

Auditory Neuroscience: Sound Segregation in the Brainstem?

Separating a mixture of sounds into its constituent parts is a complex process likely to involve many processing stages. A new study suggests that the first steps in that process may occur already at the level of the first auditory processing centre in the brainstem.

Jan W.H. Schnupp

In our modern world, full of bustling activity and noisy contraptions, silence is becoming an increasingly rare commodity. But even in a natural environment, sources of either continuous or intermittent ‘background noise’ are quite common. Sound, as we all know, is a vibration of the air that surrounds us, but our auditory system is not really interested in these vibrations *per se*. Rather, it tries to reconstruct objects and events in our environment from the received vibration patterns. Would that particular high-pitched squeak be the sound of a rusty bicycle, or the mating call of a love-struck starling? If you happened to be a receptive female starling, getting that distinction wrong would be rather embarrassing. But even small bird brains rarely make such mistakes, even though, computationally, recognizing sounds is hard, particularly if we do not know in advance whether the sounds arriving at the ears have come from one source or many. Starlings have to be able to identify the songs of other starlings even if these occur against the background noise of buzzing insects, barking dogs or human traffic. These various sound waves are mixed together before they even arrive at the ears. How can our brains ever ‘unmix’ them? This unmixing is a poorly understood process which is likely to require processing at many levels, but new research by Pressnitzer *et al.* [1], published recently in *Current Biology*, suggests that the first steps in that process may already occur at the very first auditory processing stage, the ventral cochlear nucleus.

The science of decomposing sounds is known as auditory scene analysis, and one of the pioneers of that discipline, AI Bregman [2], popularized the idea that sounds might be separated into auditory ‘streams’ even before our brains attempt to identify the sound sources responsible for each

stream. Bregman devised numerous simple artificial sound sequences, and asked listeners whether they perceived them as one auditory ‘stream’ or several.

One of the simplest examples is the so-called ABA sequence: it consists of tones of two different frequencies, A and B, which follow each other rapidly, as illustrated schematically in Figure 1. This is normally heard as a single, simple melody with a galloping rhythm: “da di da - da di da - da di da -...”, and if the frequencies A and B are chosen close together, then this percept of a single melody tends to be quite stable. If A and B are further apart, however, then the percept may suddenly flip from that of a single melody with a galloping rhythm to that of two parallel, independent regular beats, one going “da - da - da - da...” the other “- di - - - di - - - di ...”. The larger the frequency separation, the sooner the percept is likely to break from one stream into two.

However, this break-up into two auditory streams need not be permanent: if the ABA sequence continues on for long enough, listeners often hear it flip back and forth between a single galloping rhythm or two independent beats. The ABA sequence

is therefore a sort of ‘Necker cube of auditory scene analysis’. It is perceptually ambiguous, as it has two interpretations which are equally valid but mutually exclusive. (The Necker cube, shown in Figure 2, is a well-known ambiguous perspective line drawing. Most observers will interpret this figure as that of a three-dimensional cube, but the figure gives no cue as to whether the corner on the top right of the drawing should be seen as part of the front or the back of the cube, and the figure occasionally seems to flip from one configuration to the other.)

The way we perceive ambiguous stimuli, like the ABA sequence or the Necker cube, suggests that there must be several different ‘brain states’, each responsible for one of the alternative perceptions. Presumably, competitive interactions between these states ensure that only one of the alternatives dominates perception at any one time, but as the ensemble of neurons responsible for the dominant percept fatigues, alternative interpretations of the stimulus can gain the upper hand, and the percept flips. While studying the perception of ABA sequences and a different class of ambiguous visual percept, known as ‘binocular rivalry’, Pressnitzer and Hupe [3] observed that this flipping between alternative interpretations occurs in very similar ways in the visual and the auditory domain. In binocular rivalry, incompatible images are shown to each eye, and as they cannot be merged into a single coherent picture, the perception flips between that of the left eye image and that of the right. Other research groups, including those

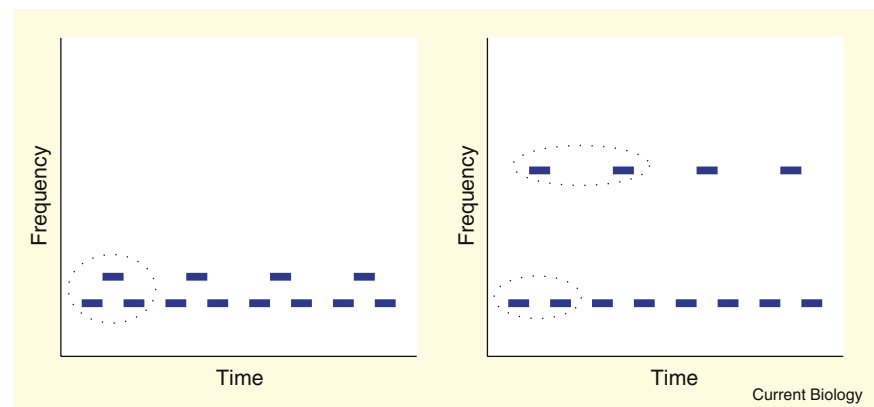


Figure 1. Bistable sound perception.

‘ABA’ sound sequences consist of tones of two different frequencies, A and B, and can be heard as a galloping rhythm (left) or as two parallel, independent regular beats (right).

