

Neural processing of natural sounds

Frédéric E. Theunissen and Julie E. Elie

Abstract | We might be forced to listen to a high-frequency tone at our audiologist's office or we might enjoy falling asleep with a white-noise machine, but the sounds that really matter to us are the voices of our companions or music from our favourite radio station. The auditory system has evolved to process behaviourally relevant natural sounds. Research has shown not only that our brain is optimized for natural hearing tasks but also that using natural sounds to probe the auditory system is the best way to understand the neural computations that enable us to comprehend speech or appreciate music.

Neural tuning

This term refers to the response property of brain cells by which they selectively represent a particular type of sensory, motor or cognitive information.

Until the late 1990s, auditory neuroscientists could be divided into two camps. In one camp, following the tradition of the great physicist and sensory physiologist Hermann von Helmholtz¹, classical auditory neurophysiologists used simple synthetic sounds such as pure tones (sound waves generated by a perfect sinusoidal oscillator) to probe the nature of neural responses in the auditory system. Indeed, just as the Helmholtz resonator separated multitone sounds into their frequency components, the principal role of the auditory portion of the inner ear, the cochlea, is to decompose the sound waveform into separate frequency bands². Thus, it is not surprising to learn that auditory neurons, at least at the lower levels of the auditory system, have been described and understood in terms of their responses to pure tones of a given frequency³ (FIG. 1a). In this classical approach, the frequency tuning curve of auditory neurons takes on a central role and more complex responses are described in terms of specific deviations from the linear summation rule; these responses are known as non-linear responses or contextual effects (reviewed in REF. 4).

In the other camp, following the tradition of the great ethologist Konrad Lorenz, auditory neuroethologists studied how natural sound stimuli that lead to specific behaviours are represented in the auditory system. One of the key findings from the neuroethologists' camp was the discovery of neurons that responded very strongly to natural and behaviourally significant sounds but not necessarily to their simpler components⁵⁻⁷ (FIG. 1b). In other words, the stimulus–response function describing the neural tuning in these neurons is dominated by the non-linear or context effects and not by the frequency tuning curve⁸. Moreover, it seemed that the appropriate 'auditory context' to probe the neural system was the natural one. Note that these two 'camps' were not antagonistic, and we use this term to stress the differences in

the two approaches and the contrasting shortcomings of each. However, there was relatively little discussion between researchers in each camp.

Each of these approaches has distinct merit from a methodological viewpoint. The reductionist approach of classical auditory physiologists enables a systematic parameterization of sound stimuli and therefore a clear method for synthesizing stimuli to explore specific mechanistic hypotheses. However, the relevance of results obtained from sounds that an animal rarely hears could always be questioned. Thus, conclusions about the implications of the results for processing behaviourally relevant complex sounds could be criticized as being post-hoc explanations that lack the strength of experimental predictions. By contrast, the behavioural relevance of the neuroethological approach was less problematic, and results showing that behaviourally relevant stimuli yield the largest neural responses⁷ (or the most informative^{9,10}) support evolutionary arguments and explanations, which state that the auditory system evolved to optimally process the sounds that matter the most for the survival of the species. However, the lack of a reductionist methodology in the neuroethologists' approach limited the exploration of underlying mechanisms.

One of the recent advances in auditory sciences has been in the merging of these two camps. This merging has been facilitated both by advances in computational approaches used for both sound and neural data analyses and by advances in experimental techniques. We review these recent developments. In the first section, we summarize what we have learned from the statistical analyses of natural sounds. Describing these statistics is important not only to define what is unique about natural sounds but also because this knowledge is needed to analyse the neural responses to these sounds and to determine whether the auditory system has evolved to

Department of Psychology
and Helen Wills Neuroscience
Institute, University of
California, Berkeley,
California 94720, USA.
e-mails: theunissen@berkeley.edu;
julie.elie@berkeley.edu
doi:10.1038/nrn3731

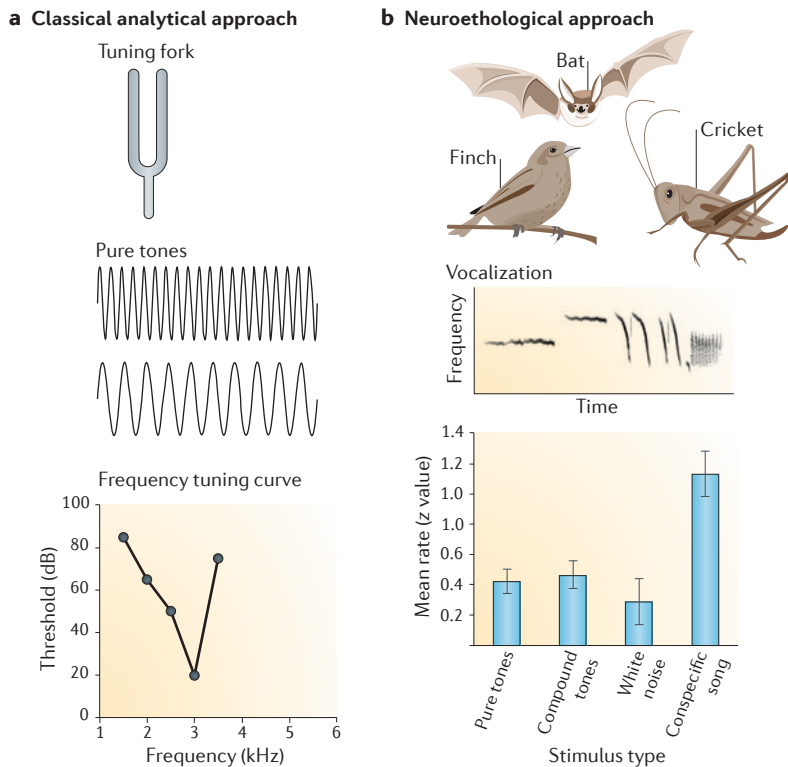


Figure 1 | Historical approaches to auditory neurosciences. Both the classical analytical approach and the neuroethological approach are based on the analysis of neural responses to sounds produced by particular sound sources. **a** | In the classical analytical approach, the sound sources are synthesizers or computers (symbolized by a man-made tuning fork in the top panel), the sounds are often pure tones (the sine waves shown in the middle panel), and neural responses are often described as a function of frequency such as in the frequency tuning curve of a neuron (bottom panel). The frequency tuning curve shows the minimum sound level of pure tones needed to elicit threshold responses. Here, we show the tuning curve of a narrow-tuned neuron from the avian inferior colliculus, the MLd (data replotted from REF. 3). This particular neuron is tuned to detect a frequency of 3 kHz and the response threshold increases sharply on either side of this frequency (hence, ‘narrow tuned’). **b** | In the neuroethological approach, the sound sources are often animals vocalizing or generating sounds by other means for communication. These natural sounds are complex signals that are best represented in a time–frequency plot such as the spectrogram of a bird song shown in the middle panel. Responses to these complex sounds are compared with responses to synthetic sounds, such as pure tones, compound tones (combination of pure tones), white noise or manipulated versions of the species vocalization. The neural data shown in the graph in the bottom panel are from single neurons in the avian primary auditory areas (data replotted from REF. 82). The average spike rates of these neurons, represented here as a z score (the deviation from the rate obtained in absence of a sound stimulus in units of the SD), show that the natural sound — here, conspecific song — is the stimulus type that best excites the neurons. The graph in part **a** is adapted with permission from REF. 3, The American Physiological Society.

process such sounds in an optimal manner. In the second section, we explain how the use of more recent machine learning techniques has enabled researchers to take into account this statistical structure when estimating neural tuning functions from responses to natural sounds. The third section focuses on the processing of communication calls, the vocalizations emitted by animals in the context of information exchange that are particularly well represented in the auditory system. In the last section, we review how progress in experimental methods has also enabled researchers to study hearing processes

Machine learning

A branch of computer science that combines statistical algorithms and artificial intelligence to extract information from complex data sets.

in more natural listening conditions. Our Review does not cover the extensive body of research that specializes in human speech processing and its neural correlates except when general principles are considered and clear parallels can be made.

Statistics of natural sounds

What is a natural sound? The question is particularly relevant given the increased prevalence of anthropomorphic noise in our daily environments that was absent during much of evolution. Natural sounds can be defined as: environmental sounds that are not generated by human-made machines, such as the sounds of footsteps, wind, fire and rain; all animal vocalizations, including human speech; and other sounds generated for communication by animals, such as stridulation in crickets¹¹, buttress drumming by chimpanzees¹² and instrumental music by humans. We first investigate the properties of isolated sounds and then briefly touch on the statistics of sound mixtures.

Perceptually relevant physical characteristics of isolated natural sounds follow a power law. Although natural sounds defined in this broad sense have heterogeneous properties, they share a structure that can be quantified by ensemble statistical analyses (FIG. 2). Specifically, it has been observed that the frequency spectra of certain fluctuating physical characteristics of natural sounds follow a $1/f$ relationship (where f is frequency) or, more generally, a power law relationship. In other words, some physical characteristic of natural sounds (ϕ) varies as the power of the frequency such as $\phi(f) = \alpha f^{-\kappa}$ with positive constants α and κ . It should be clearly noted that this relationship does not hold for the sound spectrum itself but, instead, for slower-varying structures, such as loudness, measured in the temporal envelope of the sound (FIG. 2) or in time-varying pitch (the ‘height’ of the sound) profiles^{13–15}. The power law relationship also holds for the power spectrum of the log of the sound spectrum¹⁵. This transformation of the sound waveform, called the cepstrum¹⁶, is used to extract spectral structures in the sounds (structures in the frequency domain) such as speech formants. Moreover, it has been shown that the frequencies of temporal and spectral modulations in the spectrogram, known as the modulation power spectrum (FIG. 2), exhibit specific dependencies beyond those expected from the time–frequency trade-off¹⁷. Natural sounds and vocalizations in particular have higher power at joint frequencies of low-temporal and high-spectral modulations than expected from the product of the marginals: the average power for the same temporal modulation (averaged over all spectral modulation frequencies) multiplied by the average power for the same spectral modulation (averaged over all temporal modulation frequencies)¹⁵. In other words, many animal vocalizations are dominated by relatively slow sounds with fine harmonic structure.

