



OPEN ACCESS

EDITED BY

Conor J. Houghton,
University of Bristol, United Kingdom

REVIEWED BY

Zhongtao Hu,
Beihang University, China

*CORRESPONDENCE

Anton Rogachev
✉ aorogachev@gmail.com

RECEIVED 02 July 2025

ACCEPTED 14 August 2025

PUBLISHED 03 September 2025

CITATION

Rogachev A and Sysoeva O (2025) A functional systems view on neural tracking of natural speech. *Front. Syst. Neurosci.* 19:1658243. doi: 10.3389/fnsys.2025.1658243

COPYRIGHT

© 2025 Rogachev and Sysoeva. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A functional systems view on neural tracking of natural speech

Anton Rogachev^{1*} and Olga Sysoeva^{1,2}

¹Center for Cognitive Sciences, Sirius University of Science and Technology, Sirius, Russia, ²Laboratory of Higher Nervous Activity of Human, Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Moscow, Russia

KEYWORDS

neural tracking, neural entrainment, natural speech, functional systems, brain activity, rhythmic activity

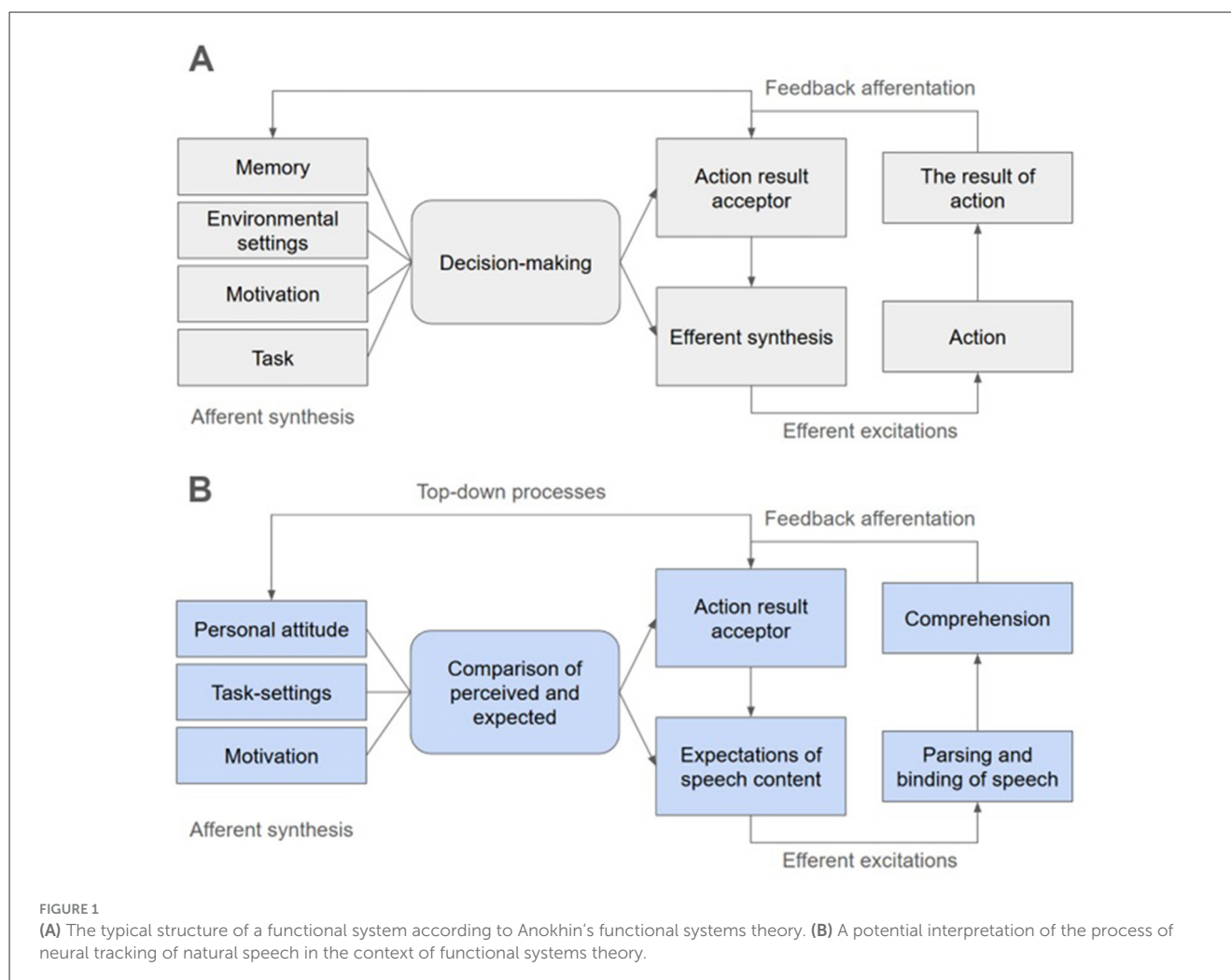
1 Introduction

Speech is a core component of human life underlying communication with the world and inner mental life. For many years, human ability to comprehend and produce speech has been widely studied in psychology and neuroscience. Neuroscience uses a wide range of invasive and non-invasive neuroimaging techniques to study brain responses to acoustic, phonetic, lexical and semantic features of speech, and how the speech acoustic (physical signal) is mapped to words, sentences and texts meanings (non-material semantics) through cerebral cortex.

An important way of neuroscience of speech research in the last decade is the focus on naturalistic speech stimuli (Hamilton and Huth, 2020; Zoefel and Kösem, 2024). In this context, studies of *neural tracking of natural speech* have a special place in literature. Neural tracking, or neural entrainment, could be defined as a phenomenon of synchronization between brain electrical activity and continuous or discrete changes in the speech stimulus components during its active, attentive perception (Lalor and Foxe, 2010; Ding and Simon, 2014). Many studies revealed that neural tracking is functionally related to speech comprehension (Ahissar et al., 2001; Broderick et al., 2022), the level of attention directed to speech (Crosse et al., 2016; Sánchez-Costa