



## Review

**Cite this article:** Itatani N, Klump GM. 2017  
Animal models for auditory streaming. *Phil.  
Trans. R. Soc. B* **372**: 20160112.  
<http://dx.doi.org/10.1098/rstb.2016.0112>

Accepted: 31 July 2016

One contribution of 15 to a theme issue  
'Auditory and visual scene analysis'.

### Subject Areas:

behaviour, neuroscience

### Keywords:

auditory scene analysis, subjective measures,  
objective measures, behaviour, neuronal  
representation

### Author for correspondence:

Georg M. Klump  
e-mail: [georg.klump@uni-oldenburg.de](mailto:georg.klump@uni-oldenburg.de)

# Animal models for auditory streaming

Naoya Itatani and Georg M. Klump

Cluster of Excellence Hearing4all, Animal Physiology and Behaviour Group, Department of Neuroscience, School of Medicine and Health Sciences, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany

GMK, 0000-0003-2501-8090

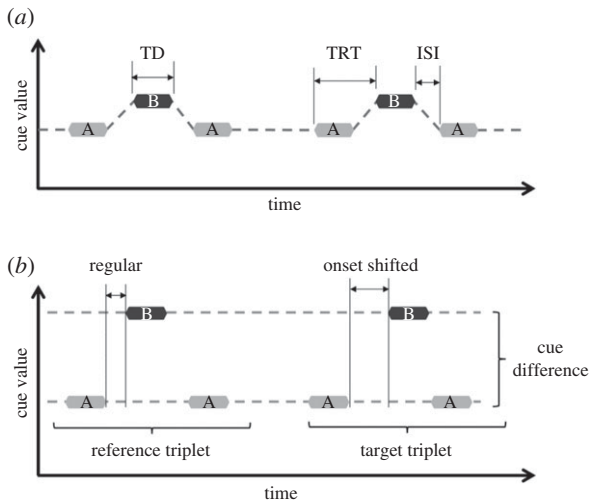
Sounds in the natural environment need to be assigned to acoustic sources to evaluate complex auditory scenes. Separating sources will affect the analysis of auditory features of sounds. As the benefits of assigning sounds to specific sources accrue to all species communicating acoustically, the ability for auditory scene analysis is widespread among different animals. Animal studies allow for a deeper insight into the neuronal mechanisms underlying auditory scene analysis. Here, we will review the paradigms applied in the study of auditory scene analysis and streaming of sequential sounds in animal models. We will compare the psychophysical results from the animal studies to the evidence obtained in human psychophysics of auditory streaming, i.e. in a task commonly used for measuring the capability for auditory scene analysis. Furthermore, the neuronal correlates of auditory streaming will be reviewed in different animal models and the observations of the neurons' response measures will be related to perception. The across-species comparison will reveal whether similar demands in the analysis of acoustic scenes have resulted in similar perceptual and neuronal processing mechanisms in the wide range of species being capable of auditory scene analysis.

This article is part of the themed issue 'Auditory and visual scene analysis'.

## 1. Introduction

Similarly to human subjects, animals often face the 'Cocktail Party Problem', i.e. listeners need to process signals from individual sources in a complex acoustic environment with many simultaneously active signallers [1]. Thus, animals should have evolved processing mechanisms that allow the grouping of components of sounds from a specific source and the segregation of components of sounds from different sources as has been described for human auditory scene analysis (ASA) [2]. Auditory streaming paradigms denote a set of experimental procedures that test the ability for ASA. Given that similar demands for ASA are on many animals and humans, animal studies can demonstrate whether human and animal auditory streaming shares many features and can possibly reveal the neural mechanisms underlying ASA on a cellular basis.

In human psychophysics, auditory streaming of simultaneous sounds and sequential sounds has been studied involving segregation of components from the different sources, and integration of components from the same source [2,3]. In this review, we will focus on the auditory streaming of sound sequences. Segregation of sounds from simultaneously active sources may involve a different set of mechanisms from streaming of sequences and a different time frame. Simultaneous source segregation will operate on short-term comparison of stimulus features such as harmonicity of frequency components or the coherent pattern of rapid modulation of signal components [1]. Auditory streaming of sequential sounds covers much longer time periods and integrates stimulus features over a range of seconds for forming hypotheses about the composition of acoustic scenes from different sources. Sequential auditory streaming has been described as a dynamic process that develops over time which has been demonstrated by the build-up of stream segregation [4,5] and the switching between integration and segregation of streams over a period of minutes [6]. We will concentrate on paradigms involving sequentially presented sounds as in the classical psychophysical study by van Noorden [7]



