How the brain discovers structure in sound sequences

Maria Chait*

Ear Institute, University College London, London WC1X 8EE, United Kingdom

Abstract: Sensitivity to patterns is fundamental to sensory processing, in particular in the auditory system, and a major component of the influential ‘predictive coding’ theory of brain function. Supported by growing experimental evidence, the ‘predictive coding’ framework suggests that perception is driven by a mechanism of inference, based on an internal model of the signal source. However, a key element of this theory — the process through which the brain acquires this model, and its neural underpinnings — remains poorly understood. Here I review recent brain imaging and behavioural work which focuses on this missing link. Together these emerging results paint a picture of the brain as a regularity seeker, rapidly extracting and maintaining representations of acoustic structure on multiple time scales and even when these are not relevant to behaviour.

Keywords: Auditory scene analysis, Functional brain imaging, Predictive coding, Surprise, MMN

PACS number: 43.64.Ri [doi:10.1250/ast.41.48]

A growing body of work demonstrates that our perception of the present does not only depend on current sensory input, but is largely driven by our previous experiences in the environment [1–4]. In doing so the brain capitalizes on the fact that the world is not random. Our environments, and specifically the animate agents which operate within them, are governed by their various physical constraints which give rise to mostly statistically structured, predictable, sensory signals. The ability to discover regularities in the sensory input therefore allows the observer to efficiently interact with their surroundings including optimizing behaviour, facilitating scene analysis [4–7] and enabling rapid detection of changes in one’s surroundings [8].

The crucial role of sensitivity to patterning is perhaps most intuitive in the context of hearing. Because of the inherently dynamic nature of sound, key attributes of acoustic sources, including their identity and state, are conveyed as a pattern across time. For example, the sound of a footstep, in isolation, is often meaningless but a sequence of these sounds is readily identified and analysed for key information (where and how quickly somebody is moving, even their approximate body size). The detection of patterns also allows the system to extrapolate from past experience to form predictions about the most likely nature of sounds to follow, facilitating effective interaction with our environment (e.g. in the case of foot-steps, we automatically anticipate when and where the next sound will occur, and approach or escape as the situation dictates). In noisy environments the detection of patterns can provide the anchor that enables an observer to identify and track a potentially relevant ‘signal’ within a noisy dynamic scene [4,5].

The mismatch negativity (MMN) paradigm [9–11] has been instrumental in revealing listeners’ exquisite sensitivity to patterns in sound. The MMN is an auditory-evoked response generated by sounds violating some regular aspect of the prior sequence and is hypothesized to reflect a discrepancy between the memory trace, or expectations, produced by the standard stimulus, and the deviant information. Accumulating MMN work has demonstrated that listeners are sensitive to the violation of a variety of patterns, including very complex regularities [9–11]. This has been interpreted as indirect evidence for remarkable sensitivity to acoustic patterning. However a crucial missing link, also characterising the predictive coding [3,11] and statistical learning [12,13] literature more broadly, pertains to understanding how patterns are detected in the first place. How does the brain acquire an internal model of regularities in the environment? Recent work is beginning to reveal these processes.

In my laboratory, the general approach has been to model complex listening tasks using simple, well-controlled signals that allow us to systematically manipulate statistical complexity. We have used two types of stimuli to investigate the process of discovering structure in rapid sound sequences. One approach involves modelling the challenges which might be faced by the auditory system
in a complex multi-sourced scene. The other involves modelling the process of discovering structure amidst a noisy background.

1. HOW DOES THE BRAIN DISCOVER THE EMERGENCE OF STRUCTURE WITHIN A RAPIDLY UNFOLDING SOUND SEQUENCE?

We use stimuli which contain transitions from random to regularly repeating frequency structure [14–20]. Stimuli consist of sequences of brief (50 ms) tone-pips arranged in a random (RAND) or regular (REG) pattern, or transitioning from random to regular (RAND-REG) as in Fig. 1. RAND sequences are generated by randomly selecting frequencies from a pre-specified frequency pool. REG sequences are generated by selecting a certain number ($R_{cyc}$) of random frequencies from the pool, permuting their order and then arranging in repeating succession. Novel patterns are generated on each trial and the complexity of the regularity can be manipulated by varying $R_{cyc}$. To detect the regular pattern, the brain must possess mechanisms that keep track of the ongoing sequence, maintain a certain portion of the past in some form of memory, compare incoming information to this representation, and decide at which point there is sufficient evidence (i.e. repeating tones) to indicate that the sequence has changed statistics (i.e. is no longer random). Tones are too short and sequences too rapid to allow any form of deliberate reasoning regarding pattern emergence. The fact that the regularity appears to perceptually ‘pop out’ from within the ongoing sequence suggests a mechanism of ‘automatic’ evidence accumulation decoupled from conscious reasoning.

