"Acoustic-driven oscillators as cortical pacemaker": a commentary on Meyer, Sun & Martin (2019)

Oded Ghitza

Published online: 10 Apr 2020.

Submit your article to this journal

Article views: 197

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

COMMENTARY

"Acoustic-driven oscillators as cortical pacemaker": a commentary on Meyer, Sun & Martin (2019)

Oded Ghitza[a,b]

aDepartment of Biomedical Engineering & Hearing Research Center, Boston University, Boston, MA, USA; bDepartment of Neuroscience, Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany

ABSTRACT

This is a commentary on a review article by Meyer, Sun & Martin (2019), "Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing", doi:10.1080/23273798.2019.1693050. At the heart of this review article is the language comprehension process. Anchored at a psycho- and neurolinguistic viewpoint, the article argues for the centrality of endogenous cortical rhythms, not only as the facilitators of processes that generate abstract representations and predictions of language but also of processes that establish intrinsic synchronicity with the acoustics, with the priority to override processes realized by acoustic-driven, exogenous cortical rhythms. In this commentary I propose that the scaffold for the speech decoding process – through parsing – is an acoustic determinant. Whether oscillation driven or not, the decoding process is paced by a hierarchical cortical clock, realized by oscillators locked to the input rhythm in multiple Newtonian-time scales, keeping the decoding process in sync with the linguistic information flow. Only if such a lockstep is secured can reliable decoding proceed.

1. Prelude

The review article "Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing" (Meyer et al., 2019) examines the possible role of cortical rhythms in the *language comprehension process*, end-to-end. This process encompasses two distinct processes: (i) *a speech process*, which maps the acoustics into abstract representation of linguistic units, and (ii) *a language process*, which uses these units to derive language features, including syntax and sentence-level semantics. The authors argue for the centrality of endogenous cortical oscillators, not only at the core of the language process but also with the priority to override processes with acoustic-driven cortical oscillators at their core. My commentary concludes that, for reliable language comprehension, both the speech process and the language process *must* operate within cortical time units (CTUs) determined by the acoustics. How did I arrive to this conclusion?

2. Role of oscillators – current view

Speech (*everyday* speech, in particular) is inherently a quasi-rhythmic phenomenon in which the talker's linguistic information is transmitted in "packets", manifested in the

Glossary table.

Speech process	Input = acoustics, output = phrase constituent candidates (PCCs)
Language process	Input = PCCs, output = syntax and sentence-level semantics
Parsing	The exhaustive division of the incoming speech into candidate constituents
Segmentation	The function of setting a time-varying window structure synchronised with the input.
Flexible oscillators	e.g. the VCO component in the classical phase-locked-loop (PLL) circuit, with quasi-periodic oscillations that are locked to the input quasi rhythm. Both theta and delta are flexible.
Syllable chunk	An acoustic chunk, with location and duration defined by the theta cycle
Phrase chunk	An acoustic chunk, with location and duration defined by the delta cycle
Oscillators **in sync** with input	When the acoustic chunks are aligned with proper linguistic units
Phrase constituent candidates (PCCs)	Generated by parsing, per one phrase chunk
Cortical time units (CTUs)	A theta CTU = one theta cycle (in Newtonian time)
	A delta CTU = one delta cycle

