The what, where and how of auditory-object perception

Jennifer K. Bizley[1] and Yale E. Cohen[2]

Abstract | The fundamental perceptual unit in hearing is the 'auditory object'. Similar to visual objects, auditory objects are the computational result of the auditory system's capacity to detect, extract, segregate and group spectrotemporal regularities in the acoustic environment; the multitude of acoustic stimuli around us together form the auditory scene. However, unlike the visual scene, resolving the component objects within the auditory scene crucially depends on their temporal structure. Neural correlates of auditory objects are found throughout the auditory system. However, neural responses do not become correlated with a listener's perceptual reports until the level of the cortex. The roles of different neural structures and the contribution of different cognitive states to the perception of auditory objects are not yet fully understood.

Hearing and communication present various challenges for the nervous system. To be heard and to be understood, an auditory signal must first be transformed from a time-varying acoustic waveform into a perceptual representation (FIG. 1). This is then converted to an abstract representation that combines the extracted information with information from memory stores and semantic information[1]. Last, this abstract representation must be interpreted to guide the categorical decisions that determine behaviour. Did I hear the stimulus? From where and whom did it come? What does it tell me? How can I use this information to plan an action?

There is broad agreement that the ventral auditory pathway — a pathway of brain regions that includes the core auditory cortex, the anterolateral belt region of the auditory cortex and the ventrolateral prefrontal cortex — has a role in auditory-object processing and perception[2–5]. However, no consensus has been reached on either the roles of different regions in this pathway in specific elements of auditory-object processing and perception or the contributions of particular cognitive states (such as attention) to the differential modulation of activity along this pathway. Here, we discuss how the brain transforms an acoustic-based representation of a stimulus into one that is object-based. We consider how object-related neural activity might emerge and how attention and behavioural state influence perception and neural activity. We also review what is known and, more importantly, what is unknown regarding the hierarchical flow and transformation of information along the ventral pathway. Finally, we focus on studies that relate neural activity to behaviour; reviews of work underlying perceptual correlates of audition in non-behaving animals can be found elsewhere[5–9].

[1]*University College London Ear Institute, 332 Grays Inn Road, London, WC1X 8EE, UK.*
[2]*Departments of Otorhinolaryngology, Neuroscience and Bioengineering, 3400 Spruce St – 5 Ravdin, Uuniversity of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.*
e-mails: j.bizley@ucl.ac.uk; ycohen@mail.med.upenn.edu
doi:10.1038/nrn3565

What is an auditory object?

The precise definition of an auditory object has been the subject of considerable debate[1,10–17]. Intuitively, we understand an auditory object to be the perceptual consequence of the auditory system's interpretation of acoustic events and happenings. For example, when sitting outside a café, we might hear a bird sing, a car passing, the hiss of a coffee machine or the voice of our friend. Each of these different and discrete sounds can be described as an auditory object[11–14]. More formally, auditory objects are the computational result of the auditory system's ability to detect, extract, segregate and group the spectrotemporal regularities in the acoustic environment into stable perceptual units[1,11,12]. Thus, we define an auditory object as a perceptual construct, corresponding to the sound (such as the hiss) that can be assigned to a particular source (the coffee machine).

Auditory objects have several general features and characteristics[11]. First, acoustic stimuli are emitted from or by things, as a consequence of actions or events. Some acoustic stimuli, such as human speech, are emitted with a clear intention, whereas others, such as environmental sounds, are not. In either case, we rarely hear sounds in isolation. Therefore, an auditory object spans multiple acoustic events that unfold over time, and a sequence of objects forms a 'stream'. For example, when a person is walking, each step is a unique acoustic event

© 2013 Macmillan Publishers Limited. All rights reserved

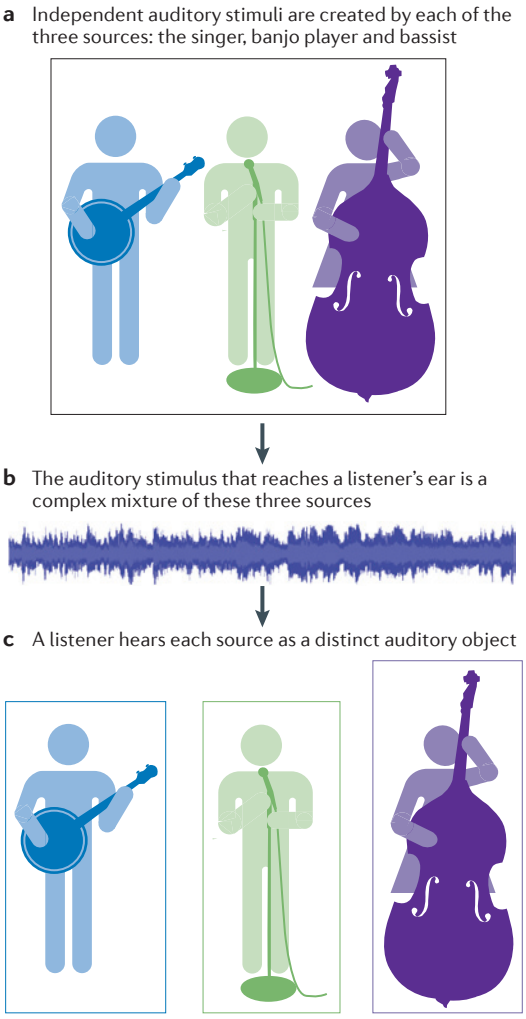


Figure 1 | **The transformation of an acoustic stimulus into a perceptual representation of a sound.** The fundamental problem that is solved by the auditory system is the need to transform an acoustic stimulus into a perceptual representation of one or more auditory objects. Typically, various independent sound sources contribute to the creation of a soundscape. **a** | In the example shown, there are three sound sources (a banjo player, a singer and a bassist), each of which is producing an acoustic stimulus with unique spectrotemporal features. **b** | The auditory stimulus that reaches a listener's ear will be a complex mixture of the stimuli produced by these three sources. **c** | However, the listener hears each source as a distinct auditory object. BOXES 1,2 discuss the grouping cues that underlie this capacity to segregate a stimulus into unique sound sources.

Pitch
The attribute of a sound that enables it to be ordered from high to low on a musical scale. The perceived pitch for a periodic sound is determined by its fundamental frequency (F0), usually the lowest frequency component.

Timbre
The quality of a sound that is determined by its spectral or temporal envelope. Timbre allows a listener to differentiate between a violin and a banjo despite the fact that the two instruments may be producing a sound that has the same pitch.

Harmonicity
A harmonic sound contains frequency components at integer multiples of the fundamental frequency (see the definition for 'pitch'). Many vocalizations and other pitch-evoking sounds have a harmonic structure.

or object. However, our auditory system groups these separate stimuli together into a temporal sequence of 'footsteps'. A stream of objects can, itself, be termed an object[1,15]. Second, we can parse the soundscape into its constituent objects. Therefore, one auditory object has spectrotemporal properties that make it separable from other auditory objects[11–15]. As a consequence, we can detect our friend's voice among myriad other sounds in the café. Third, as with a visual object, a listener can readily describe an auditory object by the combination of its features: it might have a high or low pitch, a rich timbre or a characteristic loudness. However, the same listener would find it very difficult to describe the underlying acoustic features that give rise to these percepts, such as the harmonicity of the sound or the timing difference between our ears[15]. Fourth, like vision, auditory-object recognition is invariant to various changes to its spectrotemporal properties, which result from the context in which the object is perceived. For example, a violin still sounds like a violin regardless of whether a single high note or a rapid melody is played, whether it is played loudly or softly or whether it is played alone or as part of an orchestra. As in the visual system, we must be capable of generalizing across the different ways in which an object or event occurs[1,18–20]. Last, we expect object representations to predict parts of the object for which no input is currently available. For example, Jan can still understand Jenny's speech despite the fact that Yale's sneezing has masked certain acoustic features of her speech by rendering them inaudible[11,21–25].

How are auditory objects formed? Our ear receives a composite waveform comprised of all of the acoustic stimuli in the environment. The brain's job is to appropriately group these acoustic features into perceptual features and then to group these to form a representation of discrete objects that can be further analysed (FIG. 1). An auditory stimulus comes into our awareness as an auditory object by means of the simultaneous and sequential principles that group acoustic features into stable spectrotemporal entities (BOXES 1,2). Although attention is not always necessary for auditory-object formation[26], our awareness of an object can be influenced by attention[14,17]. For example, we can choose whether to listen to — or ignore — the first violin, the strings or the whole orchestra. Likewise, we can selectively attend to the features of a person's voice that allow a listener to identify the speaker.

Hierarchical processing in the cortex

Visual information processing is thought to take place in two parallel pathways that independently analyse the identity and location of objects within the visual scene[27]. Initially, on the basis of theoretical and anatomical studies, a similar processing scheme was proposed for the auditory cortex[2–5] whereby information is processed in parallel hierarchical pathways specialized for the extraction of spatial ('where is the sound?') and non-spatial ('what is the sound?') components of an auditory stimulus. These computations occur in the so-called 'dorsal' and 'ventral' pathways, respectively. As we discuss in detail below, both functional imaging studies in humans and single-unit neurophysiology in non-human animals provide evidence in favour of a division of labour between spatial and non-spatial processing. Conversely, other studies using the same methods suggest that rather than two hierarchically organized parallel pathways, distributed, dynamically organized processing networks are likely to support auditory perception. According to this theory, feedback between brain areas would facilitate object selection.

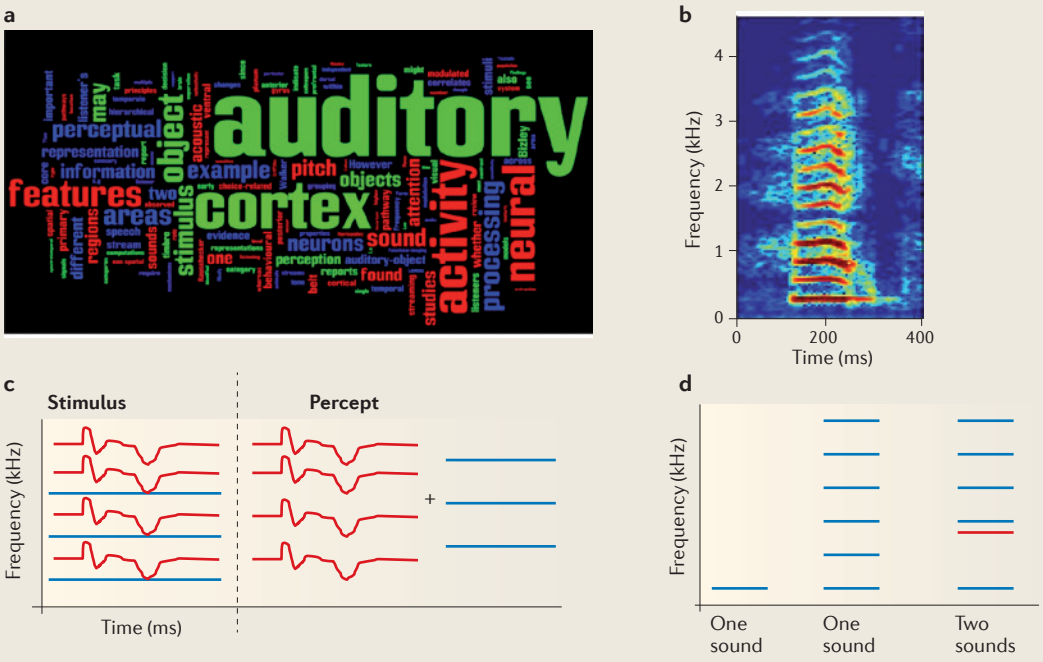
© 2013 Macmillan Publishers Limited. All rights reserved

Box 1 | **Analysing the soundscape: simultaneous grouping cues**

Identifying an auditory object involves assigning elements of the incoming sensory input into one or more sources. Several of the cues that are used to group auditory stimuli into objects can be classified as 'simultaneous cues' (REF. 11). We automatically group the elements of a visual scene, such as that shown in panel **a** of the figure into distinct objects (in this case, on the basis of the colour of the letters, the proximity and orientation of adjacent letters, the size and letter font). Similarly, in audition, the brain groups together stimuli associated with acoustic cues — such as pitch, harmonicity, timbre, common onset or modulation time and spatial location — that can be quickly derived from a sound's spectral features[72].