The onset of regularity is detectable only after the first cycle has elapsed, at what we call the ‘effective transition’ (Fig. 1). Up to that point the stimulus is indistinguishable from a random sequence. The paradigm measures how much information beyond this point is required by listeners in order to determine that a regularity is present. This process can be formalized using a variable-order Markov model [21] that allows us to determine the point in time at which an ideal observer can detect the transition. Model predictions are then compared to the behavioural response time (RT; the latency at which participants detect the transition, corrected by their response latency to a simple acoustic change). Comparing behavioural RT to the time course of brain responses at the transition allows us to probe the neural mechanisms sub-serving pattern detection. Note that detection of regularity can also be measured (behaviourally and with brain imaging) by recording responses to the onset of REG stimuli (Fig. 2(C), (D), (E), below) and comparing them to the onset of RAND stimuli: During the initial cycle, RAND and REG signals are identical with differences emerging as soon as the auditory system discovers that the pattern is repeating. This variant of the paradigm has higher yield because stimuli
can be shorter, and may be preferable under certain conditions. In Barascud et al. [17] we report data for both stimulus types.

In our experiments to date, behaviour and brain responses (measured with MEG and EEG) were recorded in different groups of subjects. In the MEG and EEG experiments participants were naïve to the stimuli and engaged by an incidental visual task. The results demonstrate that: (1) We can follow the process by which the brain discovers regularity in ‘real time’ and with fine temporal resolution (2) Brain responses to changes in statistics appear to be automatic — they are seen for naïve, non-behaving (passively listening) subjects. (3) Behaviourally, listeners are remarkably quick in detecting the emergence of regular patterns within rapidly unfolding random tone-pip sequences. For patterns of up to 10 tones (as in Fig. 1) listeners required an average of 1.5 cycles (less than one repetition!) to detect the regularity, performing on par with an ideal observer model. Brain responses measured from naïve listeners were equally rapid. Interestingly, the onset of regularity is manifest as a substantial increment in sustained potential (or magnetic field in MEG), that adds to the initial increment at stimulus onset (Fig. 2(A),(C)). Conversely, the offset of regularity is associated with a drop in DC (not shown here, but see [17]). This finding is interesting for a number of reasons: firstly it demonstrates that the brain appears to encode the state (RAND vs. REG) rather than the transition (as in e.g. MMN). Secondly, the amplitude pattern [(REG) > (RAND)] is not easily interpretable in terms of simple physical attributes of the signal. Adaptation for example would be expected to result in the opposite pattern. Instead, as discussed in Barascud et al. [17] the DC shift appears to vary consistently with the predictability (negentropy) of the ongoing stimulus pattern such that increased predictability is systematically associated with higher sustained
responses. This effect, underpinned by increased activity in a network of temporal, frontal and hippocampal sources [17,18], may reflect a mechanism which tracks the context-dependent reliability of sensory streams.

The emerging view is that the brain continually tracks and maintains a detailed representation of the structure of the unfolding sensory input and that this representation shapes the processing of incoming information. Indeed, we have recently shown [20] that frequency deviants within REG sequences evoke a substantially larger response than matched deviants in RAND sequences (over and above the sustained response difference between conditions). These results are consistent with the general view of the brain as a statistical learning machine, implementing perception as a process of prior-knowledge-driven inference [3].

2. THE ROLE OF SENSITIVITY TO STRUCTURE IN AUDITORY SCENE ANALYSIS

Most of what we know about the role of sensitivity to regularity in sensory perception in general and acoustic scenes in particular, is based on experiments using very simple stimuli, consisting of just one or two concurrent sequences [5,10,17,22,23]—a far cry from the complex multi-objected scenes we face in the natural environment. It is thus unclear whether sensitivity to regularity plays a role in more complex situations (e.g. a busy street), where the scene consists of many simultaneous objects. In my lab, we have modelled such environments by creating artificial ‘scenes’ which consist of many concurrent, uniquely patterned, streams. Sensitivity to patterning is then quantified in the context of a change detection task [24].