**Figure 1.** ABA-triplet stimuli. (a) Regular ABA-triplet sequence. A and B represent sounds differing, for example, in frequency, location, temporal structure or any other auditory cue. Additional parameters such as tone duration (TD), tone repetition time (TRT) and inter-stimulus interval (ISI) are also indicated. The dotted line indicates the linking of the stimuli during a one-stream percept by which listeners perceive a ‘galloping’ rhythm. (b) ABA-triplet sequence with a temporal irregularity in the second triplet. An onset shift of the B tone in the target triplet is introduced and the response to the irregularity is used for obtaining an objective measure of stream segregation. The dotted line indicates the linking of the stimuli while listeners perceive two separate auditory streams, each with an isochronous rhythm. The salience of the cue difference will determine whether listeners have a one-stream or a two-stream percept.

forming the foundation for the wide field of research on auditory streaming (for an example see figure 1). Here, we will first introduce the cues and conditions affecting auditory streaming and provide an overview of behavioural studies in different vertebrate classes. Then, we will compare the evidence in the electrophysiological studies regarding the effect of different cues on stream segregation and the temporal dynamics of the neurons’ response to the evidence from the behavioural studies. In the end, experiments involving simultaneous recording of neuronal response pattern and evaluation of the percept revealed by the animal’s behaviour are described that provide the most direct approach for unravelling the mechanisms underlying stream segregation.

## 2. Auditory cues affecting stream segregation

The auditory streaming percept is thought to rely on two general types of mechanisms, so-called ‘primitive’ and ‘schema-based’ mechanisms [2]. Schema-based mechanisms involve top-down processing relying, for example, on inherited or previously acquired templates for signals. Schema-based mechanisms may be modulated by attention. Primitive mechanisms involve pre-attentive processing in which a stimulus-driven representation of auditory cues provides the basis for stream segregation. Previous studies in humans revealed that differences in the spatial pattern of excitation in the ear provide potent cues for auditory streaming, but also streams of signals with overlapping patterns of excitation can be segregated by processing other cues such as temporal or spatial cues on higher levels of the auditory pathway [5]. Frequency separation (e.g. [7]), intensity differences (e.g. [8]) and spectral differences of sounds (e.g. [9]) may already serve segregating streams

based on peripheral channelling representing the spatial pattern of excitation in the ear [9]. Spatial separation of competing sounds can serve to segregate streams with humans exploiting interaural time differences (ITDs) below 1600 Hz and interaural intensity differences (ILDs) above 4 kHz (e.g. [10]). The temporal structure of sounds can also be exploited by humans for stream segregation as was demonstrated by stream segregation based on the difference of temporal envelope (e.g. [11]), phase spectrum (e.g. [12]) and fundamental frequency of harmonic tone complexes (e.g. [13,14]). Stream segregation can be abolished by cues that support the integration of sounds from the different streams. For example, common onset of sounds is a grouping cue that can completely override frequency cues that would result in perceiving segregated streams of tones [15]. This indicates that auditory streaming depends on the weighting of different cues supporting the segregation or integration of sounds into streams. A number of the cues relevant to stream segregation in human subjects have also been shown to be relevant for animals, and a number of animal studies have demonstrated neural correlates of auditory stream segregation by specific cues used by human subjects.

## 3. Behavioural studies of auditory streaming

In human psychophysical research, generally two types of experimental paradigms are used: one involving subjective and one involving objective measures [16,17]. For obtaining a subjective measure of stream segregation, listeners are required to directly indicate whether ongoing sound sequences are heard as separate streams or as one concurrent stream. For obtaining an objective measure of stream segregation, listeners do not directly indicate the streaming percept, but are performing a task in which sensitivity is enhanced or decreased by the streaming percept and, therefore, different perceptual thresholds can be observed [16].

In human studies involving subjective measures, subjects are often instructed how a stream can be identified by a typical patterning of the perceived sounds [5]. ‘Instruction’ in animal studies can be achieved by providing the animal subject with a template that is learned to identify a pattern or by relying on an innate template that is employed in recognizing species-specific signals. The former requires a carefully designed learning paradigm that supplies the animal subject with an acquired reference pattern whereas the latter requires an excellent knowledge of the natural behavioural context in which an innate reference pattern is usually involved. For example, in the goldfish a learned template can be generated by classical conditioning of a physiological response elicited by an aversive stimulus that then obtains a behavioural significance [18]. In songbirds, the species-specific song template is learned in a developmental process [19]. Templates can also be innate as, for example, is demonstrated by genetically determined release mechanisms that trigger gray treefrogs’ mating behaviour and that requires no previous learning [20]. The behavioural response relying on the template can then be employed to conclude whether the animal subject perceived signal components as belonging to one stream or to separate streams.