Physical, behavioural and neural implications of the power law structure. What are the physical, behavioural and neural implications of this naturally occurring acoustical structure? First, in terms of physical

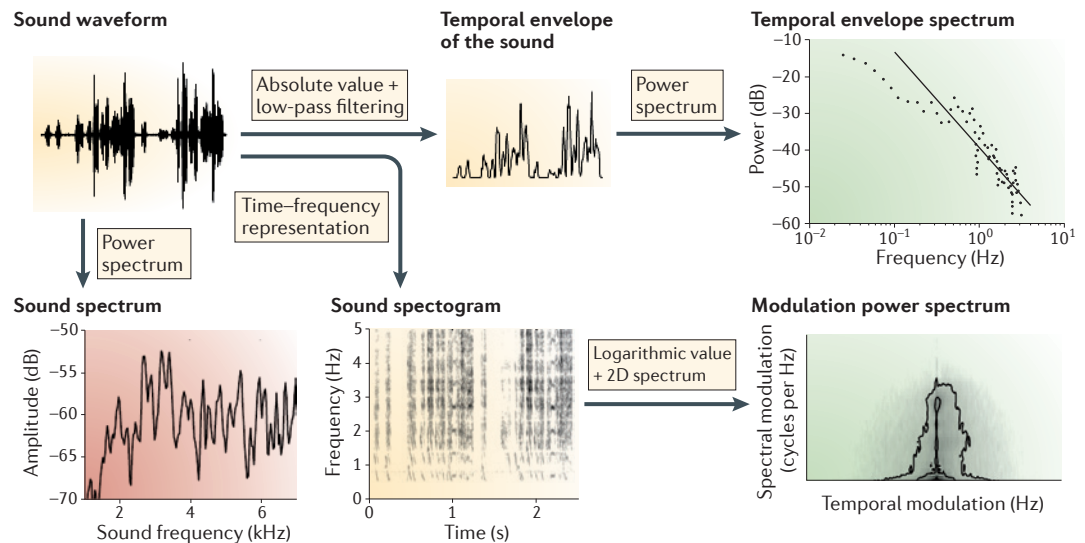


Figure 2 | Natural sound statistics. Various statistical measurements can be obtained from distinct physical characteristics of sounds. This figure illustrates some of these measurements for a zebra finch song and highlights the measures that reveal common characteristics of natural sounds (green graphs) and those that are specific to each sound class (red graph). The sound spectrum is the power of the sound pressure waveform as a function of frequency. This basic spectrum (that is, obtained without any transformations) shows unique shapes depending on species and call types. Power–frequency curves of natural sounds (sound spectra) do not obey universal relationships that would be characteristic of all natural sounds. Conversely, the temporal envelope spectrum obtained by calculating the power spectrum of the temporal envelope obeys a $1/f$ (where f is frequency) relationship (solid line) that is characteristic of all natural sounds¹³: natural sounds are dominated by low frequencies of amplitude modulation. The sound spectrogram is a more intuitive representation obtained by decomposing the sound into time and frequency bins: at each given time point (x axis), the sound is represented in terms of the amplitude of its frequency components (y axis). Just as for the basic spectrum, measures on the spectrogram are unique for each natural sound, but the modulation power spectrum obtained from a two-dimensional (2D) spectral analysis of the logarithmic values of the sound spectrogram shows a coarse shape that is characteristic of all animal vocalizations¹⁵. The bottom right panel is reprinted with permission from Singh, N. C. & Theunissen, F. E. Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* **114**, 3394–3411 (2003)¹⁵ © 2003, Acoustic Society of America.

Sound spectrum

(Also called the frequency spectrum). The representation of a waveform in its frequency domain. The sound spectrum is short for the frequency power spectrum of that sound. The sound spectrum is obtained by taking the amplitude square of the Fourier transform of a waveform and shows the energy in a signal as a function of frequency.

Envelope

The smooth representation (low-pass filtering) of the amplitude of a signal as a function of frequency (spectral envelope) or time (temporal envelope).

Formants

The peaks in the spectral envelope that correspond to a resonance in the sound source. In animal vocalizations and human speech, formants are resonance in the upper vocal tract. Formants are also the distinguishing or meaningful frequency components of voiced human speech.

Spectrogram

The representation of the sound in the time and frequency domain that shows how the sound spectrum varies over time. This representation is often obtained by calculating the sound spectrum in a short-windowed section of sound and repeating this calculation by moving the window in time. This method of calculation is also called the short-time Fourier transform.

properties, the power law relationship for time-varying signals implies that natural sounds have correlations over multiple timescales, including very long ones, as reflected by the large energies at low frequencies. In this sense, natural sounds are clearly different from signals that are completely random or uncorrelated, such as white-noise signals with flat spectral and temporal envelopes power spectra or, at the other end of this spectrum of correlations, signals that are dominated by a single-correlation time, such as those created by a perfect oscillator (for example, a sound with sinusoidally varying amplitude, as in some car alarms). It has also been argued that neither white noise nor a pure sine wave can qualify as complex and thus as information-rich or perceptually sophisticated^{13,18}. Second, in terms of behaviour, it is interesting to note that the fluctuating physical sound characteristics that show the power law characteristics in natural sounds are those that are directly linked to perceptual attributes. Whereas we are unable to perceive the details of the sound pressure waveform, the time-varying amplitude yields a percept of intensity fluctuations, rhythm and timbre; the time-varying pitch profile carries the melody in a musical phrase and the spectral envelope contains critical information for other timbral qualities of sound, including speech formants¹⁹. Last, these

observed natural statistical structures have implications in terms of neural coding. For example, sound stimuli that have such natural statistics elicit higher information rates (measured in auditory neurons) relative to matched synthetic sounds that lack some of the natural statistics^{9,10,20}. Interestingly, the spatial and temporal luminance contrasts in natural visual scenes also obey power law relations that have also been related to complexity but that are primarily the result of scale invariance^{21,22}. This power law relation ($1/f$) implies that visual scenes have stronger correlations at low spatial and temporal frequencies than at high frequencies. It has also been shown that early processing in the visual system can reverse this relationship by attenuating low frequencies and boosting high frequencies, effectively removing the correlations that are present in the stimulus images²³. Such decorrelation is useful for maximizing information transmission through a bottleneck such as the optic nerve. Although the physical causes of the power law relationship observed in natural images and in natural sounds are unrelated, it is highly probable that similar neural efficiency principles apply to both sensory systems. Indeed, a similar decorrelation has been observed in the inferior colliculus, where the gain of auditory neurons emphasizes higher temporal and spectral modulation, which effectively counterbalances

Modulation power spectrum

The spectrum obtained from the log of the amplitude of a sound spectrogram. The name comes from the fact that each 'row' in a spectrogram pictorially represents in a narrow frequency channel the amplitude envelope of the sound in that channel. These envelopes 'modulate' the intensity of the signals, and the modulation power spectrum shows the power of these modulations for different rates (that is, the temporal modulation frequencies on the x axis). Similarly, a column in the spectrogram shows the spectral envelope at a particular time point. The power spectrum of this spectral envelope shows the power at particular spectral modulation frequencies (that is, on the y axis). More generally, a spectrogram jointly tracks amplitude envelopes in both time and in frequency and the full modulation power spectrum shows the power of both spectral and temporal modulations.

Harmonic

A complex sound with a strong pitch percept that is made of a tone at a fundamental frequency or first frequency (f_0 , corresponding to the perceived pitch) and a series of overtones at multiples of this fundamental frequency.

Timbre

The quality of a sound (also known as tone colour or tone quality from psychoacoustics) that can be used to distinguish different types of sound production (voices or various musical instruments) and that is determined by certain physical properties, such as the sound spectrum.

Invariance

The property of something that does not change under a transformation (for example, in rotation-invariant face neurons, a particular face will elicit very similar responses irrespective of the face orientation).

Decorrelation

The process used to reduce the autocorrelation or the redundancy of information within a signal.

the $1/f$ relationship observed in the modulation power spectrum (and not the sound spectrum) of natural sounds²⁴. At higher levels of the auditory processing stream, it has been shown that the population of neurons has a maximum gain at intermediate modulation frequencies in a region that is particularly useful for distinguishing among different natural sounds²⁵.

The sound spectrum is idiosyncratic for each natural sound class. As mentioned above, natural sounds exhibit a power law relationship in the spectrum of particular time-varying features of sounds such as intensity, and this relationship has physical, perceptual and neural implications. This power law relationship does not exist for the sound spectrum itself as each natural sound class has an idiosyncratic sound spectrum. However, this does not mean that auditory systems are not sensitive to particular shapes of the sound spectrum of behaviourally relevant sounds. On the contrary, frequency tuning sensitivity has been shown to be one of the major factors ensuring the sender-receiver match. The neuroethological basis of this matched-frequency tuning has been particularly well documented in insects and anurans²⁶. Furthermore, more strikingly perhaps, this frequency tuning adaptation has even been observed in the cochlea of owls²⁷ and bats²⁸, in which the region of the cochlea that is mechanically tuned to frequencies that are particularly relevant for the animal is expanded in what has been called an auditory fovea. As we discuss in more detail in the section on animal vocalizations, the adaptation of the auditory system to the specific structure of conspecific communication calls might be equal to or maybe even greater than putative adaptations to more general natural sound statistics.

