NEUROSCIENCE

# What auditory cortex is waiting for

To make sense of what we hear, the brain must integrate information over time. How is this temporal integration orchestrated in human auditory cortex? A new study by Norman-Haignere and colleagues[1] introduces a promising method to estimate neural integration windows and demonstrates that category-specific versus more generic computations operate on distinct timescales.

Lea-Maria Schmitt and Jonas Obleser

When we turn on the radio, we can almost immediately tell from the last chord that we have just missed out on our favourite music programme — and are now in for the weather forecast. Yet it is only after hearing a few sentences that we can also comprehend the weatherman's utterances ultimately promising a sunny weekend. Why is the time that goes by before a sound feature is recognized so closely related to its complexity?

A first clue to answering this question is that sounds are processed hierarchically in human auditory cortex. In particular, generic filters in primary auditory cortex have been shown to extract relatively low-level acoustic features such as pitch, while filters selective for specific sound categories extract more complex features, such as phonemes, in non-primary auditory cortex[2,3]. However, this leaves open how these neural filters operate to derive such diverse sound features. One longstanding hypothesis asserts that these filters integrate across increasingly long timescales. In a study in *Nature Human Behaviour*, Norman-Haignere and colleagues[1] now bridge the gap between the computations that underlie feature abstraction and temporal integration along the auditory pathway.

Auditory cortex is sensitive to sound features that evolve over the range of milliseconds. How much time does a group of neurons in auditory cortex take into account when computing the identity of the object these sound features carry — a piano tone, a baby's cry or a shutting door? Put more technically, how long are the time windows that neural filters integrate across to extract these features? This has proven notoriously difficult to tackle experimentally.

To estimate such integration windows, Norman-Haignere and colleagues[1] built on a recently developed paradigm of temporal context invariance[4]: imagine that we listen to the very same weather forecast once after a pop song and once after a classical piano piece. As the integration window of neural filters overlaps more and more with the shared weather forecast, neural responses to the two radio excerpts will gradually fall into alignment (and off, again, when following context diverges).

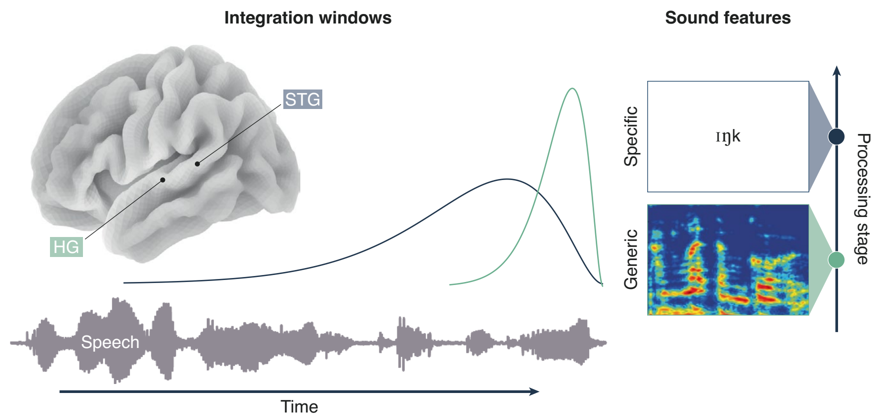


**Fig. 1 | Hierarchical organization of temporal integration windows in auditory cortex.** Norman-Hagniere et al.[1] identified temporal integration windows with the method of temporal context invariance. When listening to speech (grey waveform), primary regions in auditory cortex such as Heschl's gyrus (HG) integrate across brief periods of time (green function) to extract generic and low-level acoustic features (spectrogram), while non-primary regions at higher processing stages, such as superior temporal gyrus (STG), integrate across longer periods of time (blue function) to extract more complex and speech-specific feature categories such as phonemes.

The key trick of the present study[1] was, then, to predict the moment-by-moment alignment of neural responses to shared stimuli from models of integration windows with varying parameter values. This approach allowed the authors to infer the length and delay of integration windows in specific brain regions from the estimated parameter values of those models that fit best to the observed neural alignment patterns. A strength of this approach is that it can be used, in principle, with any stimulus and recording modality across the whole brain.