For obtaining objective measures of auditory streaming, a behavioural task is chosen in which the streaming percept influences the subject’s performance and, thereby, the animal’s performance indirectly will reveal its streaming percept. If

both animals and humans can perform a specific perceptual task, this objective measure can be used to directly compare results from behavioural experiments in animals and results from human psychophysical experiments. By designing tasks in which segregation improves perception of a signal modification or in which integration improves perception of a signal modification, either the perception of signals being segregated into different streams or of signals being integrated into one stream can be studied [21].

### (a) Auditory streaming in frogs

Frogs form noisy aggregations during breeding seasons. Male frogs produce loud calls to attract females ('advertisement call'). Where choruses of different species coexist, frogs must discriminate their conspecific vocalizations from that of other species and hence frogs have been used as animal model of auditory streaming (for review, see [22,23]). Frogs apply innate templates to recognize conspecifics, and this has been used to infer the integration of sounds into one stream.

Studies in frogs make use of characteristics of natural calls eliciting a phonotactic approach behaviour indicating their auditory streaming. The advertisement calls of gray treefrogs consist of periodic pulses with the pulse rate being an indicator of the species. The advertisement calls of the eastern gray treefrog (*Hyla versicolor*) and Cope's gray treefrog (*Hyla chrysoscelis*) have overlapping spectra with a frequency range from 1 to 2.8 kHz, but the two species differ in the pulse rate of their calls with Cope's gray treefrog having double the pulse rate (35–50 pulses  $s^{-1}$ ) as the eastern gray treefrog (18–24 pulses  $s^{-1}$ ) [23]. Investigating eastern gray treefrogs, Schwartz & Gerhardt [24] constructed pulse trains with a pulse rate similar to that of natural calls. They presented calls from two separate sources with alternating pulses and observed how the phonotactic approach was affected by features that may affect stream segregation. If the frogs would integrate the pulse trains from the two sources, they would perceive the pulse rate typical for the sibling species (Cope's gray treefrog) and should not approach the loudspeaker. If they would segregate the two sources, they should be attracted to either source because the segregated sources presented the conspecific pulse rate. The result showed that eastern gray treefrogs segregated the sources if these were separated by 120° whereas the attractiveness dropped considerably if the spatial separation was reduced to 45°. If the pulses from the two sources differed in loudness, stream segregation was improved [24]. To investigate stream segregation by frequency difference in Cope's gray treefrogs, Nityananda & Bee [25] used stimuli consisting of target pulses presented at the natural rate that were interleaved with distractor pulses of the same or a different frequency. The response towards the series of target pulses was increased if the frequency separation between target and distractor pulses was increased allowing segregating target and distractor pulses. Those studies clearly demonstrate stream segregation in gray treefrogs based on spectral, intensity and spatial cues (for review of ASA in frogs see [22,23]). The role of a template in grouping sequential sounds has even been more clearly demonstrated in the Tungara frog (*Physalaemus pustulosus*) that has a complex call of a 'wine' and 'chucks' that are produced with a specific temporal relation [26]. If the natural temporal relation is not given, this may be an indication that wines and chucks are from two different sources (i.e. calling males) and should not be integrated into an