acoustic signal in the form of temporal "chunks". Oscillation-based models of the speech process postulate a cortical computation principle by which, the *decoding* process is performed on acoustic chunks defined by a time-varying window structure synchronised with the input on multiple time scales. In the following we shall exemplify this computation principle with TEMPO (Ghitza, 2011), a model which epitomises recently proposed oscillation-
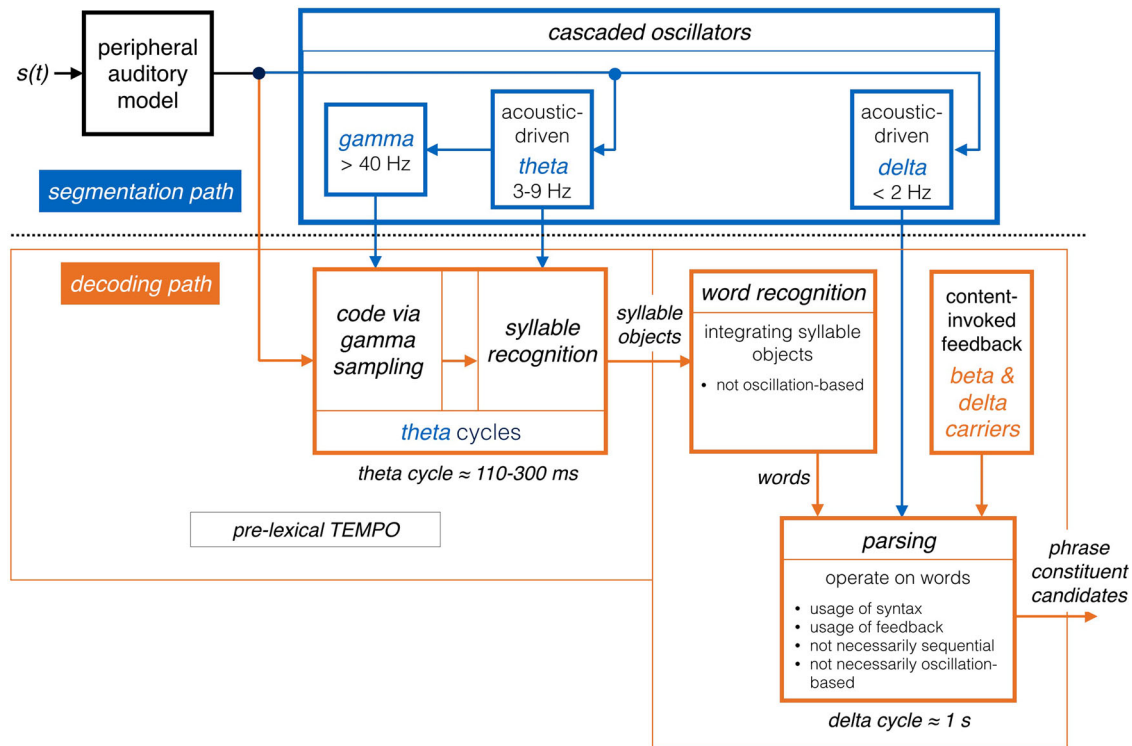
Figure 1. TEMPO. (i) The segmentation path. The theta and delta oscillators are flexible, e.g. the VCO component in the classical PLL circuit (Viterbi, 1966; Ahissar et al., 1997; see also the biophysical computational model by Pittman-Polletta et al., 2020), with quasi-periodic oscillations that are locked to the quasi-rhythmic acoustic syllable- and phrase-chunks. (ii) The decoding path. Decoding is steered by segmentation: the decoding process evolves within the theta/delta cycles. See Figures 2 and 3 for the sequence of operations on the syllable and phrase levels, respectively.

based models of speech perception (e.g. Ahissar & Ahissar, 2005; Ding & Simon, 2009; Ghitza & Greenberg, 2009; Giraud & Poeppel, 2012; Gross et al., 2013; Peelle & Davis, 2012; Poeppel, 2003).

The model is shown in Figure 1. The sensory stream (generated by a model of the auditory periphery, e.g. Chi et al., 1999; Messing et al., 2009) is processed, simultaneously, by a *segmentation* path and a *decoding* path (upper and lower paths of Figure 1, respectively). Conventional models of speech perception assume a strict decoding of the acoustic signal.[1] The decoding path of TEMPO, which links acoustic chunks of different durations with stored linguistic memory patterns, conforms to this notion. Not present in conventional models is the segmentation path, which determines the acoustic chunks (their location and duration) to be decoded. As it turns out, segmentation plays a crucial role in explaining a range of counterintuitive psychophysical data that are hard to explain by the conventional models (e.g. Ghitza & Greenberg, 2009; Ghitza, 2012, 2014, 2017). In TEMPO, the segmentation path is realised by an array of flexible oscillators locked to the input rhythm.

In the pre-lexical level of TEMPO, the segmentation process is realised by a *flexible theta oscillator* locked to the input syllabic rhythm, where the theta cycles constitute the syllabic windows. A theta cycle is set by an evolving phase-locking process (e.g. a PLL circuit, Ahissar et al., 1997; Viterbi, 1966), during which the code is generated. Doelling et al. (2014) provided magnetoencephalography (MEG) evidence for the role of theta, showing that intelligibility is correlated with the existence of acoustic-driven theta neuronal oscillations.

In the phrase level, the segmentation process is realised by a *flexible delta oscillator* locked to the input phrase-chunk rhythm, where the delta cycles constitute the phrase-chunk windows. A delta cycle is set by an evolving phase-locking process, during which contextual parsing proceed. Rimmele et al. (2020) provided MEG evidence for the role of acoustic-driven delta, showing that the accuracy of digit retrieval is correlated with the existence of acoustic-driven delta neuronal oscillations.

