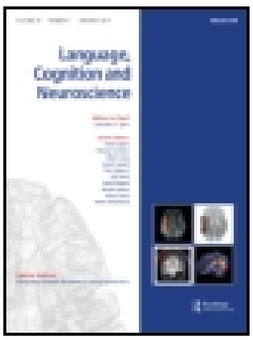

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Hierarchical oscillators in speech comprehension: a commentary on Meyer, Sun, and Martin (2019)

Laura Gwilliams

To cite this article: Laura Gwilliams (2020): Hierarchical oscillators in speech comprehension: a commentary on Meyer, Sun, and Martin (2019), Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2020.1740749

To link to this article: https://doi.org/10.1080/23273798.2020.1740749

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2020.1740749

Routledge
Taylor & Francis Group

COMMENTARY

Hierarchical oscillators in speech comprehension: a commentary on Meyer, Sun, and Martin (2019)

Laura Gwilliams [a,b]

[a]Psychology Department, New York University, New York, NY, USA; [b]NYU Abu Dhabi, Abu Dhabi, UAE

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

ARTICLE HISTORY
Received 10 February 2020
Accepted 2 March 2020

Entrainment and synchrony

Oscillatory responses are an emergent property of neuronal population firing (Börgers & Kopell, 2003; Wallace et al., 2011; Wilson & Cowan, 1972). 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 & 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 (Ding & Simon, 2012, 2014; Ghitza, 2012; Howard & Poeppel, 2010; Oganian & Chang, 2019). Second, the parsing hypothesis: entrainment is a mechanism which parses acoustic input into higher-order linguistic units such as syllables (Ding et al., 2016; Ghitza, 2013; Giraud & Poeppel, 2012). Finally, the hypothesis that entrainment serves in domain-general processes, such as the allocation of attention (Schroeder & 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 (2019) is that entrainment *proper* should only be used to describe the 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 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).

CONTACT Laura Gwilliams leg5@nyu.edu

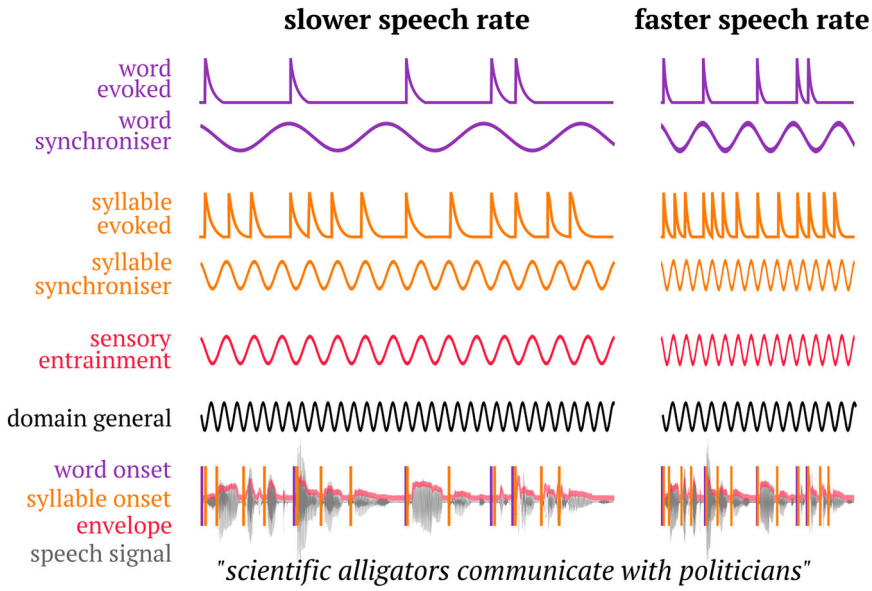

Figure 1. Schematic of different oscillatory and evoked components in response to two different speech rates. Below is the waveform for the sentence "scientific alligators communicate with politicians". The sentence was chosen to contain multi-syllabic words that deconfound syllabic and lexical units. 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.

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. (Note that a "word synchroniser" is yet to be reported – I include it for illustrative purposes). 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). 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 & Johnsrude, 2007; Gwilliams et al., 2018; Sohoglu et al., 2012), 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 (<1 Hz) activity that operates on larger, more abstract units, could be directly assimilated with the higher frequency bottom-up information encoded in the envelope.

