Neural tracking of speech

Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see http://jneurosci.org/content/jneurosci-journal-club.

Neural Tracking of Speech: Top-Down and Bottom-Up Influences in the Musician's Brain

Katerina D. Kandylaki* **and Antonio Criscuolo***
Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, 6229 ER, Maastricht, The Netherlands
Review of Puschmann et al.

A long-standing debate centers on whether the specialized skills acquired by music experts transfer to other cognitive domains. Musical training is a multisensory experience, and long-term practice has been found to be associated with neuroanatomic and neurofunctional changes (Criscuolo et al., 2021), as well as enhanced high-order cognitive functions (Criscuolo et al., 2019). This, together with the substantial overlap between the neural resources engaged for music and speech processing (Peretz et al., 2015), led to the hypothesis that music practice can foster the neural encoding of speech (Patel, 2011).

In a recent article published in *The Journal of Neuroscience*, Puschmann et al. (2021) provided support for this hypothesis by showing a link between the length of musical training and speech-related neural activity within a bilateral network. Twenty young participants with different musical backgrounds participated in a naturalistic listening experiment. Brain activity was recorded with magnetoencephalography (MEG) and then coregistered to subjects' individual anatomic brain maps acquired with MRI. After reconstructing which brain areas contribute to the MEG signal, the authors used one area as a seed for neural activity in another area and analyzed how strongly the second area aligns its oscillations with the first one [phase-locking value (PLV)], while the subjects were listening to speech. The intersubject PLV is a metric that allows researchers to isolate neural activity related to the processing of speech signals across a group of individuals. Analyses were then extended to quantify inter-regional communication within bilateral speech networks and to correlate the PLV with the duration of musical training.

In line with previous evidence (Park et al., 2015), the study by Puschmann et al. (2021) showed that speech processing engages widely distributed bilateral neural activity, which dynamically tracks distinct linguistic components at multiple spatial and temporal scales (Poeppel, 2003). The results highlighted robust patterns of speech-driven brain synchronization between bilateral auditory regions and frontotemporal and premotor networks, with oscillations in the alpha (8–13 Hz), theta (4–8 Hz), and delta (1–4 Hz) frequency bands. Moreover, the duration of musical training was found to be positively associated with inter-regional PLV in the alpha band. Accordingly, the authors support the hypothesis that musical practice can augment fine-grained auditory-processing abilities that transfer to speech processing (Patel, 2011; Peretz et al., 2015).

It is possible that brain dynamics in the alpha, theta, and delta ranges described by Puschmann et al. (2021) reflect the tracking of linguistic components at multiple time scales (Poeppel, 2003; Bornkessel-Schlesewsky et al., 2015; Kandylaki, 2015), ranging from phoneme (in this study: 11.2 Hz; range, 9.8–12.8 Hz), to syllable (4.5 Hz; range, 3.7 frequency 4.9 Hz), and to word (3.4 Hz; range, 2.8–3.9 Hz) rates. The results reveal fine-grained auditory-processing mechanisms and support the earlier formulation of the "asymmetrical sampling in time" theory (Poeppel, 2003), which states that symmetrical brain regions engage in asymmetrical processing of acoustic features (by means of high- and low-frequency oscillatory components). In this view, linguistic and acoustic components may drive a hierarchical organization featuring fast events at the bottom (e.g., phonemes; alpha–beta activity) and slower events at the top (e.g., words, sentences; delta activity) of the hierarchy.

