

INVITED TUTORIAL

Does auditory processing rely on encapsulated, or domain-general computational resources?

Katharine Molloy[1,2], Nilli Lavie[2] and Maria Chait[1,*]

[1]*Ear Institute, University College London, London WC1X 8EE, United Kingdom*
[2]*Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom*

Abstract: The extent to which auditory processing depends on attention has been a key question in auditory cognitive neuroscience, crucial for establishing how the acoustic environment is represented in the brain when attention is directed away from sound. Here I review emerging behavioural and brain imaging results which demonstrate that, contrary to the traditional view of a computationally encapsulated system, the auditory system shares computational resources with the visual system: high demand on visual processing (e.g. as a consequence of a task with high perceptual load) can undercut auditory processing such that both the neural response to, and perceptual awareness of, non-attended sounds are impaired. These results are discussed in terms of our understanding of the architecture of the auditory modality and its role as the brain's early warning system.

Keywords: Audio-visual, Auditory scene analysis, Load, Attention, Brain imaging

PACS number: 43.66.Ba [doi:10.1250/ast.41.13]

Beyond its key role in supporting communication, the auditory system is hypothesized to serve as the brain's 'early warning system.' Hearing is sensitive to a wider space than the other senses (above, below, behind, in the dark...) and is therefore often theorized to function as a monitor — automatically scanning the environment for relevant events [1]. The 'early warning' view of the auditory system suggests that it should operate automatically, independently of the focus of attention or perceptual goals so that potentially critical events in one's surroundings can be rapidly discovered and brought to awareness.

This hypothesis is indeed supported by multiple demonstrations that complex auditory processing, including figure-ground segregation [2,3] and pattern learning [4,5], take place automatically, even when outside of the focus of attention. For example, Masutomi *et al.* [6] used a complex 'repetition based segregation' stimulus [7] and a dual task design to show that listeners can extract a sound source from a dynamically varying background even when their processing resources are engaged by a competing high load task. Listeners were exposed to sound-mixture sequences while performing a "decoy" task that required continuous monitoring of a separate, concurrent stimulus stream. Attentional load was manipulated by using both a high demand condition and a control condition (Low demand) of each decoy task. When the decoy task was a visual multiple object tracking (MOT) task — participants were required to simultaneously track several rapidly moving 'target' dots among many more identical moving dots — performance on the auditory task revealed no effect of load: listeners performed well, independently of whether the MOT task was minimally demanding or when it was specifically adjusted to capacity.

For two other tasks: an RSVP digit encoding task (that required visual monitoring but also involved auditory working memory), and a challenging auditory monitoring task, there *was* a significant, but very small, effect of attention on segregation [6]. In all cases segregation performance remained high irrespective of the perceptual demand of the competing task. The authors interpreted the results as indicating that auditory scene analysis is principally an automatic process. However, that a significant (albeit small) effect of load was observed, can be taken as evidence that 'hearing' *does* draw on resources associated with visual processing. Namely, auditory processing is not encapsulated and independent but rather depends on shared computational capacity. Understanding which resources are shared, and under what conditions, is important both for basic research into perception and for revealing the implications of the depletion of computational capacity (e.g. during aging) on sensory processing. For instance, a key issue relates to whether competition for resources emerges early in the processing hierarchy or is

*e-mail: m.chait@ucl.ac.uk

introduced at later stages such as those linked to conscious awareness.

Dual task paradigms like the one used in Masutomi *et al.* [6] suffer from the inherent drawback that the task instructions involve the requirement to monitor both the 'primary' and 'secondary' stimulus streams. Attention is therefore never fully withdrawn from the investigated stimulus. In Masutomi *et al.*, though participants were encouraged to attend to the 'decoy' tasks they were also tested (and therefore incentivized) to attend to the auditory stream. Brain imaging can overcome this challenge by recording brain responses to ignored and entirely task irrelevant stimuli while participants are fully engaged in a primary task. Emerging evidence indeed demonstrates that visual load can affect auditory processing from relatively early cortical (but likely not subcortical; see [8]) processing stages [9–11].

Molloy *et al.* [10] instructed participants to perform a low- or high-load visual search task. Occasionally, sounds (simple pure tones) were presented together with the search arrays. Conditions of high visual load resulted in substantially decreased onset responses to the tones from about 120 ms after sound onset. Later responses, at about 300 ms post onset, hypothesized to reflect conscious awareness of a stimulus, were completely abolished under high load. This was interpreted as indicating that high visual load reduced the gain on early auditory responses by depleting resources critical for auditory sensory processing. Molloy *et al.* [11] (see also [12]) further showed that high load can severely impair computations associated with auditory figure-ground segregation, with effects observed from the earliest stages of processing in auditory cortex (see also [13]). Overall, these results point to the fact that what we may commonly consider as inherently auditory processes [14,15] actually draw on general computational capacity such that they can be affected by visual perceptual load. If we dare extrapolate to real world listening this might mean that when we focus hard on a visual task (e.g. reading) auditory processing may be affected. Consequences may be especially severe in life- or mission-critical situations that depend on visual processing such as driving in severe weather or for surgeons performing complex medical procedures.