Natural sound statistics and the frequency tuning of mammalian auditory nerve fibres. Looking beyond the matched-tuning found in auditory specialists such as bats and owls, could general natural sound statistics also explain the prototypical frequency tuning that has been observed in the peripheral mammalian auditory system? Primarily as a result of its mechanical properties, the cochlea decomposes sounds into a set of signals that are centred at increasing frequencies by applying filters of different shapes: narrow-band frequency filters for low frequencies and large-band frequency filters for high frequencies². As the frequency power spectrum of specific natural sounds is idiosyncratic, a simple frequency matched-tuning or decorrelation argument cannot be used to provide an adaptive explanation for this relationship. Instead, however, an examination of both the temporal and spectral statistics of different classes of natural sounds can provide an explanation. In particular, environmental sounds and animal vocalizations make two well-defined groups of sounds that have different statistical structures: animal vocalizations are dominated by sustained harmonic sounds, whereas environmental sounds are dominated by transient sounds²⁹. It has been argued that the shape of the mammalian auditory frequency filters measured at the

level of the auditory nerve is optimal at representing the independent components of combinations of animal vocalizations and environmental sounds: the lower-frequency narrow-band filters efficiently represent the relatively long but spectrally sharp animal vocalizations, and the higher-frequency broad-band filters efficiently represent the short but broad environmental sounds. The human speech signal is particular in that it combines sounds from these two classes. One can thus postulate that the physical characteristics of human speech have evolved to be optimally represented at the auditory periphery (while taking into account other constraints)²⁹.

Statistics of sound mixtures. Isolated sounds have interesting properties, but our brains are more often exposed to complex auditory scenes. Sound mixtures, such as those created by a chorus of insects or a crowd in a loud restaurant, also have their own statistical signature: specifically, the structure that is present in the modulation power spectrum of isolated vocalizations is washed out in sound mixtures, whereas the long-time average sound spectrum of isolated sound signals and their mixtures remains similar. Given that the modulation power spectra of background sounds differ from those of foreground sounds, a modulation filter bank — a set of filters in the spectral and temporal amplitude modulation domains — that is tuned to these differences could be used to separate signals from noise resulting from sound mixtures, and such a mechanism might be in place in secondary auditory cortical areas³⁰. Because the sound spectra of mixtures and signals are similar, this task would be impossible with a simple frequency filter bank — a set of filters in the sound frequency domain. Sound mixtures also seem to be processed separately from isolated sound signals: whereas the short time detail of isolated sound signals is perceived with high accuracy (enabling, for example, rather extreme rates of phoneme perception), sound mixtures are perceived and categorized in terms of their long-term statistical properties, yielding percepts of sound 'texture', which is defined as the collective result of many similar acoustic events (for example, rainstorms and insect swarms)³¹.

In summary, natural sounds have characteristic statistical properties that can be measured at different levels. All natural sounds have particular slow physical properties, such as loudness profiles that obey power law relationships. The sound spectrum does not obey this law, but its shape is nevertheless an important property for the specialized processing of behaviourally relevant sounds. Natural sounds are easily categorized as either animal vocalizations or environmental sounds based on differences in terms of relevant time-frequency scales. Sound mixtures lose the fine spectrotemporal modulations seen in isolated natural sounds and are better characterized and perceived in terms of long-term statistical properties. Both the nature of our perception of sounds and neural responses in the auditory system are sensitive to these natural sound statistics.

Stimulus–response characterizations

As explained above, natural sounds are clearly both relevant and efficient stimuli to drive auditory neurons. Moreover, both theoretical arguments used to model processing in the auditory periphery^{29,32} and information theoretic measures of empirical data^{9,10} have shown that the auditory system seems to have evolved for optimal processing of sounds with such statistical properties. These studies, however, shed little light on the actual underlying mechanisms when compared with the explanations provided by the characterization of the neuronal stimulus–response functions; that is, the mathematical formulation that describes how single neurons or neuronal ensembles respond to any given stimulus.

Estimating the stimulus–response function using synthetic sounds. Traditionally, stimulus–response characterizations had been performed with synthetic sounds that would enable the systematic probing of the effect of a single acoustical parameter (for example, frequency) on neural responses. System identification analysis, the functional description of any arbitrary input–output system, also heavily relied on the use of synthetic sounds and primarily Gaussian white noise³³. Noise-like stimuli enable not only an efficient exploration of a large set of possible sounds (for example, all frequencies within the noise band) but also facilitate the estimation of a neuron's stimulus–response function; with white noise, the average stimulus before each action potential or spike (the spike-triggered average (STA)) yields the impulse response of the neuron or, when stimuli are represented in spectrographic form, the neuron's spectrotemporal receptive field (STRF)^{34,35}. For neurons that respond linearly to sound features as represented in a spectrogram, the STRF shows the spectrotemporal pattern that would result in the highest firing rates. When the STRF is used as a model, the convolution (a mathematical operation akin to a running-time correlation) between the STRF and the sound spectrogram yields a predicted neural response. The STRF model can be generalized to model the response of any neuron by incorporating non-linear components, as we describe in more detail below. At lower levels of the auditory processing stream, in which neurons are less sensitive to contextual effects, the white noise approach can yield accurate estimations of the stimulus–response function^{36–38}. In these cases, white noise analyses are used to estimate stimulus–response functions, and these functions in turn can explain selectivity for specific vocalizations or the efficient representation of natural sounds in general²⁹. However, at higher levels of the auditory system, neural responses can be dominated by contextual effects^{8,39,40}. Although sound features that drive neurons might be present in white noise, they might only elicit responses when they are presented in a natural acoustic context: for example, following silence, or following a sequence of specific sounds or presented jointly with other specific sounds. In other words, neurons become tuned to more complex spectrotemporal patterns that are characteristic of natural sounds but that are poorly sampled in white noise. In those cases, stimulus–responses functions can only be estimated using the appropriate context: behaviourally relevant natural sounds.

Methods for estimating STRFs using natural sounds.

Fortunately, advances in regression techniques and machine learning have enabled the estimation of STRFs using natural sounds (FIG. 3). Great progress has been made on four critical issues. First, natural sound ensembles occupy a limited region of the entire space of possible sounds. One must therefore be aware that the shape of the estimated STRF will depend on the subset of sounds being sampled and is only valid for sounds sharing the same characteristics as the sampled one. In this case, this issue is simply solved by clearly describing the sampled subset in the space that is relevant for STRFs. For example, if the STRFs are based on spectrograms, the phase and amplitude of the modulation spectrum (that is, the spectrum of the spectrogram) will describe how the selected natural sounds sample the space¹⁵. In addition, when one compares STRFs estimated with two distinct sound ensembles (whether they are natural or synthetic), one needs to estimate the STRFs carefully by only using sound features that are found with sufficient frequency in both sound subspaces⁴¹.

Second, for natural sounds, the time-averaged energy of sound features (or, equivalently, the average intensity and frequency of occurrence of these features) is not uniform throughout the subset of sounds being sampled. For this reason, simple estimation techniques such as the STA, which is a straight averaging operation, will yield biased estimates of the STRF. This bias can be removed by using the appropriate normalization techniques. These normalization techniques can be thought of as a weighted average operation in which sound features that are sampled more infrequently are given more weight to compensate for this undersampling^{42,43}.

Third, again, as natural sounds might only effectively span a small subset of possible sounds, one must carefully match the effective dimensionality of this sampling to the dimensionality of the sound representation. For example, an STRF operating on the spectrographic representation of sounds might have 100 slices in time (for example, 100 ms window with 1 ms sampling) and 100 slices in frequency (for example, 100 Hz bands between 0 Hz and 10 kHz) for a total of 10,000 time–frequency ‘pixels’, which constitute the parameters of the STRF model (FIG. 3). Natural sounds might sample these 10,000 dimensions very sparsely and thereby yield very poor estimates for all 10,000 STRF parameters. This is a well-known problem in statistics: if a model has too many parameters (here, the number of time–frequency pixels of the STRF) relative to the number of observations (here, the number of natural sounds and corresponding neural responses), then the model risks fitting not only the underlying relationship between the stimulus and the neural response but also random fluctuations of the particular data set. To prevent this phenomenon, which is known as overfitting, regularization techniques must be used⁴⁴. Regularization adds constraints in the form of priors on the model parameters that effectively impose a penalty on model complexity: for example, principal component regression (also called subspace regression) and ridge regression implement zero-mean Gaussian priors on the STRF

Inferior colliculus

The principal auditory nucleus of the mammalian midbrain, which receives inputs from several peripheral brainstem nuclei, including direct and indirect inputs from the cochlear nucleus.

Gain

For neurons, the gain is the sensitivity of a neural output to the input signal.

Frequency filters

Devices for selectively transmitting specified frequencies of the input signal by attenuating, or filtering out, unwanted frequencies.

