Maastricht University

Distinct cortical rhythms in speech and language processing and some more

Please check the document version of this publication:

Routledge
Taylor & Francis Group

COMMENTARY

Distinct cortical rhythms in speech and language processing and some more: a commentary on Meyer, Sun, & Martin (2019)

Katerina D. Kandylaki and Sonja A. Kotz

Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, Netherlands

This is a commentary on Meyer et al. (2019), "Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing." DOI: 10.1080/23273798.2019.1693050

In this commentary we discuss some ideas put forward in Meyer et al. (2019) and propose to extend them for future investigations. The authors propose a new terminological dissociation on how to interpret neural oscillations in language comprehension by distinguishing two types of cortical oscillations: (i) the phase-locking of neural oscillations to the timing of phonemes and syllables in speech termed "entrainment proper" or "exogenous cortical rhythm", and (ii) the internal generation of syntactic, semantic, and discourse representations, coded as "intrinsic synchronicity" or "endogenous cortical rhythm". The authors rectify the need for this differentiation referring to temporal regularity in speech and syntax processing mainly investigated by event related potentials (ERPs).

We see lots of merit in defining such functionality of neural oscillations in speech and language comprehension. We also agree that the phase-locking of neural oscillations to exogenous speech signals is well documented and promising when investigating the neurocognitive basis of speech and language comprehension (e.g. Di Liberto et al., 2015). Speech is a temporally unfolding signal and is well suited to investigate neural oscillations, which, in turn, are relevant for cognitive processes such as auditory language comprehension (Peelle & Davis, 2012). Note that we are referring to auditory language comprehension as one process comprising different basic mechanisms, such as auditory perception, beat extraction, semantic role assignment to name a few. With regard to hierarchical processing in auditory language comprehension, Gross et al. (2013) proposed that the hierarchical organisation of speech aligns with cortical oscillations in a symmetrically hierarchical manner. Just as phonemes are nested in syllables and syllables are nested in words and phrases, neural oscillations entrain to corresponding nested frequencies (theta, delta, and gamma), when sampling the continuous auditory signal into linguistic chunks.

We agree that in speech and language processing there is need to specify if and how the phase-locking mechanism is linked to the structure and combination of phrases in sentences. Naturalistic studies on hierarchical speech processing provide evidence that EEG signatures relate to syntactic and semantic surprisal values. These effects were found within a range of 200–600 ms after word onset (Brennan & Hale, 2019). In the frequency domain, surprisal values based on neural network trained language models revealed EEG activity in the beta and gamma ranges (Weissbart et al., 2020). Please note that semantic surprisal in these two studies is not only specific to lexical semantics but is enriched by the semantic context starting with the sentence onset (n-grams in the case of Brennan & Hale, 2019) or, even richer, by the semantic context from the beginning of a story (neural networks in Weissbart et al., 2020). This surprisal could be termed "contextual semantic surprisal"; at the same time the different instantiations in the different computational algorithms employed to model surprisal should be acknowledged.

Further specification on how neural oscillations correlate with the hierarchical structure of language would foster a better understanding of how far speech/language comprehension relies on predictive processing *mechanisms* (Friston, 2010), heuristics (Karimi & Ferreira, 2016), dynamic attending (Jones, 1976; Jones & Boltz, 1989), or even combinations thereof as suggested by Callaway et al. (2017). For example, dynamic attending could explain how the continuous speech signal is chunked into phrases based on roving attention to linguistic features (Kotz & Schwartze, 2010) while predictive processing mechanisms could be used to understand how thematic roles and syntactic structures are

predicted and how these predictions are matched to chunking.

Language in context resolves ambiguity

Meyer et al. (2019) introduce an endogenous level of description in speech/language comprehension. This level includes syntactic predictions and assumes that these may be "disguised" as entrainment at the phrase/sentence level, even though such predictions are internally generated. To support these assumptions, the authors refer to their own previous findings. Meyer et al. (2016) aimed at disentangling prosodic and syntactic effects, by testing classical ambiguous sentences such as *the man shot the woman with the gun*, where the phrase *with the gun* can be attached to the verb thereby specifying *how* the man shot the woman, or it can be attached to the woman specifying *which woman* was shot. This type of ambiguity is indeed present when the sentences are taken out of context, but it is rarely found in everyday language use. Normally, visual (Knoeferle et al., 2005) or preceding linguistic context (e.g. a question on "Which woman did the man shoot?"), allows disambiguating such an ambiguous utterance.

The influence of *context in speech/language comprehension needs to be considered* in future attempts to relate neural entrainment to auditory language comprehension. The authors refer to studies on "sentence processing" (MacDonald et al., n.d.; Spivey-Knowlton & Sedivy, 1995; Swets et al., 2007) that provide informative but not exclusive sources for investigating the neurobiology of language. Rather, speech and language should be tested in context such as in stories, poems, dialogues, and dialogues with intention. Further, naturalistic language studies should consider how situational contexts and multimodality (e.g. visual context) might influence phase-locking to phrases and sentences.