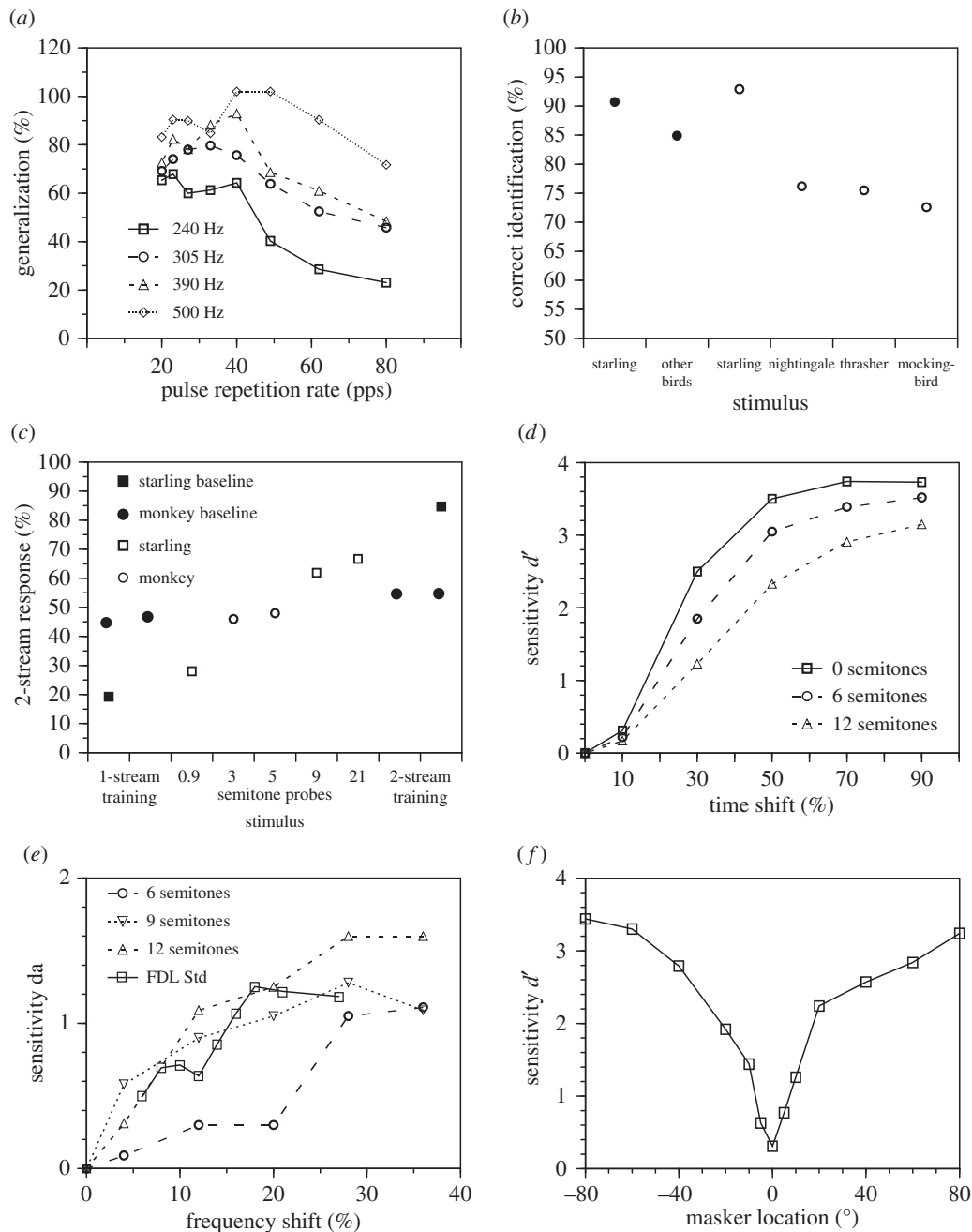
auditory stream. Thus, the relative timing of these two call components is crucial for eliciting a phonotactic response by the females. Also in this species, spatial separation has found to provide cues relevant for segregation [26].

### (b) Auditory streaming in fish

Although fish hear with a different part of the inner ear than humans (i.e. the sacculus), fish are also able to segregate signals and assign these to different sources for analysis. Fay [18,27] investigated stream segregation in the goldfish (*Carassius auratus*) by observing respiratory water flow that was modified by classical conditioning. In an initial training, a train (or a mixture of trains) of tone pulses was presented with a specific pulse rate and carrier frequency and the animal was classically conditioned to respond with a reduction of gill movement upon hearing the training sequence. In subsequent tests, Fay [18] varied the pulse rate and observed how the fish's response was related to the response conditioned with a specific pulse rate. If it was the same despite differences in the stimulus carrier frequency, the animals must have generalized. This implies that the fish heard segregated trains during conditioning with a mixture because they now responded to isolated components of the mixture. Segregation of the pulse trains by frequency was further enhanced by onset asynchrony of two trains. To specifically investigate the effect of spectral overlap on stream segregation in the goldfish, Fay [27] used a sequence of ABAB pulse trains as conditioning stimuli in which the A pulse centre frequency was fixed at 625 Hz and the B pulse centre frequency varied between 240 and 500 Hz and the overall pulse rate was held constant at 40 pulses  $s^{-1}$  (A and B rates were 20 pulses  $s^{-1}$  each; figure 2a). The fish were then tested for the generalization of the response to a pulse train composed of 625 Hz tone pulses only varying in rate between 20 and 80 pulses  $s^{-1}$ . If the fish had integrated the pulses of different frequency in the training phase, they would show the strongest generalization to 40 pulses  $s^{-1}$ . If they had segregated the pulse trains in the training, the strongest generalized response would be expected at 20 pulses  $s^{-1}$ . The goldfish's response indicated integration if the carrier frequency of the alternating pulses was 500 and 625 Hz, and increasing segregation was observed if the separation of centre frequency of the two pulse trains was increased. These results suggest that the goldfish has similar auditory stream segregation to that observed in human perception.