Natural sounds, such as speech, are often harmonic: that is, they have energy at integer multiples of the lowest (or fundamental) frequency. This is illustrated in panel **b** of the figure, which shows a spectrogram of a human speech sound in which horizontal bands of energy are visible. Importantly, individual harmonics change coherently over time, and harmonic frequencies that change coherently are grouped together. This shown schematically in panel **c**: sound elements that change coherently are grouped together such that the red and blue sound elements form two separate auditory objects. Pitch is another important grouping cue that allows a listener to identify and track simultaneous speakers. Panel **d** of the figure shows a related cue, harmonicity. Here, a single pure tone or a harmonic series of pure tones (blue) are both perceived as a single sound. However, the introduction of a 'mistuned' harmonic — that is, a harmonic at a frequency that is not an integer of the fundamental frequency (red) — results in the perception of an additional separate sound. Differences in timbre are used to identify different vowel sounds or different musical instruments even when the instruments are playing the same note.

Sound components with a common onset time are likely to be perceived as originating from the same object. In natural listening conditions, onset time is one of the more important grouping cues. Spatial location provides relatively weak grouping[72,166,167], but when a listener attends to a particular location, attentional resources can facilitate the distinction between simultaneous speech sounds[14].



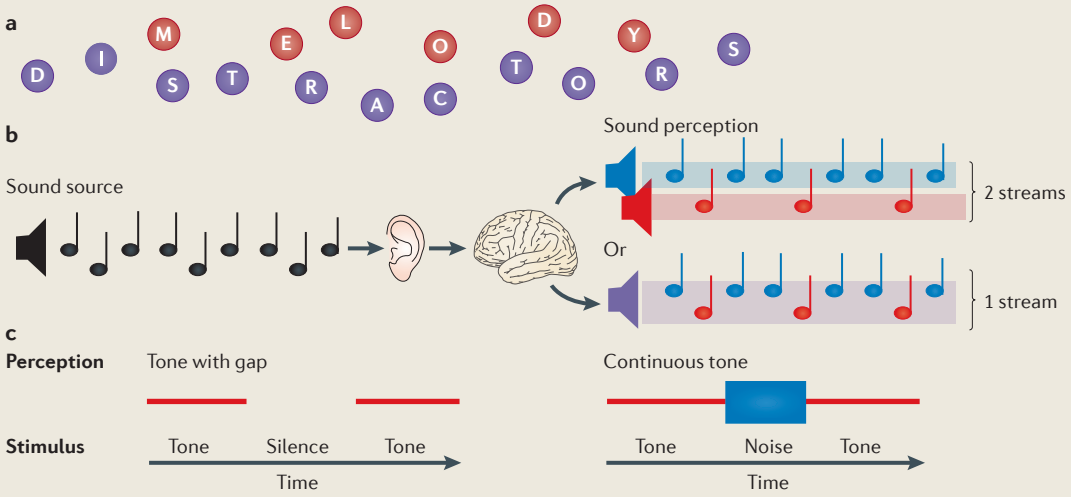
Processing strategies within auditory cortex. Under a hierarchical-processing model, auditory-object extraction occurs in the ventral processing pathway, and we might expect to see, as we move along the pathway, a transition from the representation of acoustic features to perceptual features and finally to objects or category-specific representations at the highest stages — computations perhaps analogous to those that are well described in higher visual areas[28–31]. At least in non-human primates, the ventral pathway begins in the core auditory cortex — specifically, the primary auditory cortex and the rostral field (FIG. 2). These core areas project to the anterolateral belt region of the auditory cortex. In turn, this belt region projects to the ventrolateral prefrontal cortex.

There are several pieces of evidence suggesting that auditory-object and spatial processing occurs in separate, parallel pathways (FIG. 2). Some of the first physiological evidence for a separation of spatial and non-spatial processing was provided by a study[32] that investigated neural sensitivity to sound location and identity using a series of monkey vocalizations presented at different spatial locations. This study found that belt regions in the ventral auditory pathway were more sensitive to vocalization type, whereas belt regions in the dorsal pathway were modulated more by the location of a stimulus. Similarly, early human imaging data supported a division of spatial and non-spatial processing[33,34]. Furthermore, a meta-analysis of functional imaging data showed that spatial tasks almost always activate

© 2013 Macmillan Publishers Limited. All rights reserved

Box 2 | **Analysing the soundscape: sequential grouping cues**

Auditory stimuli can be grouped into objects using what are known as sequential grouping cues[11]. Sequential grouping cues enable temporal sequences of sounds to be assigned to a common source: panel **a** of the figure shows a visual analogy in which the sets of letters are grouped into two words because they form a sequence from left to right. As shown in panel **b** of the figure, these cues have been studied using repeating patterns of pure tones in which the patterns are separated perceptually into two or more streams[168]. Two factors determine most stream segregation: frequency separation (a bigger difference in the frequency of the tones makes it more likely that two streams will be perceived) and speed (if the presentation rate of the tones is increased, a listener is more likely to hear two streams). A hallmark of such streaming is that listeners find it hard to make inter-stream judgements, such as judging the order of two sounds that are in separate streams. Such percepts can be 'bistable': at intermediate frequency separations (such as 3–7 semitones), the perception of 'one stream' and 'two streams' alternates over time. However, with increased listening time, a stable two-stream percept is developed. Panel **c** illustrates another example of sequential integration that is called 'amodal completion' (the continuity illusion). Here, a discontinuous tone is heard as continuous when a noise burst occurs during the gap.



the posterior auditory cortex (part of the dorsal stream), whereas non-spatial activity is observed across the temporal lobe[35]. Finally, other findings have shown that the ventral stream is involved in the categorization of speech sounds[36–38], which is an important component of auditory-object processing[1]. Preferential spatial and non-spatial processing is also found outside the auditory cortex: for example, in the prefrontal cortical regions that are part of these hypothesized dorsal and ventral pathways[39–41].

Nevertheless, substantive auditory-object processing has been identified in the dorsal pathway, and substantial information about auditory space has been found in the ventral pathway[42–48]. Such findings suggest that a model of parallel hierarchical processing might be too simplistic and that a mixture of spatial and non-spatial auditory information might be useful for those computations that create the consistent perceptual representations that guide goal-directed behaviour. For example, spatial information can act as a grouping cue to enable auditory stream formation. When a rhythmic sequence of identical sound bursts is presented from a single location, it is perceived as one source by human observers. However, such a sequence is perceived as two sources, each with a distinct rhythm, when the sound sequences are presented from two spatially separated locations[49]. Neural correlates of this paradigm are observed in the auditory cortex of anaesthetized cats[50]. Likewise, non-spatial (object) information processed in the dorsal stream might contribute to computations that involve target selection, the online computational processing of dynamic auditory information, audiomotor processing and other computations that involve organization of the auditory scene (see REFS 42,43,51–54 for reviews of hierarchical processing of speech in both the posterior and anterior auditory cortex). However, as most, if not all, studies have asked listeners to attend to either spatial or non-spatial features of a sound but not to both simultaneously, the interaction between these two pathways has not been fully resolved within either the auditory or visual systems[55].

Within the ventral and dorsal processing pathways, both single-neuron studies[32,56–59] and functional imaging studies[60–64] indicate that the perceptual features of a sound might be localized and organized in a hierarchical manner. Pitch is probably the most widely studied perceptual feature; below, we use it to explore findings that support both hierarchical and distributed organizational schemes.

Pitch processing: hierarchical or distributed? Several important studies indicate that pitch-selective neurons are localized to specific cortical areas. For example, in non-human primates, pitch-selective neurons are found at the border between the core and belt auditory