The ability to detect changes in our surroundings is a fundamental requisite for survival and it is widely assumed that the auditory system plays a critical role in this process, serving as an ‘early warning system’ which continuously scans the unfolding acoustic scene for behaviourally-relevant events (e.g. the approach of predators or prey). A change detection task [24] is thus an ecologically relevant means by which the role of sensitivity to temporal regularity in the course of auditory scene analysis can be studied.

Signals are artificial ‘scenes’ (Fig. 3) populated by multiple streams of pure tones that are designed to model acoustic sources. Each source is characterized by a different carrier frequency (drawn from a pool of widely spaced fixed values; to minimize inter-component masking), and is furthermore amplitude modulated (AM) by a square wave (the source can be seen as a stream of tone pips) at a distinct rate. Choice of frequencies and AMs is random for each trial. The AM mimics temporal properties found across many natural sounds and ensures that, similarly to natural scenes, the stimuli are perceived as a composite ‘sound-scape’ that is perceptually separable, so that each stream can be attended to individually (as verified in a control experiment). Scene size (number of source in the scene) can be systematically varied to create increasingly complex scenes. 50% of scenes contain no change.

Fig. 3 (A) RAND and REG scene stimuli consisted of several concurrent tone-pip streams each with a unique frequency and tone-pip duration. In REG scenes each stream is characterized by a unique, fixed, rate. In RAND scenes inter-tone intervals are random. Scenes were 2,500–3,500 ms in duration. (B) MEG responses from Sohoglu & Chait [8]. The horizontal green line indicates the time intervals where significant differences were observed between REG and RAND conditions. A significant difference between conditions emerges from ~400 ms after scene onset.
(each source is active throughout the stimulus; NC). In the remaining scenes, a single component is removed partway through the scene (‘change-disappear,’ CD, stimuli; 25%), or added to the scene (‘change-appear,’ CA, stimuli; 25%). The timing of change and the identity of the changing component vary randomly. In REG (Regular) scenes inter-tone intervals are fixed within each stream, resulting in a regular rate. In RAND scenes inter tone intervals are random (Fig. 3(A)).

Using this paradigm in an MEG study [8] (Fig. 3(B)) (see also behavioral results in [25]) we demonstrated a pattern of brain responses that is generally remarkably similar to that described for the frequency patterns, above: Brain responses, from naïve distracted listeners, revealed increased sustained responses to the REG scenes relative to RAND scenes from 400 ms post onset. Over and above this difference, brain responses to scene changes (appearance of a new stream) evoked larger responses from 70 ms post onset. The same effects were mirrored in behaviour — when required to respond to the scene changes, subjects were faster and more accurate in detecting changes in REG relative to RAND scenes. Notably the difference between REG and RAND scenes occurred despite the fact that scenes were otherwise spectrally matched. This was interpreted as reflecting the operation of mechanisms which infer the precision (predictability) of sensory input and upregulate responses to reliable sensory information, such that violations of these patterns (e.g., in the form of an appearing sources) evoke higher prediction errors. Overall, these results are consistent with an account according to which the auditory system rapidly discovers regular structure in the unfolding sensory input. Transients (onsets of individual tones) in regular streams are therefore predictable and easier to ignore rendering un-expected events as more salient [5,19]. In contrast in RAND scenes effectively every transient was somewhat unpredictable, reducing the ‘surprisal’ associated with the appearing component.

3. OUTLOOK

Understanding sensitivity to patterning is essential for understanding fundamental aspects of listening. Together the results reviewed here paint a picture of the brain as a regularity seeker, extracting and maintaining representations of acoustic structure on multiple time scales and even when these are not relevant to behaviour. A brain network of auditory cortical frontal and parietal sources is implicated in the course of extracting the regularity and maintaining a top down predictive model. The similarity in brain responses across different instantiations of predictability suggest that they may be reflecting general processes linked to predictive perception. Exactly what computations these brain responses reflect, what acoustic patterns are learnable, etc. remain to be discovered.