et al., 2025), and the development of speech and language in children (Rogachev and Sysoeva, 2024). Moreover, neural tracking is disturbed in some clinical cases of developmental language disorders (Nora et al., 2024) or post-stroke aphasia (De Clercq et al., 2025). Thus, neural tracking research has both fundamental and applied significance and potential implications.

In the literature, there is a substantial discussion on the nature and functional roles of neural tracking of speech. Summarizing, this discussion raises several questions. What is neural tracking physiologically: endogenous oscillatory activity associated with processing the rhythm of speech, or the sum of evoked activity on the acoustic edges of a speech audio signal (Meyer et al., 2020; Lalor and Nidiffer, 2025)? Are the so-called speech rhythms [e.g., electromagnetic rhythmic brain activity related to the averaged rhythms of phonemes, words and phrases in a natural speech (Poeppel and Assaneo, 2020)] indeed linked to the rhythms of speech through endogenous oscillators tuned to the rhythm of the native language in ontogenesis, or are they also only represent evoked activity (Kazanina and Tavano, 2023a,b)? What is the contribution of top-down processes, i.e., the influence of prior experience and predictions about speech content, on neural tracking (Broderick et al., 2019; Gwilliams, 2020; Klimovich-Gray and Molinaro, 2020; Klimovich-Gray et al., 2023)?

We propose that these questions can be narrowed to the one: is neural tracking an active process, based on endogenous oscillators activity which is based on their inner state and predictions about speech content, or is it a passive process of cortical tuning to the rhythm of speech stimulus? Of course, the answer to this question requires a lot of



empirical work. However, an attempt to address this question can be given from a theory-driven point of view, which may provide a new perspective on the problem of the nature of neural tracking. Here, we provide a possible theoretical interpretation of neural tracking of speech from the perspective of functional systems theory to raise questions regarding the construction of experiments on natural speech perception and data analysis approaches, empirical answers to that can help solve common neural tracking issues.