© 2013 Macmillan Publishers Limited. All rights reserved

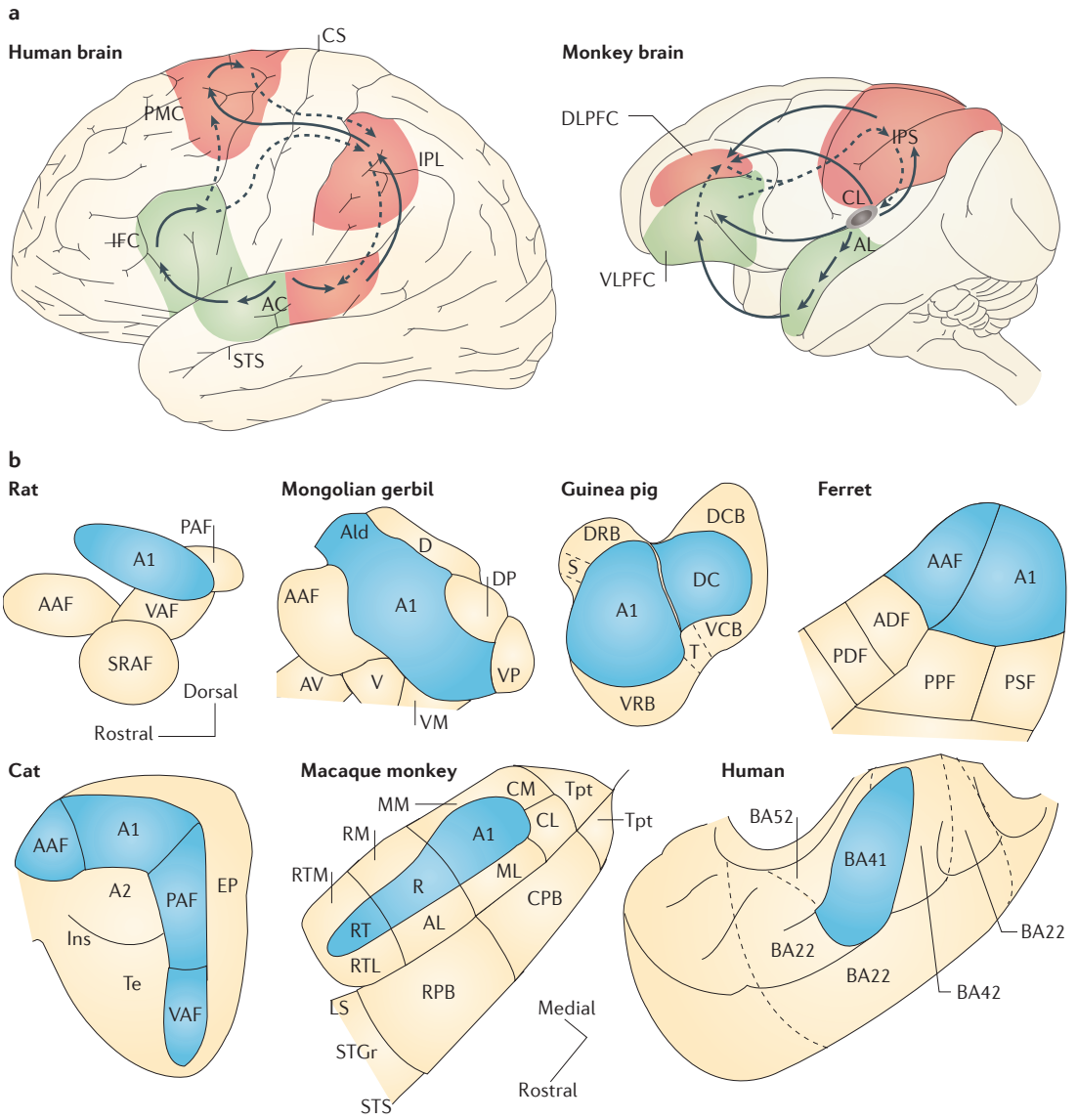


Figure 2 | **Dual pathways of information flow in the auditory system and the organization of the auditory cortex. a** | Information processing in the primate auditory system is hypothesized to occur in two streams. Neurons in the 'dorsal' stream (red), which may preferentially analyse space and motion, are involved in audiomotor processing, whereas those in the 'ventral' stream (green) are preferentially involved in auditory-object processing[6]. Solid arrows indicate feedforward projections, and dashed arrows indicate feedback projections. **b** | A schematic representation of the organization of the auditory cortex (AC) in different species[73]. The lemniscal auditory thalamocortical projection terminates in the 'core' regions of the AC (blue shading), including the primary auditory cortex (A1). In humans, this core region is in Brodmann area 41 (BA41). From these core areas, there is both serial and parallel processing in the surrounding 'belt' regions (such as the anterolateral (AL) and middle-lateral (ML)) regions in the macaque monkey or the secondary AC (A2) in the cat) and from there to the 'parabelt' regions (such as the rostral parabelt (RPB) in the macaque; see REF. 73 for more details). Although this organization was originally described in non-human primates, it appears to be a general organizational scheme in a variety of primate and non-primate species. Solid lines indicate boundaries between auditory fields, and dashed lies indicate anatomical boundaries**.** AAF, anterior auditory field; ADF, anterior dorsal field; Ald, dorsal region of the primary auditory field; AV, anteroventral field; CL, caudolateral belt region of the AC; CM, caudomedial area; CPB, caudal parabelt; CS, central sulcus; D, dorsal field; DC, dorsocaudal field; DCB, dorsocaudal belt; DLPFC, dorsolateral prefrontal cortex; DP, dorsoposterior field; DRB, dorsorostral belt; EP, ectosylvian posterior auditory region; IFC, inferior frontal cortex; Ins, insula; IPL, intraparietal lobule; IPS, intraparietal sulcus; LS, lateral sulcus; MM, mediomedial belt; PAF, posterior auditory field; PDF, posterior dorsal field; PMC, premotor cortex; PPF, posterior pseudosylvian field; PSF, posterior suprasylvian field; RM, rostromedial belt; RPB, rostral parabelt; RTL, rostrotemporal lateral belt; RTM, rostrotemporal medial belt; SRAF, suprarhinal auditory field; STGr, superior temporal gyrus rostral to the parabelt; STS, superior temporal sulcus; T, transitional belt area; Te, temporal; Tpt, temporal lobe association cortex; V, ventral field; VAF, ventral auditory field; VCB, ventrocaudal belt; VLPFC, ventrolateral prefrontal cortex; VM, ventromedial field; VP, ventroposterior field; VRB, ventrorostral belt. Part **a** is modified, with permission, from REF. 6 © (2009) Macmillan Publishers Ltd. All rights reserved. Part **b** is modified, with permission, from REF. 73 © (2011) Elsevier.

© 2013 Macmillan Publishers Limited. All rights reserved

cortex[56]. Similarly, in humans, a pitch-sensitive area has been identified anterior to Heschl's gyrus[60–62]. Moreover, whereas neural activity throughout the auditory cortex correlates better with changes in a listener's reports of features such as pitch than with changes in the stimulus features, activity recorded in the low-frequency core and belt regions of the auditory cortex predicts both pitch and listeners' reports of pitch better than activity recorded in other regions[65].

However, many of the same studies also provide evidence that a broader network of brain areas may subserve pitch perception. For example, pitch-related activity has also been reported in both the core[66] and the non-core[63,67,68] auditory cortex in humans. Similarly, pitch-sensitive neurons are broadly distributed in core and non-core regions of the ferret auditory cortex, and neural responses in multiple regions of the auditory cortex correlate with pitch-perception judgements in this species[44,65,69].

The difficulties inherent in comparing data derived using different experimental methods (and often in different species) limit a comprehensive understanding of the neural correlates underlying pitch perception. For example, comparing studies using single-unit recordings and those using functional imaging is difficult as both are subject to different methodological constraints[70]. Functional MRI (fMRI) experiments, for example, usually compare the activity elicited by a pitch-evoking stimulus with that evoked by a control sound without pitch. By contrast, single-neuron studies present a particular class of pitch-evoking stimuli and test for a neuron's tuning to a specific fundamental frequency. Also, studies rarely attempt to map a neuron's pitch tuning while also using a number of spectrally different sounds in order to explore pitch constancy (although see REFS 56,68,71 for exceptions). Finally, it has proven difficult to identify individual brain regions or neurons that respond to a pitch irrespective of the stimulus' spectral properties[68,71].

Consequently, further studies (such as experiments in which particular neurons or brain areas are inactivated) will be required to determine whether putative pitch-selective areas have a causal role in auditory perception and to determine how these areas function interdependently of one another. Neurophysiological experiments would additionally benefit from exploring neural tuning using various pitch-evoking stimuli[68,71] to test for neural representations that can abstract pitch. Performing such studies in animals that are actively discriminating sounds on the basis of their pitch is essential to determine the response properties underlying pitch perception.

We predict two broad outcomes of such sets of experiments. It is possible that activity in a specialized area underlies pitch perception[9] but that broadly distributed pitch sensitivity enables pitch to be used for making sense of the auditory scene — for example, by enabling common pitch to be used as a grouping cue[72]. Alternatively, a distributed network of pitch-activated areas might form a processing hierarchy[70]. For example, pitch processing within the primary auditory cortex could depend on the listening context, whereas pitch processing in extra-core regions (such as the planum temporale[73,74]) might be context-independent. In other words, there might be an invariant representation of pitch in the planum temporale but not in core areas such as Heschl's gyrus[63], which is consistent with the idea of a pitch-processing hierarchy.

Timbre: explicit and implicit representations. Similar principles can be drawn from the study of other perceptual dimensions. Another important perceptual feature of a sound is timbre. The neural representation of timbre is broadly distributed: in both core and belt regions of the auditory cortex, both single-neuron[44,75,76] and functional imaging[64,77] studies have shown that neurons are sensitive to the timbre of a sound. However, this neural representation of timbre is not invariant, as neural sensitivity to timbre is modulated by other sound features, such as pitch or spatial location[78]. Despite this, neural activity might represent different stimulus features unambiguously at different time points: when responding to a stimulus, single-unit spiking activity is initially tuned for the sound's timbre but later becomes tuned for its pitch[79]. The core auditory cortex might thus contain an 'implicit' representation of both an object and identity-preserving transformations of the object (such as changes in location or loudness) in a manner that may be analogous to the different types of visual representation contained in visual area V4 (REF. 28).

However, an explicit or invariant representation of timbre does seem to emerge in later stages of processing, at least in humans. For example, neural responses to vowel sounds represent stimulus acoustics at the level of the brainstem but represent perceptual mappings at the level of the cortex[80]. Functional imaging studies indicate that neurons in the planum temporale encode an invariant representation of a sound's spectral envelope, one of the key determinants of timbre[64]. Indeed, dynamic causal modelling has directly identified a serial-processing architecture in which timbre information originates in Heschl's gyrus, is transmitted to the planum temporale and then to the superior temporal gyrus; according to this model, spectral envelope extraction is complete by the time the information reaches the planum temporale[64]. Such a hierarchical-processing scheme might underpin a representation of sound timbre that allows us to perceptually recognize and identify a music instrument as a bassoon or a violin across different pitches and melodies.

In summary, although single neurons in the early core and belt auditory cortex of non-human animals show broad sensitivity to a number of perceptual features, there is good evidence for specialized processing of some of these features within particular areas. Whether these areas form a linear, hierarchical processing stream or a more dynamic, distributed assembly remains a matter of debate. To advance our understanding of the mechanisms underlying timbre perception, it may prove beneficial to carry out single-unit recording studies to test predictions derived from computational modelling techniques[64].

Spectral envelope
This term refers to the distribution of power across frequency in a sound. For a harmonic sound, this equates to the relative power across harmonics.

Dynamic causal modelling
A computational approach that performs Bayesian model comparisons in order to infer the organizational structure of processing within different brain regions.

© 2013 Macmillan Publishers Limited. All rights reserved

From stimulus to perception

Studies in behaving animals offer the potential to observe neural correlates of perception, as indexed by changes in neural activity as a function of an animal's behavioural choice during a listening task. That is, by holding a stimulus constant and testing whether neural activity is modulated by the animal's behavioural responses or choices (such as an animal indicating whether a target pitch is perceived as higher or lower than a reference pitch), neural activity that is associated with the stimulus itself can be dissociated from neural activity associated with the sensory decision. Choice-related activity (that is, activity that represents the animal's behavioural choice rather than the stimulus)[81] is thought to arise owing to correlations in the noise structure of neurons contributing to a sensory decision[82]. By examining how choice-related activity and other behaviourally related signals are modulated in different cortical areas, we can gain insight into how the nervous system transforms a sensory signal into a decision variable[81,83,84].

Recent investigations in behaving primate and non-primate species have found that neural activity is significantly correlated with a listener's behavioural reports[65,85]. For example, in core and non-core regions of the auditory cortex, local-field potentials and spiking activity are modulated more by ferrets' decisions regarding the pitch of a target sound than by the actual pitch category[65]. Similarly, in macaque monkeys, single- and multiunit recordings during an amplitude-modulation detection task reveal that activity in neurons in the primary auditory cortex is, once again, correlated with an animal's behavioural reports[85]. Last, blood-oxygen-level-dependent (BOLD) signals measured in early belt regions (areas adjacent to Heschl's gyrus and in the planum temporale) with fMRI can be decoded to predict a human listener's percept of an ambiguous speech sound[86]. These findings suggest that a population of core auditory cortical neurons contribute to or reflect the computations that underlie perceptual decision-making.

However, not all studies have found choice-related activity in the core auditory cortex[87–91]. For example, in an auditory flutter experiment, choice-related activity was not found in the auditory cortex but appeared in the ventral premotor cortex[90,91]. Similarly, in macaques that were discriminating between two phonemes and morphs of these phonemes, choice-related activity was not present in the auditory cortex but was found in the ventrolateral prefrontal cortex[87–89].