### (c) Auditory streaming in birds

Songbirds typically use vocal signals to communicate and doing so in the cacophony of the dawn chorus requires that birds solve the stream segregation problem. Wisniewski & Hulse [34] reported that the European starling (*Sturnus vulgaris*) could learn to discriminate song segments from two different starling individuals, even when the segments were embedded in a background of additional conspecific songs from a third individual. This suggests that the birds perceive the songs of each singer as separate auditory streams. Hulse *et al.* [28] further investigated the capability of stream segregation in the European starling by adding song segments of different bird species or a dawn chorus to starling song segments while they learned to discriminate snippets of bird call sounds that either included or did not include starling song. Starlings could pick out the snippet with the conspecific song segment even if the song of different bird



**Figure 2.** Behavioural measures of stream segregation. (a) Stream segregation in the goldfish [27]: the fish learned to associate a specific rate of tone pulses (20 or 40 pulses per second) with an electric shock as indicated by the gill response. In training, the pulses had a carrier frequency alternating between 625 Hz and the frequency indicated in the legend. In testing, only 625 Hz pulses were presented. If the two pulse frequencies are perceived as one stream, 100% generalization is expected. (b) Discrimination of songs presented in a multispecies song background by the European starling [28]. Baseline shows discrimination with a training set, probe shows the generalization to novel songs. Chance performance is 50%. (c) Perceptual segregation of two sequences of  $a$  and  $b$  tones differing in frequency by the European starling (squares [29]) and by the Rhesus macaque monkey (circles [30]). Baseline training stimuli should evoke a clear one-stream- or two-stream-like percept.  $A$  and  $B$  tone in probe sequences differed in frequency as indicated on the horizontal axis. Chance performance is 50%. (d) Objective behavioural measure of stream segregation in the European starling [31]. Sensitivity for detecting the onset time shift of the middle tone in an ABA triplet (frequency difference between  $a$  and  $b$  indicated by the legend) is expected to be reduced if  $A$  and  $B$  tones are processed in separate streams. (e) Objective measure of stream segregation in ferrets [32]. Sensitivity data for detecting a frequency shift of  $B$  tones in ABAB sequences (frequency difference between lower  $A$  and higher  $B$  tone indicated by the legend) are expected to be better if  $A$  and  $B$  tones are processed in separate streams. The FDL Std graph shows the ferret's discrimination ability for a single sequence of tones. (f) Sensitivity for discriminating two different rhythms in the cat presented from the front ( $0^\circ$ ) in relation to angle of sound incidence from which a masker rhythm is presented (averaged over six animals [33]). Sensitivity is expected to be reduced when signal and masker are processed within a single stream.

species or the dawn chorus were simultaneously presented. They even could generalize features from the songs to novel song segments that the starlings were able to discriminate without additional training. After having been trained with the mixtures, they were also able to classify individual single-species song segments as being part of the starling or non-starling containing mixtures (figure 2b). All these results demonstrate that

the starlings possibly segregate the mixed sounds and perceive them as different auditory streams. Dent and co-workers [35] investigated effects of simpler features of song elements on auditory streaming. They initially trained budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*) to discriminate a sequence of song syllables from a sequence in which one of the syllables was missing. They then presented probe



stimuli in which that syllable was presented from another spatial source location, with another sound level, time shifted, with a filtered frequency spectrum, or in which the syllable was replaced by another syllable (even from another species) to evaluate how this modified syllable was integrated into the auditory stream of song elements. The birds were expected to show fewer ‘whole song’ responses if the feature-changed syllable was segregated from the rest of the syllables (i.e. appeared not to belong to the song). Location and intensity had larger effects on the response than the other features that were tested.

Stream segregation in birds was also systematically investigated using artificial stimuli similar to those used in human psychophysical studies. MacDougall-Shackleton *et al.* [29] observed perceptual stream segregation in European starlings for series of tone pulses. Using a two-alternative forced-choice (2AFC) paradigm, they initially trained birds to discriminate a galloping rhythm produced by triplets of tones separated by a pause from two types of rhythm patterns that either presented tones at a slow or a fast isochronous rate (one-stream and two-stream baseline training; figure 2c). After initial training on that task, they presented the birds with unrewarded ABA-triplet sequences (A and B represented tones of different frequencies) and observed the birds’ choice probabilities for isochronous rhythm or galloping rhythm, respectively. They demonstrated that the birds chose the isochronous response in more than 60% of the probe trials if A and B tones differed by nine semitones or more, whereas the isochronous response was shown in only about 30% of probe trials if the frequency separation was only 0.9 semitones. Their response to the unrewarded probe triplets suggests that the birds have a subjective one-stream percept if the frequency difference between A and B is small and a subjective two-stream percept if the frequency difference between A and B is 9 semitones or above. Itatani & Klump [31] also investigated stream segregation in European starlings with ABA-triplet stimuli using an objective measure (figure 2d). Similar to the previous study, ABA-triplet sequences of tones were used with the frequency difference being 0, 6 or 12 semitones. The starling had to detect a deviant triplet in which the middle B tone onset time was slightly shifted. Birds were required to report the time shift, which has been shown to be more difficult to detect by human subjects if A and B tones are represented by separate streams [16]. Similar to human subjects, increasing frequency separation between A and B tones reduced the starlings’ sensitivity to perceive the time shift. These findings suggest that the starling perceives auditory streams of tone sequences similar to humans.