Several other studies have failed to reveal effects of load [9,16]. There could be many reasons for this including the specific paradigms used. Indeed, Molloy *et al.* [10,11] used paradigms in which the presentation of the visual and auditory stimuli were precisely aligned such that sounds were presented at the point within the visual trial where the demands on computational resources were highest. One might argue that most natural tasks do not involve such precisely timed auditory and visual stimulation and hence the actual effect of load on hearing in realistic situations may be modest (e.g. such as the one reported behaviorally in Masutomi *et al.* [6]).

Importantly, the consistent presence of load effects across diverse paradigms, points to an inter-dependence between modalities that affects auditory processing from early stages within the cortical hierarchy. This is also in line with other work e.g. [17,18] which is beginning to reveal shared computational resources across hearing and vision. The emerging understanding is that the parcellation of sensory perception into separate systems of 'hearing' and 'vision' which has characterized much of the research into systems and cognitive neuroscience may in fact be limiting our understanding of the underlying computational organization. Instead, it is becoming increasingly clear that hearing and sight share computational machinery with profound consequences for perception.

REFERENCES

[1] S. Murphy, N. Fraenkel and P. Dalton, "Perceptual load does not modulate auditory distractor processing," *Cognition*, **129**, 345–355 (2013).

[2] J. A. O'Sullivan, S. A. Shamma and E. C. Lalor, "Evidence for neural computations of temporal coherence in an auditory scene and their enhancement during active listening," *J. Neurosci.*, **35**, 7256–7263 (2015).

[3] S. Teki, N. Barascud, S. Picard, C. Payne, T. D. Griffiths and M. Chait, "Neural correlates of auditory figure-ground segregation based on temporal coherence," *Cereb. Cortex*, **26**, 3669–3680 (2016).

[4] E. Sohoglu and M. Chait, "Detecting and representing predictable structure during auditory scene analysis," *eLife*, doi:10.7554/eLife.19113 (2016).

[5] N. Barascud, M. T. Pearce, T. D. Griffiths, K. J. Friston and M. Chait, "Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns," *Proc. Natl. Acad. Sci. USA*, **113**, E616–E625 (2016).

[6] K. Masutomi, N. Barascud, M. Kashino, J. H. McDermott and M. Chait, "Sound segregation via embedded repetition is robust to inattention," *J. Exp. Psychol. Hum. Percept. Perform.*, **42**, 386–400 (2016).

[7] J. H. McDermott, D. Wrobleski and A. J. Oxenham, "Recovering sound sources from embedded repetition," *Proc. Natl. Acad. Sci. USA*, **108**, 1188–1193 (2011).

[8] E. Holmes, D. W. Purcell, R. P. Carlyon, H. E. Gockel and I. S. Johnsrude, "Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates," *J. Assoc. Res. Otolaryngol.*, **19**, 83–97 (2018).

[9] B. J. Dyson, C. Alain and Y. He, "Effects of visual attentional load on low-level auditory scene analysis," *Cogn. Affect. Behav. Neurosci.*, **5**, 319–338 (2005).

[10] K. Molloy, T. D. Griffiths, M. Chait and N. Lavie, "Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses," *J. Neurosci.*, **35**, 16046–16054 (2015).

[11] K. Molloy, N. Lavie and M. Chait, "Auditory figure-ground segregation is impaired by high visual load," *J. Neurosci.*, **39**, 1699–1708 (2019).

[12] D. Raveh and N. Lavie, "Load-induced inattentional deafness," *Atten. Percept. Psychophys.*, **77**, 483–492 (2015).

[13] Z. Xie, R. Reetzke and B. Chandrasekaran, "Taking attention

away from the auditory modality: Context-dependent effects on early sensory encoding of speech," *Neuroscience*, **384**, 64–75 (2018).

[14] M. Elhilali, L. Ma, C. Micheyl, A. J. Oxenham and S. A. Shamma, "Temporal coherence in the perceptual organization and cortical representation of auditory scenes," *Neuron*, **61**, 317–329 (2009).

[15] S. A. Shamma and C. Micheyl, "Behind the scenes of auditory perception," *Current Opin. Neurobiol.*, **20**, 361–366 (2010).

[16] C. Alain and A. Izenberg, "Effects of attentional load on auditory scene analysis," *J. Cogn. Neurosci.*, **15**, 1063–1073 (2003).

[17] S. W. Michalka, L. Kong, M. L. Rosen, B. G. Shinn-Cunningham and D. C. Somers, "Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks," *Neuron*, **87**, 882–892 (2015).

[18] D. Fougnie, J. Cockhren and R. Marois, "A common source of attention for auditory and visual tracking," *Atten. Percept. Psychophys.*, **80**, 1571–1583 (2018).