and some other crustaceans, visual feedback somehow influences hormone systems based in either the eyestalks or main body, affecting changes in chromatophore cells [13–15]. However, how exactly this works remains unclear. Finally, Abram *et al.*'s [5] biochemical analyses raise questions about what the pigments are that cause changes in egg brightness. Clearly, we have much left to discover regarding both the functions and mechanisms of colour change and egg coloration in nature.

REFERENCES

1. Wallace, A.R. (1867). Mimicry and other protective resemblances among animals. Westminster Rev. (London ed.) *1 July*, 1–43.
2. Cott, H.B. (1940). Adaptive Coloration in Animals (London: Methuen & Co. Ltd.).
3. Kilner, R.M. (2006). The evolution of egg colour and patterning in birds. Biol. Rev. *81*, 383–406.
4. Cherry, M.I., and Gosler, A.G. (2010). Avian eggshell coloration: new perspectives on adaptive explanations. Biol. J. Linn. Soc. *100*, 753–762.
5. Abram, P.K., Guerra-Grenier, E., Després-Einspenner, M.-L., Ito, S., Wakamatsu, K., Boivin, G., and Brodeur, J. (2015). An insect with selective control of egg coloration. Curr. Biol. *25*, 2007–2011.
6. Moreno, J., and Osorno, J.L. (2003). Avian egg colour and sexual selection: does eggshell pigmentation reflect female condition and genetic quality? Ecol. Lett. *6*, 803–806.
7. Soler, J.J., Navarro, C., Pérez-Contreras, T., Avilés, J.M., and Cuervo, J.J. (2008). Sexually selected egg coloration in spotless starlings. Am. Nat. *171*, 183–194.
8. Reynolds, S.J., Martin, G.R., and Cassey, P. (2009). Is sexual selection blurring the significance of eggshell coloration hypotheses? Anim. Behav. *78*, 209–215.
9. Winters, A.E., Stevens, M., Mitchell, C., Blomberg, S.P., and Blount, J.D. (2014). Maternal effects and warning signal honesty in eggs and offspring of an aposematic ladybird beetle. Funct. Ecol. *28*, 1187–1196.
10. Dreon, M.S., Ituarte, S., and Heras, H. (2010). The role of the proteinase inhibitor ovorubin in apple snail eggs resembles plant embryo defense against predation. PLoS One *5*, e15059.
11. Keeble, F.W., and Gamble, F.W. (1900). The colour-physiology of *Hippolyte varians*. Proc. R. Soc. B *65*, 461–468.
12. Sumner, F.B., and Keys, A.B. (1929). The effects of differences in the apparent source of illumination upon the shade assumed by a flatfish on a given background. Physiol. Zoo. *2*, 495–504.
13. Fingerman, M. (1973). Behavior of chromatophores of the fiddler crab *Uca pugilator* and the dwarf crayfish *Cambarellus shufeldti* in response to synthetic *Pandalus* red pigment-concentrating hormone. Gen. Comp. Endocrin. *20*, 589–592.
14. Fingerman, M., and Yamamoto, Y. (1967). Daily rhythm of melanophoric pigment migration in eyestalkless fiddler crabs, *Uca pugilator* (Bosc). Crustaceana *12*, 303–319.
15. Shibley, G.A. (1968). Eyestalk function in chromatophore control in a crab, *Cancer magister*. Physiol. Zool. *41*, 268–279.

Auditory Perception: Attentive Solution to the Cocktail Party Problem

Simon Carlile
Starkey Hearing Research Centre, 2110 Shattuck St #408, Berkeley, CA 94704 USA and School of Medical Sciences, University of Sydney, NSW 2006, Australia
Correspondence: simon_carlile@starkey.com
http://dx.doi.org/10.1016/j.cub.2015.07.064

A recent study has demonstrated how the focus of auditory attention can rapidly shift to follow spectrally dynamic speech-like sounds in the presence of a similar interferer. This requires multidimensional variation in sound features and a minimum spacing in spectral feature space.

Enquiries, directions, an invitation or warning, a plea, a command, a heated brainstorming or a convivial cocktail party: all important pieces in the way in which humans interact with each other. In fact, any animal that enjoys hearing shares some aspects of this communication banquet. Evolution has had plenty of time to fine-tune this interactive channel, which is not a bad thing as it presents the nervous system with, in computational terms, a very ill-formed problem. Essentially we have one receptor surface (the inner ear) that receives the sounds from many concurrent sources, such as the chorus around the pond at night, and 'multiplexes' all this information into a single channel (the auditory nerve). The computational challenge then is to sort out which parts of the encoded sound belong to which source and then group them together in a way that allows the nervous system to extract the information of interest against the background of other sounds [1]. The most interesting sounds, especially speech, vary rapidly over time so that this problem begins to look like a Rasta dreadlock! How does the system track the rapid dynamic variations in the distinguishing features? What are the critical acoustic features that enable this process? What is the frequency-temporal resolution of such a system? These are the questions that Woods and McDermott [2] have addressed in their study published in this issue of *Current Biology*, using a simple but highly innovative perceptual experiment with human listeners.

