An ear for statistics

Israel Nelken & Alain de Cheveigné

A study finds that sound textures are stored in auditory memory as summary statistics representing the sound over long time scales; specific events are superimposed, forming a ‘skeleton of events on a bed of texture’.

What do we retain of the incessant streams of sound that reach our ears? Potentially, there is a lot to remember: the popular MP3 format, designed on the basis of human perceptual limits, stores anywhere from 32 to 320 kbits for each second. Remembering sensory input is useful, for example to spot recurring patterns that might be of survival value, such as the soft footsteps of a predator. And yet we cannot retain it all because our memory would fill up, and patterns would be hard to find. Besides, most of the detail of our sensory experience is of little use, as argued eloquently by Jorge Luis Borges in the short story “Funes, the Emoirous”. So what should we keep and what should we discard? In this issue, McDermott et al. make a major contribution to this question, showing that for a certain class of sounds that they call “textures”, all that we remember is a small set of summary statistics.

Psychophysicists like to annoy their subjects by presenting them over and over (and over) with pairs of stimuli to be compared. (“Which sound had the higher pitch?” “Which picture contained a faint change in contrast?”) A constant observation in such studies is that the longer the stimulus is, the better perceptual invariance in audition. For example, Ohm’s acoustic law states that sounds that only differ in the relative phases of their sinusoidal components sound alike. Ohm’s law can be traced to a large extent to the function of the peripheral auditory system. It may be useful in reducing sensitivity to phase distortion due to propagation and reverberation in the environment. Another, more complex perceptual invariance is categorical perception of speech sounds, according to which sounds that fall in the same phonemic category (for example, “e”) are harder to distinguish than sounds that fall on either side of a boundary between phonemes (for example, “e” and “o”). Different tokens of the same phoneme differ acoustically and at the level of the early auditory system and yet are perceptually very
similar. As in the case of textures, discrimination efficiency of speech in noise may be reduced when the acoustic-to-phonetic transformation of the sound signal is engaged.

Nevertheless, the auditory system can store, under the right conditions, more than just statistics. As an example, a faint bird song over the texture-like sound of a stream is somehow perceived separately from the texture, although it may barely disturb the summary statistics. This implies that extra storage space is available. Why is it not used to store details of other exemplars of textures, possibly improvised? For example, individual chunks of white noise can be stored precisely in memory and retrieved weeks later. Thus, the picture that emerges is that of a flexible sensory memory mechanism that stores not only the slowly varying statistics of ongoing textures but also the detailed patterns of temporally localized events. Applied to textures, statistics serve both to reduce storage costs and to generalize over irrelevant detail.

These findings raise many questions. Statistics that describe longer-term patterns would be expected to be remembered longer than the more detailed information that characterizes shorter events. We could conjecture, therefore, that the existence of yet higher level statistics applicable to even longer sounds, describing, for example, the pattern of variation over time of the statistics proposed by McDermott et al. (‘meta-textures’). Next, are all the necessary statistics produced in parallel, directly from incoming sensory information? Or are higher-order statistics elaborated from lower-level representations according to a hierarchical compaction process? What triggers their calculation? Are they extracted only for textures, or are they extracted generally but are the only information that is stored in the case of textures? How does the auditory system decide that a texture is a texture?

This study opens new perspectives on sound and hearing. In the past, countless studies have surveyed our ability to detect subtle differences between sounds, such as the presence of a weak target, a minute temporal gap, or a small increment in frequency or amplitude, leading to the notion that every detail of a sound counts. We now discover that there exists a class of sounds, textures, for which radically different waveforms evoke the same percept if they share a minimum set of statistical features. This finding has implications both for the processing taking place in the auditory system and for the general design principles of sensory systems. It suggests that statistics may be applicable to elements of the scene and that the auditory system may maintain a multi-thread representation involving events and statistics of multiple sources: a ‘skeleton of events on a bed of texture’.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.