in language (Gordon et al, 2015b; Jung et al, 2015). Relatedly, the syntactic deficits observed in Parkinson's Disease patients may actually be more to do with a deficit in the ability to process the meter of language than a deficit to syntax directly (Kotz & Schmidt-Kassow, 2008).

These findings are surprising: why should meter support syntax? While it is conceivable that the dynamic allocation of attention could support the processing of speech under noisy conditions (where signal and noise are time delimited) such as in 'cocktail party' paradigms where this oscillatory entrainment mechanism is implicated (Zion Golumbic et al, 2013). It is not clear how this mechanism would support syntactic processing specifically. Some recent work, however, has started to provide clues. Rimmele et al (2018b) have shown that delta is involved in chunking auditory short-term memory. And relatedly, the BUMP model (Hartley et al, 2016) has provided a mechanism by which entrained oscillators support auditory short-term memory for serial order, further supported by Gilbert et al (2017) who provided empirical support for a shared resource underpinning this aspect of short-term memory and temporal precision. In summary, metrical structure (especially in the delta-range) may support aspects of short-term memory, which would then in turn, support syntactic processing.

Another not mutually exclusive possibility is suggested by Nelson and colleagues (2017). Using intracranial electrophysiological recordings, they observed fine-grained neural dynamics of syntactic structure building. Specifically, they observed a monotonic ramping of activity for each new word presented in a sentence until a syntactic constituent could be formed, at which time there is a spike of activity proportional to the number of words then a sudden decrease of activity reflecting the freeing of working-memory resources. They interpreted this in terms of a Chomskian/Minimalist merge operation (see Friederici et al, 2017), however, these observations can also be interpreted in less theoretically committal 'chunking' terms. Regardless, this result captures real-time dynamics of processing demands that relate to syntactic structure building, and that these demands stack-up toward ends of phrases where ramping of activity reaches its summit and where there is a 'spike' of activity that merges/chunks the information. And importantly, these demands are time localized. Therefore, if the 'strong' and 'weak' of meter relate to oscillatory fluctuations in neural excitability then perhaps meter may also function to temporally align neural resources with these processing demands.

Some evidence linking delta-oscillations with such an idea is suggested by Meyer & Gumbert (2018), who found that the phase of delta-oscillations tends to align excitability with phrase-endings (however, they interpreted this as aligning delta with syntactic informativeness, nonetheless, their data are consistent with our idea here).

In summary, a more general way to make sense of this relationship between meter and syntax is in terms of prediction and efficiency of processing (Gibson et al, 2019), and how this is constrained by the oscillatory nature of the brain (Rimmele et al, 2018a). Syntax gives top-down prediction of "what next" (Levy, 2008) and meter/neural-resonance gives bottom-up prediction of "when next" (Large & Kolen, 1994). However, together they are more flexibly able to entrain to not just low-level acoustics but also higher-level structures, and thus enable more efficient processing of syntactically structured sequences as in language and music.

We now explore this idea in two experiments that manipulate the alignment of meter and syntax and measure the effect of this alignment on comprehension.

Experiment 1

Method

Our central hypothesis that is tested in both of the following experiments is that comprehension (measured by probe accuracy and response-times) is highest when the strong-beat of meter aligns most often with phrase-boundaries (see Figure 1). We also manipulate syntactic complexity by using both subject-extracted and object-extracted relative-clause sentence structures. The difference in complexity here is defined in terms dependency locality theory (Gibson, 1998), where object-extracted sentences require integrating over a greater number of words and thus pose a greater strain on resources. We predict an interaction between sentence complexity and congruency on the grounds that better-aligned resource allocation may be more needed in sentences that integrate over more words. Thus, this yields a 2 X 2 factorial design, manipulating syntactic complexity (subject- vs object-extracted relative-clause) and congruency (congruent, incongruent). While we also manipulate which clause of the sentence is probe (main or relative), we have no theoretical prediction about this other than the main clause would have higher accuracy.