For example, it is quite possible that a shared intention to find a way out of an "escape room", relies on previous knowledge, objects already discovered, riddles already solved and cues already used to create a situational context that allows local or global ambiguities to be instantly resolved. Specific prosodic accentuation embedded in the speech signal might trigger rhythmic processing e.g. to segment the continuous speech stream not only into syllables but linguistic chunks of information that lead to specific ways of combining and memorising information. For example, in an utterance such as "*This* jar contains *three* numbers which relate to *three* cues on *this* board" the speaker summarises, draws attention to cues and asks the listeners to combine this information to solve a problem. Further,

context might influence the rhythmicity of the spoken utterance in a way that old information is not accentuated but new information is. Processing of such an utterance might manifest entrainment or synchronisation of the neural oscillations between interlocutors or between the rhythmic features of the utterance and the neural oscillations.

In the same situation when a different player explains his/her arguments to the other players, these might phase-lock to this speaker's speech rate. Additionally, some speakers might add gestures or over-articulate the important information, thereby providing further contextual cues, to which neural oscillations could be phase-locked.

We therefore would like to put forward an extended perspective of entrainment and synchronisation to auditory language comprehension by adding supra-segmental features that might indicate speaker intention and context use.

Speech comprehension uses existing neurobiology for the processing of continuous signals

Speech and spoken language evolves continuously and in time. This is quite different from written language where language is segmented into identifiable elements in space. We note that even though written language also unfolds in time, it does so more incrementally and discrete. Current accounts of auditory signal processing relate auditory processing to primate audition (Rauschecker & Scott, 2009) and follow ideas on predictive coding (Friston, 2010). Further, Dynamic Attending Theory provides a strong foundation of how a continuous signal is attended to and chunked based on its rhythmical properties (Henry & Herrmann, 2014). Thirdly, in the context of a task, listeners have been shown to adopt "good enough" parsing strategies by binding phrases locally and neglecting the global interpretation of a sentence (Ferreira & Patson, 2007). For example, after the sentence "While Anna dressed the baby played in the crib," people answered "Yes" to whether Anna dressed the baby (Christianson et al., 2006). It is therefore possible that, when extracting a message from an utterance, listeners process this message in a time-restricted manner. For example if greeting your neighbour in the morning always takes the same amount of time, then processing it might rely on heuristic timing strategies of greeting from a distance. Consequently, the use of such heuristic strategies in the processing of language cannot be missing from neurobiologically based accounts of neural oscillations.

We consider that any neurolinguistic hypothesis (or level of cognitive neuroscience of language) must be rooted in the underlying neurobiology of perceptual systems. We therefore suggest that any attempt in interpreting the neural oscillations of speech and language should regard the brain as economic machinery, which uses its resources efficiently. For example, whenever strategies have been developed to parse a continuous signal (e.g. music), the same strategies are to be applied, at least to some extent, also in the parsing of speech. Specifically, the neural correlates underlying the rhythm network (Kotz et al., 2018) confirm an interplay of auditory processing and the linguistic composition of speech in a similar fashion as in music.

Argument structure processing guiding neural oscillations

We previously drew attention to two possible foci regarding the neural oscillations of speech and language: (i) the use of context in language experiments and, (ii) the rooting of hypotheses and interpretations in the existing neurobiological architecture for processing continuous signals. In addition to these more general propositions we introduce a third parameter that neural oscillations might phase-lock to. Referring to VanRullen's (2016) perceptual cycles, we propose that (iii) building-up an argument structure of who-does-what (-to-whom) could be one of the cyclic processes that oscillate in the low delta and theta frequencies. Specifically, we consider the timing of argument structure, prosody as a bottom-up cue influencing top-down hierarchical predictions, and working memory.

Firstly, the chunking of the speech signal into argument structure chunks might depend on the timing of extracting who-does-what (-to-whom) in an utterance and on the timescales in which these chunks fluctuate. We can therefore ask the question: how long does it take to understand who-does-what (-to-whom) in speech/language comprehension. Further, if such a temporal grid were used, how would it be oscillating? Previous attempts to connect predicate argument structure building and its neural underpinnings have been heavily debated, because many parameters need to be understood and taken into account, such as the neural pathways involved (visual and auditory dorsal and ventral streams), situation-based language evolution, and embodiment of perception (for the debate see Hurford (2003) and commentaries).

We would nevertheless like to estimate how long such a temporal grid could be. For the German language such an estimation was reported in Kandylaki et al. (2016), where sentences were presented in the active and passive voice. Sentence length varied between 4 and 8 words, resulting in a duration range of 2–5 s. This duration was dependent on the length of words and the speaking rate of the speaker. If argument structure takes about 2–3 s to build-up in spoken language, this would go along with Gross et al.'s (2013) findings that phrases are processed in rhythms of delta oscillations, following the unfolding rhythm of phrases.