It is not clear why some studies have found choice-related activity in the primary auditory cortex, whereas others have only found such activity in more anterior areas. One important consideration might be the task itself. For example, whether an animal is engaged in a single- or multiple-interval forced-choice task, the task design or the animal's strategy to solve the task might determine the location of choice-related activity: a brain area that encodes the stimulus in a multiple-interval choice task is also unlikely to perform the comparison of the two stimuli[87,90]. In such a task, choice-related activity would first be observed in more anterior processing areas, such as the ventrolateral prefrontal cortex or the premotor cortex[87,90]. By contrast, when a task can be solved on the basis of listening during a single interval, that interval could also code a sensory decision. Therefore, differences in the level of abstraction required by the animals might determine whether choice-related activity is observed within the auditory cortex: a categorical 'same' versus 'different' task[88,89] necessitates a higher level of abstraction than does a high or low pitch judgement[65] or detection of a particular stimulus feature (such as modulation[85] or frequency change[92]).

Nevertheless, the finding of such signals in any brain region does not indicate that a particular cortical area is a locus for decision-making. A decision outcome is thought to require the accumulation of sensory evidence into a decision variable[93]. It seems likely that the neural correlates of perception that are observed in the early auditory cortex represent the sensory evidence that is needed to form a perceptual decision, which is then fed forward to other areas of the ventral pathway. Alternatively, this choice-related activity could reflect feedback signals from higher areas[82,94]. Finally, the time when choice-related activity appears during the temporal evolution of a task is an important consideration. For example, if choice-related activity appears before the stimulus that forms the basis of the animal's decision (such as the second stimulus in a paradigm requiring an animal to compare two sequentially presented sounds), this activity should be considered to be reflective of the listener's bias in making one alternative (choice) more favourable than the other[65,85,95].

To identify the neural mechanisms underlying auditory decision-making, scientists must systematically study changes in neural representations throughout a circuit of cortical areas to determine whether such signals reflect sensory evidence or a true decision variable. Such work has proven to be fruitful in the visual and somatosensory systems[83,96] but has yet to be applied broadly to the auditory system[38,87,88]. Additionally, formal computational models of perceptual decision-making that incorporate psychophysical and neurophysiological predictions need to be introduced into auditory studies[83].

Grouping features into objects

As described above, evidence suggests that the transformation from sound-source acoustics into perceptual features such as pitch and timbre, which are used to describe an object, occurs in the early auditory cortex, where, in some instances, neural activity correlates with an animal's behavioural report. It is worth repeating that these perceptual features are components of an auditory object rather than the object themselves. For example, a cat's meow has a higher pitch when someone stands on its tail than when the cat wants to be fed. Other studies have focused on how and where features are bound together to allow extraction of auditory objects.

Auditory scientists test where and how objects are extracted by analysing how the sequential and simultaneous grouping principles (BOXES 1,2) that bind perceptual features into a unified auditory object are represented in the cortex. For example, in one set of studies, fMRI data

Auditory flutter
The sensation produced by a periodic stimulus in which a listener can hear the sound as being intermittent. At higher frequencies, the sound is fused into one with a continuous melodic pitch. The border between being heard as intermittent or continuous is the flicker–fusion limit.

© 2013 Macmillan Publishers Limited. All rights reserved

were recorded while human listeners judged whether a target sound was continuous or discontinuous[97,98] (the illusion that a discontinuous sound is continuous is called amodal completion; see BOX 2). These studies found that physically identical acoustic stimuli elicited different BOLD signals in the primary auditory cortex depending on whether a listener reported a continuous or a discontinuous percept. The fact that listeners did not report a discontinuous percept suggests that, in this case, the auditory object itself, rather than the low-level spectrotemporal details, determined the listener's percept. Consistent with the idea that central brain regions are responsible for this illusion[99], computational simulations predict that cortical activity should correlate with the identity of the object and not its spectrotemporal components. Finally, single-neuron correlates of amodal completion have been found in the primary auditory cortex of rhesus macaques[22]. However, because behavioural reports and neural data were not gathered simultaneously, it is not clear whether this activity was related to the primitive grouping principles that are needed to form an auditory object or to the object itself.

The 'ABA streaming' paradigm is commonly used to test sequential grouping. In this paradigm, two interleaved sequences of tone bursts at two different frequencies (frequency A and B) are presented to a listener. At slow rates, a listener is more likely to hear a single stream of alternating tones (FIG. 3a). When the semitone separation between frequency A and B is small (0.5 semitones), listeners are likely to report hearing one auditory stream (FIG. 3b). When this separation is large (>10 semitones), listeners reliably report hearing two auditory streams. At intermediate semitone separations, listeners hear one or two auditory streams on alternate trials. This type of stimulus is called a 'bistable' stimulus because the listener's perceptual report may alternate between the two possibilities; therefore, neural activity related to the perceptual report can be disassociated from neural activity related to the stimulus. These auditory bistable stimuli might be analogous to visual bistable stimuli[84,100,101]. The tone-burst duration, listening duration, repetition rate and other factors can also modulate a listener's reports[102].

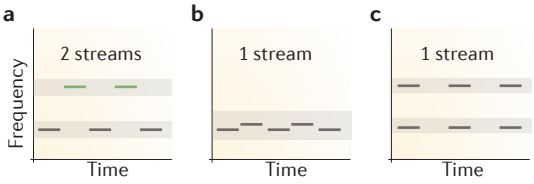


Figure 3 | **Auditory streaming.** In a classic paradigm of auditory streaming, two sequences of tone bursts are presented in an alternating fashion[11,109]. **a** | When the frequency separation between the tone bursts in the two sequences is large, listeners typically hear two streams. **b** | By contrast, when the frequency separation between the two sequences is small, listeners typically report hearing one stream. However, at intermediate frequency separations, the listener's report is bistable over time: they alternate between perceiving one or two streams (not shown). With longer listening times, this report stabilizes and listeners reliably report two streams. **c** | In addition to parameters such as listening duration and other parameter manipulations[168], the temporal relationship between the two sequences is critical. When the two sequences are presented concurrently, listeners consistently report hearing one stream. This observation suggests that the temporal coherence between different neural populations is the critical mechanism for the determination of whether a listener hears one or two streams. See REFS 104,169 for more details on the role that temporal coherence has in auditory streaming. Figure is modified, with permission, from REF. 169 © (2011) Elsevier.

What neural computations underlie a listener's perception of one or two auditory streams? Correlates of the grouping principles thought to underlie ABA streaming can be observed as early as the cochlear nucleus[103]. One reasonable hypothesis is that neurons downstream from the core auditory cortex, such as those in the belt cortex or even the frontal and parietal lobes[54,104–109], read out the topographic distribution of activity in the core auditory cortex. That is, if the semitone separation is small, there would be one peak of activity, which downstream neurons — as a proxy for a listener's behavioural reports — would decode as one stream. By contrast, if the semitone separation was large, there would be two peaks of activity, which would be decoded as two streams. At intermediate separations, the number of peaks would be unclear and trial-by-trial neural noise would alternate the readout between one and two peaks of activity. Importantly, however, temporal parameters also influence both listeners' reports and neural activity. For example, when the intervals between tones are short, listeners are more likely to report hearing one stream. The mechanism of this bias, which is likely to be partly inherited from earlier parts of the processing pathway[103], might be forward masking, which would 'eliminate' or minimize the second peak of activity[104,110]. However, as streaming can occur in response to various sounds, including noises and harmonic sounds, that would elicit overlapping spectral representations, this topographic readout explanation is probably too simplistic.

Indeed, recent work has proven that a topographic readout is insufficient to explain auditory streaming, at least in the ABA paradigm. If spatially segregated populations of neurons are necessary for streaming to occur, then the relative timing of tone A and tone B should be inconsequential because the only factor that would be important is the topographic representation of neural activity in the auditory cortex. In an elegant series of experiments, this hypothesis was explored by testing how the timing of tone A and tone B affected a listener's behavioural reports. These authors found that, independent of semitone separation, when tone A and tone B were presented simultaneously, listeners reliably reported one stream[111] (FIG. 3c). Thus, the relative timing of these peaks of activity is critical: when the two peaks are in phase, listeners report one stream but when they are out of phase, they are reported as two streams. This neural mechanism of temporal synchrony might also be involved in grouping of other cues such as harmonic stimuli and stimulus onset and offset. A strict interpretation of the temporal coherence model has itself recently

Forward masking
A process by which a sound is obscured by a masker (for example, a noise burst) that precedes the sound.

© 2013 Macmillan Publishers Limited. All rights reserved

been challenged by the finding that although temporal coherence is an important factor in the formation of perceptual streams, temporally coherent sounds can be streamed[112]. Unfortunately, the specific neural readout mechanisms that are sensitive to such timing information are not known. Future work in which large groups of neurons are recorded simultaneously while temporal synchrony is parametrically alternated are essential for addressing this question.

Whereas single-neuron recording studies in the cochlear nucleus indicate that, in principle, the information in activity patterns of neurons in the cochlear nucleus are sufficient to support streaming[103], evidence from the functional imaging literature suggests that the perception of streaming occurs in or beyond the auditory cortex[113]. Unfortunately, despite the apparent elegance and simplicity of the ABA-stimulus paradigm, the role of different cortical areas in this streaming percept has been difficult to resolve. However, whereas the auditory cortex seems to be important for constructing the stream and the perceptual organization of the auditory scene, activity in regions in the frontal and parietal lobes appear to be correlated with a listener's reports[54,104–109].

Key to the grouping principles underlying both streaming paradigms and amodal completion is the idea of predictability: the auditory system must generate some sort of prediction from current and previously present sounds to build a model of what is likely to occur next[12]. Neural activity in early auditory areas seems to represent the prediction of a regular sequence of sounds: if a sound is omitted from a fully predictable sequence of sounds, auditory cortex activity will respond to this omission as if the sound was actually presented[23]. Activity that precedes this omission-related response arises from sources within and beyond the primary auditory cortex and is thought to be the best candidate for a signal that represents a violation of ongoing predictions[12].

Assigning objects to categories

Neural correlates of categorical perception have been found in both the core and belt regions of the auditory cortex. For example, in one study[92], monkeys participated in a task in which the correct response depended on whether the frequency of a series of tone bursts was increasing or decreasing independent of the start and end frequencies. This revealed two classes of cells in the core and early belt auditory cortex (specifically, area A1 and the caudomedial belt region of the auditory cortex): the first showed phasic responses that discriminated between the two categories (increasing versus decreasing), whereas the second class showed tonic firing that, at the population level, correlated with the monkey's behavioural response.

Similarly, in another study[88], monkeys made a 'same or different' judgement based on the sequential presentation of two speech sounds ('dad' versus 'bad') or a series of morphed versions of these sounds (FIG. 4). The behavioural data showed that monkeys perceived these morphed stimuli categorically; that is, despite the fact that the acoustic stimulus varied smoothly, the monkeys consistently assigned the morphs to one of the two categories, with a sharp transition between morphed sounds being perceived as 'dad' rather than 'bad'. Neurons in the belt region of the auditory cortex likewise responded in a categorical fashion. Interestingly, the degree of neural categorization depended on the type of recorded neuron: fast-spiking neurons (putative interneurons) responded more categorically. That is, they showed greater invariance across morphs that were categorized behaviourally to be the same than did slow-spiking neurons (putative pyramidal neurons)[114].

Studies using fMRI indicate that there are categorical representations of speech sounds in both the posterior and anterior auditory cortex[36,42,115,116]. Fewer studies have investigated category selectivity with non-speech stimuli. These studies are important because they allow researchers to investigate more abstract categories that are not based on similarities between stimulus features. For example, category specificity for musical and human-speech sounds is found in the anterior superior temporal cortex[117]. By contrast, no such specificity is seen for songbird or 'other animal' vocalizations, although this might be because vocalization-specific clusters are inter-digitated among other category-sensitive regions or are simply so small that they cannot be resolved by fMRI. An alternative interpretation is that object recognition might not require segregated, category-specific cortical subregions to represent different classes of objects.