In opposition to the asymmetrical sampling in time theory, when analyzing the directionality of inter-regional interplay within the speech network, a reversed hierarchical functional organization emerged (Park et al., 2015). This organization features the interaction between bottom-up sensory processing (from auditory-to-frontal regions) traveling along

Received Apr. 9, 2021; revised May 16, 2021; accepted May 24, 2021.
*K.D.K. and A.C. contributed equally to this work.
For the initial discussion, which brought up many of the points discussed in the current article, we thank all the members of the BAND-laboratory group at Maastricht University, composed of Prof. Sonja Kotz, Dr. Michael Schwartze, Xanthate Duggirala, Alexandra Emmendorfer, Pia Brinkman, Joseph Johnson, Lisa Goller, Hanna Honcamp, and the students who participated in the meeting.
Correspondence should be addressed to Katerina D. Kandylaki at katerina.kandylaki@maastrichtuniversity.nl.
https://doi.org/10.1523/JNEUROSCI.0756-21.2021

delta/theta frequency bands, and top-down modulatory influences (from frontal and motor regions) mediated via beta oscillations (13–30 Hz). Hence, according to this hypothesis, slower components (delta/theta) are at the bottom of the hierarchy, while beta is at the top. Park et al. (2015) further showed that synchronization between speech components and auditory cortex activity was associated with speech intelligibility and was significantly modulated by top-down signals. Similarly, a separate series of studies demonstrated that speech intelligibility was modulated by speech–brain entrainment in the delta/theta frequency band (Riecke et al., 2015a,b, 2018).

How can we reconcile these contrasting views and interpret the results from Puschmann et al. (2021)? The PLV metric used by the authors may reflect the neural tracking of speech, yet the results on the inter-regional communication are difficult to interpret in the absence of information about directionality. It is therefore still unclear whether the observed effects stem from a bottom-up process (i.e., signals traveling from auditory to frontal and motor regions) or whether they also involve top-down influences (e.g., high-order modulatory effects on auditory regions by frontal cortex)? Puschmann et al. (2021) interpret the positive association between musical training and inter-regional alpha band PLV in the study as an enhanced ability to process fine-grained auditory information. By doing so, however, one risks limiting the current understanding of speech processing to a bottom-up process while downplaying the complexity of the underlying cognitive processes. Instead, future investigations should also account for top-down modulations in speech processing.

In this regard, there is a paucity of research investigating the representational nature of top-down influences on speech comprehension. Some authors associate top-down information with predictions, but it still remains unclear how, when, and what they predict (Heilbron et al., 2020). If the brain is an ecological system that optimizes the dynamic allocation of resources while processing speech, it may exploit learned regularities to predict linguistic components, acoustic features, and the timing of sensory events, in parallel. These processes would engage at least three networks: the speech and language network (Hickok and Poeppel, 2007; Bornkessel-Schlesewsky et al., 2015), which processes acoustic and articulatory components; the sensorimotor network (Hickok et al., 2011), which features sensorimotor predictions from motor centers; and the rhythm and speech networks (Kotz and Schwartze, 2010; Kotz et al., 2018), which subserve temporal processing in speech and music. Finally, predictions may travel along parallel top-down streams in the beta/alpha frequency range (Park et al., 2015) and interact with bottom-up signals featured in slower rhythms (delta, theta). By doing so, these predictions modulate the neural tracking of speech (Park et al., 2015) and, in turn, speech intelligibility (Riecke et al., 2015a,b, 2018). Such a comprehensive view, however, requires future investigations.

Finally, how does musical training influence speech processing? Puschmann et al. (2021) interpreted their findings in relation to music-induced neuroanatomical and functional changes (Criscuolo et al., 2021) and similarities in speech and music networks (Peretz et al., 2015). Here, we would like to extend this argument, highlighting that specific brain regions may play a pivotal role for the observed transfer effects, and that these regions are not necessarily left lateralized.

In the article by Puschmann et al. (2021), the musicians' neural activity showed a significant degree of right lateralization originating from both the left and the right auditory cortices in the alpha band (8–12 Hz) compared with the neural activity of nonmusicians. The authors acknowledged that the effects of musical training extended beyond the predominantly left-lateralized sensorimotor network of speech processing (Hickok and Poeppel, 2007) and interpret the right lateralization as musician expertise in the processing of spectral cues (e.g., pitch, timbre). In our opinion, the influence of musical training extends beyond an enhanced ability to process acoustic information. Rather, we suggest that the bilateral involvement of the sensorimotor and rhythm networks in combination with the hierarchical processing architecture of the dorsal stream may play a pivotal role in the musician's speech processing.