Norman-Haignere et al.[1] used intracranial recordings of high-frequency activity ('gamma'; 70–140 Hz) to estimate the integration windows used in auditory cortex while patients undergoing neurosurgery listened to sequences of natural sounds. They found a hierarchical organization of integration windows from primary to non-primary auditory regions, with integration windows being early (about 70 ms after sound onset) and narrow (about 70 ms, or less than a tenth of a second) in Heschl's gyrus but delayed (around 200 ms) and wide (about 270 ms, or about a quarter of a second) in superior temporal gyrus.

Next, the authors made a theoretically interesting connection between these integration windows and the representation of certain sound features. In line with hypotheses on the generic nature of computations in early cortical regions, electrodes with short integration windows were predominantly sensitive to the frequency-decomposed output (as provided by the ascending auditory pathway after the inner ear). By contrast, electrodes with long integration windows were more sensitive to the category labels of sounds and showed spatially more-separable responses to different categories such as speech, music and nature sounds.

In sum, this study makes the case that primary auditory cortex responds quickly to acoustic features extracted from a brief time interval of any incoming sound, whereas non-primary regions respond to more complex features of specific sound categories extracted over longer time intervals and with a corresponding delay (Fig. 1).

How can the neural representation of sound at multiple scales be explained computationally? These results integrate well with general accounts of cortical processing, which state that the hierarchical organization of neural computations per se enables the extraction of complex features from sensory input[5]. In such a scenario, higher levels in the hierarchy would directly operate on the output from lower levels to combine, for example, specific sequences of temporal and spectral (that is, frequency) modulations into a phoneme representation (but see Hamilton and colleagues[6] for a more nuanced perspective on the neuroanatomical hierarchies involved). In this sense, the length and delay of integration windows might be inherently tied to such hierarchical organization principles.

An open question is how such a ‘temporal-windowing’ regime, as a parsimonious computational principle in audition, would scale up to more complex levels of representation in speech and language. For instance, in natural soundscapes, listeners are confronted with temporal dependencies extending well beyond a few hundred milliseconds. Even though higher-order brain regions such as parietal cortex have been shown to code for such long dependencies in a hierarchical fashion[7], transferring the mechanistic principles of rigid time windows (as put forward in the present study[1]) to more realistic scenarios poses a challenge. There are studies that suggest that the semantic context of continuous speech is segmented into increasingly coarse representations according to its underlying event structure[8], but the boundaries of these events — as with the end of the weather forecast — are set independent from the length of any fixed integration window proposed in the present study.

An interesting avenue for future research will thus be to focus on the computations that equip the brain beyond auditory cortex with the flexibility to extract those features from natural sounds that are meaningful at multiple levels of abstraction. ❐

**Lea-Maria Schmitt**[1,2] **and Jonas Obleser**[1,2] ✉

*[1]Department of Psychology, University of Lübeck, Lübeck, Germany. [2]Center of Brain, Behavior and Metabolism, University of Lübeck, Lübeck, Germany.*

✉*e-mail: jonas.obleser@uni-luebeck.de*

Published online: 10 February 2022
https://doi.org/10.1038/s41562-021-01262-x

## References

1. Norman-Haignere, S. V. et al. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-021-01261-y (2022).
2. Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V. & McDermott, J. H. *Neuron* **98**, 630–644 (2018).
3. Leaver, A. M. & Rauschecker, J. P. *J. Neurosci.* **30**, 7604–7612 (2010).
4. Chien, H.-Y. S. & Honey, C. *J. Neuron* **106**, 675–686 (2020).
5. Himberger, K. D., Chien, H.-Y. & Honey, C. J. *Neuroscience* **389**, 161–174 (2018).
6. Hamilton, L. S., Oganian, Y., Hall, J. & Chang, E. F. *Cell* **184**, 4626–4639 (2021).
7. Lerner, Y., Honey, C. J., Silbert, L. J. & Hasson, U. *J. Neurosci.* **31**, 2906–2915 (2011).
8. Schmitt, L.-M. et al. Predicting speech from a cortical hierarchy of event-based timescales. *Sci. Adv.* (in the press).

## Competing interests

The authors declare no competing interests.