Spike-triggered average

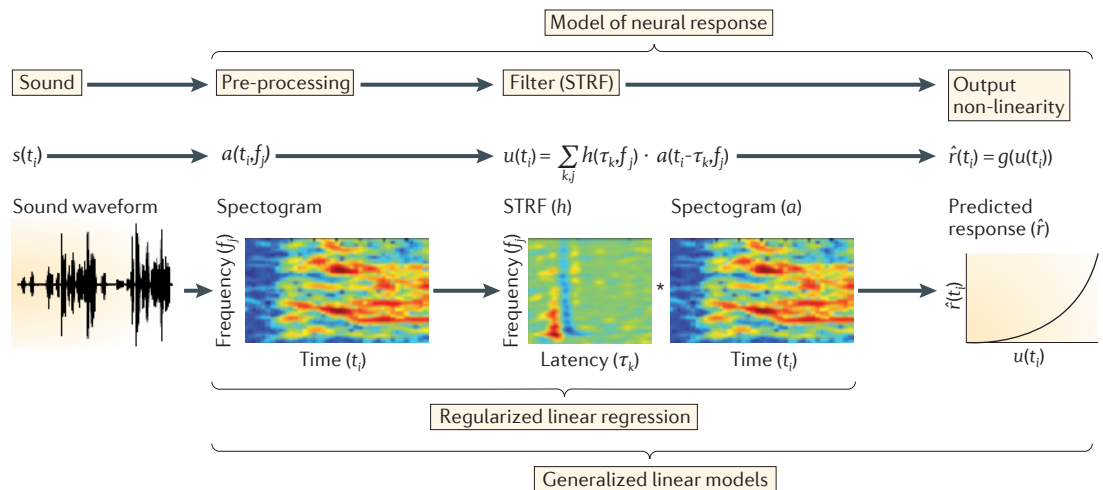
(STA). A tool for estimating the stimulus–response function of a neuron by using the average stimulus before each spike. For linear neurons and for white noise stimuli, the STA will yield the neuron's receptive field.

Spectrotemporal receptive field

(STRF). In the most general sense, the STRF is used to label the stimulus–response function of a neuron that uses any spectrotemporal description of the sound as the stimulus (for example, the spectrogram). That general STRF can be a linear or non-linear model.

Regularization

In statistics and machine learning, regularization methods are used for model selection, in particular to prevent overfitting by penalizing models with extreme parameter values or extreme number of parameters. For example, principal component analysis applied on a dataset can be used to reduce the dimensionality of the dataset. The resulting model is a principal component regression or subspace regression.



Priors

Probability distributions attached to parameters before certain data are observed. Prior is short for 'prior probability'.

Principal component regression

The regularization procedure that uses principal component analysis in conjunction with a linear regression analysis to reduce the dimensionality of a data set.

Ridge regression

The combination of linear regression with a regularization procedure that assumes a priori that the coefficients of the regression are normally distributed around zero. By varying the spread of this normal distribution, one can constrain the coefficients to be very close to zero unless there is strong evidence against it in the data. Ridge regression is therefore useful when there are too many parameters in the linear regression to be fitted with limited data (that is, to prevent overfitting). In those cases, most parameters will be equal to zero and only the few that really matter will have non-zero values. It is in this sense, that ridge regression reduces the dimensionality of the problem. Ridge regression is also known as L2 regularization and statisticians also call regularization procedures of this type shrinkage.

Zero-mean Gaussian

A quality of a distribution that is normal with a mean of zero.

Generalized linear models

(GLMs). A generalization of ordinary linear regression that allows for the random aspects of response variables (that is, the noise) to have distributions other than the normal distribution (for example, a Poisson distribution).

Figure 3 | Stimulus–response characterization. The stimulus–response function of auditory neurons can be estimated using natural sounds and advanced techniques in regression and machine learning. The neural response, $r(t_i)$ is modelled as a multistep transformation of the sound stimulus $s(t_i)$, yielding a predicted neural response $\hat{r}(t_i)$. The three steps of this neural model include a pre-processing step, a filtering step and an output non-linearity. The parameters of the three steps are first estimated using a data set of sound stimuli and their corresponding known neural responses, and then the model can be used to predict the neural responses to new sound stimuli ('sound waveform' in the figure). In the pre-processing step, the sound pressure waveform is transformed into a new representation, such as the spectrogram in which the amplitude a is expressed as a function of time t_i and frequency f_j (shown here as an example in the 'pre-processing' column; $a(t_i, f_j)$), a cochleogram (not shown) that models the filtering and processing occurring at the level of the cochlear nuclei (brainstem nuclei that receive inputs from the cochleae)⁴⁹ or higher-level processing such as that based on probabilistic expectations (not shown)⁵¹. The next step involves the estimation of a linear filter ($h(\tau_k, f_j)$ here). Because the new sound representation obtained in the pre-processing step can have many dimensions, regularization regression techniques must be used when estimating the filter to prevent overfitting^{43,44,46}. When a time–frequency representation of sound is used, the linear filter $h(\tau_k, f_j)$ obtained is called the spectrotemporal receptive field (STRF h , shown here). When the x axis of the STRF is set up to indicate increasing delay τ_k from the beginning of a stimulus (shown here), then the STRF represents the neural response obtained to a theoretical impulse stimulus: the so-called impulse function. When the x axis is set up to indicate the time preceding a spike, equivalent to a vertical reflection of the previous matrix, then the STRF represents the spectrotemporal features that most drive the neuron (shown in FIG. 4). More advanced methods can yield multicomponent linear filters (not shown)⁵⁵. In the last step, the output of the linear filter $u(t_i)$ is transformed into the predicted response using a static non-linearity $g(u(t_i))$. Generalized linear models can be used to simultaneously estimate both the STRF and this non-linear output function for different noise distributions⁵².

coefficients with a variable variance. By setting a small variance on this prior, STRF parameters will be estimated to be very close to zero during the model-fitting procedure, unless there is robust evidence that they significantly contribute to the prediction of the neural response. Principal component regression and ridge regression also have analytical solutions that are computationally very efficient (solutions that can be found by solving a mathematical equation)^{43,45}. Regularization can also be implemented by using other priors on STRF coefficients and iterative algorithms^{44,46}.

Fourth, the stimulus–response functions of high-level auditory neurons are often dominated by non-linearities that are not captured in the STRF, which, in its simplest form as a model, predicts neural responses from a linear combination of spectrotemporal features. Estimating the nature of the non-linearities is not only important to fully capture the computations performed by the system but also to avoid errors induced by non-linearities for the estimation of the STRF obtained with natural stimuli^{47,48}. There are many approaches to this problem. Input non-linearities can be incorporated in

the chosen representation for the sound stimuli. For example, sound representations can include known non-linearities such as adaptive mechanisms^{49,50} or probabilistic expectations⁵¹. Output non-linearities, such as those produced by a spiking threshold, can be very efficiently modelled using the generalized linear framework⁵² even in combination with input non-linearities⁵³. Finally, dynamical second-order or higher-order non-linearities have been estimated using techniques that yield multicomponent STRFs^{47,54–57}.

The computations in the auditory processing stream revealed by the STRFs. These methodological advances have enabled auditory neuroscientists to make considerable progress in understanding the nature of the auditory computations that are found in the ascending processing stream of both birds^{30,41,48,51,58–61} and mammals^{38,62–65}. Selectivity for natural sounds is already present at the level of the inferior colliculus in the sense that inferior colliculus STRFs show spectrotemporal features that are found in behaviourally relevant sounds^{38,41,66–68}. Then, novel types of STRFs appear at