2 Functional systems theory

Functional systems theory (FST) is a psychophysiological theoretical framework, developed by Russian and Soviet physiologist Pyotr Anokhin, describing the mechanisms of interaction between a person (and, in general, any organism) and the external environment (Anokhin, 1971; Egiazaryan and Sudakov, 2007; Rusalov, 2018; Vityaev and Demin, 2018; Shadrnikov, 2019). Functional system (FS), a core concept of FST, is the dynamic integration of different body and brain systems for the realization of behavior aimed at achieving a specific adaptive goal. For example, if an organism begins to experience hunger, its

physiological systems (sensory, central and motor) will be tuned to solve a sequence of tasks: searching for opportunities to get food, deciding on a specific search strategy, performing these actions, etc., with possible correction of decisions during the execution. According to FST, any organism is an active subject whose behavior is determined by an adaptive behavioral goal in the future rather than a set of environmental or intrinsic stimuli in the past.

The typical structure of FS is presented in Figure 1A. In FS, there are four units connected by feedback loops. The first unit, afferent synthesis, synthesizes information preceding a current behavioral act (i.e., motivation, previous experience, external environment state, task specifications, etc.). Based on afferent synthesis, a decision is made, which is also based on the image of the result, which is located in the action result acceptor. This resulted in efferent synthesis, in which mental and somatic systems are programmed to perform a behavioral act to achieve the result. Next, the behavioral act itself is realized. During its realization, there is a constant comparison of current results with expected ones by the action result acceptor. Thus, actions, decisions and images of results can be adjusted during the implementation of the behavioral act in accordance with the available opportunities for its realization. Notably, FSs are “dimensionless”: they can encompass behavioral

acts of different duration and scope, be nested within each other, and flow one into the other.

It can be seen that Anokhin's FST operates with abstract concepts describing any behavior as a whole, without reference to psychological or physiological processes, but capturing both psychological and physiological processes in a single united system. According to the theory, the components of the FS cannot be localized in specific organs or brain areas, but they are dynamically built during the realization of behavior based on the functional predispositions of organs and brain areas. Continual activity of the organism can reflect the dynamics of the FS.

Psychologically and physiologically, FST provides, first, a new perspective on behavior (with determination through goal and future outcome rather than past stimulation), and second, a possible solution to the brain-mind problem (mental processes cannot be localized in the brain, but components of the FS directed to the realization of behavioral acts can be observed using neuroimage methods). It can be observed that the FST shares conceptual similarities with other "active" neurocognitive theories. For instance, the predictive coding approach explains behavioral and brain activity by means of the continuous generation of hypotheses regarding the environment and their subsequent verification (Friston and Kiebel, 2009; Shipp, 2016). Additionally, both theories highlight the importance of predictions in brain activity and behavior, as well as the significance of error correction in this process. However, FST incorporates both physiological and psychological elements, combining them into a unified, irreducible system, while the predictive coding approach is better grounded into neurophysiological processes.

It is important to note that FST is a high-order abstract theory which describes holistic processes related to the interactions between biological organisms and the environment. Neural tracking, on the contrary, is an empirical phenomenon of electromagnetic brain activity. We believe that the FST framework can be applied to interpretations of neural tracking research results since it is already implemented in some neuroscience studies of cognition, perception, learning and emotions (Alexandrov and Sams, 2005; Alexandrov, 2015; Alexandrov and Pletnikov, 2022), it has rarely been compared with current trends in experimental studies. However, it can offer new perspectives on the planning and analysis of experiments.

3 A potential interpretation of neural tracking of speech using FST

It is noteworthy to briefly discuss that neural tracking of speech can be thought of as a part of an active process of speech comprehension. Indeed, usually studies provide active tasks to human participants (e.g., "please listen to the audio to further answer questions about its content"). In such task-setting a person is given a task with the final goal of answering questions, and the person's activity during the listening session will be determined by this goal (however, if the person has the internal or external motivation to complete the task). This is supported by studies of dichotic listening, in which only attended stimulus shows its neural tracking (Crosse et al., 2016; Broderick et al., 2018). The task of listening to only one channel determines the specific alignment

of physiological systems, which in turn also determines where attention will be directed.