However, another recent study suggests that the anterior areas might not be uniquely specialized for auditory-category information[118]. This study used a heterogeneous set of natural sounds to explore the representation of stimulus categories for non-speech stimuli. The authors carried out a variance decomposition analysis that enabled them to differentiate variability due to low-level stimulus features from variability due to category specificity. Consistent with results from studies of animals[78], large areas of the human cortex were sensitive to low-level stimulus features. In addition, posterior areas of the auditory cortex (such as the planum temporale) can encode the abstract categories of living sounds and human sounds[118]. Such findings suggest that there might be an increase in information abstraction as the cortical hierarchy ascends from the primary cortex in both anterior and posterior directions[64]. In support of this notion, category representation for pitch-matched stimuli was seen in the anterolateral Heschl's gyrus, the planum temporale and the posterior superior temporal gyrus. Areas showing category specificity and specificity for acoustic information (in this case, pitch contrast) overlapped and included areas of both the lower and higher auditory cortex[67].

This abstraction of categorization continues beyond the auditory cortex and into the prefrontal cortex regions of the ventral auditory pathway. For example, neurons in the rhesus prefrontal cortex do not differentiate between vocalizations that transmit the same type of information despite the fact that these vocalizations have different acoustic features. That is, these neurons code the 'meaning' of vocalizations[119] (FIG. 4).

Categorical perception
The experience of perceiving a stimulus as being the same (that is, invariant) despite the fact that the physical properties of the stimulus have changed smoothly along a specific axis or continuum. A characteristic of categorical perception is that for a continuously changing stimulus dimension, subjects generalize across changes, with a sharp change in the perception from one class to another at the position of the boundary of the stimulus identity.

© 2013 Macmillan Publishers Limited. All rights reserved

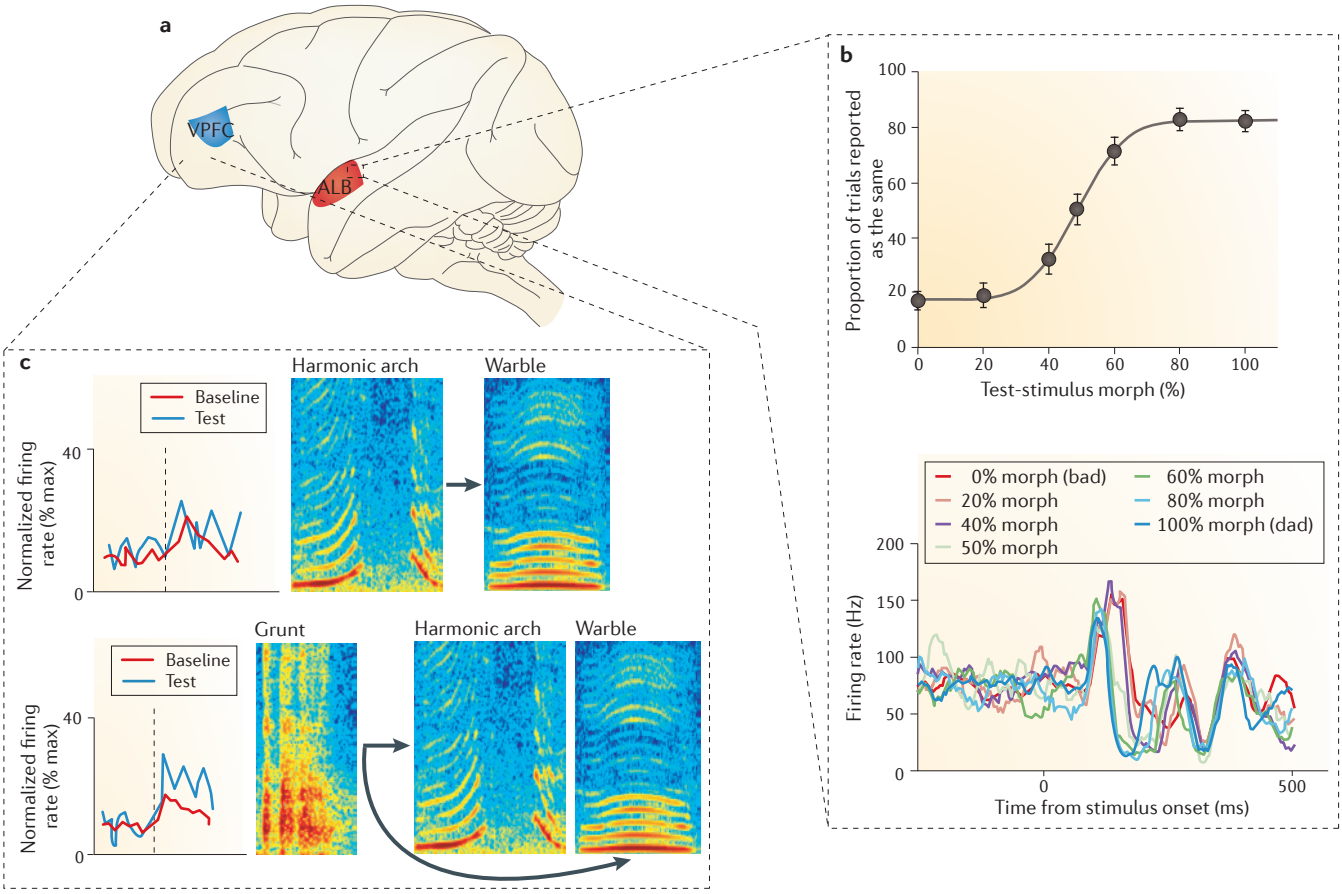


Figure 4 | **Categorization in the ventral auditory pathway.** **a** | The involvement of two key regions of the ventral auditory pathway, the anterolateral belt (ALB) and the ventral prefrontal cortex (VPFC), in assigning auditory objects to categories has been demonstrated in a series of experiments. **b** | In the experiment illustrated, monkeys participated in a task that required them to discriminate between a reference stimulus and a test stimulus. The reference sound was 'dad', a different sound, 'bad', or an acoustic morph of these two sounds. The 0% stimulus is the sound 'bad', and the 100% stimulus is the sound 'dad'. Intermediate morph values have proportional values of the two stimuli; for example, an 80% morph has 80% of the acoustic features of 'bad' and 20% of 'dad'. Data were reported in terms of the proportion of trials in which the monkeys reported that the reference and test stimuli were the same (upper panel). As can be seen, the monkeys' behavioural reports are categorical. They treat sounds less than 50% morph stimuli as one category and those greater than 50% morph stimuli as a second category. Similarly, when recording ALB neurons during such categorization, neural activity also responds in a categorical fashion (lower panel). That is, ALB neurons respond similarly to all less than 50% morph stimuli and respond in a different manner to greater than 50% morph stimuli. **c** | In rhesus monkeys, VPFC neurons encode the membership of a particular type of call in response to food to an abstract category. The two categories are calls that transmit information regarding low-food quality (a grunt) and calls that transmit information about high-quality food (a harmonic arch or a warble). Population VPFC activity is shown for a baseline condition and in response to a test vocalization. The presentation of the test vocalization (at the time indicated by the position of the dashed line) was preceded by repeated presentations of a different reference vocalization. Also shown are the spectrograms for the different types of vocalization. VPFC activity preferentially codes transitions between food calls that belong to different abstract categories independently of differences between acoustics of the vocalizations (lower panels). By contrast, VPFC neurons do not code transitions between acoustically distinct stimuli that transmit the same information (upper panels). Part **b** (upper panel) is modified, with permission, from REF. 88 © (2011) The American Physiological Society. Part **b** (lower panel) is modified, with permission, from REF. 114 © (2012) The Physiological Society. Part **c** is modified, with permission, from REF. 119 © (2005) MIT Press Journals.

How does learning shape neural category representation? In one study[120], gerbils were trained to categorize frequency-modulated tones as 'upwards' or 'downwards' regardless of the starting frequency, the ending frequency or the rate of the frequency modulation. During the task, epidural-evoked potentials were recorded from multiple sites over the auditory cortex. An analysis of these recordings demonstrated that over time, as the gerbils acquired the categorization rule, the neural activity patterns changed. Initially, neural activity reflected the acoustical properties of the frequency-modulated tones. After learning, neural activity reflected the categorical membership of the frequency-modulated tones independently of their properties. This transformation of information

© 2013 Macmillan Publishers Limited. All rights reserved

representation might be mediated through feedback projections between the prefrontal cortex and auditory cortex that modulate task-relevant information[121].

Neural computations underlying object recognition are thought to require selectivity for object-specific features, invariance across identity-preserving changes and generalization to enable categorization[29]. Whereas studies looking at hierarchical processing in the auditory system have sought increasing levels of selectivity and, as discussed above, some studies have looked for category-specific neural firing, the question of invariance remains underexplored[122].

This crucial, but unresolved, question in auditory neuroscience is particularly pertinent to our understanding of how auditory objects are formed: to perform scene analysis, we must be able to generalize across identity-preserving changes. The continuity illusion discussed earlier can be seen as a very basic form of invariance, but the ability to generalize across multiple stimulus dimensions in order to assign a particular acoustic event to the right auditory object is more computationally challenging. This task requires selectivity for certain stimulus parameters, a tolerance for differences in other parameters and ultimately the ability to generalize across features to assign a sound to a more general category, or class, of sounds[18].

There are two contrasting models of how neurons might represent the identity of an object (FIG. 5). Distributed-coding models postulate that ensembles of neurons represent object identity. By contrast, sparse-coding models suggest that only a small number of neurons are activated by a given stimulus, so that these neurons explicitly represent the to-be-identified object[29]. Although sparse codes are energetically efficient and easy to read out, taken to extremes such a theory would predict the existence of grandmother cells, which would require an intractable number of neurons to represent all possible objects. Experimental evidence from the visual system also suggests that the increasing selectivity that one would expect to see at each hierarchical stage in a sparse-coding model is not observed and that accurate object identification is apparently achieved through a population code[123].

To formally understand the mechanisms underlying auditory-object formation and recognition, as has been done for the visual system[18], we need to develop computational models to generate testable hypotheses as to how population activity in higher auditory areas creates explicit, implicit and tolerant representations of auditory objects. However, to date, such models have not been identified for the auditory system, and this remains an important issue in auditory neuroscience.

The role of attention in object perception

Simultaneous grouping principles and their neural correlates, such as object-related negativity, can operate independently of the listener's attentional state[124]. Attention is not required for a person to detect changes in a stimulus feature. For example, oddball paradigms, in which a rare (deviant) sound is interposed into a stream of repeating standard sounds, show that deviance-detection mechanisms operate automatically and do not require a subject to overtly attend to the stimulus[125–128]. Other studies indicate that the continuity illusion does not require attention[26]. Together, these findings support the idea that the auditory cortex automatically generates and monitors predictions about the current sound-scene[12].

