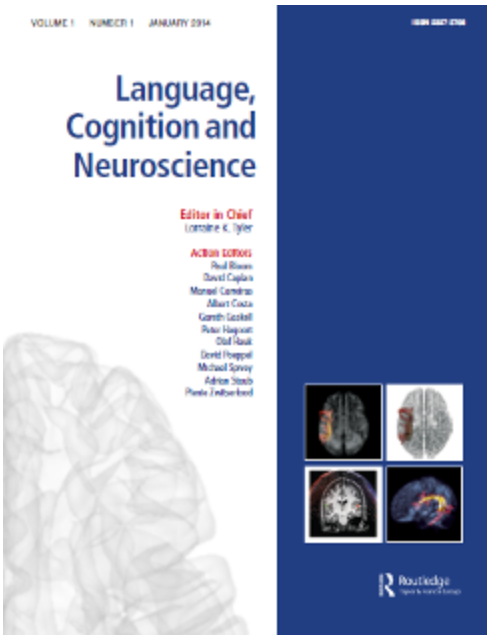

Hierarchical oscillators in speech comprehension: A commentary on Meyer, Sun & Martin (2020)

Journal:	*Language, Cognition and Neuroscience*
Manuscript ID	Draft
Manuscript Type:	Response Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Gwilliams, Laura; New York University,
Keywords:	not required, NA

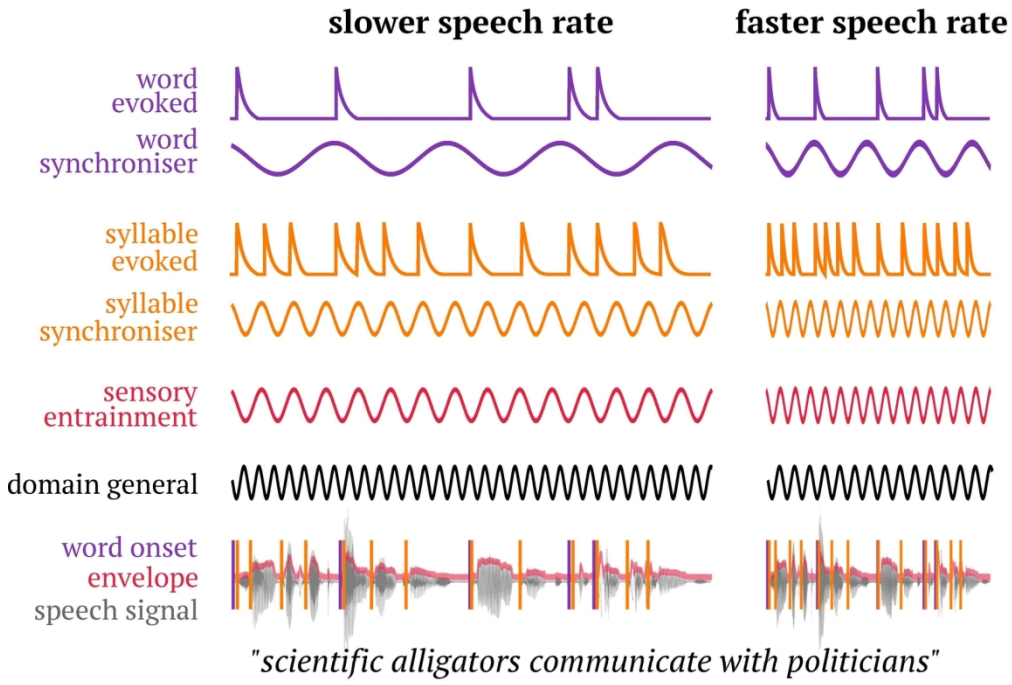

Figure 1. Schematic of different oscillatory and evoked components in response to two different speech rates. Below is the waveform for the sentence "scientific alligators communicate with politicians". The sentence was chosen to contain multi-syllabic words that de-confound syllabic and lexical units. The syllable onsets are marked in orange; the word onsets in purple. The speech envelope is shown in red. The row labelled "domain general" refers to non-speech specific processes that reside inherently within a particular frequency band (e.g. attention in alpha band), and the frequency does not change with speech rate. Sensory entrainment refers to the oscillatory tracking of the speech envelope. Syllable synchroniser corresponds to an oscillator that aligns to the syllable rate. This rate scales with the rate of the speech input. Syllable evoked is the alternative hypothesis that activity occurs in response to a syllable onset but is not generated by an oscillator. The syllable rate and the envelope rate are identical in the example, making them impossible to distinguish. Finally, the word synchroniser is a higher-level oscillator that tracks word boundaries. The word evoked model simulates neuronal firing every time a word boundary is recognised, but is not an oscillator per se.