Thus, neural tracking can be represented as a *physiological part* of task-specific FS aimed at speech comprehension (Figure 1B). It can be assumed that in order to achieve the goal of speech comprehension, bottom-up components of FS are formed, aimed at decoding linguistic information from the acoustic speech signal, parsing linguistic units, their binding into higher order meaningful units (words, word combinations, phrases, sentences, narratives), and mapping these units to corresponding semantic content (Ding et al., 2016). It is essential that at this stage, a person's individual attitude to the listened speech is also developed, and emotional responses are formed. This can be attributed to the units of afferent synthesis and action (which has perhaps implicit nature) to perceive the linguistic unit that is now the focus of attention.

It is also known that speech perception is based on the principle of predictive coding of further content of speech (Hovsepian et al., 2020; Heilbron et al., 2022); this can be attributed to the action result acceptor. The effects of semantic mismatch expressed in negative brain activity in response to the mismatch (the so-called N400 effect (Broderick et al., 2018; Nour Eddine et al., 2024)) can be interpreted as a correction of the action result acceptor via feedback afferentation. In this way, top-down regulation of neural tracking of speech based on its content and changes in this content can occur, which is confirmed by some studies (Broderick et al., 2019; Klimovich-Gray et al., 2023). In the structure of FS, this can be referred to action result acceptor, loops of efferent feedback, orienting feedback and decision-making (implicit decision-making about the congruence of the perceived and expected).

Interestingly, some EEG studies of speech perception show that pre-stimulus (before a sound or a word presentation) activity correlates with and predicts post-stimulus activity (Barraza et al., 2016; Hansen et al., 2019; Liu and Liu, 2025). We would like to propose that this phenomenon could be an indicator of the FS functioning through mechanisms of efferent and orienting feedback, and also the correction of the action results acceptor state.

In general, the outcome and positive adaptive effect of the considered FS is speech understanding at the psychological level and neural tracking at the physiological level.

4 Future directions

Why could FST be useful to interpret findings of neural tracking of speech studies? It is a theoretical framework that allows us to consider the psychological and physiological sides of cognitive processes as a whole. It allows these processes to be viewed as active, aimed at achieving an adaptive beneficial outcome. It is likely that this interpretation can help in furthering our understanding of brain functioning in different conditions, as well as in understanding the relationship between physiological and psychological processes.

Considering neural tracking of speech as an active process of speech comprehension raises new questions for empirical testing. How do experimental task peculiarities affect not only behavioral but also physiological outcomes? How can the content of speech stimuli and a person's attitude toward them influence neural tracking? What psychological factors, other than attention level, influence neural tracking?

Another line of work that is spawned by the FST approach is changing approaches to data analysis. According to the theory, system components cannot be localized to specific brain regions. It can be useful to apply network analysis methods and analyze global brain activity and its dynamics when listening to speech stimuli. It may be also of interest to investigate the relationship between pre-stimulus and post-stimulus EEG activity in response to words in the context of natural speech perception. Additionally, new predictive linguistic measures can be developed to analyze neural tracking of natural speech data. Current and widely used measures (e.g., semantic dissimilarity or word frequency) reflect the processing of words that have already been heard or are currently being heard; new measures should consider a subject's expectations regarding the upcoming content of speech (e.g., emotional attitudes, perception of the entire narrative, etc.).

The interpretation of neural tracking using FST has certain methodological limitations. While the key components of FST mentioned above do not necessarily have unambiguous neurobiological counterparts, their neurophysiological foundation requires further empirical verification. Additionally, it is necessary to investigate the extent to which neural tracking can truly be linked to the goal of the system. This can be achieved through experiments that involve goal-setting inconsistencies. Furthermore, FST offers a qualitative rather than quantitative description of processes, necessitating the development of mathematical tools. However, despite these limitations, the heuristic significance of FST should not be disregarded. Its emphasis on purposeful and flexible nature provides a framework for exploring the context-specific dynamics of neural tracking that other models may overlook.