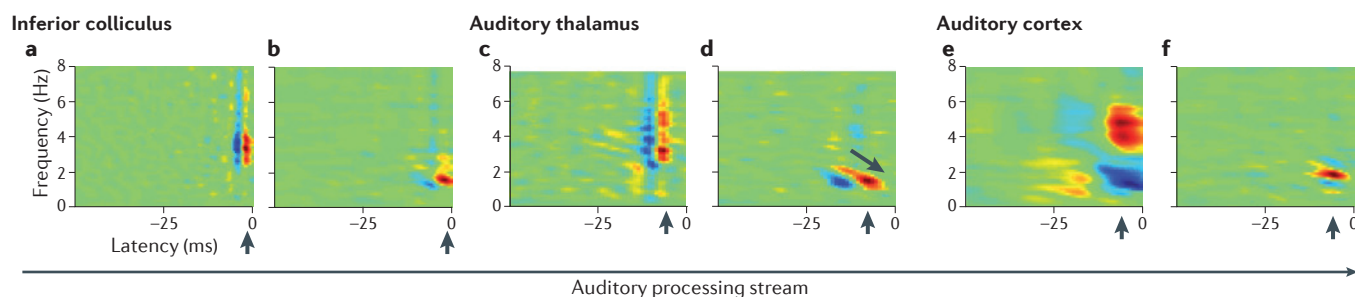


Figure 4 | STRFs at different levels of the auditory system. In each region of the auditory system, one finds multiple types of spectrotemporal receptive fields (STRFs), and in each region some of these types efficiently extract spectrotemporal features of natural sounds. The figure shows pairs of illustrative STRFs (shown in pseudocolour in the figure, where red represents the most intense response, and blue the lowest response) found at different levels of the avian auditory system, which includes: the inferior colliculus (also known as MLd), the auditory thalamus (also known as nucleus ovoidalis) and the avian auditory cortex (also known as field L). Note that for each STRF, the x axis is the time preceding the response and that therefore the sound features that excite the neuron are read from left to right, whereas the impulse function is read from right to left. As one follows the auditory processing stream, neurons become tuned to slower and more complex features. **a** | In the inferior colliculus, some neurons show STRFs with a brief (narrow in time) and large frequency band of inhibition (blue) followed by a brief and large frequency band of excitation (red); such fast broad-band neurons will effectively detect the onset of song syllables and encode the temporal rhythm of a song. **b** | The narrow-band neuron shown on the right is also selective for the onset of sound but at a particular frequency, around 2.5 kHz. **c,d** | The auditory neurons in the thalamus exhibit greater latencies than inferior colliculus neurons: they respond 10–15 ms after the peak energy in the STRF (indicated by the arrows), whereas inferior colliculus neuron responses have latencies of around 5–10 ms. Auditory thalamic neurons also show greater sensitivity to slower features. The narrow-band STRF shown in part **d** is more complex than the one found in the inferior colliculus (part **b**), with frequency tuning that goes down with time (indicated by the arrow). This neuron is sensitive to down-sweeps that are common in zebra finch song syllables. **e,f** | Much slower and more complex STRFs appear at the level of the auditory cortex. The broad-band neuron shown in part **e** not only decodes spectral shape at the coarse scale that is useful for the representation of structures such as formants but is also sensitive to a combination of low-frequency sounds (<3 kHz) followed by high-frequency sounds (>3 kHz). The narrow-band neuron in part **f** shows a sharp excitatory region that is flanked by two inhibitory regions. Such narrow-band neurons are exquisitely tuned to notes of a particular pitch either as pure tones or as harmonic complexes. In the avian auditory system, STRFs that combine excitatory and inhibitory regions at the same time point (as shown in these two examples) appear only at the level of the cortex. Additional avian STRF types and examples can be found in REFS 51, 58, 59. Examples in the mammalian auditory system can be found in REFS 24, 62–64. Parts **a**, **b**, **e** and **f** republished with permission of Society for Neuroscience, from Woolley, S. M., Gill, P. R., Fremouw, T. & Theunissen, F. E. Functional groups in the avian auditory system. *J. Neurosci.* **29**, 2780–2793 (2009)⁵⁸; permission conveyed through Copyright Clearance Center, Inc. Part **c** is based on data from REF. 59. Part **d** adapted with permission from REF. 59, The American Physiological Society.

the level of the primary auditory cortex, and these STRF models can be considered to achieve selectivity for more complex and slower acoustical features compared with the simpler STRFs found in the inferior colliculus and thalamus (the principal relay of sensory inputs from the sensory periphery to the cortex)^{59,62,63} (FIG. 4). These changes in the STRF go hand in hand with increased selectivity for natural sounds^{9,20,25}. Contextual effects also become more important at the higher levels of the auditory system^{40–42,48,51,69,70}; these contextual effects manifest themselves as changes in the selectivity for spectrotemporal features owing to the presence of particular sounds ‘outside’ a classically estimated STRF^{48,71}, changes due to expectations about stimulus statistics⁵¹, changes in the correlated properties measured in ensemble neurons⁴¹ and changes due to learning and behavioural relevance^{69,72}. Again, it is postulated or shown that these auditory contextual effects increase the efficiency of the neural representation for behaviourally relevant natural sounds either at the single^{51,69,72} or population^{41,73} level. Finally, researchers have begun to understand how complex stimulus–response functions found at higher levels

of the auditory processing stream could be used to achieve complex auditory tasks that go beyond ‘template-matching’ between an STRF and an acoustical feature that is present in natural sounds. For example, the multiscale time–frequency modulation tuning of the auditory cortex can be used to separate bird song or speech from non-speech signals or noise^{30,74,75}.

As an alternative to the estimation of linear and non-linear STRFs from responses to natural sounds, researchers have also used synthetic sounds that are designed to have particular natural statistics. Families of such synthetic natural-like sounds can then be used to isolate the specific natural feature that is particularly important for understanding behavioural or neural responses. This approach has been used, for example, to elucidate the natural sound features that are crucial for phonotaxis in crickets¹¹, sound texture perception in humans^{31,76} and selectivity for conspecific songs in songbirds^{9,15}.

In summary, analytical and computational advances have enabled auditory researchers to use natural sounds or synthetic natural-like sounds to estimate the stimulus–response functions of high-level auditory neurons.

In doing so, they have been able not only to extract these functions for neurons that do not respond to white noise or other synthetic stimuli but they have also been able to investigate auditory contextual effects and the nature of the computations that generate selective responses for natural sounds.

Animal vocalizations

Animal vocalizations as a class of natural sounds have played and continue to play an important part in auditory neuroscience. Historically, the first use of natural sounds in auditory neuroscience came from neuroethologists who investigated how conspecific vocalizations or communication signals were selectively processed in the auditory systems of auditory specialists. These investigations in model systems led to the discovery of cricket song-selective neurons and their contribution to female phonotaxis behaviour⁷⁷, of call-selective neurons in frogs⁷⁸ and guinea pigs⁷⁹, of song-selective neurons in songbirds^{7,80–82}, of neurons selective for the echolocation signal in bats⁵ and of brain regions selective for conspecific calls in primates⁸³. Selectivity for conspecific communication calls can be reflected not only in the mean rate of single-neuron responses but also (and sometimes only) in time-varying responses^{84,85} or ensemble responses^{86,87}. Thus, the auditory system is not only selective for natural sounds in a broad sense but also seems to exhibit specialized circuitry for the sole purpose of detecting and processing conspecific communication calls. One of the striking results from this line of research has been the relatively high degree of selectivity that has been measured in these vocalization-selective neurons^{6,88}: systematic manipulations of bird song syllables and bat echolocation calls have shown that this selectivity is achieved by non-linear mechanisms that detect a specific temporal or spectral combination of unique sound features that are present in specific conspecific vocalizations^{5,6,89}.

Although such acute selectivity might be useful for auditory tasks that require high fidelity, such as the processing of echolocation pulses or guiding vocal commands in song learning, its utility for processing sounds in terms of their communicative value is more problematic. For example, both primates and songbirds produce alarm calls that need to be correctly categorized in order to guide the appropriate behaviour. Such categorization requires invariant responses to all communication calls that belong to the same category as well as recognition of category boundaries⁹⁰. Thus, auditory processing for communication purposes might require not only low-level feature detection processing but also categorization of higher-level structure. Such high-level categorization might involve hierarchical processing steps, such as the representation of particular sound features (for example, formant frequency) that are resistant to variation in other physical parameters of the sound (for example, azimuthal location), and such representations have been found in secondary mammalian and avian auditory areas^{91,92}. In terms of higher-level categorization, research in

starlings points to a role for the caudomedial nidopallium in classifying behaviourally relevant classes of songs⁹³ and research in primates suggests that both the superior temporal gyrus and the ventrolateral prefrontal cortex could be involved in semantic discrimination^{94–98}. Similar cortical areas have been shown in a large body of research to be crucial for human speech processing⁹⁹. However, it is fair to state that our understanding of the neural mechanisms that generate such high-level categorization of sounds is still in its infancy. Songbirds that have a large repertoire of communication calls that are used in distinct behavioural contexts could also become a powerful animal model to study the neural computations involved in the categorization that is needed to extract meaning from variable communication sounds^{100,101}.

The ontogeny of selective neural responses for vocalizations has also been extensively studied. Although many animal communication calls are innate or have innate characteristics, neural selectivity in the perceptual system for innate calls could arise during development simply as a result of experience and repeated exposure. Moreover, vocalizations show learning components in both production, as is the case for song in songbirds¹⁰², and perception, as is the case for the interpretation of alarm calls in primates¹⁰³, the interpretation of pup calls in mothers versus virgin mice¹⁰⁴, the discrimination of familiar versus unfamiliar contact calls in zebra finches¹⁰⁵ and the recognition of individual songs in starlings¹⁰⁶. Not surprisingly, selective neural responses for natural sounds have been shown to have strong developmental and environmental components; this is true both for the low level of selectivity such as that found in primary auditory cortical areas^{73,107–109} as well as for the high level of selectivity found in sensorimotor areas of songbirds^{110–112}. Experience during development can also affect perceptual boundaries and their putative neural correlates¹¹³.

In summary, on one hand, the initial study of the neural representation of conspecific vocalizations in the auditory system has played a crucial part in advancing our understanding of the nature of the non-linear neural responses that are found in the higher auditory areas and in establishing the need to use behaviourally relevant sounds to decipher these computations. On the other hand, research on the nature of invariant representations for vocalization classes and on the link between sound and the perception of meaning is still in its infancy, and research in this area could further increase our understanding of the neural mechanisms involved in human speech processing. For example, categorization of sounds for lexical retrieval or for voice recognition requires a combination of filtering (to ignore irrelevant features) and grouping (to allow for variation in the coding features) that only more complex and non-linear STRFs could achieve. Finally, it is clear that selectivity for natural sound features and vocalizations has both innate and learned components, and the relative importance of each factor is an active area of research.

Towards natural hearing

Most of the neurophysiology research described above relied on the passive playback of isolated sounds to animals that were either anaesthetized or restrained, but natural hearing often involves attention and action on the part of the sender and receiver, such as in bat echolocation¹¹⁴, the interpretation of alarm calls originating from different individuals¹¹⁵ or communication between mates in bonding behaviours¹¹⁶. Moreover, natural hearing also involves the processing of complex auditory scenes. Until recently, the natural sounds that have been analysed or used in laboratory experiments have been mostly free of natural noise or natural degradation. In the real world, communication signals are most often perceived in unfavourable listening conditions, in which there is background noise, distortions owing to propagation and echoes¹¹⁷, and superimposition from other potential acoustical signals¹¹⁸. Vocal communication and auditory perception are also affected by the social context, such as in the audience effect¹¹⁹, or by internal states, such as stress levels. These social and emotional cues can also be mediated by other sensory modalities.