3. Role of oscillators – a broader look

As seen in Section 2, the functional role of the acoustic-driven theta and delta oscillators is to facilitate a time-varying window structure, synchronised with the input, where the theta/delta cycles determine the syllable/phrase chunks to be decoded. In this Section, a broader functional role for the acoustic-driven theta and delta is
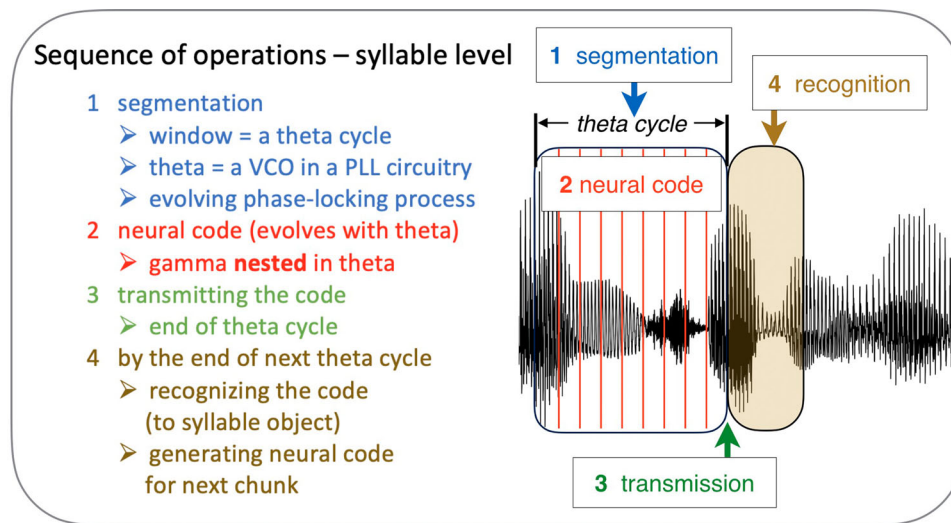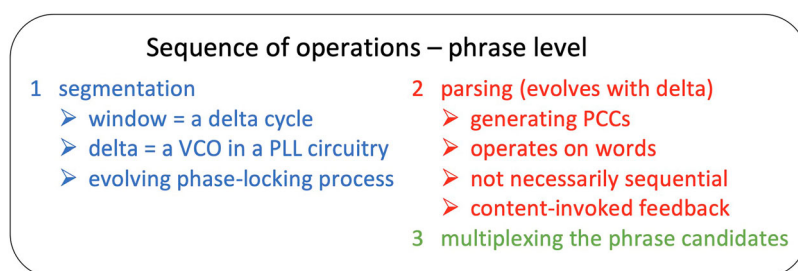
Figure 2. The sequence of operations at the syllable level. See text for details.

postulated, namely, they constitute an internal, hierarchical clock that pace the speech decoding process to stay in sync with the linguistic information flow, via keeping the decoding process operating on acoustic chunks aligned with proper linguistic units. In the following, I shall outline the rationale for this postulate.

In Figure 2, the sequence of operations that are executed in mapping the acoustic stream onto a series of syllable objects is outlined in more detail. First is the segmentation process, in the form of acoustic-driven theta cycle, set by an evolving phase-locking process[2] (step 1 in Figure 2). While the theta cycle is evolving, a neural code for the syllable chunk is generated throughout the theta cycle, e.g. in the form of gamma *nested* in theta[3] (step 2). The code is transmitted at the end of the theta cycle (step 3), then recognised (i.e. a working memory storage is activated) during the next theta cycle (step 4). An additional functional role of theta – beyond the setting of the theta window – emerges: the end-time of the theta cycle marks the moment at which the code is transmitted, i.e. it marks the moment by which the code generation *must* end. This is a necessary condition because, beyond this moment, the code-generation circuitry should already be occupied with the generation of the code for the next theta chunk.