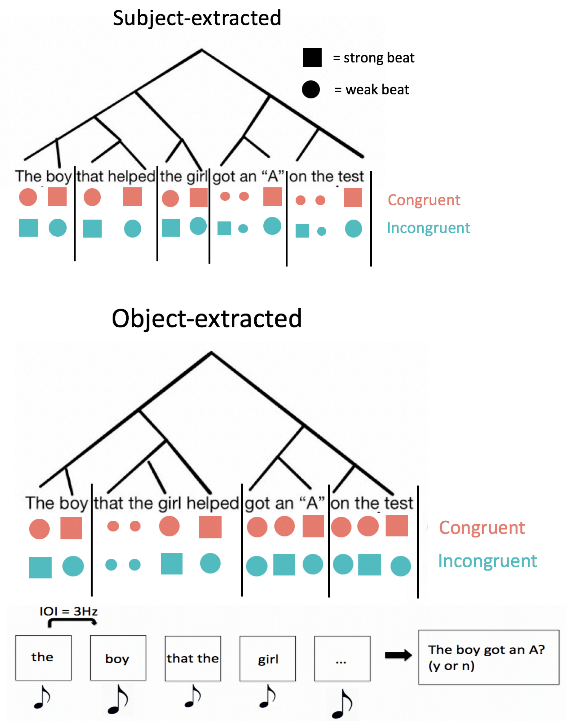

Figure 1, Top & Middle: examples of syntactic tree-structure for a sentence and possible alignment with meter defining congruencies. Bottom: trial presentation schematic

Participants 40 native English speakers (20 female) from the Sydney area took part in this study and were naive to its purpose.

Materials The language materials consisted of 48 sentences composed of largely monosyllabic words, each 12 words long. Each sentence had subject- or object-extracted versions. We used an additional 25 filler-sentences of assorted structure and length. Each sentence has an accompanying comprehension probe, which was balanced within participants as to probing either the main- or relative-clause and whether the correct answer was “yes” or “no”. For example, if the sentence was “The boy that the girl helped got an A on the test”, the probe was either “The boy/girl got an A?” or “The boy/girl helped the boy/girl?”. The congruency manipulation was achieved by shifting the phase of the metrical pattern relative to the language presentation such that the strong beats fell on different positions in each phrase, e.g. “the BOY” or “THE boy” (see Figure 1). To fit the structure of the subject or object extracted forms, these sentences appeared in either binary or ternary meters respectively. Conditions were randomized over the sentences for each participant and presented in a random order.

The auditory materials were generated using a Python script and consisted of a 333Hz pure tone in which a 3Hz beat was induced by amplitude-modulating the signal with an asymmetric Hanning window with 80% depth and a 19:1 ratio of rise-to-fall time. Metrical accents were then applied by a 50% volume increase every 2 (binary) or 3 (ternary) tones.

Procedure The experiment was self-paced, and after an initial practice block, was completed in a single block where the participant was encouraged to take short breaks between trials. The experiment was run using software written in Python, using the PsychoPy library. Each trial begins with one full-bar of the meter (three strong beats) while a fixation-cross is shown center screen, after which the words begin appearing in the place of the fixation cross synchronized to the auditory tones. At the end of the sentence, the probe question appears center screen, and the participant is prompted to respond as quickly as possible with either “y” or “n” keys on a keyboard. If participants take longer than 5 seconds to respond, they will be prompted to speed up on the next trial. The participant also receives corrective feedback after each trial and is encouraged to balance speed with accuracy.

Results

Comprehension data were analyzed using a mixed-effects logistic regression including fixed-effects for congruency (congruent, incongruent), syntactic complexity (subject-RC, object-RC), probed clause (main-clause, relative-clause), and the interaction between congruency and syntactic complexity. We also included random intercepts for participants and items. Response times (RTs) were analyzed using a linear mixed-effects regression with the same structure. All analyses were done in R.