#### (d) Auditory streaming in mammals

Similar to birds, most studies on auditory stream segregation in mammals involve a learned discrimination of stimulus patterns, i.e. reflect to a larger degree the subjective perception of the pattern relying on a template. Izumi [36] trained Japanese monkeys (*Macaca fuscata*) to discriminate two target tone sequences with differing frequency contours. He then added distractor tone sequences that either spectrally overlapped or did not overlap the discriminated target tone sequences. Performance of discrimination was significantly better if distractor frequencies were non-overlapping with target tone sequences than if they were overlapping.

Christison-Lagay & Cohen [30] trained two Rhesus monkeys (*Macaca mulatta*) to report whether an ABAB tone series was

heard as one stream or two streams. To achieve a baseline discrimination, they rewarded a ‘one-stream’ response if the frequency of A and B tones differed by 1 or 0.5 semitones and rewarded a ‘two-stream’ response if the frequency of A and B tones differed by 10 or 12 semitones (figure 2c). The deviation from chance performance (50%) was not very large, but it was statistically significant. When they introduced probe stimuli with a frequency difference between A and B tones of 3 or 5 semitones that were associated with a random reward, the probability of a one-stream response differed from chance for the 3 semitones but not for the 5 semitone ABAB pattern (figure 2c). In the 10 semitone condition, the monkeys were less likely to produce a ‘two-stream’ response if A and B tones were presented synchronously rather than alternating suggesting that temporal coherence influences the streaming percept in the monkey similar to the human. However, compared with the results obtained with European starlings, the effect of the frequency difference between A and B tones on the reported one-stream or two-stream percepts was much smaller. European starlings’ responses differed considerably more from chance level (50%) than the responses of the monkeys (figure 2c).

Streaming in the rat relying on frequency cues has been studied using the ABA-triplet paradigm [37]. Rats were trained to discriminate between a slow isochronous rhythm (similar to the B-tone rhythm) and a fast isochronous rhythm (similar to the A-tone rhythm) or a galloping rhythm (i.e. presenting tone triplets followed by a pause). In this initial training, all tones had the same frequency. After reaching a baseline performance (65% report probability of the slow rhythm), ABA triplets with a frequency difference between A and B tones were introduced as unrewarded stimuli and the response probability for indicating a slow rhythm was determined. Even in the baseline training using tones of a single frequency, a high rate of false positive responses (on average 44%) was observed which was significantly lower than the hit rate upon presentation of the slow rhythm. The probability for reporting a slow rhythm if the rats were presented with ABA triplets was significantly larger at a frequency difference of 12 semitones than at 2 or 6 semitones. The hit rate in the 2 and 6 semitone conditions was in the same range (34–38%) as the rate of false positive responses in the baseline training. These results suggest that the rats may perceive two separate streams of A and B tones if these differ by 12 semitones, but not if they differ by 6 semitones or less.

Ma and co-workers [32] investigated an objective measure of stream segregation in the ferret (*Mustela putorius*) and proposed that a reduced sensitivity would be expected if stream segregation occurred (figure 2f). Ferrets were required to report the frequency shift of B tones in an ABAB sequence with A and B tones differing in frequency. Thresholds for detecting the frequency shift were higher when the frequency difference between A and B tones got larger which was similar to psychophysical results obtained in humans [21].

Auditory streaming relying on spatial cues has been evaluated in cat (*Felis catus*) [33] and humans [10] using a similar stimulus paradigm that was based on the discrimination of the rhythm of noise bursts. Subjects had to report the presentation of one of two target rhythms in the presence of a masker rhythm interleaved with the target rhythm. If the subjects were able to segregate masker and target rhythms, the sensitivity for discriminating the target rhythms should be improved. For a target being presented at 0° and the masker being presented at the same or different angles, the

cat's sensitivity monotonically increased on average with an increasing spatial separation of target and masker reaching an average sensitivity of nearly 3.5 if masker and target were spatially separated by 80° (figure 2f) [33]. This result suggests that spatial acoustic cues can strongly support auditory stream segregation.