153x102mm (300 x 300 DPI)

Hierarchical oscillators in speech comprehension: A commentary on Meyer, Sun & Martin (2020)

Laura Gwilliams
New York University & NYU Abu Dhabi

This is a commentary on Meyer, Sun & Martin (2020), Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing, <DOI>.

Entrainment and synchrony

Oscillatory responses are an emergent property of neuronal population firing (Wilson and Cowan, 1972; Börgers and Kopell, 2003; Wallace et al., 2011). The function of these rhythmic responses for cognition broadly, and for speech comprehension specifically, is an area of heated debate.

Among the different types of oscillatory behaviours involved in speech processing, *entrainment* is probably the most heavily discussed. In its general definition, entrainment refers to the phase and frequency alignment between the activity of an oscillator and its input (in this case, the inherently rhythmic speech signal). Evidence for entrainment comes from intra-cortical recordings in primates, as well as invasive and non-invasive electrophysiological recordings in humans (Buzsáki and Draguhn, 2004).

What is the role of entrainment for speech comprehension? A number of different proposals exist, which can be roughly grouped into three camps. First, the acoustic hypothesis: entrainment arises from tracking acoustic properties of the input such as acoustic edges and spectral-temporal features (Howard and Poeppel, 2010; Ding and Simon, 2012; Ghitza, 2012; Oganian and Chang, 2019). Second, the parsing hypothesis: entrainment is a mechanism which parses acoustic input into higher-order linguistic units such as syllables (Giraud and Poeppel, 2012; Ghitza, 2013; Ding et al., 2016). Finally the hypothesis that entrainment serves in domain-general processes, such as the allocation of attention (Schroeder and Lakatos, 2009) and environmental sampling (Schroeder et al., 2010). Of course, these functions significantly differ from each other but are not mutually exclusive; it is possible that entrainment serves as an instrument in one or all of these processes.

One of the main claims of Meyer, Sun and Martin (2020) is that entrainment *proper* should only be described relative to processing of the non-speech-specific sensory input (i.e. the acoustic hypothesis). The authors suggest that complementary to and separate from entrainment is *synchronisation*: the frequency-coupling between neural responses and the regular generation, or recognition, of abstract linguistic units. Evidence in favour of this cognitively-driven (rather than sensory-driven) process is that neural synchrony has been reported for aspects of the speech input that are experimentally absent from the acoustic signal (Ding et al., 2016). And, in natural speech, linguistic units such as morphemes, lexemes and phrases, as well as their semantic and syntactic content, are not straightforwardly aligned to prevalent features of the speech signal itself. This posits the existence of two distinct rhythmic processes: entrainment to the sensory input (red