Advances in chronic neural recording techniques have enabled researchers to begin to examine neural processing in these more natural scenarios. Researchers have shown how responses in the primary auditory cortex are influenced by both expectations of natural structure in the sound and behavioural relevance, both of which might involve top-down modulations⁷². Chronic recordings in awake and vocalizing animals have also been used to obtain neural responses in auditory areas to the animal's own vocalizations. Such experiments have been carried out in bats^{120,121}, primates¹²² and birds¹²³. The experiments in bats were crucial for understanding how the pulse–echo pair was processed by the auditory system and are landmark experiments in that field. In primates and birds, these awake recordings gave us unique insights into how self-vocalizations are processed for self-monitoring and, in birds, potentially for guiding vocal learning. However, neural recordings in both senders and receivers in the midst of vocal communication bouts, such as antiphonal calling in marmosets¹²⁴ or duets in social songbirds^{116,125}, have not yet been performed. Such experiments could be carried out in the near future and are needed to increase our understanding of the computations performed by the auditory system for extracting the information content of communication calls.

The auditory processing of communication sounds in noisy backgrounds or in complex auditory scenes is also an active research area¹²⁶. For example, noise-invariant neurons — that is, neurons that respond to a given stimulus without being influenced by the presence of background noise — have been described in the secondary auditory areas of songbirds^{30,127} and in primary auditory areas in humans¹²⁸. Noise-invariance has also been shown to emerge in the auditory processing stream as a result of adaptive mechanisms for particular stimulus statistics¹²⁹. Similarly, responses in the human auditory cortex use a gain control to emphasize the temporal modulations that are characteristic of speech.

Neurophysiological studies in primates^{130,131} and birds¹³² have also begun to unravel how multiple auditory streams could be represented in the auditory system.

In summary, auditory neuroscientists have mostly focused their attention on understanding the computations that are needed to recognize and categorize natural sounds passively, and much more research is needed to understand how acoustical signals are processed in active communication and in natural soundscapes. Neurophysiological research in this area is in its infancy but, given the increase in our knowledge achieved from classical playback experiments and the technical advances in chronic recordings, natural hearing research is poised to make giant leaps in the near future.

Conclusion

The use of natural sounds (and in particular conspecific sounds) has had a long tradition in neuroethological research, and the findings in these model systems have inspired the more recent development of analytical techniques for both sound analysis and neural data analysis. These developments have enabled auditory neuroscientists to use natural sound stimuli to describe and understand in much greater detail the neural responses of higher-level auditory neurons in both specialists such as crickets, bats and songbirds and in generalists such as guinea pigs, cats, ferrets and non-human primates. Sounds with natural statistics seem to be optimally represented at multiple levels of the auditory system, and stimulation with natural sounds has facilitated the characterizations of the stimulus–response functions for neurons that respond poorly to white noise or other simple synthetic sounds. Thus, for the systematic characterization of stimulus–response function, the use of simple synthetic sounds is no longer required and should even be discouraged. Conversely, complex synthetic sounds that preserve particular natural statistics and that are designed to systematically investigate the importance of natural statistics provide an additional and powerful insight. Auditory neuroscientists have also been able to begin to relate auditory representations to specific computations that are needed for recognizing and categorizing behaviourally relevant sounds, such as communication calls.

These past successes will facilitate the design and interpretation of even more naturalistic experiments. In the near future, we see five areas of promising scientific explorations: non-linear computations for invariant representation of communication calls, neurophysiological research in humans, auditory scene analysis, social and multimodal effects, and investigations of the ensemble neural code. First, neural recordings in animals that actively communicate with other animals will permit both the natural investigation of robust neural representation for call types and a direct assessment of the relationship between sound and meaning. Second, advances in both invasive^{133,134} and non-invasive¹³⁵ neurophysiological recordings in humans will further enable researchers to make links between animal work and human work. Given the wealth of knowledge in speech and music processing in humans, these links will greatly help with the

challenge of understanding the sound-to-meaning transformations that occur in the auditory system. Third, it is still unclear how the auditory system detects, recognizes and classifies behaviourally relevant signals with degraded signals and multiple sound sources; neural recordings not only with natural sounds but in the natural environment (that is, in the field) could be performed to study how naturally propagated and corrupted signals are represented. However, this line of research will also require the statistical characterization of natural auditory scenes, which is a particularly challenging problem. Fourth, we know that communication behaviour and auditory perception depend on the social and emotional context and that the physiology of the auditory system can be modulated by hormones¹³⁶. But how the neural code for natural sounds is affected by naturalistic stimuli from other modalities such as vision or self-motion¹³⁷ or by modulatory effects from brain systems involved in emotional or stress responses remains in large part unexplored. Last, although neurophysiologists are now regularly recording the simultaneous activity of many neurons, the role of correlated activity in the ensemble neural code is still unknown¹³⁸. One apparently insurmountable difficulty for studying the ensemble neural code is the explosion in the dimensionality of the problem as a result of combinatorial effects: the number of

potential neural activity patterns across neurons becomes so large that investigating the potential role of such patterns becomes impossible. For example, if a single neuron can reliably represent information with 10 different patterns, the code from two such neurons could represent 100 patterns, the code from three neurons 1,000 patterns, and so forth. For these combinatorial ensemble patterns to carry unique information about the stimulus (that is, information beyond that obtained from the individual responses), the response in one neuron must be correlated with the response in another neuron. A recent statistical analysis of neural patterns recorded in the visual system under natural stimulation showed that ensemble neural responses are indeed correlated but very sparse¹³⁹. In other words, natural scenes seemed to be represented with very few response patterns from all possible combinations that could be possible. These experiments and analyses suggest that using natural stimuli might be the only way to resolve this dimensionality curse. Although this position might be extreme, given the important role that natural sounds have already had in understanding the auditory system and that questions in natural hearing will require further investigations with natural sounds, auditory neuroscientists might also be well placed to elucidate the nature of the ensemble neural code in sensory systems.

1. Darrigol, O. Number and measure: Hermann von Helmholtz at the crossroads of mathematics, physics, and psychology. *Stud. Hist. Philos. Sci.* **34**, 515–573 (2003).
2. Robles, L. & Ruggero, M. A. Mechanics of the mammalian cochlea. *Physiol. Rev.* **81**, 1305–1352 (2001).
3. Woolley, S. M. & Casseday, J. H. Response properties of single neurons in the zebra finch auditory midbrain: response patterns, frequency coding, intensity coding, and spike latencies. *J. Neurophysiol.* **91**, 136–151 (2004).
4. Eggermont, J. J. Between sound and perception: reviewing the search for a neural code. *Hear. Res.* **157**, 1–42 (2001).
5. Suga, N., O'Neill, W. E. & Manabe, T. Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the moustache bat. *Science* **200**, 778–781 (1978).
6. Margoliash, D. & Fortune, E. S. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J. Neurosci.* **12**, 4309–4326 (1992).
7. Margoliash, D. & Konishi, M. Auditory representation of autogenous song in the song system of white-crowned sparrows. *Proc. Natl Acad. Sci. USA* **82**, 5997–6000 (1985).
8. Nelken, I., Rotman, Y. & Bar Yosef, O. Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* **397**, 154–157 (1999). **This study reports a very surprising result: auditory cortical neurons can be more sensitive to the natural auditory context than to the main component of the sound signal.**
9. Hsu, A., Woolley, S. M., Fremouw, T. E. & Theunissen, F. E. Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *J. Neurosci.* **24**, 9201–9211 (2004). **Using estimations of mutual information, this study shows that sounds that are otherwise identical in intensity and frequency but have a modulation power spectrum with natural distributions are more efficiently encoded in the avian auditory cortex.**
10. Rieke, F., Bodnar, D. A. & Bialek, W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B* **262**, 259–265 (1995).
11. Hedwig, B. Pulses, patterns and paths: neurobiology of acoustic behaviour in crickets. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **192**, 677–689 (2006).
12. Arcadi, A. C., Robert, D. & Boesch, C. Buttress drumming by wild chimpanzees: temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* **39**, 505–518 (1998).
13. Voss, R. F. & Clarke, J. *1/f* noise in music and speech. *Nature* **258**, 317–318 (1975). **In this first study on the statistical structure in music and speech sounds, Voss and Clarke show that the amplitude envelope and other features of these natural sounds follow a power law relationship and discuss the significance of this relationship.**
14. Attias, H. & Schreiner, C. E. in *Advances in Neural Information Processing Systems* (eds Mozer, M. C., Jordan, M. I. & Petsche, T.) 27–33 (MIT Press, 1997).
15. Singh, N. C. & Theunissen, F. E. Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* **114**, 3394–3411 (2003). **This paper introduces the joint temporal and spectral modulation power spectrum of sounds and shows that natural sounds have a characteristic signature in this representation.**
16. Chen, J. D., Paliwal, K. K. & Nakamura, S. Cepstrum derived from differentiated power spectrum for robust speech recognition. *Speech Comm.* **41**, 469–484 (2003).
17. Cohen, L. in *Time-Frequency Analysis* Ch. 3 50–52 (Prentice Hall, 1995).
18. Bialek, W., Nemenman, I. & Tishby, N. Predictability, complexity, and learning. *Neural Comput.* **13**, 2409–2463 (2001).
19. Elliott, T. M., Hamilton, L. S. & Theunissen, F. E. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *J. Acoust. Soc. Am.* **133**, 389–404 (2013).
20. Garcia-Lazaro, J. A., Ahmed, B. & Schnupp, J. W. H. Emergence of tuning to natural stimulus statistics along the central auditory pathway. *PLoS ONE* **6**, e22584 (2011).
21. Srivastava, A., Lee, A. B., Simoncelli, E. P. & Zhu, S. C. On advances in statistical modeling of natural images. *J. Math. Imaging Vis.* **18**, 17–33 (2003).
22. Ruderman, D. L. Origins of scaling in natural images. *Vis. Res.* **37**, 3385–3398 (1997).
23. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
24. Rodriguez, F. A., Chen, C., Read, H. L. & Escabi, M. A. Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. *J. Neurosci.* **30**, 15969–15980 (2010). **In this study, Rodriguez et al. show that the modulation tuning of neurons in the inferior colliculus balances out the 1/f relationship found in the modulation power spectrum of natural sounds to transmit modulation information efficiently, with reduced redundancy.**
25. Woolley, S. M., Fremouw, T. E., Hsu, A. & Theunissen, F. E. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neurosci.* **8**, 1371–1379 (2005).
26. Gerhardt, H. C. & Huber, F. (eds) *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions* Ch. 4 82–128 (University of Chicago Press, 2002).
27. Koppl, C., Gleich, O. & Manley, G. A. An auditory fovea in the barn owl cochlea. *J. Comp. Physiol.* **171**, 695–704 (1993).
28. Bruns, V. & Schmieszek, E. Cochlear innervation in the greater horseshoe bat: demonstration of an acoustic fovea. *Hear. Res.* **3**, 27–43 (1980).
29. Lewicki, M. S. Efficient coding of natural sounds. *Nature Neurosci.* **5**, 356–363 (2002). **Lewicki shows that the particular decomposition of sound into frequency channels found at the level of the auditory nerve results in an efficient representation of a mixture of animal communication sounds and environmental sounds.**
30. Moore, R. C., Lee, T. & Theunissen, F. E. Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. *PLoS Comput. Biol.* **9**, e1002942 (2013). **This study demonstrates the presence of neurons that robustly encode song in noise in higher levels of the avian auditory system and, inspired by the spectrotemporal filtering of these neurons, proposes an algorithm for real-time noise reduction.**