By applying a paradigm building upon audio-visual (AV) integration between a regularly timed series of light flashes and a series of tones, Selezneva and co-workers [38] investigated the perception of the tone series (ABAB pattern, A and B tones differing in frequency). Generally, both in humans and crab-eating macaque monkeys (*Macaca fascicularis*) responses to the cessation of the light flash presentation were faster if the presentation of the tone stimuli stopped before the point in time before the next flash could be expected compared with the response time for purely visual stimulus series. This result indicates that humans and monkey showed AV integration. In the AV condition, light flashes were presented in two different temporal relations with the auditory stimulus. In the first condition (flash synchronized with every second tone), it was expected that perceptual segregation has little effect on the response time, whereas in the second condition (flash synchronized with every third tone), it was expected that separating the A and B tones by frequency would increase the response latency. As expected, in human subjects there was no effect of frequency in the first AV condition on the visually evoked response latency while in the second AV condition an increase in the frequency difference between A and B tones resulted in an increase of response latency to the cessation of the visual stimulus suggesting impaired AV integration. The results in the two monkeys that Selezneva and co-workers [38] tested pointed in the same direction showing clear effects of frequency difference in the second AV condition. Thus, also AV integration can provide an objective measure of auditory streaming.

#### 4. Electrophysiological correlates of streaming

Only in animal studies can we use electrophysiological methods to record localized neuronal responses (e.g. local field potentials, multi-unit activity, single-unit activity) that are associated with auditory streams whereas human studies generally rely on non-invasive methods [3]. Two important features of the electrophysiological response patterns have been related to auditory streaming. Firstly, it has been proposed that different auditory streams with segregated sequences of sounds are represented by separate populations of neurons [39,40]. As correlated activity across many neurons (also being related to oscillatory activity of the cortex) has been associated with perceptual binding, it has been proposed that temporally coincident activity across different populations of neurons will lead to a perceptual integration of signals into a single stream whereas temporally anti-coincident activity results in perceptual stream segregation [15]. So far the hypothesis relating to the representation of streams by separate populations has been well studied in animal experiments while the aspect of temporal coherence across populations has rarely been investigated [15,41].

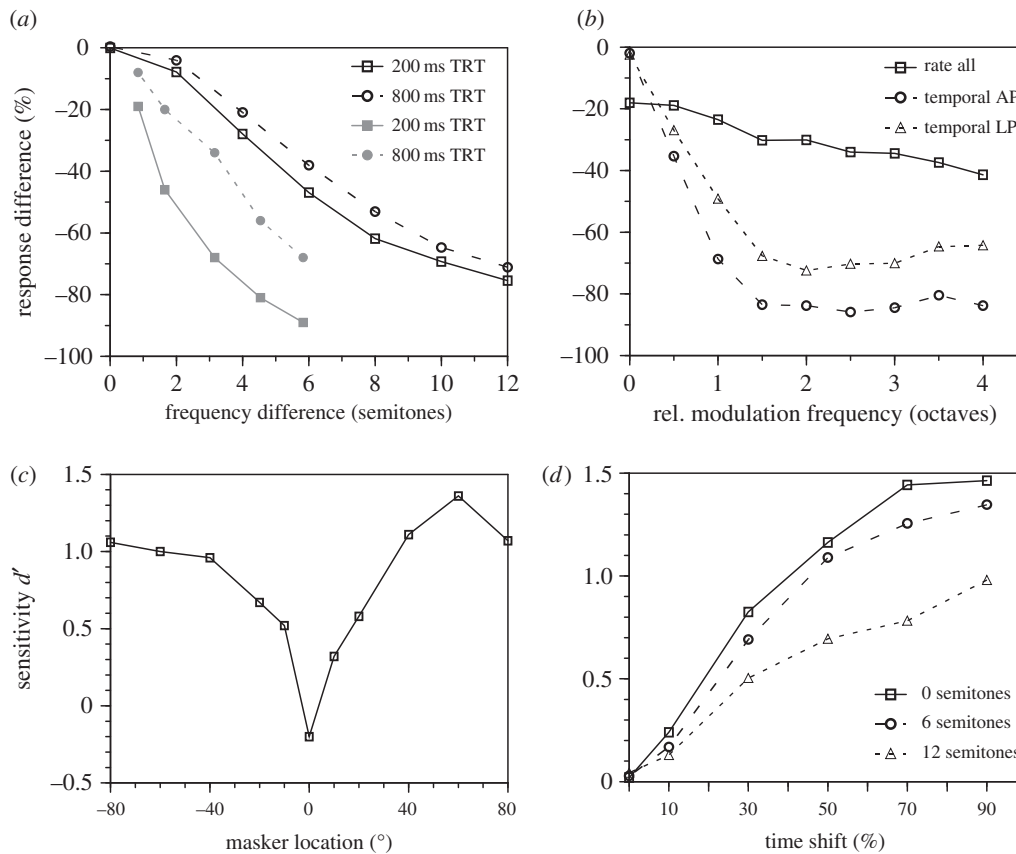
As neurons often will be tuned to specific physical characteristics of sounds (e.g. frequency, intensity, amplitude modulation frequency, interaural time and intensity differences), differences between sound signals associated with different streams are likely to be represented by different populations of neurons,