REFERENCES

[1] M. F. S. Rushworth and T. E. J. Behrens, “Choice, uncertainty and value in prefrontal and cingulate cortex,” Nat. Neurosci., 11, 389–397 (2008).
[2] K. P. Körding and D. M. Wolpert, “Bayesian integration in sensorimotor learning,” Nature, 427, 244–247 (2004).
[3] A. Clark, “Whatever next? Predictive brains, situated agents, and the future of cognitive science,” Behav. Brain Sci., 36, 181–204 (2013).
[4] I. Winkler, S. L. Denham and I. Nelken, “Modeling the auditory scene: Predictive regularity representations and perceptual objects,” Trends Cogn. Sci., 13, 532–540 (2009).
[5] L.-V. Andreou, M. Kashino and M. Chait, “The role of temporal regularity in stream segregation,” Hear. Res., 280, 228–235 (2011).
[6] J. Zhao, N. Al-Aidroos and N. B. Turk-Browne, “Attention is spontaneously biased toward regularities,” Psychol. Sci., 24, 667–677 (2013).
[7] M. Maravall, S. Ostojic, D. Pressnitzer and M. Chait, “More than the sum of its parts: Perception and neuronal underpinnings of sequence processing,” Neuroscience (in press).
[8] E. Sohoglu and M. Chait, “Detecting and representing predictable structure during auditory scene analysis,” eLife, doi:10.7554/eLife.19113 (2016).
[9] R. Näätänen, T. Kujala and I. Winkler, “A unique window to central auditory processing opened by the mismatch negativity and related responses,” Psychophysiology, 48, 4–22 (2011).
[10] A. Bendixen, I. SanMiguel and E. Schröger, “Early electrophysiological indicators for predictive processing in audition: A review,” Int. J. Psychophysiol., 83, 120–131 (2012).
[11] M. Heilborn and M. Chait, “Great expectations: Is there evidence for predictive coding in auditory cortex? — A critical review,” Neuroscience (in press).
[12] B. Wilson, Y. Kikuchi, L. Sun, D. Hunter, F. Dick, K. Smith, A. Thiele, T. D. Griffiths, W. D. Marslen-Wilson and C. I. Petkov, “Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans,” Nat. Commun., 6, 8901 (2015).
[13] A. C. Schapiro, E. Gregory, B. Landau, M. McCloskey and N. B. Turk-Browne, “The necessity of the medial temporal lobe for statistical learning,” J. Cogn. Neurosci., 26, 1736–1747 (2014).
[14] M. Patel and M. Chait, “Retroactive adjustment of perceived time,” Cognition, 119, 125–130, doi:10.1016/j.cognition.2010.10.011 (2011).
[15] Z. Jaunmahomed and M. Chait, “The timing of change detection and change perception in complex acoustic scenes,” Front. Psychol., 3, doi:10.3389/fpsyg.2012.00396 (2012).
[16] M. Chait, C. C. Ruff, T. D. Griffiths and D. McAlpine, “Cortical responses to changes in acoustic regularity are differentially modulated by attentional load,” Neuroimage, 59, 1932–1941, doi:10.1016/j.neuroimage.2011.09.006 (2012).
[17] 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).
[18] R. Aukstzulewicz, N. Barascud, R. Cooray, A. C. Nobre, M. Chait and K. Friston, “The cumulative effects of predictability on synaptic gain in the auditory processing stream,” J. Neurosci., 37, 6751–6760, doi:10.1523/JNEUROSCI.0291-17 (2017).
[19] R. Southwell, A. Baumann, C. Gal, N. Barascud, K. Friston and M. Chait, “Is predictability salient? A study of attentional
capture by auditory patterns,” *Philos. Trans. R. Soc. B Biol. Sci.*, **372**(1714) (2017).

[20] R. Southwell and M. Chait, “Enhanced deviant responses in patterned relative to random sound sequences,” *Cortex*, **109**, 92–103 (2018).

[21] M. T. Pearce, M. H. Ruiz, S. Kapasi, G. A. Wiggins and J. Bhattacharya, “Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation,” *NeuroImage*, **50**, 302–313 (2010).

[22] T. Baldeweg, “Repetition effects to sounds: Evidence for predictive coding in the auditory system,” *Trends Cogn. Sci.*, **10**, 93–94 (2006).

[23] N. Barascud, T. D. Griffiths, D. McAlpine and M. Chait, “Change deafness arising from inter-feature masking within a single auditory object,” *J. Cogn. Neurosci.*, **26**, 514–528 (2014).

[24] F. C. Constantino, L. Pinggera, S. Paranamana, M. Kashino and M. Chait, “Detection of appearing and disappearing objects in complex acoustic scenes,” *PLoS ONE*, **7**, e46167 (2012).

[25] L. Aman, L.-V. Andreou and M. Chait, “Listeners track the temporal statistics of multiple concurrent auditory sources,” https://www.biorxiv.org/content/early/2017/04/11/126763 (in review).