Author contributions

AR: Writing – review & editing, Conceptualization, Funding acquisition, Project administration, Writing – original draft. OS: Conceptualization, Writing – review & editing.

References

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci.* 98, 13367–13372. doi: 10.1073/pnas.201400998
- Alexandrov, Y. I. (2015). "Cognition as systemogenesis," in *Anticipation: Learning from the Past*, ed. M. Nadin (Cham: Springer International Publishing), 193–220. doi: 10.1007/978-3-319-19446-2_11
- Alexandrov, Y. I., and Pletnikov, M. V. (2022). Neuronal metabolism in learning and memory: the anticipatory activity perspective. *Neurosci. Biobehav. Rev.* 137:104664. doi: 10.1016/j.neubiorev.2022.104664
- Alexandrov, Y. I., and Sams, M. E. (2005). Emotion and consciousness: ends of a continuum. *Cogn. Brain Res.* 25, 387–405. doi: 10.1016/j.cogbrainres.2005.08.006
- Anokhin, P. K. (1971). Philosophical aspects of the theory of a functional system. *Sov. Stud. Philos.* 10, 269–276. doi: 10.2753/RSP1061-19671003269
- Barraza, P., Jaume-Guazzini, F., and Rodríguez, E. (2016). Pre-stimulus EEG oscillations correlate with perceptual alternation of speech forms. *Neurosci. Lett.* 622, 24–29. doi: 10.1016/j.neulet.2016.04.038
- Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.* 28, 803–809.e3. doi: 10.1016/j.cub.2018.01.080
- Broderick, M. P., Anderson, A. J., and Lalor, E. C. (2019). Semantic context enhances the early auditory encoding of natural speech. *J. Neurosci.* 39, 7564–7575. doi: 10.1523/JNEUROSCI.0584-19.2019
- Broderick, M. P., Zuk, N. J., Anderson, A. J., and Lalor, E. C. (2022). More than words: neurophysiological correlates of semantic dissimilarity depend on comprehension of the speech narrative. *Eur. J. Neurosci.* 56, 5201–5214. doi: 10.1111/ejn.15805
- Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10:604. doi: 10.3389/fnhum.2016.00604
- De Clercq, P., Kries, J., Mehraram, R., Vanthornhout, J., Francart, T., and Vandermosten, M. (2025). Neural tracking of natural speech: an effective marker for post-stroke aphasia. *Brain Commun.* 7:fcaf095. doi: 10.1093/braincomms/fcaf095
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Russian Science Foundation, grant number 25-28-01349, <https://rscf.ru/en/project/25-28-01349/>.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Egiazaryan, G. G., and Sudakov, K. V. (2007). Theory of functional systems in the scientific school of P.K. Anokhin. *J. Hist. Neurosci.* 16, 194–205. doi: 10.1080/09647040600602805
- Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300
- Gwilliams, L. (2020). Hierarchical oscillators in speech comprehension: a commentary on Meyer, Sun, and Martin (2019). *Lang. Cogn. Neurosci.* 35, 1114–1118. doi: 10.1080/23273798.2020.1740749
- Hamilton, L. S., and Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. *Lang. Cogn. Neurosci.* 35, 573–582. doi: 10.1080/23273798.2018.1499946
- Hansen, N. E., Harel, A., Iyer, N., Simpson, B. D., and Wisniewski, M. G. (2019). Pre-stimulus brain state predicts auditory pattern identification accuracy. *NeuroImage* 199, 512–520. doi: 10.1016/j.neuroimage.2019.05.054
- Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., and de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proc. Natl. Acad. Sci.* 119:e2201968119. doi: 10.1073/pnas.2201968119
- Hovsepian, S., Olasagasti, I., and Giraud, A.-L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nat. Commun.* 11:3117. doi: 10.1038/s41467-020-16956-5