The dorsal stream for language comprehension (Bornkessel-Schlesewsky et al., 2015), the sensorimotor network (Hickok et al., 2011), and the rhythm network (Kotz et al., 2018) overlap anatomically in the premotor cortex, the supplementary motor area, and the left inferior parietal cortex. These regions are implicated in the processing of music and speech rhythms and in predicting the beat in music, and they further allow dynamic action synchronization (Kotz et al., 2018). Additionally, they exhibit anatomic and functional differences between expert musicians and nonmusicians (Criscuolo et al., 2021), pointing toward mechanisms of functional specialization. In this view, rhythm processing and sensorimotor predictions (Kotz and Schwartze, 2010), which are strongly involved during musical training, may be exploited to support speech processing in the musician brain. This hypothesis, however, requires further investigation.

In conclusion, the results from the study by Puschmann et al. (2021) reveal that the neural tracking of speech is instantiated in a bilateral network and can be modulated by musical training. This work stimulates new and vibrant research questions regarding the potential hierarchical functional organization of speech processing, the role of bottom-up and top-down dynamic interactions, and the role of specific neural oscillations during speech comprehension. Last, it remains largely unclear how speech processing may benefit from music training. We suggest that a potential answer may lie in internetwork interactions among bilateral speech, sensorimotor, and rhythm networks. To answer these questions, multimodal investigations addressing the causal role of a hierarchical functional organization of speech processing may prove critical.

References

Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP (2015) Neurobiological roots of language in primate audition: common computational properties. Trends Cogn Sci 19:142–150.

Criscuolo A, Bonetti L, Särkämö T, Kliuchko M, Brattico E (2019) On the association between musical training, intelligence and executive functions in adulthood. Front Psychol 10:1704.

Criscuolo A, Pando-Naude V, Bonetti L, Vuust P, Brattico E (2021) Rediscovering the musician' s brain: a systematic review and meta-analysis. BioRxiv. doi: 2021.03.12.434473.

Heilbron M, Armeni K, Schoffelen JM, Hagoort P, de Lange FP (2020) A hierarchy of linguistic predictions during natural language comprehension. BioRxiv. doi: 2020.12.03.410399.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69:407–422.

Kandylaki KD (2015) Put it in context. the neurobiology of language explored with controlled stimuli in naturalistic auditory stories. PhD thesis, Philipps-Universität Marburg.

Kotz SA, Schwartze M (2010) Cortical speech processing unplugged: timely subcortico-cortical framework. Trends Cogn Sci 14: 392–399.

Kotz SA, Ravignani A, Fitch WT (2018) The evolution of rhythm processing. Trends Cogn Sci 22:896–910.

Park H, Ince RAA, Schyns PG, Thut G, Gross J (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr Biol 25:1649–1653.

Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front Psychol 2:142.

Peretz I, Vuvan D, Lagrois M-É, Armony JL (2015) Neural overlap in processing music and speech. Philos Trans R Soc Lond B Biol Sci 370:20140090.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time". Speech Commun 41:245–255.

Puschmann S, Regev M, Baillet S, Zatorre RJ (2021) MEG inter-subject phase-locking of stimulus-driven activity during naturalistic speech listening correlates with musical training. J Neurosci 41:2713–2722.

Riecke L, Formisano E, Herrmann CS, Sack AT (2015a) 4-Hz transcranial alternating current stimulation phase modulates hearing. Brain Stimul 8:777–783.

Riecke L, Sack AT, Schroeder CE (2015b) Endogenous delta/theta sound-brain phase entrainment accelerates the buildup of auditory streaming. Curr Biol 25:3196–3201.

Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E (2018) Neural entrainment to speech modulates speech intelligibility. Curr Biol 28:161–169.e5.