However, whether auditory streaming requires attention is a more controversial matter. Whereas attention is not always required for streams to form[129], attention can heavily influence a listener's perception, and switching attention 'resets' streaming[130]. It seems likely that attention is required to resolve or select representations in an ambiguous auditory scene. Compatible with the concept of a two-stage process is the finding that when listeners are presented with ABA tone sequences, two distinct event-related potential (ERP) components are evoked with different latencies[131–133]. The first component is thought to be the initial representation of two alternative interpretations of the sound (one stream versus two streams), whereas the later component reflects the listener's decision (one stream)[131]. In natural listening

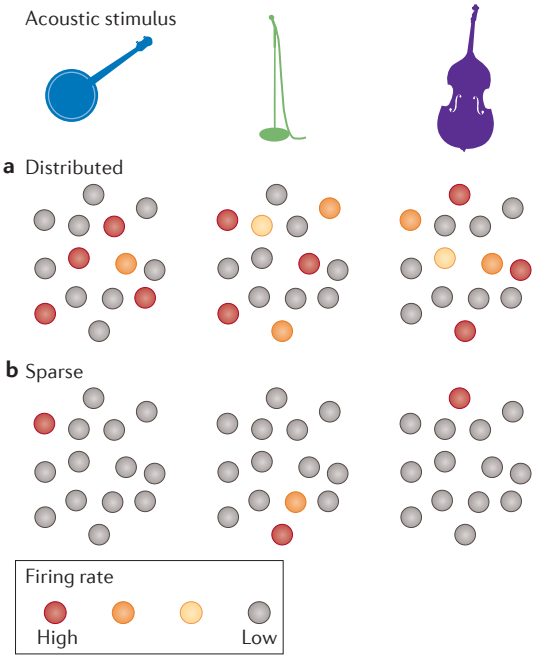


Figure 5 | **Strategies for coding auditory object identity.** Two neural coding strategies might hypothetically underlie how information is represented in a cortical field: distributed coding or sparse coding. **a** | Information about the nature of an auditory object (in this case the identity of a musical instrument in a situation in which all three instruments play the same note) could be represented by the pattern of activity across the neural ensemble. Here, each sound category elicits activity in many neurons, with any individual neuron potentially increasing its firing rate to multiple sound categories. Nevertheless, each sound category elicits a unique pattern of activity across the network. **b** | By contrast, in a sparse representation, each neuron in the array is tuned to a single sound category such that each musical instrument elicits activity from only a very small number of neurons.

Scene analysis
The process by which the brain organizes and segregates acoustic stimuli into meaningful elements or objects.

Grandmother cells
Hypothetical cells that represent a very specific complex object or concept — such as one's grandmother.

Object-related negativity
An evoked-potential component that is elicited when two concurrently presented sounds are perceived as originating from different sources based on simultaneous grouping cues.

© 2013 Macmillan Publishers Limited. All rights reserved

conditions, when there are almost always multiple competing sources, auditory-scene analysis is likely to be heavily influenced by attention and the behavioural goals of the listener[14].

Once an auditory scene has been parsed into its component objects, selective attention can operate on these components to facilitate further processing and resolve competition between multiple sources[134,135]. Attention operates at the level of objects[17,136,137], and even when attention is focused on a low-level stimulus feature (such as the pitch of someone's voice), there is enhanced sensitivity to other features of that source (such as its location)[138]. Failures of object formation impair the ability to analyse a sound source[139–141], and attention itself influences perception of the auditory scene[142]. Selective attention to a particular object in the visual scene is thought to be essential as the brain has limited resources. As a result of these limited resources, there is a biased competition between objects[136,143]. As in vision, both bottom-up and top-down cues can direct auditory attention to a particular object[135,144], and thus one of the hallmarks of an 'object-based' neural representation is that it is modulated by behavioural demands. Indeed, highly skilled listeners have enhanced neural-processing mechanisms for particular object-based listening tasks. For example, regions in the left anterior superior temporal gyrus are modulated by a listener's expertise in perceiving and producing a given sound class: actors have greater neural activation in response to speech compared to music, whereas violinists have the opposite pattern[145].

Attentional signals are found throughout the auditory cortex. In the early auditory cortex, attention can modify the tuning properties of neurons in the primary auditory cortex[146–149] and can increase the magnitude of ERPs and fMRI signals[150–155]. In later parts of the auditory cortex, such as the posterior auditory cortex, which roughly corresponds to the planum temporale, neural signals reflect the listener's perception of a particular auditory object[156,157]. For example, when a listener is asked to attend to one of two spectrotemporally overlapping speech signals, the attended signal preferentially modulates neural activity in this region of the auditory cortex[156]. Similarly, in experiments conducted using surface electrodes in human patients, neural responses to irrelevant sounds are suppressed relative to those that are attended[157].

Attention is not mediated by a simple feedforward network. Instead, attention is mediated by a complex network that has distinct activity patterns for spatial versus non-spatial auditory attention[39,40]. Differential activity patterns have been found in auditory regions of the superior temporal gyrus[137,158–160] as well as the superior temporal sulcus and the inferior parietal sulcus; these latter regions exhibit more attention-related modulation when listeners are asked to attend to a sound that is embedded within a complex and realistic listening environment[39]. It seems likely that these networks may provide feedback activity to early sensory areas, enabling the selection of activity related to the object of interest[161].

Synthesis and discussion

We have discussed and reviewed how the auditory system represents the perceptual features and grouping principles that underlie the creation of auditory objects. We have also highlighted several important principles, such as the hierarchical processing of information and the role of the ventral stream in auditory-object processing. However, we believe that two fundamental issues remain to be investigated. First, beyond the 'classical' auditory cortex, a network of areas subserves the functions associated with processing auditory objects. For example, neural activity in the prefrontal cortex[162–164] and hippocampus[165] interacts with auditory cortex activity to process auditory memory and the meaning and emotional content of sounds. We do not fully understand the roles of these brain regions in auditory cognition, or the neural mechanisms that underlie these roles. Second, it is unclear which cortical areas have causal roles in auditory-object processing and perception. Thus, to drive our understanding of how the auditory cortex parses the auditory scene into recognizable objects, researchers must exploit techniques that enable perception and neural activity to be studied simultaneously in combination with methods that perturb neural activity to provide causal evidence for the contribution of particular brain areas to defined functions, and design computational models that generate testable hypotheses.

1. Griffiths, T. D. & Warren, J. D. What is an auditory object? *Nature Rev. Neurosci.* **5**, 887–892 (2004).
2. Rauschecker, J. P. Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol. Suppl.* **532**, 34–38 (1997).
3. Kaas, J. H. & Hackett, T. A. Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl Acad. Sci. USA* **97**, 11793–11799 (2000).
4. Romanski, L. M. *et al.* Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neurosci.* **2**, 1131–1136 (1999).
5. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11800–11806 (2000).
6. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neurosci.* **12**, 718–724 (2009).
7. Recanzone, G. H. & Cohen, Y. E. Serial and parallel processing in the primate auditory cortex revisited. *Behav. Brain Res.* **206**, 1–7 (2010).
8. Sharpee, T. O., Atencio, C. A. & Schreiner, C. E. Hierarchical representations in the auditory cortex. *Curr. Opin. Neurobiol.* **21**, 761–767 (2011).
9. Bendor, D. & Wang, X. Cortical representations of pitch in monkeys and humans. *Curr. Opin. Neurobiol.* **16**, 391–399 (2006).
10. Fishman, Y. I. & Steinschneider, M. in *The Oxford Handbook of Auditory Science: the Auditory Brain* (ed. Rees, A.) 215–245 (Oxford Univ. Press, 2010).
11. Bregman, A. S. *Auditory Scene Analysis* (MIT Press, 1990).
12. Winkler, I., Denham, S. L. & Nelken, I. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* **13**, 532–540 (2009).
13. Kubovy, M. & Van Valkenburg, D. Auditory and visual objects. *Cognition* **80**, 97–126 (2001).
14. Shinn-Cunningham, B. G. Object-based auditory and visual attention. *Trends Cogn. Sci.* **12**, 182–186 (2008).
15. Schnupp, J. W., Nelken, I. & King, A. J. *Auditory Neuroscience: Making Sense of Sound* (MIT Press, 2012).
16. Miller, C. T. & Cohen, Y. E. in *Primate Neuroethology* (eds Ghazanfar, A. & Platt, M. L.) 237–255 (Oxford Univ. Press, 2010).
17. Alain, C. & Arnott, S. R. Selectively attending to auditory objects. *Front. Biosci.* **5**, D202–D212 (2000).
18. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
19. Ding, N. & Simon, J. Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl Acad. Sci. USA* **109**, 11854–11859 (2012).
20. Reddy, L. & Kanwisher, N. Coding of visual objects in the ventral stream. *Curr. Opin. Neurobiol.* **16**, 408–414 (2006).