As seen in Figure 2, participants made fewer comprehension mistakes in the congruent conditions ($\chi^2 = 7.99$, p = .005), fewer mistakes for the subject-RC sentences over the object-RC ones ($\chi^2 = 26.21$, p = <.001) and fewer mistakes when the main clause is probed rather than the relative clause ($\chi^2 = 40.03$, p = <.001). There was, however, no significant interaction between congruency and syntactic complexity ($\chi^2 = 0.43$, p = .513). There was also no significant effect of congruency on reaction times ($\chi^2 = 1.20$, p = 0.273). However, there were significant effects of syntactic complexity and probed-clause on RTs ($\chi^2 = 16.314$, p = <.001; $\chi^2 = 35.796$, p = <.001).

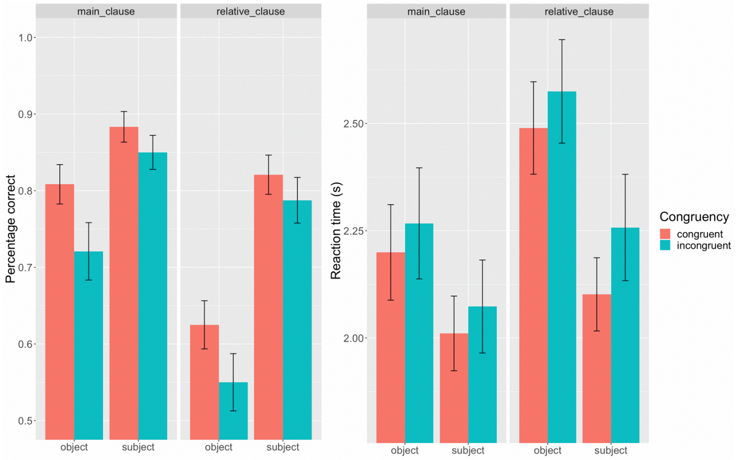

Figure 2, Experiment 2 results. Left: accuracy results as percentage correct. Right: response-time results in seconds.

Discussion

In line with our main prediction, congruency affected comprehension accuracy, however, there was no significant difference for response times. Against our initial prediction, there was also no significant interaction between congruency and syntactic complexity.

One limitation of the design was that the meter was induced passively with the auditory stimuli. Thus, we do not know to what degree participants actually interpreted the stimuli according to this meter. And although there was no significant interaction between congruency and syntactic complexity, one potential problem of the design was that object-RC sentences always had a ternary meter and subject-RC sentences always had a binary meter. This issue is an inevitable consequence of the phrase lengths in these respective sentence types, however, it may complicate the interpretation of an interaction. Further, reading sentences presented in an RSVP format is not a naturalistic way of processing language.

Experiment 2

Method

In our second experiment we wanted to build on the results of the first, replicate the congruency effect on accuracy, and address some of its limitations. Notably, we presented the sentence stimuli as auditory speech to make it more naturalistic and to check the robustness of the congruency effect to stimulus modality. As opposed to Experiment 1, where meter was induced *passively,* in Experiment 2 we induced meter *actively* by asking participants to tap on a drum-pad in time with the strong metric-beats while they listen to the speech stimuli (see for how tapping/motor actions entrain auditory attention: Morillon & Baillet, 2017). This also allows us to use tapping consistency as a DV, thus, we add the prediction that tapping will be most consistent in congruent trials (consistency being defined as the standard deviation of their accuracy). Finally, both subject and object extracted RC sentences were presented to the same ternary meter. This allows us to discount the possible meter by syntactic-complexity confound. Although, in order to make the subject-RC sentence fit a ternary meter, we had to introduce a new potential confound of inserting silences as in Figure 3. The main reason why we opted for a ternary meter, however, was to enable us to have three levels of the congruency condition for each sentence type (congruent, incongruent-1, incongruent-2). This came with the additional hypothesis that incongruent-1 would be the most incongruent metric alignment. That is, according to our delta-oscillation hypothesis, while incongruent-1 & 2 are both equidistant from the 'merge' position of the phrase, for incongruent-2, the merge would occur while attentional resources are rising, whereas for incongruent-1, the merge would occur while this attentional energy is falling (Figure 5).