Turning to the phrase level, the sequence of operations that take place in mapping the stream of syllable objects onto phrase constituent candidates (PCCs) is shown in Figure 3. Segmentation comes first, in the form of acoustic-driven delta set by an evolving phase-locking process[4] (step 1 in Figure 3). While the delta cycle is evolving, PCCs are obtained by a parsing process that take place throughout the delta cycle (step 2). The PCCS are multiplexed at the end of the delta cycle (step 3). An additional functional role of delta emerges, analogous to that of theta: the end-time of the delta cycle marks the moment by which the PCCs *must* be delivered. Three points merit discussion. First, while we find cortical oscillations with cycle durations that correspond to syllables and phrases (theta and delta), we do not have oscillations that correspond to words. Indeed, there is no compelling linguistic evidence that words are regular enough for phase locking. Therefore, in TEMPO, the lexical access process operates on the syllable stream without any segmentation-based supervision (see, for example, the model TRACE, Luce & McLennan, 2005). Second, in generating the PCCs, numerous computation strategies can be considered (e.g. template matching; statistical pattern recognition; predictive coding; inference Bayesian approach; analysis-by-synthesis; relations via correlations; statistical learning). The

Figure 3. The sequence of operations at phrase level. PCCs = Phrase constituent candidates. See text for details.
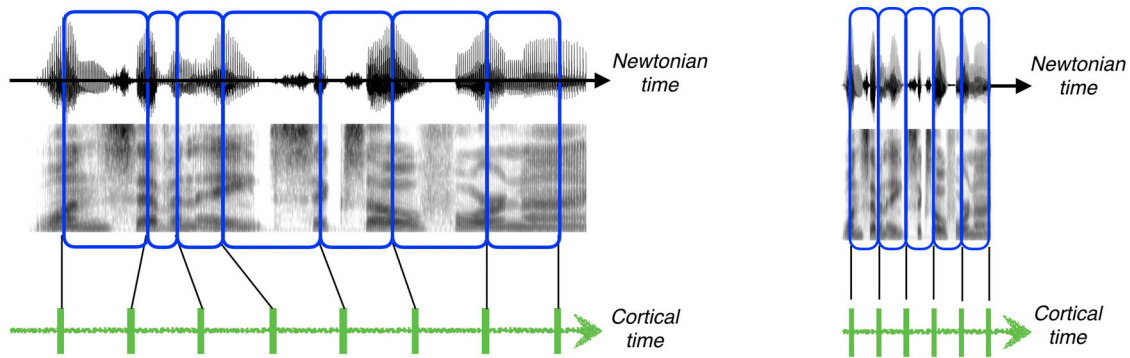
Figure 4. Newtonian time and cortical time, illustrated at the syllable level for normal rate (left) and fast speech (right). In both speeds, decoding proceeds uniformly in cortical time and syllable objects are transmitted one per theta CTU tick. In normal rate (left), the theta tracking is successful ⇒ a syllable chunk associated with a theta CTU is aligned with a syllabic unit. However, when the input rate is too fast (right, speech is time-compressed by 3) theta is "stuck" at upper frequency range ⇒ loss of tracking ⇒ acoustic chunks associated with the theta CTUs are no longer aligned with syllabic units.

parsing process operates on words throughout the delta window and is not necessarily sequential, nor it is necessarily oscillation-based.[5] Important to our discussion, regardless of the computation strategy, in order to stay in lockstep with the input information flow the derivation of the PCCs *must* be concluded by the end of the delta window. And third, in deriving language features, including syntax and sentence-level semantics, the language process operates on a *sequence* of PCCs that span a few delta cycles. A few questions – beyond the scope of this commentary – remain open, e.g.: how the duration of the language-process "window" is determined? Is it formed by a segmentation process realised by ultra-slow oscillators locked to the sentence-level information flow?

4. Cortical time units

The sequence of operations described in Section 3, in the syllable level and in the phrase level, *is repetitive*, irrespective of the theta/delta window durations. Functionally, therefore, the speech decoding process can be viewed as a process paced by an internal clock with *uniform* cortical time units (CTUs): (i) a theta CTU, with duration – in Newtonian time[6] – of one theta cycle, and (ii) a delta CTU, with duration of one delta cycle. The CTUs are set by oscillators that are in sync with the input. As such, the CTUs, uniform in the internal domain, span non-uniform durations in Newtonian time (Figure 4, left). Crucially, the CTUs have a limited range, bounded in Newtonian time by the upper frequency range of the oscillators. Hence, the shortest duration of a theta CTU is about 125 ms (for $theta_{max} = 8$ Hz), and the shortest duration of a delta CTU is about 0.5 s (for $delta_{max} = 2$ Hz). Speech decoding, therefore, is viewed as a process that proceeds in uniform cortical-time ticks: at the syllable level, the

entire sequence of operations in Figure 2 is executed in one theta CTU; at the phrase level, the entire sequence of operations in Figure 3 is executed in one delta CTU.