In solving this problem, one advantage for the auditory system is that it has evolved in a world of physically sounding objects, and the patterns of sound energy from individual sources conform to simple

acoustic rules. The physical structure of each sound source establishes clear statistical regularities in the sound waves it emits that are characteristic of that structure. For example: a resonating body produces frequencies that are harmonically related to its fundamental resonant frequency. Likewise, the onsets and offsets of the different frequency components from a single source will come on and go off at roughly the same time and their amplitude and frequencies will also vary coherently (for a recent review see [3]). On a short time scale (tens of milliseconds), the auditory system uses these acoustic rules to group the different components into separate 'chunks' and, on a longer time scale (seconds), uses similar rules of plausibility to stream these chunks over time to generate the auditory objects of our perception. Auditory research to-date has demonstrated that distinguishing features of each 'chunk' play an important role in establishing and maintaining the stream — such differences as pitch, timbre or spatial location. One significant analytical problem is that most communication sounds, and many other sounds of biological significance, vary dramatically over time so that, in the presence of similar concurrent sounds, the distinguishing features can intertwine and intersect (a Gordian knot indeed!).

Over the last decade or so the important role of attention in the formation of auditory objects has become better understood — notwithstanding the fact that Colin Cherry [4] pointed out that this was a critical piece more than half a century earlier! Like vision, auditory attention works on perceptual objects that are represented in working memory [5,6]. The focus of attention likely increases the neural representation of the attended-to object, possibly by enhancing the preconscious processing at the cortical (or lower) levels [7,8].

An acoustic signal, speech can be characterised as a combination of time-varying, harmonically related and broadband sounds (the source) that are shaped by the physical dimensions of the vocal apparatus (the filter). Much of the information in speech is contained in variations in the fundamental frequency (F0) and the first (F1) and second (F2) formants produced by the prominent resonances of the vocal tract. In a lovely illustration of these dynamic changes, Woods and McDermott [2] plot this information for two concurrently spoken sentences demonstrating how these two streams of information intertwine in the three-dimensional feature space of F0, F1 and F2 (see Figure 1 in [2]).

We know from personal experience that it is relatively straightforward to listen to one talker in the presence of another concurrent talker. There are a range of different cues we can use including difference in the location of the talkers, difference in voice quality and the semantic content of the speech [1,3]. To eliminate many of these cues and to focus on the frequency variations, Woods and McDermott [1] synthesized artificial 'voices' from a smoothly time varying harmonic series (like the complex sound from a trombone played glissando) which were then filtered in a manner that resembles the formant filtering by the vocal apparatus. They first presented a short sample (500 ms) of the onset of a target sound (the cue) and then played the whole sound in the presence of another different 'voice'. The subject's task was to follow the cued sound and then to say if a subsequent short probe sound came from the end of the target sound or not. Although effortful, most subjects did quite well on this streaming task, suggesting that the focus of attention could be rapidly and dynamically varied to follow the trajectory of the target sound in the frequency feature space.

To demonstrate that this was actually due to a focus of attention, a second experiment required listeners to also detect if one of the voices contained a brief (200 ms) period of vibrato. For those subjects who performed well on the streaming task, when the vibrato occurred in the cued voice, detection was significantly higher than when in the uncued voice. In two other experiments, the authors also demonstrated that that vibrato detection performance did not vary significantly over the length of the stimulus and that temporal discontinuities, similar to those found in natural speech, did not degrade performance. Both findings have significant implication for the understanding of natural speech with competing talkers. To probe the underlying mechanisms, they also examined what happens when the 'voices' cross in feature space, or at least become quite close or when only one frequency feature in each voice varies. The former caused a graceful degradation in performance as frequency spacing decreased from around 4–5 semitones and the latter basically eliminated the ability to do the streaming task. This hints at the resolution and the multidimensional nature of the inputs to the attentional tracking system.

This experiment [2] extends the growing body of evidence that attention plays a key role in the streaming of an auditory object by demonstrating how this occurs for stimuli with distinguishing features that are highly dynamic in the frequency feature space. Masking interactions between both speech and non-speech stimuli has been previously characterised as energetic or informational. Energetic masking representing a swamping of the target sound by the energy from the masker, whilst informational masking was initially (and rather unhelpfully!) characterised as everything else (review [3]). It is unlikely that energetic masking is playing a key part in the interactions between these stimuli, with the exception of when the stimulus feature trajectories were in close proximity. Informational masking has been attributed to a failure of attention in selecting or sustaining the focus on the correct target over time — a particularly top-down view of the processes that requires that the auditory object is in working memory and an object of perception [6].