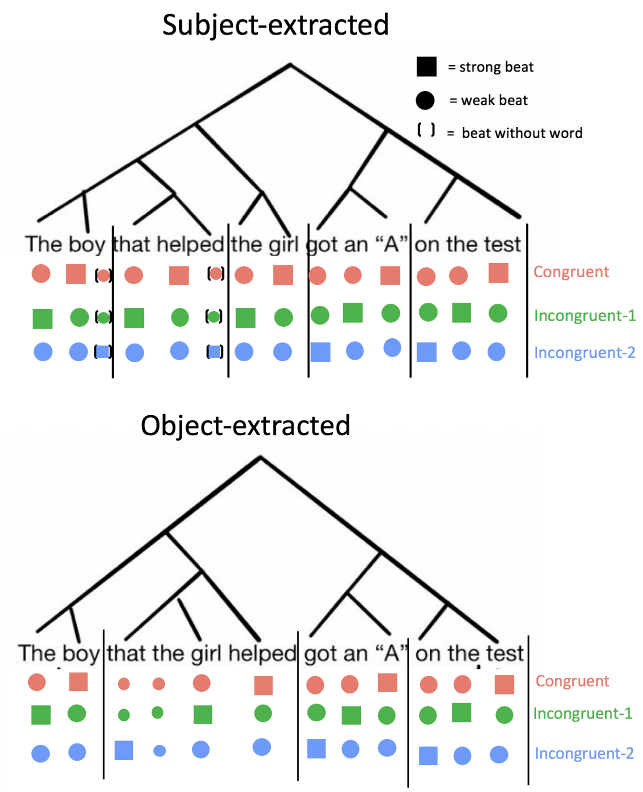

Figure 3, Congruency examples Experiment 4: syntactic tree diagrams for subject and object extracted sentences with accompanying alignments to meter defining the congruencies.

Participants (same specification as Experiment 1)

Materials Extending the 48 sentences and probes from Experiment 1, we created a further 24 of the same constraints, yielding a total of 72 sentences and probes. Speech stimuli were then generated and preprocessed from these sentence materials using a custom Python script, using Google's text-to-speech API to generate audio-files for each word individually. These stimuli were then volume normalized and cut and stretched to 2.5Hz (this new presentation-rate was based on piloting), then assembled into the sentences. Like Experiment 1, each trial starts with one full-bar of the tones to set the metric context. In a departure from Experiment 1, however, the tones drop-out when the speech stimuli start.

Procedure To ensure an active percept of the meter in Experiment 2, participants were required to tap on a drumpad in time with the strong-beats while listening to the speech (tapping once every three words). Participants used their right index finger to tap on a pressure sensitive MIDI drumpad. Before the main section of the experiment, participants completed a 'tapping-only' trial-block which estimated their tapping consistency without any language stimuli. This was then followed by practice trials for the language section and then the main trial block. Otherwise, the trial design followed that of Experiment 1.

Results

Comprehension and response-time data were analyzed using logistic and linear mixed-effects models as in Experiment 1, with the only difference being three levels of the congruency fixed-effect (congruent, incongruent1, incongruent2).

The results replicate the main effect from Experiment 1, showing that congruency significantly affected comprehension (incongruent1: χ^2 = 13.23, p = <.001, incongruent2: χ^2 = 8.30, p = .004). While the incongruent2 condition had a smaller cost on comprehension that incongruent1 (as predicted), the difference between these predictors was not significant (χ^2 = 0.533, p = 0.465). As before, there was also a significant difference between which clause is probed (χ^2 = 24.641, p = <.001). Surprisingly, however, there was no significant effect of syntactic-complexity (χ^2 = 0.101, p = .750). We believe this is a likely consequence of the added rhythmic complexity required to make subject-RC sentences fit a ternary meter.