31. McDermott, J. H., Schemitsch, M. & Simoncelli, E. P. Summary statistics in auditory perception. *Nature Neurosci.* **16**, 493–498 (2013).
McDermott et al. show that the acoustical statistical structure perceived depends on the length of the sound and that, for longer, complex sounds, only averaged statistics are perceived, leading to percepts of sound texture.
32. Smith, E. C. & Lewicki, M. S. Efficient auditory coding. *Nature* **439**, 978–982 (2006).
33. Marmarelis, P. Z. & Marmarelis, V. Z. *Analysis of Physiological Systems: The White Noise Approach* (Plenum, 1978).
34. Aertsen, A. M. & Johannesma, P. I. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol. Cybern.* **42**, 133–143 (1981).
35. Eggermont, J. J., Aertsen, A. M. & Johannesma, P. I. Quantitative characterisation procedure for auditory neurons based on the spectro-temporal receptive field. *Hear. Res.* **10**, 167–190 (1983).
36. Eggermont, J. J., Aertsen, A. M. & Johannesma, P. I. Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field. *Hear. Res.* **10**, 191–202 (1983).
37. Schafer, M., Rubsamen, R., Dorrscheidt, G. J. & Knipschild, M. Setting complex tasks to single units in the avian auditory forebrain. II. Do we really need natural stimuli to describe neuronal response characteristics? *Hear. Res.* **57**, 231–244 (1992).
38. Andoni, S., Li, N. & Pollak, G. D. Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *J. Neurosci.* **27**, 4882–4893 (2007).
39. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nature Neurosci.* **6**, 391–398 (2003).
40. Asari, H. & Zador, A. M. Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J. Neurophysiol.* **102**, 2638–2656 (2009).
41. Woolley, S. M., Gill, P. R. & Theunissen, F. E. Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *J. Neurosci.* **26**, 2499–2512 (2006).
42. Theunissen, F. E., Sen, K. & Doupe, A. J. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* **20**, 2315–2331 (2000).
This methods paper describes the analytical solution for the simplest normalized and regularized linear regression solution for the estimation of STRFs from responses to natural stimuli.
44. Sahani, M. & Linden, J. in *Advances in Neural Information Processing Systems* (eds Becker, S., Thrun, S. & Obermeyer, K.) 301–308 (MIT Press, 2003).
45. Hoerl, A. E. & Kennard, R. W. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **42**, 80–86 (2000).
46. David, S. V., Mesgarani, N. & Shamma, S. A. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* **18**, 191–212 (2007).
47. Christianson, G. B., Sahani, M. & Linden, J. F. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J. Neurosci.* **28**, 446–455 (2008).
48. Schneider, D. M. & Woolley, S. M. N. Extra-classical tuning predicts stimulus-dependent receptive fields in auditory neurons. *J. Neurosci.* **31**, 11867–11878 (2011).
49. Gill, P., Zhang, J., Woolley, S. M., Fremouw, T. & Theunissen, F. E. Sound representation methods for spectro-temporal receptive field estimation. *J. Comput. Neurosci.* **21**, 5–20 (2006).
50. David, S. V. & Shamma, S. A. Integration over multiple timescales in primary auditory cortex. *J. Neurosci.* **33**, 19154–19166 (2013).
51. Gill, P., Woolley, S. M., Fremouw, T. & Theunissen, F. E. What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. *J. Neurophysiol.* **99**, 2809–2820 (2008).
52. Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L. & Woolley, S. M. N. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE* **6**, e16104 (2011).
53. McFarland, J. M., Cui, Y. W. & Butts, D. A. Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Comput. Biol.* **9**, e1003143 (2013).
In this study, Atencio et al. use the maximally informative dimension algorithm to extract multicomponent STRFs and show that the number of components needed to characterize auditory neurons is greater in the cortex than in the thalamus.
54. Sharpee, T., Rust, N. C. & Bialek, W. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput.* **16**, 223–250 (2004).
55. Atencio, C. A., Sharpee, T. O. & Schreiner, C. E. Receptive field dimensionality increases from the auditory midbrain to cortex. *J. Neurophysiol.* **107**, 2594–2603 (2012).
56. Depireux, D. A. & Elhilali, M. (eds) *Handbook of Modern Techniques in Auditory Cortex* (Nova Science, 2014).
57. Ahrens, M. B., Linden, J. F. & Sahani, M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J. Neurosci.* **28**, 1929–1942 (2008).
58. Woolley, S. M., Gill, P. R., Fremouw, T. & Theunissen, F. E. Functional groups in the avian auditory system. *J. Neurosci.* **29**, 2780–2793 (2009).
This study describes the STRFs found in the avian inferior colliculus and avian auditory cortex and shows how they form unique functional groups that extract distinct features of natural sounds that are in turn important for distinct percepts.
59. Amin, N., Gill, P. & Theunissen, F. E. Role of the zebra finch auditory thalamus in generating complex representations for natural sounds. *J. Neurophysiol.* **104**, 784–798 (2010).
60. Nagel, K. I. & Doupe, A. J. Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron* **58**, 938–955 (2008).
61. Kim, G. & Doupe, A. Organized representation of spectrotemporal features in songbird auditory forebrain. *J. Neurosci.* **31**, 16977–16990 (2011).
62. Miller, L. M., Escabi, M. A., Read, H. L. & Schreiner, C. E. Functional convergence of response properties in the auditory thalamocortical system. *Neuron* **32**, 151–160 (2001).
This study compares the STRFs in the mammalian auditory thalamus with those in the auditory cortex, illustrating the hierarchical processing found in the ascending auditory processing stream.
63. Miller, L. M., Escabi, M. A., Read, H. L. & Schreiner, C. E. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J. Neurophysiol.* **87**, 516–527 (2002).
64. Depireux, D. A., Simon, J. Z., Klein, D. J. & Shamma, S. A. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.* **85**, 1220–1234 (2001).
65. Chi, T., Ru, P. & Shamma, S. A. Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**, 887–906 (2005).
66. Escabi, M. A. & Schreiner, C. E. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J. Neurosci.* **22**, 4114–4131 (2002).
67. Escabi, M. A., Miller, L. M., Read, H. L. & Schreiner, C. E. Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J. Neurosci.* **23**, 11489–11504 (2003).
68. Carlson, N. L., Ming, V. L. & Deweese, M. R. Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. *PLoS Comput. Biol.* **8**, e1002594 (2012).
69. 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).
70. David, S. V., Fritz, J. B. & Shamma, S. A. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc. Natl Acad. Sci. USA* **109**, 2144–2149 (2012).
71. Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H. & King, A. J. Spectrotemporal contrast kernels for neurons in primary auditory cortex. *J. Neurosci.* **32**, 11271–11284 (2012).
72. Mesgarani, N., David, S. V., Fritz, J. B. & Shamma, S. A. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J. Neurophysiol.* **102**, 3329–3339 (2009).
73. Amin, N., Gastpar, M. & Theunissen, F. E. Selective and efficient neural coding of communication signals depends on early acoustic and social environment. *PLoS ONE* **8**, e61417 (2013).
74. Mesgarani, N., Slaney, M. & Shamma, S. A. Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations. *IEEE Trans. Speech Audio Process.* **14**, 920–930 (2006).
75. Mesgarani, N., David, S. V., Fritz, J. B. & Shamma, S. A. Phoneme representation and classification in primary auditory cortex. *J. Acoust. Soc. Am.* **123**, 899–909 (2008).
76. McDermott, J. H. & Simoncelli, E. P. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* **71**, 926–940 (2011).
77. Libersat, F., Murray, J. A. & Hoy, R. R. Frequency as a releaser in the courtship song of 2 crickets, *Gryllus bimaculatus* (de Geer) and *Teleogryllus oceanicus*: a neuroethological analysis. *J. Comp. Physiol. A* **174**, 485–494 (1994).
78. Feng, A. S., Hall, J. C. & Gooler, D. M. Neural basis of sound pattern-recognition in anurans. *Prog. Neurobiol.* **34**, 313–329 (1990).
79. 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).
80. McCasland, J. S. & Konishi, M. Interactions between auditory and motor activities in an avian song control nucleus. *Proc. Natl Acad. Sci. USA* **78**, 7815–7819 (1981).
81. Doupe, A. J. & Konishi, M. Song-selective auditory circuits in the vocal control system of the zebra finch. *Proc. Natl Acad. Sci. USA* **88**, 11339–11343 (1991).
82. Grace, J. A., Amin, N., Singh, N. C. & Theunissen, F. E. Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophysiol.* **89**, 472–487 (2003).
83. Newman, J. & Wollberg, Z. Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res.* **54**, 287–304 (1978).
84. Huetz, C., Gourevitch, B. & Edeline, J. M. Neural codes in the thalamocortical auditory system: from artificial stimuli to communication sounds. *Hear. Res.* **271**, 147–158 (2011).
85. Ter-Mikaelian, M., Semple, M. N. & Sanes, D. H. Effects of spectral and temporal disruption on cortical encoding of gerbil vocalizations. *J. Neurophysiol.* **110**, 1190–1204 (2013).
86. Suta, D., Popelar, J. & Syka, J. Coding of communication calls in the subcortical and cortical structures of the auditory system. *Physiol. Res.* **57** (Suppl. 3), 149–159 (2008).
87. Wallace, M. N., Grimsley, J. M., Anderson, L. A. & Palmer, A. R. Representation of individual elements of a complex call sequence in primary auditory cortex. *Front. Syst. Neurosci.* **7**, 72 (2013).
88. Theunissen, F. E. & Doupe, A. J. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J. Neurosci.* **18**, 3786–3802 (1998).
89. Lewicki, M. S. & Konishi, M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl Acad. Sci. USA* **92**, 5582–5586 (1995).
90. Marler, P. R. Avian and primate communication: the problem of natural categories. *Neurosci. Biobehav. Rev.* **6**, 87–94 (1982).
91. Walker, K. M. M., Bizley, J. K., King, A. J. & Schnupp, J. W. H. Multiplexed and robust representations of sound features in auditory cortex. *J. Neurosci.* **31**, 14565–14576 (2011).
92. Meliza, C. D. & Margoliash, D. Emergence of selectivity and tolerance in the avian auditory cortex. *J. Neurosci.* **32**, 15158–15168 (2012).
93. George, I., Cousillas, H., Richard, J. P. & Hausberger, M. A potential neural substrate for processing functional classes of complex acoustic signals. *PLoS ONE* **3**, e2203 (2008).
94. Petkov, C. I. et al. A voice region in the monkey brain. *Nature Neurosci.* **11**, 367–374 (2008).
95. Perrodin, C., Kayser, C., Logothetis, N. K. & Petkov, C. I. Voice cells in the primate temporal lobe. *Curr. Biol.* **21**, 1408–1415 (2011).
96. 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).