© 2013 Macmillan Publishers Limited. All rights reserved

21. Miller, C. T., Dibble, E. & Hauser, M. D. Amodal completion of acoustic signals by a nonhuman primate. *Nature Neurosci.* **4**, 783–784 (2001).
22. Petkov, C. I., O'Connor, K. N. & Sutter, M. L. Encoding of illusory continuity in primary auditory cortex. *Neuron* **54**, 153–165 (2007).
23. Bendixen, A., Schroger, E. & Winkler, I. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *J. Neurosci.* **29**, 8447–8451 (2009).
The authors propose a key role for the auditory cortex in the generation of predictions about sequences of ongoing sounds. ERP recordings demonstrate that the neural response to a predictable but omitted sound looks very similar to the neural response to the tone when actually present.
24. Shinn-Cunningham, B. G. & Wang, D. Influences of auditory object formation on phonemic restoration. *J. Acoust. Soc. Am.* **123**, 295–301 (2008).
25. Warren, R. M., Obusek, C. J. & Ackroff, J. M. Auditory induction: perceptual synthesis of absent sounds. *Science* **176**, 1149–1151 (1972).
26. Micheyl, C. *et al.* The neurophysiological basis of the auditory continuity illusion: a mismatch negativity study. *J. Cogn. Neurosci.* **15**, 747–758 (2003).
27. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J.) 549–586 (MIT Press, 1982).
28. Rust, N. C. & Stocker, A. A. Ambiguity and invariance: two fundamental challenges for visual processing. *Curr. Opin. Neurobiol.* **20**, 382–388 (2010).
29. Ison, M. J. & Quiroga, R. Q. Selectivity and invariance for visual object perception. *Front. Biosci.* **13**, 4889–4903 (2008).
30. Riesenhuber, M. & Poggio, T. Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.* **12**, 162–168 (2002).
31. Riesenhuber, M. & Poggio, T. Models of object recognition. *Nature Neurosci.* **3**, 1199–1204 (2000).
32. Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J. P. Functional specialization in rhesus monkey auditory cortex. *Science* **292**, 290–293 (2001).
33. Alain, C., Arnott, S. R., Hevenor, S., Graham, S. & Grady, C. L. "What" and "where" in the human auditory system. *Proc. Natl Acad. Sci. USA* **98**, 12301–12306 (2001).
34. Maeder, P. P. *et al.* Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* **14**, 802–816 (2001).
35. Arnott, S. R., Binns, M. A., Grady, C. L. & Alain, C. Assessing the auditory dual-pathway model in humans. *Neuroimage* **22**, 401–408 (2004).
36. Obleser, J. *et al.* Vowel sound extraction in anterior superior temporal cortex. *Hum. Brain Mapp.* **27**, 562–571 (2006).
37. Chang, E. F. *et al.* Categorical speech representation in human superior temporal gyrus. *Nature Neurosci.* **13**, 1428–1432 (2010).
38. Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A. & Ward, B. D. Neural correlates of sensory and decision processes in auditory object identification. *Nature Neurosci.* **7**, 295–301 (2004).
The authors attempt to identify both sensory and decision-making activity in the human brain using fMRI. They demonstrate a functional distinction between sensory and decision mechanisms underlying auditory-object identification.
39. Hill, K. T. & Miller, L. M. Auditory attentional control and selection during cocktail party listening. *Cereb. Cortex* **20**, 583–590 (2010).
40. Lee, A. K. *et al.* Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Front. Neurosci.* **6**, 190 (2012).
The authors combined magnetoencephalography recordings and structural MRI data to map the attentional networks involved in selectively attending to either spatial or non-spatial features of a sound. Left frontal eye fields were activated by spatial attention, whereas lateral posterior superior temporal sulcus was activated by attention to pitch.
41. Cohen, Y. E. *et al.* A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. *Proc. Natl Acad. Sci. USA* **106**, 20045–20050 (2009).
42. Obleser, J. & Eisner, F. Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* **13**, 14–19 (2009).
43. Rauschecker, J. P. Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci.* **4**, 7 (2012).
44. Bizley, J. K., Walker, K. M., Silverman, B. W., King, A. J. & Schnupp, J. W. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J. Neurosci.* **29**, 2064–2075 (2009).
45. Miller, L. M. & Recanzone, G. H. Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc. Natl Acad. Sci. USA* **106**, 5931–5935 (2009).
The authors measured neural responses to sounds that varied in spatial location and used optimal decoding strategies to assess whether neural responses could support behavioural localization abilities. Although neural populations throughout the auditory cortex contained spatial information in their responses, only those in the caudolateral field had sufficient information to account for behaviour.
46. Stecker, G. C. & Middlebrooks, J. C. Distributed coding of sound locations in the auditory cortex. *Biol. Cybern.* **89**, 341–349 (2003).
47. Harrington, I. A., Stecker, G. C., Macpherson, E. A. & Middlebrooks, J. C. Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. *Hear. Res.* **240**, 22–41 (2008).
48. Cloutman, L. L. Interaction between dorsal and ventral processing streams: where, when and how? *Brain Lang.* http://dx.doi.org/10.1016/j.bandl.2012.08.003 (2012).
49. Middlebrooks, J. C. & Onsan, Z. A. Stream segregation with high spatial acuity. *J. Acoust. Soc. Am.* **132**, 3896–3911 (2012).
50. Middlebrooks, J. C. & Bremen, P. Spatial stream segregation by auditory cortical neurons. *J. Neurosci.* **33**, 10986–11001 (2013).
51. Rauschecker, J. P. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* **271**, 16–25 (2011).
52. Teki, S., Chait, M., Kumar, S., von Kriegstein, K. & Griffiths, T. D. Brain bases for auditory stimulus-driven figure–ground segregation. *J. Neurosci.* **31**, 164–171 (2011).
53. Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R. & Rauschecker, J. P. Brain activation during anticipation of sound sequences. *J. Neurosci.* **29**, 2477–2485 (2009).
54. Cusack, R. The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* **17**, 641–651 (2005).
55. Rao, S. C., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824 (1997).
56. Bendor, D. & Wang, X. The neuronal representation of pitch in primate auditory cortex. *Nature* **436**, 1161–1165 (2005).
The authors demonstrate that a subset of neurons — specifically in the low-frequency border of area A1 and the rostral field in the marmoset — respond to sounds with a fundamental frequency that matches their characteristic frequency regardless of whether the fundamental frequency is present or not.
57. Lee, C. C. & Middlebrooks, J. C. Specialization for sound localization in fields A1, DZ, and PAF of cat auditory cortex. *J. Associ. Res. Otolaryngol.* **14**, 61–82 (2013).
58. Camalier, C. R., D'Angelo, W. R., Sterbing-D'Angelo, S. J., de la Mothe, L. A. & Hackett, T. A. Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proc. Natl Acad. Sci. USA* **109**, 18168–18173 (2012).
59. Grimsley, J. M., Shanbhag, S. J., Palmer, A. R. & Wallace, M. N. Processing of communication calls in guinea pig auditory cortex. *PloS ONE* **7**, e51646 (2012).
60. Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. & Griffiths, T. D. The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767–776 (2002).
The authors present evidence for the hierarchical processing of pitch by performing fMRI on human listeners using sounds that are matched in spectral content but that either did or did not evoke a pitch percept.
61. Penagos, H., Melcher, J. R. & Oxenham, A. J. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* **24**, 6810–6815 (2004).
62. Warren, J. D. & Griffiths, T. D. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *J. Neurosci.* **23**, 5799–5804 (2003).
63. Garcia, D., Hall, D. A. & Plack, C. J. The effect of stimulus context on pitch representations in the human auditory cortex. *Neuroimage* **51**, 808–816 (2010).
64. Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J. & Griffiths, T. D. Hierarchical processing of auditory objects in humans. *PLoS Computat. Biol.* **3**, e100 (2007).
The authors present evidence for the hierarchical processing of spectral timbre in human listeners. The use of dynamic causal modelling techniques indicated that processing was both serial and hierarchical.
65. Bizley, J. K., Walker, K. M., Nodal, F. R., King, A. J. & Schnupp, J. W. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr. Biol.* **23**, 620–625 (2013).
The authors recorded neural responses in the auditory cortex of ferrets performing a pitch-direction discrimination task. Neural activity was modulated more by the ferrets' decision regarding the pitch of a target sound than by the actual pitch category.
66. Griffiths, T. D. *et al.* Direct recordings of pitch responses from human auditory cortex. *Curr. Biol.* **20**, 1128–1132 (2010).
67. Staeren, N., Renvall, H., De Martino, F., Goebel, R. & Formisano, E. Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.* **19**, 498–502 (2009).
68. Hall, D. A. & Plack, C. J. Pitch processing sites in the human auditory brain. *Cereb. Cortex* **19**, 576–585 (2009).
69. Bizley, J. K., Walker, K. M., King, A. J. & Schnupp, J. W. Neural ensemble codes for stimulus periodicity in auditory cortex. *J. Neurosci.* **30**, 5078–5091 (2010).
70. Griffiths, T. D. & Hall, D. A. Mapping pitch representation in neural ensembles with fMRI. *J. Neurosci.* **32**, 13343–13347 (2012).
71. Nelken, I. *et al.* Responses of auditory cortex to complex stimuli: functional organization revealed using intrinsic optical signals. *J. Neurophysiol.* **99**, 1928–1941 (2008).
72. Darwin, C. J. Auditory grouping. *Trends Cogn. Sci.* **1**, 327–333 (1997).
73. Hackett, T. A. Information flow in the auditory cortical network. *Hear. Res.* **271**, 133–146 (2011).
74. Dick, F. *et al. In vivo* functional and myeloarchitectonic mapping of human primary auditory areas. *J. Neurosci.* **32**, 16095–16105 (2012).
75. Schebesch, G., Lingner, A., Firzlaff, U., Wiegrebe, L. & Grothe, B. Perception and neural representation of size-variant human vowels in the Mongolian gerbil (*Meriones unguiculatus*). *Hear. Res.* **261**, 1–8 (2010).
76. Versnel, H. & Shamma, S. A. Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J. Acoust. Soc. Am.* **103**, 2502–2514 (1998).
77. Formisano, E., De Martino, F., Bonte, M. & Goebel, R. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* **322**, 970–973 (2008).
78. Bizley, J. K. & Walker, K. M. Distributed sensitivity to conspecific vocalizations and implications for the auditory dual stream hypothesis. *J. Neurosci.* **29**, 3011–3013 (2009).
79. Walker, K. M., Bizley, J. K., King, A. J. & Schnupp, J. W. Multiplexed and robust representations of sound features in auditory cortex. *J. Neurosci.* **31**, 14565–14576 (2011).
80. Bidelman, G. M., Moreno, S. & Alain, C. Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* **79**, 201–212 (2013).
81. Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
82. Nienborg, H., Cohen, M. R. & Cumming, B. G. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu. Rev. Neurosci.* **35**, 463–483 (2012).
83. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
84. Schall, J. D. & Bichot, N. P. Neural correlates of visual and motor decision processes. *Curr. Opin. Neurobiol.* **8**, 211–217 (1998).