As with Experiment 1, there was no significant effect of congruency on RTs (incongruent-1: χ^2 = 1.371, p = 0.242, incongruent-2: χ^2 = 0.837, p = 0.360) although syntactic-complexity and clause-probed had strong effects (χ^2= 19.900, p = <.001; χ^2= 49.801, p = <.001).

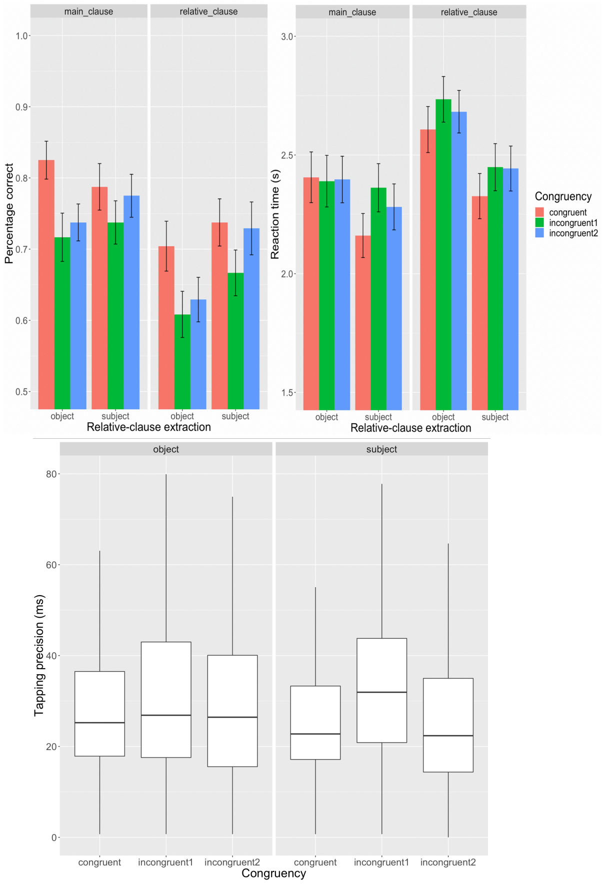

Figure 4: Comprehension results from Experiment 2. Left: accuracy. Right: response times. Bottom: tapping accuracy (standard deviation of asynchrony between tap and target tone)

For the tapping data during sentence processing, we also used a linear-mixed effects regression with congruency and sentence-extraction, and their interaction, as fixed effects, and participants as a random-effect.

Congruency had a significant effect on tapping precision in the incongruent1 conditions (χ^2= 10.27, p = .001) and only marginally significant for incongruent2 (χ^2= 3.25, p = .071). There was also a significant interaction between incongruent1 and syntactic complexity (χ^2= 12.72, p = <.001). However, it is problematic to interpret this interaction in light of the above-mentioned issues with rhythmic complexity, so we do not interpret it further.

General discussion

The results of both experiments support our hypothesis that sentence comprehension is optimal when metrical strong-beats align with phrase boundaries. We interpret this as supporting the more geneal idea that meter, and its alignment to phrase structure, plays a role in syntactic processing.

Jung and colleagues (2015) showed a similar effect of temporal expectancy on syntactic processing by having key words arrive early or late compared to an established rhythm. Kotz & Schmidt-Kassow (2015) showed that the syntactic deficit of Parkinson's Disease patients was actually due to a deficit processing the meter/timing of speech. In our study, however, each word was perfectly predictable from a rhythmic standpoint, and participants had no generalized timing deficit, however, we showed that the hierarchical distribution of attention embodied by meter, and its alignment with syntactic structure, was sufficient to show differences in comprehension.

We did not, however, find significant differences in response times. Although it is worth noting that the direction of the RT-effect was consistent with our hypothesis in both experiments. One possible explanation for this null-result is that top-down endogenous attention tends to affect accuracy, while bottom-up exogenous attention affects reaction-times (Prinzmetal et al, 2005). While lower levels of meter may be driven by bottom-up cues in the signal (Large & Kolen, 1994), it is likely that higher-levels increasingly rely on top-down phase-resetting of oscillations in response to syntactic structure or other structural cues (Rimmele et al, 2018a). Thus, this could explain why congruency had a stronger effect on accuracy over RTs. However, it may have also been that the effect was too small to detect for our sample size.