In Sections 4.1 and 4.2 we shall examine, through the internal clock prism, the resulting output of TEMPO when the input is speech at normal rate, and when it is accelerated. Recall that the intelligibility of time-compressed speech is flawless when the speech rate is inside the theta range, and is sharply deteriorated when the rate is outside theta (e.g. Foulke & Sticht, 1969; Garvey, 1953; Ghitza, 2014). As we shall see, as long as the input is at normal rate, the CTUs are aligned with acoustic chunks associated with syllables and phrases in their primitive sense, hence the internal clock and the linguistic information flow are in lockstep. When the input speech rate is too fast, the CTUs are no longer aligned with proper linguistic units, hence synchronisation is lost.

4.1. Input rate inside theta range (*Figure 4, left*)

In cortical time, syllabification and parsing proceed uniformly: a syllable object is generated and transmitted every theta CTU tick, and the PCCs are generated and multiplexed every delta CTU tick. Importantly, the derivation of the PCCs is concluded within one delta CTU, regardless of computation strategy.

4.2. Input rate too fast (*Figure 4, right*)

Two scenarios are considered: (i) the syllable-chunk rate is outside the theta range, but the phrase-chunk rate is inside the delta range, and (ii) the syllable-chunk rate is inside but the phrase-chunk rate is outside. In both scenarios there is a mismatch between the linguistic

information flow and the internal clock, resulting in a deterioration in performance.

In scenario (i), viewed in Newtonian time, since the syllable-chunk rate is outside theta range, the synchronisation between the acoustic stream and the theta oscillator is disrupted because the oscillator reaches its upper boundary. The oscillator is stuck at frequency theta$_{max}$ ⇒ erroneous segmentation, in both the location and the duration of the theta window ⇒ the acoustic chunk is no longer aligned with a syllabic unit ⇒ the stream of syllable objects is corrupted. Consequently, the resulting PCCs are in error. Viewed in Cortical time, objects are transmitted per theta CTU tick (each spans a duration of one theta$_{max}$ cycle, Newtonian time) but with error. The error in the syllable-objects stream affect parsing: indeed, the PCCs are emitted per delta CTU tick, in sync with the phrase-chunk rate but with a compromised accuracy due to the erroneous syllable-objects stream.

In scenario (ii), if the syllabic rate is inside the theta range, synchronisation on the syllable level is maintained and syllable objects are correctly recognised, one per theta CTU. However, synchronisation between the acoustic stream and the delta oscillator is disrupted because the oscillator is stuck at frequency delta$_{max}$, resulting in erroneous segmentation in both the location and the duration of the delta window. Consequently, the PCCs – emitted one per delta CTU tick (each spans a duration of delta$_{max}$ cycle, in Newtonian time) – are in error.

4.3. Partial restoration of intelligibility

As we see, for fast speech the deterioration in intelligibility is the result of a mismatch between the internal clock and the information stream, such that the acoustic chunks associated with CTUs are no longer aligned with proper linguistic units. In order to restore intelligibility, the speech acoustics should be modified, in order to bring the input rate back inside the range of the internal clock. Two studies examined this approach: (i) in the syllable level, it has been shown that intelligibility is improved as a result of "repackaging" – a process of dividing the time-compressed waveform into fragments, called packets, and delivering the packets in a prescribed rate determined by insertion of gaps in-between the packets (Ghitza & Greenberg, 2009; Ghitza, 2014; see Christiansen & Chater, 2016). The insertion of gaps is, in fact, a procedure of tuning the packaging rate in a search for a better synchronisation between the input information flow and the cortical clock, resulting in improvement in intelligibility. And (ii) in the phrase level, it has been shown that performance is impaired when the phrase-chunk presentation rate is outside the delta range, and that performance is restored by bringing the chunk rate back inside the delta range via inserting gaps in-between the chunks (Ghitza, 2017; Rimmele et al., 2020).

5. Summary

We claim that, from a functional role perspective, speech decoding is a process paced by an internal, hierarchical clock with uniform CTUs, a theta CTU with duration of one theta cycle in Newtonian time, and a delta CTU with duration of one delta cycle. The CTUs are synchronised with the input. A necessary condition emerges according to which, the sequence of operations to decode one syllable *must* be performed within one theta CTU and the sequence of operations to parse one phrase *must* be performed within one delta CTU. Importantly, these necessary conditions hold for any decoding computation strategy that may be in place, whether context-invoked or not, whether sequential or not, or whether oscillation driven or not. Hence, the scaffold for the speech decoding process is an acoustic determinant, realised by acoustic-driven theta and delta oscillators.