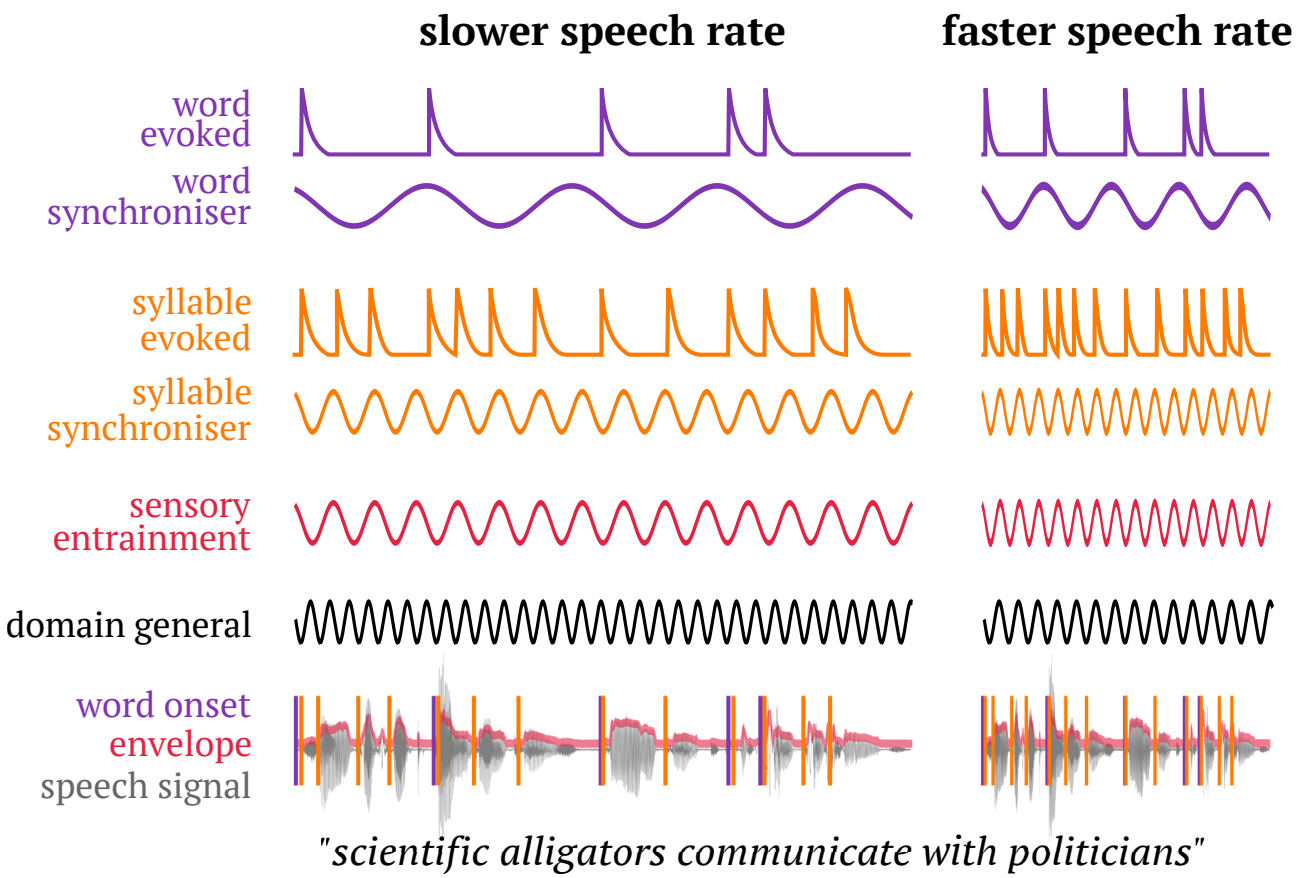

Figure 1. **Schematic of different oscillatory and evoked components in response to two different speech rates**. Below is the waveform for the sentence "scientific alligators communicate with politicians". The sentence was chosen to contain multi-syllabic words that de-confound syllabic and lexical units. The syllable onsets are marked in orange; the word onsets in purple. The speech envelope is shown in red. The row labelled "domain general" refers to non-speech specific processes that reside inherently within a particular frequency band (e.g. attention in alpha band), and the frequency does not change with speech rate. Sensory entrainment refers to the oscillatory tracking of the speech envelope. Syllable synchroniser corresponds to an oscillator that aligns to the syllable rate. This rate scales with the rate of the speech input. Syllable evoked is the alternative hypothesis that activity occurs in response to a syllable onset but is not generated by an oscillator. The syllable rate and the envelope rate are identical in the example, making them impossible to distinguish. Finally, the word synchroniser is a higher-level oscillator that tracks word boundaries. The word evoked model simulates neuronal firing every time a word boundary is recognised, but is not an oscillator per se.

sinusoid in Figure 1), and synchrony to the higher-order (language dependent) linguistic features (orange and purple sinusoids in Figure 1).

Meyer et al. (2020) stress that these two processes have been largely confounded in the literature, and what has been previously described as entrainment may be more accurately described as synchrony. Part of the problem in distinguishing them, though, is that the acoustic signal is temporally correlated with abstract linguistic units. In English, for example, most morphemes are mono-syllabic (e.g. bake, -er, pre-, war, dark, -ness…), and the syllabic structure is firmly encoded in the speech envelope. Therefore, the neural response that entrains to the envelope is difficult to

de-couple from a response that is synchronised to the processing of morphological units (because, both are correlated with the syllabic structure).

Disassociating acoustic processes from truly linguistic ones calls for carefully controlled experiments that have sufficient variability between the acoustic and the linguistic levels of description. A cartoon example of this is shown in Figure 1, where synchrony to syllabic structure is confounded with entrainment to the acoustic envelope, but not to the lexical structure. It is also notable that linguistic units are linked to acoustic features of speech to varying degrees, across different unit types, depending on the language (see (Gwilliams, 2020) for an overview). Combining controlled experimental paradigms that artificially vary linguistic and acoustic structure, with naturalistic studies that capitalise on intrinsic cross-linguistic variability, may be a powerful way to untangle the relative contribution of entrainment and synchrony.