Comparison across the two studies also shows that this congruency effect is robust to modality (visual presentation in Experiment 1 and auditory presentation in Experiment 2), and thus is not specific to speech rhythms. A hypothesis that would need further experimentation to explore would be that any sequential stimulus that must incrementally form hierarchical structures would be influenced by this metric congruency effect. If this hypothesis were confirmed, it would suggest that meter is part of a more general cognitive strategy for the timely allocation of resources to process structural relations, and that music and language are just the most prominent domains in which this plays out.

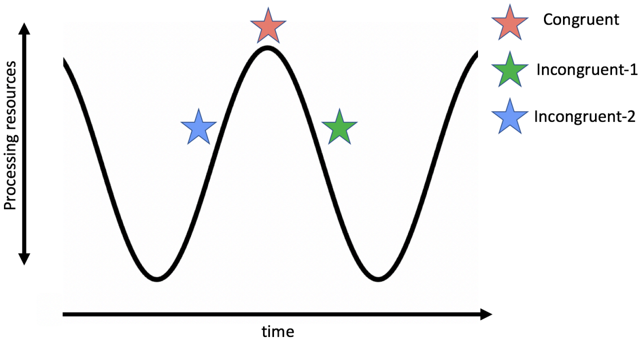

Figure 5: hypothetical oscillation of attentional resources. Stars represent the position of the phrase-ending word relative to this oscillation in each congruency condition.

Experiment 2 showed suggestive behavioral evidence of behavioral oscillations at the delta rate. Specifically, incongruent-1 conditions had a larger effect on comprehension than incongruent-2, in line with the differences in neuronal excitability predicted if the meter did phase-align delta to strong-beats (Figure 5). Although it is important to stress that this difference was not statistically significant, it was however consistent in direction across all conditions, including in the tapping data. It is also important to acknowledge that preferences for the alignment of meter to phrase structure may also be mediated by cultural factors relating to differences in language structure and preferences for grouping (Iversen et al, 2008). Future research is planned to further test these possibilities, including with more sensitive electrophysiological paradigms.

More generally, the idea that the analysis of syntactic phrases is somehow constrained by oscillatory processes is in line with the average phrase duration of speech at 2-3seconds (Vollrath, 1992), fitting within the delta-range. This may also be a neuronal constraint that results in Uniform Information Density (Levy & Jaeger, 2007), which stipulates that we, as rational communicators, attempt to spread information across a signal in a uniform way, and in a way that makes the most of our capacity (i.e. not undershooting). In other words, part of what defines our capacity to process linguistic information is these rhythmic processing constraints and the extent to which we can optimally entrain our internal rhythms to the rhythms of the information in the signal. Thus, delta sampling of syntactic phrases may define a crucial biologically grounded bottleneck that explains these patterns in human communication.

It is also likely that meter serves a similar function for the processing of harmonic syntax in music to the function articulated here for linguistic syntax (Patel, 2003). This would be consistent with some recent studies showing that harmonic structure is a strong cue for metrical strength (White, 2017). This would also accord with data showing an interaction between the processing of linguistic and musical syntax (Fedorenko et al, 2009).

Conclusion

We have shown that the alignment of meter to syntactic structure influences sentence comprehension. We have also discussed possible mechanisms from which this effect arises, namely, how delta-oscillations facilitate aspects of short-term memory processing that in turn allow for syntactic structure-building. These results imply that entraining the 'inside to the outside' may allow for more efficient processing of syntactic sequences. Future work will be required to further pick-apart the details of these ideas, ground them in neural measurement, and to explore their generality cross-linguistically and to other syntactically structured domains such as music and mathematics. In general, this work supports a co-dependency of "what" and "when" predictions, and grounds this in the biological implementational constraints of the rhythmic brain.

References

Bourguignon, M., De Tiège, X., De Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., … Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314–326.