Notes

1. In conventional models of speech perception phones are identified first, and the ordered sequence of identified phonemes results in a pointer to the word lexicon (e.g. Marslen-Wilson, 1987; Luce & McLennan, 2005; Stevens, 2005).
2. The acoustic cues to which the theta oscillator is locked to are still under debate (acoustic edges? vocalic nuclei?). Here, the theta cycle is locked to vocalic nuclei, hence the syllable objects are in the form of VCVs (Ghitza, 2013).
3. A possible mechanism to generate the neural code is via gamma sampling (Shamir et al., 2009; Ghitza, 2011).
4. The delta oscillator is locked to *accentuation* attributes; the acoustic cues that form accentuation are still under pursuit.
5. The role of endogenous oscillations in generating abstract linguistic predictions (e.g. Meyer & Gumbert, 2018) is still under debate.
6. Newtonian time, in seconds. See Chapter "Newtonian and Bergsonian Time," in Wiener, 1948.

Disclosure statement

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

Funding

References

Ahissar, E., & Ahissar, M. (2005). Processing of the temporal envelope of speech. In R. Konig, P. Heil, E. Bundinger, & H. Scheich (Eds.), *The auditory cortex. A synthesis of human and animal research Ch 18.* (pp. 295–314). Lawrence Erlbaum. https://doi.org/10.4324/9781410613066.

Ahissar, E., Haidarliu, S., & Zacksenhouse, M. (1997). Decoding temporally encoded sensory input by cortical oscillations and thalamic phase comparators. *Proceedings of the National Academy of Sciences*, *94*(21), 11633–11638. https://doi.org/10.1073/pnas.94.21.11633

Chi, T., Gao, Y., Guyton, M., Ru, P., & Shamma, S. A. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, *106*(5), 2719–2732. https://doi.org/10.1121/1.428100

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, https://doi.org/10.1017/S0140525X1500031X

Ding, N., & Simon, J. Z. (2009). Neural representations of complex temporal modulations in the human auditory cortex. *Journal of Neurophysiology*, *102*(5), 2731–2743. https://doi.org/10.1152/jn.00523.2009

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Foulke, E., & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, *72*(1), 50–62. https://doi.org/10.1037/h0027575

Garvey, W. D. (1953). The intelligibility of speeded speech. *Journal of Experimental Psychology*, *45*(2), 102–108. https://doi.org/10.1037/h0054381

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130. https://doi.org/10.3389/fpsyg.2011.00130

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

Ghitza, O. (2014). Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, *5*, 652. https://doi.org/10.3389/fpsyg.2014.00652

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*. *32*(5). https://doi.org/10.1080/23273798.2016.1232419

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1-2), 113–126. https://doi.org/10.1159/000208934

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

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., & Poeppel, D. (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

Luce, P. A., & McLennan, C. (2005). Spoken word recognition: The challenge of variation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 591–609). Blackwell. https://doi.org/10.1002/9780470757024.ch24

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Messing, D. P., Delhorne, L., Bruckert, E., Braida, L. D., & Ghitza, O. (2009). A non-linear efferent-inspired model of the auditory system; matching human confusions in stationary noise. *Speech Communication*, *51*(8), 668–683. https://doi.org/10.1016/j.specom.2009.02.002

Meyer, L., & Gumbert, M. (2018). Synchronization of delta-band oscillations to syntax benefits linguistic information processing. *Journal of Cognitive Neuroscience*, *30*(8), 1066–1074. https://doi.org/10.1162/jocn_a_01236

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

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320

Pittman-Polletta, B. R., Wang, Y., Stanley, D. A., Schroeder, C. E., Whittington, M. A., & Kopell, N. J. (2020). Differential contributions of synaptic and intrinsic inhibitory currents to parsing via excitable phase-locking in neural oscillators. *bioRxiv*. https://doi.org/10.1101/2020.01.11.902858

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time". *Speech Communication*, *41*(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Rimmele, J. M., Poeppel, D., & Ghitza, O. (2020). Acoustically driven cortical delta-band oscillations underpin perceptual chunking. Submitted.

Shamir, M., Ghitza, O., Epstein, S., & Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Computational Biology*, *5*(5), e1000370. https://doi.org/10.1371/journal.pcbi.1000370

Stevens, K. (2005). Features in speech perception and lexical access. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 125–155). Blackwell Publishing.

Viterbi, A. J. (1966). *Principles of coherent communication*. McGraw-Hill.

Wiener, N. (1948). *Cybernetics*. MIT Press.