Secondly, the building-up of an argument structure relies on top-down hierarchical combinatorics, which may be influenced by bottom-up cues such as prosodic features embedded in the auditory signal. Previous findings have shown that intonational and metric cues can guide syntactic and semantic parsing (Marslen-Wilson et al., 1992; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009; Steinhauer et al., 1999). We can then hypothesise that prosodic fluctuations would probably underline the argument structure building in spoken language. It is therefore intrinsically challenging to disentangle argument structure building from prosodic cues in natural language processing. Rather than pulling these factors apart, we think the intrinsic relationship of the two factors needs to be considered in tandem to understand how prosody might reset phase-locking of previously established phrases.

Thirdly, how we build-up argument structure must rely on agent – and object – related affordances, which relate to thematic roles and semantic composition. In order to account for this, we consider memory as an essential component influencing neural oscillations in speech and language comprehension. Such a component cannot be neglected when assuming top-down processes such as syntactic and semantic composition and how they translate into neural oscillations. It is therefore relevant to ask how such a component would oscillate. Possibly, working memory capacity (WMC) would provide a reliable time window for oscillations to entrain to (compare Nakano et al. (2010) for ERP effects based on working memory capacity). However, caution must be taken to ensure that researchers use the appropriate measurement of WMC, as for example reading span could not predict episodic long-term memory in listening (Sörqvist & Rönnberg, 2012). Therefore, the modality of WMC measurement should be considered when attempting to link WMC with interpretations of neural oscillations in speech/language comprehension.

Conclusion

We value the concept of phase shifting of neural oscillations to the external environment and we agree that an additional top-down mechanism involved in chunking the signal into phrases is needed. However, we

suggest not confining this latter mechanism to syntactic structure building as often the case in sentence processing studies. Speech/language comprehension occurs in situational contexts and should therefore be tested and understood in context. Also, any account on neural oscillations in speech/language comprehension must be rooted in the existing neural architecture of processing continuous signals and ideally draw connections to supramodal processing mechanisms of prediction, attention, and (working) memory. Lastly, the timing of argument structure building, prosody underlining this argument structure in a bottom-up manner, and a memory component supporting top-down prediction of thematic roles, are rather important features in speech and language comprehension and should be accounted for in current and future theories of neural oscillations in speech/language comprehension.

Disclosure statement

Funding

References

Brennan, J. R., & Hale, J. T. (2019). Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLOS ONE*, *14*(1), e0207741. https://doi.org/10.1371/journal.pone.0207741

Cdallaway, F., Hamrick, J., & Griffiths, T. (2017). *Discovering simple heuristics from mental simulation* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/wrqtp

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech Reflects Phoneme-level processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ferreira, F., & Patson, N. D. (2007). The?good enough? Approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Henry, M. J., & Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Perception*, *2*(1), 62–86. https://doi.org/10.1163/22134468-00002011

Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, *26*(3). https://doi.org/10.1017/S0140525X03000074

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323. https://doi.org/10.1037/0033-295X.83.5.323

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459. https://doi.org/10.1037/0033-295x.96.3.459

Kandylaki, K. D., Nagels, A., Tune, S., Kircher, T., Wiese, R., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2016). Predicting "when" in discourse engages the human dorsal auditory stream: An fMRI study using naturalistic stories. *Journal of Neuroscience*, *36*(48), 12180–12191. https://doi.org/10.1523/JNEUROSCI.4100-15.2016

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Science*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (n.d.). *Lexical nature of syntactic ambiguity resolution*. 28.

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 73–87. https://doi.org/10.1080/14640749208401316

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2016). Linguistic bias modulates interpretation of speech via neural delta-Band oscillations. *Cerebral Cortex*, cercor; bhw228v1. https://doi.org/10.1093/cercor/bhw228

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*, 1–11. https://doi.org/10.1080/23273798.2019.1693050

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886–2898. https://doi.org/10.1162/jocn.2009.21400

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00320

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

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

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a Late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, *21*(9), 1693–1708. https://doi.org/10.1162/jocn.2008.21153

Sörqvist, P., & Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: What Is the role for working memory capacity? *Journal of Speech, Language, and Hearing Research*, *55*(1), 210–218. https://doi.org/10.1044/1092-4388(2011/10-0353)

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267. https://doi.org/10.1016/0010-0277(94)00647-4

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191–196. https://doi.org/10.1038/5757

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64

VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences*, *20*(10), 723–735. https://doi.org/10.1016/j.tics.2016.07.006

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical tracking of surprisal during continuous speech comprehension. *Journal of Cognitive Neuroscience*, *32*(1), 155–166. https://doi.org/10.1162/jocn_a_01467