Breen, M. (2018). Effects of metric hierarchy and rhyme predictability on word duration in The Cat in the Hat. *Cognition*, *174*(January), 71–81.

Brown, S., Pfordresher, P. Q., & Chow, I. (2017). A musical model of speech rhythm. *Psychomusicology: Music, Mind, and Brain*, *27*(2), 95–112.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory and Cognition*, *37*(1), 1–9.

Fitzroy, A. B., & Breen, M. (2019). Metric Structure and Rhyme Predictability Modulate Speech Intensity During Child-Directed and Read-Alone Productions of Children's Literature. *Language and Speech*.

Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., & Bolhuis, J. J. (2017). Language, mind and brain. *Nature Human Behaviour*, *1*(10), 713–722.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E., Futrell, R., Piantadosi, S. T., Dautriche, I., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, 1–40.

Gilbert, R. A., Hitch, G. J., & Hartley, T. (2017). Temporal precision and the capacity of auditory–verbal short-term memory. *Quarterly Journal of Experimental Psychology*, *70*(12), 2403–2418.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin Mcauley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635–644.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in Psychology*, *6*(DEC), 1–16.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, *50*(4), 354–377.

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135–178.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, *124*(4), 2263–2271.

Jung, H., Sontag, S., Park, Y. S., & Loui, P. (2015). Rhythmic effects of syntax processing in music and language. *Frontiers in Psychology*, *6*(NOV), 1–11.

Jones, M. R. (1976). Time, out Lost dimension. *Psychological Review*, *83*(5).

Kotz, S. A., & Schmidt-Kassow, M. (2008). Event-related Brain Potentials Suggest a Late Interaction of Meter and Syntax in the P600 Impact of social interaction on second language learning. *Journal of Cognitive Neuroscience*, 1693–1708.

Kotz, S. A., & Schmidt-Kassow, M. (2015). Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*, *68*, 48–60.

LaCroix, A. N., Diaz, A. F., & Rogalsky, C. (2015). The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Frontiers in Psychology*, *6*(August), 1–19.

Large, E. W., & Kolen, J. F. (1994). *Resonance and the perception of musical meter*. *Connection Science*, 6

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159.

Large, E. W., Kim, J. C., Barucha, J. J., & Krumhansl, C. L. (2016). A Neurodynamic Account of Music Tonality. *Music Perception*, *33*(3), 319–331.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, F. (2001). The sounds ofpoetry viewed as music. Ann. N.Y. Acad. Sci. 930, 337–354.

Levy, R. and Jaeger, T.F. (2007) Speakers optimize information density through syntactic reduction. In Adv. Neural Inf. Process. Syst. (Jordan, M.I. et al., eds), Adv. Neural Inf. Process. Syst. 849–856 72.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, *27*(9), 4293–4302.

Meyer, L., & Gumbert, M. (2018). Synchronization of Electrophysiological Responses with Speech Benefits Syntactic Information Processing. *Journal of Cognitive Neuroscience*.

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences*, 201705373.

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., … Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, *114*(18), 3669–3678.

Nolan, F., & Jeon, H. S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658).

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, *31*(28), 10234–10240.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D., & Morgan, E. (2016). Exploring Cognitive Relations Between Prediction in Language and Music. *Cognitive Science*, *41*, 1–18.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. Journal of Experimental Psychology: Human Perception and Performance, 16, 564–573.

Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, *31*(3–4), 599–611.

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*, *134*(1), 73–92.

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018a). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882.

Rimmele, J. M., Poeppel, D., Ghitza, O. (2018b). Accuracy in chunk retrieval is correlated with the presence of acoustically driven delta brain waves. (poster*) Society for Neuroscience*.

Vollrath, M. (1992). A universal constant in temporal segmentation of human speech. *Naturwissenschaften*, 10.

White, C. (2017). Relationships Between Tonal Stability and Metrical Accent in Monophonic Contexts. *Empirical Musicology Review*, (1983), 2–5.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron*, *77*(5), 980–991.