Proposed mechanism

What is the point of having both sensory entrainment and higher-order synchrony? Meyer et al. (2020) suggest that these two mechanisms allow for uninterrupted segmentation of the speech signal: when the acoustic signal is noisy, top-down synchronous activity can compensate, and vice versa. In this sense, there is a dynamic trade-off between the use of bottom-up (acoustic) and top-down (abstract, linguistic) information to guide comprehension. The idea of top-down facilitation in service to speech perception is in line with a number of previous studies (e.g. (Davis and Johnsrude, 2007; Sohoglu et al., 2012; Gwilliams et al., 2018)), and indeed appears to be a pervasive observation across multiple domains of cognition (Engel et al., 2001).

The exchange of information between coupled oscillators has been widely and repeatedly established (Uhlhaas et al., 2009). This makes it easy to understand how bottom-up and top-down information may be exchanged in the case of syllable segmentation, because their frequencies are comparable. However, a bigger challenge may be to understand how low-frequency (< 1Hz) activity that operates on larger, more abstract units, could be directly assimilated with the higher frequency information encoded in the envelope. Perhaps this would instead involve a hierarchy of oscillators, which can pass information across adjacent frequency bands through phase alignment (Burgess, 2012).

Research has begun to associate particular types of abstract linguistic information with specific frequency bands. For example, the acoustic rhythm has been linked to the delta range (1-4 Hz) (Ding and Simon, 2014); phonetic features to the theta band (4-8 Hz) (Di Liberto et al., 2015); phonotactic regularity and information theoretic metrics such as surprisal and entropy have been found across both bands (1-9 Hz) (Di Liberto et al., 2019; Donhauser and Baillet, 2019). Therefore, based on prior work, it is possible that information at different levels of abstraction are generated and maintained through corresponding oscillators.

Future directions

Does this mean, then, that there exists a set of hierarchical oscillators, descending in their natural frequency rate, which synchronise to the corresponding hierarchical features of language? Oscillations have been reported between 0.05 - 500 Hz (Buzsáki and Draguhn, 2004), which comfortably cover the size of primitive linguistic structures that may need to be parsed from the speech signal (i.e. from duration 20 s to 2 ms); so, it is certainly possible. This is a provocative proposal, and one that I suggest should be tested against two alternatives.

The first possibility is that the observed responses are actually not oscillatory, but instead reflect the rhythmic production of evoked responses (as shown in Figure 1). In the specific case of speech perception, any unit which is important to the language system (e.g. morpheme, lexeme, phrase…) will elicit an event-locked neural response. Because the timing of these units is sufficiently regular, these periodic responses to the recognition or generation of abstract linguistic features become difficult to distinguish from the temporal dynamics of an oscillation. This is hard to falsify given that an oscillator and rhythmic evoked responses can elicit very similar outputs. Recent studies have gone to great lengths to show that for the case of acoustic envelope tracking, both evoked and oscillatory responses are necessary to account for observed sensory entrainment (Doelling et al., 2019). I suggest that the so-called synchronous activity under discussion here should be subject to the same scrutiny. Although the data would look very similar under each account, the differences in interpretation have substantial consequences for the neural architecture supporting speech comprehension. If synchronous activity comes from an oscillator, perhaps it can be understood as the instantiation of a *mechanism* that generates linguistic properties. However, a set of periodic evoked responses would be better described as a *reflection* of the true neural mechanism: simply following the computations rather than performing the computations itself.