© 2013 Macmillan Publishers Limited. All rights reserved

85. Niwa, M., Johnson, J. S., O'Connor, K. N. & Sutter, M. L. Activity related to perceptual judgment and action in primary auditory cortex. *J. Neurosci.* **32**, 3193–3210 (2012).
The authors recorded single- and multiunit activity in the auditory cortex of animals performing an auditory modulation detection task. In addition to acoustic information, neural activity was informative about both motor actions and the animals' behavioural choice.
86. Kilian-Hutten, N., Valente, G., Vroomen, J. & Formisano, E. Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J. Neurosci.* **31**, 1715–1720 (2011).
87. Russ, B. E., Orr, L. E. & Cohen, Y. E. Prefrontal neurons predict choices during an auditory same-different task. *Curr. Biol.* **18**, 1483–1488 (2008).
The authors recorded from neurons in the ventrolateral prefrontal cortex of monkeys performing a non-spatial same–different task. Neural activity predicted animals' behavioural choices, demonstrating a direct link between single neurons and behavioural choice.
88. Tsunada, J., Lee, J. H. & Cohen, Y. E. Representation of speech categories in the primate auditory cortex. *J. Neurophysiol.* **105**, 2634–2646 (2011).
89. Russ, B. E., Ackelson, A. L., Baker, A. E. & Cohen, Y. E. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J. Neurophysiol.* **99**, 87–95 (2008).
90. Lemus, L., Hernandez, A. & Romo, R. Neural encoding of auditory discrimination in ventral premotor cortex. *Proc. Natl Acad. Sci. USA* **106**, 14640–14645 (2009).
91. Lemus, L., Hernandez, A. & Romo, R. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *Proc. Natl Acad. Sci. USA* **106**, 9471–9476 (2009).
92. Selezneva, E., Scheich, H. & Brosch, M. Dual time scales for categorical decision making in auditory cortex. *Curr. Biol.* **16**, 2428–2433 (2006).
93. Gold, J. I. & Shadlen, M. N. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* **5**, 10–16 (2001).
94. Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J. & Desimone, R. Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl Acad. Sci. USA* **108**, 11262–11267 (2011).
95. Niwa, M., Johnson, J. S., O'Connor, K. N. & Sutter, M. L. Differences between primary auditory cortex and auditory belt related to encoding and choice for AM sounds. *J. Neurosci.* **33**, 8378–8395 (2013).
96. Romo, R. & Salinas, E. Sensing and deciding in the somatosensory system. *Curr. Opin. Neurobiol.* **9**, 487–493 (1999).
97. Riecke, L. *et al.* Hearing an illusory vowel in noise: suppression of auditory cortical activity. *J. Neurosci.* **32**, 8024–8034 (2012).
98. Riecke, L., Mendelsohn, D., Schreiner, C. & Formisano, E. The continuity illusion adapts to the auditory scene. *Hear. Res.* **247**, 71–77 (2009).
99. Riecke, L., Micheyl, C. & Oxenham, A. J. Global not local masker features govern the auditory continuity illusion. *J. Neurosci.* **32**, 4660–4664 (2012).
100. Pressnitzer, D., Suied, C. & Shamma, S. A. Auditory scene analysis: the sweet music of ambiguity. *Front. Hum. Neurosci.* **5**, 158 (2011).
101. Leopold, D. A. & Logothetis, N. K. Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* **3**, 254–264 (1999).
102. Shamma, S. A. & Micheyl, C. Behind the scenes of auditory perception. *Curr. Opin. Neurobiol.* **20**, 361–366 (2010).
103. Pressnitzer, D., Sayles, M., Micheyl, C. & Winter, I. M. Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* **18**, 1124–1128 (2008).
104. Micheyl, C. *et al.* The role of auditory cortex in the formation of auditory streams. *Hear. Res.* **229**, 116–131 (2007).
105. Gutschalk, A., Micheyl, C. & Oxenham, A. J. Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* **6**, e138 (2008).
106. Kondo, H. M. & Kashino, M. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* **29**, 12695–12701 (2009).
107. Deike, S., Gaschler-Markefski, B., Brechmann, A. & Scheich, H. Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* **15**, 1511–1514 (2004).
108. Hill, K. T., Bishop, C. W., Yadav, D. & Miller, L. M. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC Neurosci.* **12**, 85 (2011).
109. Micheyl, C., Tian, B., Carlyon, R. P. & Rauschecker, J. P. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* **48**, 139–148 (2005).
110. Fishman, Y. I., Reser, D. H., Arezzo, J. C. & Steinschneider, M. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res.* **151**, 167–187 (2001).
The authors present single-unit recordings in the auditory cortex in response to ABA tone sequences. Non-best frequency tones were suppressed at presentation rates and frequency separations in a manner that mirrored human perception.
111. Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J. & Shamma, S. A. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* **61**, 317–329 (2009).
Using psychophysical methods, the authors demonstrate that spectral components that are well separated in frequency are no longer heard as separate streams if presented synchronously rather than consecutively. The authors present a 'temporal coherence' theory of auditory streaming.
112. Micheyl, C., Kreft, H., Shamma, S. & Oxenham, A. J. Temporal coherence versus harmonicity in auditory stream formation. *J. Acoust. Soc. Am.* **133**, EL188–EL194 (2013).
113. Kashino, M. & Kondo, H. M. Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Phil. Trans. R. Soc. B* **367**, 977–987 (2012).
114. Tsunada, J., Lee, J. H. & Cohen, Y. E. Differential representation of auditory categories between cell classes in primate auditory cortex. *J. Physiol.* **590**, 3129–3139 (2012).
115. Obleser, J., Leaver, A. M., Vanmeter, J. & Rauschecker, J. P. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Front. Psychol.* **1**, 232 (2010).
116. Chevillet, M. A., Jiang, X., Rauschecker, J. P. & Riesenhuber, M. Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* **33**, 5208–5215 (2013).
117. Leaver, A. M. & Rauschecker, J. P. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* **30**, 7604–7612 (2010).
The authors used fMRI to investigate the hierarchical processing of natural sounds in the ventral pathway. Category-selective responses were identified in anterior superior temporal regions, whereas responses in the superior temporal sulcus were not category-selective but rather responded to acoustic features.
118. Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N. & Belin, P. Abstract encoding of auditory objects in cortical activity patterns. *Cereb. Cortex* **23**, 2025–2037 (2013).
The authors combined multivariate analyses of fMRI data with analysis of the low-level acoustical information to examine the abstract encoding of non-speech categories. They observed category sensitivity in the planum temporale, suggesting that object processing is not restricted to the ventral pathway.
119. Gifford, G. W., MacLean, K. A., Hauser, M. D. & Cohen, Y. E. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *J. Cogn. Neurosci.* **17**, 1471–1482 (2005).
120. Ohl, F. W., Scheich, H. & Freeman, W. J. Change in pattern of ongoing cortical activity with auditory category learning. *Nature* **412**, 733–736 (2001).
The authors recorded from the auditory cortex of gerbils while the animals learned an acoustic classification task. They demonstrate that the stimulus representation in the auditory cortex undergoes a dramatic change in its dynamic pattern at the point when animals begin to correctly classify the acoustic stimuli.
121. Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P. & Shamma, S. A. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nature Neurosci.* **13**, 1011–1019 (2010).
122. King, A. J. & Nelken, I. Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neurosci.* **12**, 698–701 (2009).
123. Hegde, J. & Van Essen, D. C. Role of primate visual area V4 in the processing of 3D shape characteristics defined by disparity. *J. Neurophysiol.* **94**, 2856–2866 (2005).
124. Alain, C. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* **229**, 225–236 (2007).
125. Naatanen, R. & Picton, T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* **24**, 375–425 (1987).
126. Kujala, T., Tervaniemi, M. & Schroger, E. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol. Psychol.* **74**, 1–19 (2007).
127. Picton, T. W., Alain, C., Otten, L., Ritter, W. & Achim, A. Mismatch negativity: different water in the same river. *Audiol. Neurootol.* **5**, 111–139 (2000).
128. Alain, C., Woods, D. L. & Ogawa, K. H. Brain indices of automatic pattern processing. *Neuroreport* **6**, 140–144 (1994).
129. Sussman, E. S., Horvath, J., Winkler, I. & Orr, M. The role of attention in the formation of auditory streams. *Percept. Psychophys.* **69**, 136–152 (2007).
130. Cusack, R., Deeks, J., Aikman, G. & Carlyon, R. P. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 643–656 (2004).
131. Winkler, I., Takegata, R. & Sussman, E. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Res. Cogn. Brain Res.* **25**, 291–299 (2005).
132. Snyder, J. S., Alain, C. & Picton, T. W. Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* **18**, 1–13 (2006).
133. Snyder, J. S., Carter, O. L., Hannon, E. E. & Alain, C. Adaptation reveals multiple levels of representation in auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 1232–1244 (2009).
134. Knudsen, E. I. Fundamental components of attention. *Annu. Rev. Neurosci.* **30**, 57–78 (2007).
135. Shinn-Cunningham, B. G. & Best, V. Selective attention in normal and impaired hearing. *Trends Amplif.* **12**, 283–299 (2008).
136. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
137. Zatorre, R. J., Mondor, T. A. & Evans, A. C. Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage* **10**, 544–554 (1999).
138. Duncan, J. E.P. S. Mid-Career Award 2004: brain mechanisms of attention. *Q. J. Exp. Psychol.* **59**, 2–27 (2006).
139. Lee, A. K. & Shinn-Cunningham, B. G. Effects of reverberant spatial cues on attention-dependent object formation. *J. Assoc. Res. Otolaryngol.* **9**, 150–160 (2008).
140. Darwin, C. J. & Hukin, R. W. Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *J. Acoust. Soc. Am.* **103**, 1080–1084 (1998).
141. Best, V., Gallun, F. J., Carlile, S. & Shinn-Cunningham, B. G. Binaural interference and auditory grouping. *J. Acoust. Soc. Am.* **121**, 1070–1076 (2007).
142. Shinn-Cunningham, B. G., Lee, A. K. & Oxenham, A. J. A sound element gets lost in perceptual competition. *Proc. Natl Acad. Sci. USA* **104**, 12223–12227 (2007).
143. Kastner, S. & Ungerleider, L. G. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* **23**, 315–341 (2000).
144. Shamma, S. On the emergence and awareness of auditory objects. *PLoS Biol.* **6**, e155 (2008).
145. Dick, F., Lee, H. L., Nusbaum, H. & Price, C. J. Auditory-motor expertise alters "speech selectivity" in professional musicians and actors. *Cereb. Cortex* **21**, 938–948 (2011).
146. Fritz, J., Shamma, S., Elhilali, M. & Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neurosci.* **6**, 1216–1223 (2003).

© 2013 Macmillan Publishers Limited. All rights reserved

147. Atiani, S., Elhilali, M., David, S. V., Fritz, J. B. & Shamma, S. A. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* **61**, 467–480 (2009).
148. Niwa, M., Johnson, J. S., O'Connor, K. N. & Sutter, M. L. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J. Neurosci.* **32**, 9323–9334 (2012).
149. Lee, C. C. & Middlebrooks, J. C. Auditory cortex spatial sensitivity sharpens during task performance. *Nature Neurosci.* **14**, 108–114 (2011).
150. Alain, C. & Woods, D. L. Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology* **34**, 534–546 (1997).
151. Woods, D. L., Alho, K. & Algazi, A. Intermodal selective attention: evidence for processing in tonotopic auditory fields. *Psychophysiology* **30**, 287–295 (1993).
152. Woods, D. L., Alho, K. & Algazi, A. Intermodal selective attention. I. Effects on event-related potentials to lateralized auditory and visual stimuli. *Electroencephalogr. Clin. Neurophysiol.* **82**, 341–355 (1992).
153. Petkov, C. I. *et al.* Attentional modulation of human auditory cortex. *Nature Neurosci.* **7**, 658–663 (2004).
154. Rinne, T. *et al.* Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *Neuroreport* **18**, 1311–1314 (2007).
155. Woldorff, M. G. *et al.* Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl Acad. Sci. USA* **90**, 8722–8726 (1993).
156. Ding, N. & Simon, J. Z. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* **107**, 78–89 (2012).
157. Mesgarani, N. & Chang, E. F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485**, 233–236 (2012). **The authors used electrocorticographic recording in human patients to investigate neural activity in listeners selectively attending to one stream of speech while ignoring a distractor stream. Neural activity represented crucial features of the attended speech while apparently suppressing the unattended stream.**
158. Degerman, A., Rinne, T., Salmi, J., Salonen, O. & Alho, K. Selective attention to sound location or pitch studied with fMRI. *Brain Res.* **1077**, 123–134 (2006).
159. Salmi, J., Rinne, T., Degerman, A. & Alho, K. Orienting and maintenance of spatial attention in audition and vision: an event-related brain potential study. *Eur. J. Neurosci.* **25**, 3725–3733 (2007).
160. Ahveninen, J. *et al.* Task-modulated "what" and "where" pathways in human auditory cortex. *Proc. Natl Acad. Sci. USA* **103**, 14608–14613 (2006).
161. Buffalo, E. A., Fries, P., Landman, R., Liang, H. & Desimone, R. A backward progression of attentional effects in the ventral stream. *Proc. Natl Acad. Sci. USA* **107**, 361–365 (2010).
162. Sugihara, T., Diltz, M. D., Averbeck, B. B. & Romanski, L. M. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J. Neurosci.* **26**, 11138–11147 (2006).
163. Romanski, L. M., Averbeck, B. B. & Diltz, M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* **93**, 734–747 (2005).
164. Gifford, G. W., Hauser, M. D. & Cohen, Y. E. Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. *Brain Behav. Evol.* **61**, 213–224 (2003).
165. Teki, S. *et al.* Navigating the auditory scene: an expert role for the hippocampus. *J. Neurosci.* **32**, 12251–12257 (2012).
166. Culling, J. F. & Summerfield, Q. Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. *J. Acoust. Soc. Am.* **98**, 785–797 (1995).
167. Darwin, C. J. & Hukin, R. W. Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *J. Acoust. Soc. Am.* **102**, 2316–2324 (1997).
168. McAdams, S. & Bregman, A. S. Hearing musical streams. *Computer Music J.* **3**, 26–43 (1979).
169. Shamma, S. A., Elhilali, M. & Micheyl, C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* **34**, 114–123 (2011).

Acknowledgements
We thank H. Hersh for a critical reading of the manuscript. J.K.B. is supported by a Royal Society Dorothy Hodgkin Research Fellowship and BBSRC grant BB/H016813/1. Y.E.C. is supported by grants from the US National Institute on Deafness and Other Communication Disorders and US National Institutes of Health.

Competing interests statement
The authors declare no competing financial interests.

© 2013 Macmillan Publishers Limited. All rights reserved