97. 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).
98. Cohen, Y. E., Theunissen, F., Russ, B. E. & Gill, P. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J. Neurophysiol.* **97**, 1470–1484 (2007).
99. 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).
100. Marler, P. Bird calls: their potential for behavioral neurobiology. *Ann. NY Acad. Sci.* **1016**, 31–44 (2004).
101. Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S. & Mooney, R. Neural correlates of categorical perception in learned vocal communication. *Nature Neurosci.* **12**, 221–228 (2009).
102. Marler, P. Birdsong: the acquisition of a learned motor skill. *Trends Neurosci.* **4**, 88–94 (1981).
103. Seyfarth, R. M. & Cheney, D. L. Vocal development in vervet monkeys. *Animal Behav.* **34**, 1640–1658 (1986).
104. Miranda, J. A. & Liu, R. C. Dissecting natural sensory plasticity: hormones and experience in a maternal context. *Hear. Res.* **252**, 21–28 (2009).
105. Menardy, F. *et al.* Social experience affects neuronal responses to male calls in adult female zebra finches. *Eur. J. Neurosci.* **35**, 1322–1336 (2012).
106. Gentner, T. Q. & Margoliash, D. Neuronal populations and single cells representing learned auditory objects. *Nature* **424**, 669–674 (2003).
107. Woolley, S. M. N., Hauber, M. E. & Theunissen, F. E. Developmental experience alters information coding in auditory midbrain and forebrain neurons. *Dev. Neurobiol.* **70**, 235–252 (2010).
108. Hauber, M. E., Woolley, S. M. N., Cassey, P. & Theunissen, F. E. Experience dependence of neural responses to different classes of male songs in the primary auditory forebrain of female songbirds. *Behav. Brain Res.* **243**, 184–190 (2013).
109. Sanes, D. H. & Bao, S. W. Tuning up the developing auditory CNS. *Curr. Opin. Neurobiol.* **19**, 188–199 (2009).
110. Doupe, A. J. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.* **17**, 1147–1167 (1997).
111. Solis, M. M. & Doupe, A. J. Contributions of tutor and bird's own song experience to neural selectivity in the songbird anterior forebrain. *J. Neurosci.* **19**, 4559–4584 (1999).
112. Volman, S. F. Development of neural selectivity for birdsong during vocal learning. *J. Neurosci.* **13**, 4737–4747 (1993).
113. Kover, H., Gill, K., Tseng, Y. T. L. & Bao, S. W. Perceptual and neuronal boundary learned from higher-order stimulus probabilities. *J. Neurosci.* **33**, 3699–3705 (2013).
114. Moss, C. F. & Surlykke, A. Probing the natural scene by echolocation in bats. *Front. Behav. Neurosci.* **4**, e00033 (2010).
115. Cheney, D. L. & Seyfarth, R. M. Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Animal Behav.* **36**, 477–486 (1988).
116. Elie, J. E. *et al.* Vocal communication at the nest between mates in wild zebra finches: a private vocal duet? *Animal Behav.* **80**, 597–605 (2010).
117. Penna, M., Llusia, D. & Marquez, R. Propagation of natural toad calls in a Mediterranean terrestrial environment. *J. Acoust. Soc. Am.* **132**, 4025–4031 (2012).
118. Aubin, T. & Jouventin, P. In *Advances in the Study of Behavior* Vol. 31 (eds Slater, P. J. B., Rosenblatt, J. S., Snowdon, C. T. & Roper, T. J.) 243–277 (Academic, 2002).
119. Vignal, C., Mathevon, N. & Mottin, S. Audience drives male songbird response to partner's voice. *Nature* **430**, 448–451 (2004).
120. Kawasaki, M., Margoliash, D. & Suga, N. Delay-tuned combination-sensitive neurons in the auditory-cortex of the vocalizing mustached bat. *J. Neurophysiol.* **59**, 623–635 (1988).
121. Suga, N. & Shimoza, T. Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science* **183**, 1211–1213 (1974).
122. Eliades, S. J. & Wang, X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* **453**, 1102–1106 (2008).
123. Keller, G. B. & Hahnloser, R. H. Neural processing of auditory feedback during vocal practice in a songbird. *Nature* **457**, 187–190 (2009).
124. Miller, C. T. & Wang, X. Q. Sensory–motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **192**, 27–38 (2006).
125. Fortune, E. S., Rodriguez, C., Li, D., Ball, G. F. & Coleman, M. J. Neural mechanisms for the coordination of duet singing in wrens. *Science* **334**, 666–670 (2011).
126. Lewicki, M. S., Olshausen, B. A., Surlykke, A. & Moss, C. F. Scene analysis in the natural environment. *Front. Psychol.* **5**, 199 (2014).
127. Schneider, D. M. & Woolley, S. M. N. Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* **79**, 141–152 (2013).
128. Ding, N. & Simon, J. Z. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* **33**, 5728–5735 (2013).
129. Rabinowitz, N. C., Willmore, B. D. B., King, A. J. & Schnupp, J. W. H. Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol.* **11**, 1710–1710 (2013).
This study shows how increasing levels of adaptation as one ascends the auditory processing stream result in noise-invariant representation of signals at the higher levels.
130. Middlebrooks, J. C. & Bremen, P. Spatial stream segregation by auditory cortical neurons. *J. Neurosci.* **33**, 10986–11001 (2013).
131. 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).
132. Bee, M. A. & Klump, G. M. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol.* **92**, 1088–1104 (2004).
133. Greenlee, J. D. W. *et al.* Human auditory cortical activation during self-vocalization. *PLoS ONE* **6**, e14744 (2011).
134. Pasley, B. N. *et al.* Reconstructing speech from human auditory cortex. *PLoS Biol.* **10**, e1001251 (2012).
135. Regev, M., Honey, C. J., Simony, E. & Hasson, U. Selective and invariant neural responses to spoken and written narratives. *J. Neurosci.* **33**, 15978–15988 (2013).
136. Al-Mana, D., Ceranic, B., Djahanbakhch, O. & Luxon, L. M. Hormones and the auditory system: a review of physiology and pathophysiology. *Neuroscience* **153**, 881–900 (2008).
137. Teramoto, W., Sakamoto, S., Furune, F., Gyoba, J. & Suzuki, Y. Compression of auditory space during forward self-motion. *PLoS ONE* **7**, e39402 (2012).
138. Beckers, G. J. L. & Gahr, M. Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *J. Neurosci.* **32**, 10594–10608 (2012).
139. Schneidman, E. *et al.* Synergy from silence in a combinatorial neural code. *J. Neurosci.* **31**, 15732–15741 (2011).

Acknowledgements

The authors thank the members of the Theunissen laboratory and the three anonymous reviewers for critical feedback and suggestions.

Competing interests statement

The authors declare no competing interests.