- Kazanina, N., and Tavano, A. (2023a). Reply to ‘What oscillations can do for syntax depends on your theory of structure building.’ *Nat. Rev. Neurosci.* 24, 724–724. doi: 10.1038/s41583-023-00735-4
- Kazanina, N., and Tavano, A. (2023b). What neural oscillations can and cannot do for syntactic structure building. *Nat. Rev. Neurosci.* 24, 113–128. doi: 10.1038/s41583-022-00659-5
- Klimovich-Gray, A., Di Liberto, G., Amoruso, L., Barrena, A., Agirre, E., and Molinaro, N. (2023). Increased top-down semantic processing in natural speech linked to better reading in dyslexia. *NeuroImage*, 273:120072. doi: 10.1016/j.neuroimage.2023.120072
- Klimovich-Gray, A., and Molinaro, N. (2020). Synchronising internal and external information: a commentary on Meyer, Sun & Martin (2020). *Lang. Cogn. Neurosci.* 35, 1129–1132. doi: 10.1080/23273798.2020.1743875
- Lalor, E., and Nidiffer, A. (2025). On the generative mechanisms underlying the cortical tracking of natural speech: a position paper. [Preprint]. doi: 10.31219/osf.io/xf8ay_v1
- Lalor, E. C., and Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.* 31, 189–193. doi: 10.1111/j.1460-9568.2009.07055.x
- Liu, S., and Liu, X. (2025). Context influence on speech perception: evidence for acoustic-level mechanism across the voice onset time continuum. *NeuroImage* 310:121140. doi: 10.1016/j.neuroimage.2025.121140
- Meyer, L., Sun, Y., and Martin, A. E. (2020). “Entraining” to speech, generating language? *Lang. Cogn. Neurosci.* 35, 1138–1148. doi: 10.1080/23273798.2020.1827155
- Nora, A., Rinkinen, O., Renvall, H., Service, E., Arkkila, E., Smolander, S., et al. (2024). Impaired cortical tracking of speech in children with developmental language disorder. *J. Neurosci.* 44:e2048232024. doi: 10.1523/JNEUROSCI.2048-23.2024
- Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., and Kuperberg, G. R. (2024). A predictive coding model of the N400. *Cognition* 246:105755. doi: 10.1016/j.cognition.2024.105755
- Poeppel, D., and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* 21, 322–334. doi: 10.1038/s41583-020-0304-4
- Rogachev, A., and Sysoeva, O. (2024). Neural tracking of natural speech in children in relation to their receptive speech abilities. *Cogn. Syst. Res.* 86:101236. doi: 10.1016/j.cogsys.2024.101236
- Rusakov, V. (2018). Functional systems theory and the activity-specific approach in psychological taxonomies. *Philos. Trans. R. Soc. B Biol. Sci.* 373:20170166. doi: 10.1098/rstb.2017.0166
- Sánchez-Costa, T., Carboni, A., and Cervantes Constantino, F. (2025). Never mind the repeat: how speech expectations reduce tracking at the cocktail party. *Cortex* 189, 1–19. doi: 10.1016/j.cortex.2025.05.003
- Shadrikov, V. (2019). The activity theory: the activity psychological functional system and abilities as a mechanism of activity implementation. *Psychol. J. High. Sch. Econ.* 16, 593–607. doi: 10.17323/1813-8918-2019-4-593-607
- Shipp, S. (2016). Neural elements for predictive coding. *Front. Psychol.* 7:1792. doi: 10.3389/fpsyg.2016.01792
- Vityaev, E. E., and Demin, A. V. (2018). Cognitive architecture based on the functional systems theory. *Procedia Comput. Sci.* 145, 623–628. doi: 10.1016/j.procs.2018.11.072
- Zoefel, B., and Kösem, A. (2024). Neural tracking of continuous acoustics: properties, speech-specificity and open questions. *Eur. J. Neurosci.* 59, 394–414. doi: 10.1111/ejn.16221