The second alternative is that the apparent carrier frequency reflects a domain-general cognitive process rather than synchronisation to the abstract unit. For example, the alpha band (~8-12 Hz) has been linked to modulation and allocation of attention (Haegens et al., 2011; Haegens and Zion Golumbic, 2018) and beta (~13-30 Hz) to top-down control (Sherman et al., 2016; Spitzer and Haegens, 2017). In a similar way, it is possible that metrics of phonological expectation in the form of phonotactics, surprisal and entropy, relate to predictive processes *in general,* and do not reflect computations on the linguistic units per se. Figure 1 shows a way to adjudicate between these alternatives, by comparing responses to different speech rates: if the carrier frequency of the effect scales with the speed of the input, it likely reflects unit processing. If, however, it remains stable, it probably reflects a more general cognitive process which is inherently tied to a particular frequency band. For example, in Ding et al., (2016) the signature of phrasal and sentential processing was found to scale with the input: it occurred at 2Hz and 1Hz (respectively) for the Chinese materials, and 1.56 Hz and 0.78 Hz for the English materials, which perfectly aligns to the rate difference in the stimuli themselves. Thus, this suggests that these responses are not due to a process which inherently resides at a particular frequency, but instead reflect synchronous activity that is aligned to the rate of phrasal and sentential processing. Similar tests could be performed across different levels of linguistic structure, to establish which are truly supported by dedicated oscillators.

Conclusion

The proposal by Meyer, Sun and Martin (2020) brings to light the important distinction between entrainment to acoustic input, which is non-speech-specific, and neural synchrony which reflects the computation of abstract linguistic units. While the mechanistic role of synchronous responses remains to be fully described for units of different sizes, and its distinction from periodic evoked responses needs to be established, the proposal offers new directions for understanding the hierarchical operations supporting speech comprehension. It is up to future study to delineate the role of oscillatory mechanisms for different higher-level linguistic processes.

Acknowledgements

This work was supported by the Abu Dhabi Institute Grant G1001 and the Dingwall Foundation.

Bibliography

Börgers C, Kopell N (2003) Synchronization in networks of excitatory and inhibitory neurons with sparse, random connectivity. Neural computation 15:509-538.
Burgess AP (2012) Towards a unified understanding of event-related changes in the EEG: the firefly model of synchronization through cross-frequency phase modulation. PLoS One 7:e45630.
Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. science 304:1926-1929.
Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 229:132-147.
Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr Biol 25:2457-2465.
Di Liberto GM, Wong D, Melnik GA, de Cheveigne A (2019) Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. Neuroimage 196:237-247.
Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. Journal of neurophysiology 107:78-89.
Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.
Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19:158-164.
Doelling KB, Assaneo MF, Bevilacqua D, Pesaran B, Poeppel D (2019) An oscillator model better predicts cortical entrainment to music. Proc Natl Acad Sci U S A 116:10113-10121.
Donhauser PW, Baillet S (2019) Two Distinct Neural Timescales for Predictive Speech Processing. Neuron.
Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top–down processing. Nature Reviews Neuroscience 2:704-716.
Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Frontiers in psychology 3:238.
Ghitza O (2013) The theta-syllable: a unit of speech information defined by cortical function. Frontiers in psychology 4:138.
Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511-517.
Gwilliams L (2020) How the brain composes morphemes into meaning. Philos Trans R Soc Lond B Biol Sci 375:20190311.
Gwilliams L, Linzen T, Poeppel D, Marantz A (2018) In Spoken Word Recognition, the Future Predicts the Past. J Neurosci 38:7585-7599.
Haegens S, Zion Golumbic E (2018) Rhythmic facilitation of sensory processing: A critical review. Neurosci Biobehav Rev 86:150-165.
Haegens S, Handel BF, Jensen O (2011) Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. J Neurosci 31:5197-5204.
Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. Journal of neurophysiology 104:2500-2511.
Oganian Y, Chang EF (2019) A speech envelope landmark for syllable encoding in human superior temporal gyrus. Science Advances 5:eaay6279.
Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9-18.
Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of active sensing and perceptual selection. Current opinion in neurobiology 20:172-176.
Sherman MA, Lee S, Law R, Haegens S, Thorn CA, Hamalainen MS, Moore CI, Jones SR (2016) Neural mechanisms of transient neocortical beta rhythms: Converging evidence from humans, computational modeling, monkeys, and mice. Proc Natl Acad Sci U S A 113:E4885-4894.
Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. J Neurosci 32:8443-8453.
Spitzer B, Haegens S (2017) Beyond the Status Quo: A Role for Beta Oscillations in Endogenous Content (Re)Activation. eNeuro 4.
Uhlhaas PJ, Pipa G, Lima B, Melloni L, Neuenschwander S, Nikolic D, Singer W (2009) Neural synchrony in cortical networks: history, concept and current status. Front Integr Neurosci 3:17.
Wallace E, Benayoun M, van Drongelen W, Cowan JD (2011) Emergent oscillations in networks of stochastic spiking neurons. PLoS One 6:e14804.
Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. Biophysical journal 12:1-24.