The focus of attention has also been shown to modulate the grouping and streaming of information relating to the attended-to auditory object [9] (in this case the cued voice). Detection performance in the current experiment [2] could well be modulated by the frequency and temporal resolution of the system that steers non-spatial attention (for review see [10]). This is consistent with the streaming errors evident when the two voices become close in frequency feature space. Recent work indicates that there are also forms of bottom-up informational masking, not directly reflecting the top-down steering of attention. In particular, unintelligible, speech-like sounds with the same modulation characteristics of speech demonstrate high levels of masking over and above

their energetic masking components [11]. Modulation masking of speech has also been demonstrated and modelled using non-speech like stimuli (for example [12,13]). It will be an important question for future work to disentangle these different top-down/bottom-up effects.

One intriguing aspect of the data of Woods and McDermott [2] is that the temporal variation of the position of the vibrato signal did not vary detection performance — there appeared to be no 'build-up' of streaming over the course of the stimulus as has been reported in many streaming experiments using sequences of tones (for example [14]). This most likely results from the very different nature of the stimuli used here and may well have been exogenously driven, but it does suggest caution in the interpretation of previous results in the context of more ecological examples of auditory streaming, as tapped into by Woods and McDermott [2]. On the other hand, being able to rapidly form streams and focus attention would be critical for good performance in cocktail party listening where there is often also little to no gap in conversational turn-taking [15]. In that context it would be most interesting to explore the use of this most elegant and simple test as a diagnostic for various attentional disorders such as attentional deficit disorder and auditory processing disorder where speech understanding is also affected. Not only might it provide a very sensitive test of disability, it might reveal more of the underlying mechanism of dysfunction in these conditions.

REFERENCES

1. Darwin, C.J. (2008). Listening to speech in the presence of other sounds. Phil. Trans. R. Soc. Lond. B *363*, 1011–1021.
2. Woods, K.J.P., and McDermott, J.H. (2015). Attentive tracking of sound sources. Curr. Biol. *25*, 2238–2246.
3. Carlile, S. (2014). Active listening: Speech intelligibility in noisy environments. Acoust. Aust. *42*, 98–104.
4. Cherry, E.C. (1953). Some experiments on the recognition of speech with one and two ears. J. Acoust. Soc. Am. *25*, 975–979.
5. Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. Trends Cogn. Sci. *12*, 182–186.
6. Knudsen, E.I. (2007). Fundamental components of attention. Annu. Rev. Neurosci. *30*, 57–78.
7. Ding, N., and Simon, J.Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. USA *109*, 11854–11859.
8. Rimmele, J.M., Zion Golumbic, E., Schroger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex *68*, 144–154.
9. Shinn-Cunningham, B.G., Lee, A.K.C., and Oxenham, A.J. (2007). A sound element gets lost in perceptual competition. Proc. Natl. Acad. Sci. USA *104*, 12223–12227.
10. Hafter, E.R., Sarampalis, A., and Loui, P. (2008). Auditory attention and filters. In Auditory Perception of Sound Sources, W.A. Yost, A.N. Popper, and R.R. Fay, eds. (New York: Springer), pp. 115–143.
11. Carlile, S., and Corkhill, C. (2015). Selective spatial attention modulates bottom-up informational masking of speech. Sci. Rep. *5*, 8662.
12. Jørgensen, S., and Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. J. Acoust. Soc. Am. *130*, 1475–1487.
13. Stone, M.A., and Moore, B.C.J. (2014). On the near non-existence of "pure" energetic masking release for speech. J. Acoust. Soc. Am. *135*, 1967–1977.
14. Cusack, R., Deeks, J., Aikman, G., and Carlyon, R.P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. Exp. Psychol. Hum. Percept. Perform. *30*, 643–656.
15. Lin, G., and Carlile, S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking. Front. Neurosci. *19*, 125.

Mammalian Evolution: A Jurassic Spark

Michael S.Y. Lee[1,2,*] and Robin M.D. Beck[3]
[1]South Australian Museum, North Terrace, Adelaide SA 5000, Australia
[2]School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia
[3]School of Environment & Life Sciences, University of Salford, Salford M5 4WT, UK
*Correspondence: Mike.Lee@samuseum.sa.gov.au
http://dx.doi.org/10.1016/j.cub.2015.07.008

There is increasing evidence that early mammals evolved rapidly into a range of body forms and habitats, right under the noses of the dinosaurs.

Mammals first appear in the fossil record at about the same time as the earliest dinosaurs (~220 million years ago), and so the first two-thirds of mammalian evolutionary history thus occurred during the Mesozoic 'Age of Dinosaurs' [1,2]. Mesozoic mammals were long portrayed as tiny, shrew-like creatures, unable to diversify due to severe competition and predation from dinosaurs and other reptiles. However, discoveries in the past two decades have greatly expanded the known diversity of Mesozoic mammals, revealing the existence of specialised gliders, climbers and burrowers, semi-aquatic forms and even badger-sized carnivores that ate small dinosaurs [1–4]. Evidence of extensive ecological differences has been found even between closely-related species [5,6], and quantitative analyses of the skulls and skeletons of Mesozoic mammals suggest a diverse range of diets and locomotor modes [4,7–9]. Although the ecological