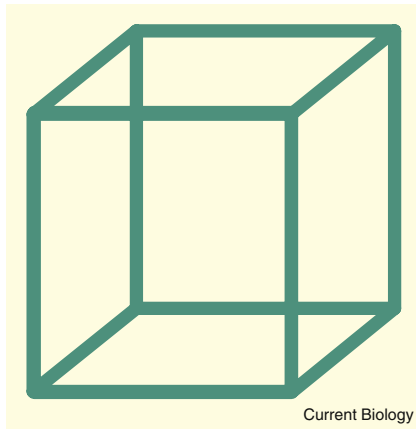


Figure 2. The Necker cube.

of Logothetis [4] and Rees [5], have started to shed some light on the physiology of the ‘competing brain states’ that are associated with alternative percepts in binocular rivalry, but particularly in the auditory domain we still know very little about the networks and mechanisms responsible for the perception of bistable stimuli, such as the ABA sequence.

The little we do know comes largely from the work of Micheyl *et al.* [6], and Fishman *et al.* [7], who recorded responses to ABA sequences in the auditory cortex of rhesus monkeys. They argued, quite plausibly, that if neurons responded both to the A-tone and the B-tone more or less equally, then their activity might support a percept which groups A and B together — the ABA galloping rhythm — but if neurons responded strongly to only one of the tones and only weakly or not at all to the other, then that activity would support the perception of independent streams (separate A-beats and B-beats). But, as mentioned earlier, listeners will normally hear ABA sequences at first as a single stream, and the percept then breaks up into that of two streams, where the frequency difference between A and B determines how long it typically takes for the galloping rhythm to turn into two separate beats.

How do we account for the perceptual flip in terms of neural responses? Neurons in auditory cortex tend to be broadly frequency tuned, so a neuron which responds strongly to tone A normally also responds to tone B, but the response to B is weaker the further the frequency of B is from A. When presented with ABA rhythms,

these neurons will initially respond to all of the stimuli, but under continuous, repetitive stimulation these responses ‘adapt’ and grow progressively weaker. In neurons tuned to frequency A, however, the responses to B already start off weaker, and Micheyl *et al.* [6] argued that adaptation might soon drive them below the ‘threshold’ required for perception. Using a very simple threshold model, they were able to show that the time course of adaptation nicely predicts the rate at which “two stream” perceptions of ABA sequences become more likely.

Auditory scene analysis therefore depends, in perhaps unexpected ways, on good-old-fashioned neurophysiological mechanisms like adaptation. And it may contribute to auditory scene analysis in more ways than one. ABA sequences are unusual, in that the two possible streams (the A and the B rhythms) both start at about the same time. Let us step out of the laboratory and back into the ‘real world’, where the little starling we met in the introduction still needs to separate out the song of another starling (an A stream) from superimposed dog barks or traffic noises (B streams). If the traffic noise and the bird song happened to start simultaneously, then it might conceivably take the starling a few moments to work out that the sudden cacophony of sound contains several separate streams. But it is much more likely that one of the two sounds will have been going for a while by the time the second sound starts. In those cases, another phenomenon known as ‘stimulus specific adaptation’ [8] may already have caused the sensation of the first stream to fade into the background by the time the second stream starts, leaving the brain ready to process the novel input undistracted.

But stimulus specific adaptation is thought to be a cortical phenomenon, as it does not happen at ‘lower’ processing stages, like the midbrain [8], and studies of adaptation in ABA sequences too have so far described this only in cortex [7,9]. For our little starling that is a problem, because birds, having diverged from mammals around 200 million years ago, strictly speaking do not have a cortex. But that anatomical shortcoming appears not to prevent starlings from being proficient auditory scene analysers [10,11]. Perhaps cortex is not required after all,

a possibility that seems considerably more likely in the light of new data from Pressnitzer *et al.* [1], which show that, at least in guinea pigs, adaptation to ABA sequences which is essentially identical to that seen in auditory cortex, and which is predictive of streaming, can already be observed at the level of the ventral cochlear nucleus, the brain’s very first auditory processing station. In fact, even ‘primary like’ bushy cells of the cochlear nucleus behave in this manner, even though these cells receive only very few but strong synaptic inputs, mostly through unusually large ‘calyceal’ synapses, directly from auditory nerve fibre afferents. The first steps toward auditory scene analysis may thus be taken already before there is much opportunity for synaptic processing of the incoming sounds.

References

1. Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I.M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128.
2. Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA; London: MIT Press).
3. Pressnitzer, D., and Hupe, J.M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* 16, 1351–1357.
4. Leopold, D.A., and Logothetis, N.K. (1996). Activity changes in early visual cortex reflect monkeys’ percepts during binocular rivalry. *Nature* 379, 549–553.
5. Haynes, J.D., and Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* 15, 1301–1307.
6. Micheyl, C., Tian, B., Carlyon, R.P., and Rauschecker, J.P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139–148.
7. Fishman, Y.I., Arezzo, J.C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am.* 116, 1656–1670.
8. Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398.
9. Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., and Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. *Hearing Res.* 229, 116–131.
10. Bee, M.A., and Klump, G.M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol.* 92, 1088–1104.
11. Klump, G. (2008). The European starling as a model for understanding mechanisms underlying auditory scene analysis. *J. Acoust. Soc. Am.* 123, 3052.

Department of Physiology, Anatomy and Genetics, Sherrington Building, University of Oxford, Parks Road, Oxford OX1 3PT, UK.
E-mail: jan@physiol.ox.ac.uk