Future directions

It is not news that language (and the units that compose it) is hierarchically structured. Does it follow then, from the line of argumentation outlined here, that for each level of

linguistic description (e.g. phoneme, syllable, morpheme …) there exists a dedicated oscillator? This would implicate an ensemble of oscillators, organised by natural frequency rate, which synchronise to the corresponding hierarchical features of language. The oscillators that process smaller units (e.g. phonemes, syllables) would have a faster natural frequency than the oscillators that process larger units (e.g. phrases, sentences). In this way, linguistically adjacent information (e.g. syllables and morphemes, putatively) could be exchanged through the phase alignment of the corresponding frequency bands (Burgess, 2012). Oscillations have been reported between 0.05 and 500 Hz (Buzsáki & Draguhn, 2004), which comfortably covers 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 alternative is that the observed responses are actually not oscillatory afterall, 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, periodic responses elicted by the recognition or generation of abstract linguistic features become difficult to distinguish from the temporal dynamics of an oscillation. 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 both an evoked and oscillatory 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. For example, it could be the mechanism by which (i) syllables are segmented; (ii) phonemes are bound into sequences; (iii) morphemes are bound into lexical items. However, a set of periodic event-locked responses would be better described as a *reflection* of the true neural mechanism: following the computations rather than performing the computations itself. In this case, rhythmic activity is simply a by-product of the existence of hierarchical units in language, not the mechanism by which those units come to exist.

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 & Zion Golumbic, 2018) and beta (~13–30 Hz) to top-down control (Sherman et al., 2016; Spitzer & Haegens, 2017). In a similar way, for instance, 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 (Di Liberto et al., 2019; Donhauser & Baillet, 2019). 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, Ding et al. (2016) found that the signature of phrasal and sentential processing scaled with the input: it occurred at 2 and 1 Hz (respectively) for the Chinese materials, and 1.56 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 input. 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 et al. (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 event-locked 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 New York University Abu Dhabi Institute Grant G1001, NIH R01DC05660 and the William Orr Dingwall Dissertation Fellowship. I thank Arianna Zuanazzi, Florencia Assaneo and Saskia Haegens for their useful feedback on this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the NYU Abu Dhabi Institute Grant G1001, NIH R01DC05660 and the William Orr Dingwall Dissertation Fellowship.

ORCID

Laura Gwilliams http://orcid.org/0000-0002-9213-588X

References

Börgers, C., & Kopell, N. (2003). Synchronization in networks of excitatory and inhibitory neurons with sparse, random connectivity. *Neural Computation*, *15*(3), 509–538. https://doi.org/10.1162/089976603321192059

Burgess, A. P. (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*(9). Article e45630. https://doi.org/10.1371/journal.pone.0045630

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929. https://doi.org/10.1126/science.1099745

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

Di Liberto, G. M., Wong, D., Melnik, G. A., & de Cheveigne, A. (2019). Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. *Neuroimage*, *196*, 237–247. https://doi.org/10.1016/j.neuroimage.2019.04.037

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. https://doi.org/10.1038/nn.4186

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311. https://doi.org/10.3389/fnhum.2014.00311

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, *116*(20), 10113–10121. https://doi.org/10.1073/pnas.1816414116

Donhauser, P. W., & Baillet, S. (2019). Two distinct neural timescales for predictive speech processing. *Neuron*. doi:10.1016/j.neuron.2019.10.019.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, *2*(10), 704–716. https://doi.org/10.1038/35094565

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. https://doi.org/10.3389/fpsyg.2012.00238

Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*, 138. https://doi.org/10.3389/fpsyg.2013.00138

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gwilliams, L. (2020). How the brain composes morphemes into meaning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791). Article 20190311. https://doi.org/10.1098/rstb.2019.0311

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *The Journal of Neuroscience*, *38*(35), 7585–7599. https://doi.org/10.1523/JNEUROSCI.0065-18.2018

Haegens, S., Handel, B. F., & Jensen, O. (2011). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of Neuroscience*, *31*(14), 5197–5204. https://doi.org/10.1523/JNEUROSCI.5199-10.2011

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

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, *104*(5), 2500–2511. https://doi.org/10.1152/jn.00251.2010

Meyer, L., Sun, Y., & Martin , A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing,. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2019.1693050

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5*(11), eaay6279. https://doi.org/10.1126/sciadv.aay6279

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–176. https://doi.org/10.1016/j.conb.2010.02.010

Sherman, M. A., Lee, S., Law, R., Haegens, S., Thorn, C. A., Hamalainen, M. S., Moore, C. I., & Jones, S. R. (2016). Neural mechanisms of transient neocortical beta rhythms: Converging evidence from humans, computational modeling, monkeys, and mice. *Proceedings of the National Academy of Sciences*, *113*(33), E4885–E4894. https://doi.org/10.1073/pnas.1604135113

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Spitzer, B., & Haegens, S. (2017). Beyond the status quo: A role for beta oscillations in endogenous content (re)activation. *eNeuro*, *4*(4). https://doi.org/10.1523/ENEURO.0170-17.2017

Uhlhaas, P. J., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolic, D., & Singer, W. (2009). Neural synchrony in cortical networks: History, concept and current status. *Frontiers in Integrative Neuroscience*, *3*, 17. https://doi.org/10.3389/neuro.07.017.2009

Wallace, E., Benayoun, M., van Drongelen, W., & Cowan, J. D. (2011). Emergent oscillations in networks of stochastic spiking neurons. *PLoS One*, *6*(5). Article e14804. https://doi.org/10.1371/journal.pone.0014804

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*(1), 1–24. https://doi.org/10.1016/S0006-3495(72)86068-5