each tuned to the characteristics of the sounds in the one or the other stream. Furthermore, the responses to the signals in the different auditory streams may suppress each other, which is another neuronal response property that has been associated with auditory stream segregation. Here, we will summarize the evidence of neuronal correlates of auditory streaming related to the different physical features of the sound and the role of neuronal adaptation and suppression.

The stimulus paradigms that have been applied in the neuronal studies of auditory streaming were similar to those applied in psychophysical studies in humans. Often the 'ABA-' or 'ABAB' sequences commonly applied in psychoacoustical studies with human subjects (e.g. [7]) have also been used as stimuli in the neurophysiological studies in which A and B signals were characterized by a difference in the physical feature of the sound. If the neural responses to A and B sounds can be well distinguished (i.e. different populations of neurons respond to each of the sounds), this is viewed as a correlate of auditory stream segregation. As tuning already is found in the auditory periphery, neural correlates of auditory streaming should be evident throughout all levels of the auditory pathway. While frequency tuning already is evident in the inner ear, tuning to other features of the sound (e.g. amplitude modulation or binaural cues) will only become evident at higher levels of the auditory pathway. Where in the auditory pathway the neural basis of auditory perceptual organization emerge is still actively debated and may strongly depend on the stimulus features upon which auditory streaming relies [3,42–45]. Such perceptual organization might be affected by the subject's state (awake or under anaesthesia) and whether it is passively listening or attentively analysing a stream [41,46]. Only few animal studies have investigated the dynamics of the neuronal response during build-up in stream segregation [47,48]. So far no neuronal correlates of perceptual switching in auditory streaming have been investigated because it requires recording of the neuronal activity during the ongoing behavioural test of the streaming percept.

##### (a) Stream segregation by frequency cues

As auditory nerve fibres are tuned to a specific frequency, stream segregation relying on frequency difference ( $\Delta f$ ) may start at the auditory periphery. This peripheral channeling hypothesis [9] was evaluated by computational modelling [49]. Neural correlates of this hypothesis were first observed by Fishman and co-workers [39]. They determined multi-unit activity and current source density derived from local field potentials elicited by alternating two tone (ABAB) sequences in the primary auditory cortex of the awake crab-eating macaque monkey. The A tone frequency was set at the best frequency (BF) of the recording site and B tone frequency was 10–50% lower or higher than the A tone frequency. Tone repetition time (TRT), i.e. the onset-to-onset asynchrony between A and B tones, was also varied. Their results indicate increasing suppression of the B tone response as  $\Delta f$  between the two tones increased. TRT also had an effect on the suppression of the B tone responses indicating forward masking. The role of forward masking becomes also evident if tone duration (TD: 25, 50, and 100 ms) is changed while the TRT is held constant (figure 3a, [40,50]). In recording responses of auditory cortex neurons in awake Rhesus monkeys stimulated with ABA-triplet sequences, Micheyl and co-workers [48] observed a



**Figure 3.** Neural measures of stream segregation. (a) Differences between the normalized rate responses to BF A tones and off-BF B tones (tone duration 25 ms, onset-to-onset inter-stimulus interval as indicated by legend, tones presented as ABA triplets). Open symbols show data from European starling primary cortical neurons of awake subjects [40], filled symbols show data from neurons in the macaque monkey primary auditory cortex of anaesthetized subjects [50]. A large negative value indicates a segregated representation of A and B tones. (b) Relative response differences to A and B tones with different modulation frequencies (ABA triplets, sinusoidally amplitude modulated (SAM) tones, modulation frequency of A tones eliciting the neuron's largest response, B tone modulation frequency was higher as indicated by the  $x$ -axis value [51]). A large negative value indicates a segregated representation of A and B tones. Solid line represents average rate responses across all recording sites. Dashed lines depict temporal responses (vector strength) observed from recording sites of two major modulation tuning types, all-pass (AP) and low-pass (LP). (c) Sensitivity of neurons in the cat primary auditory cortex for discriminating two different rhythms for signals presented from the front ( $0^\circ$ ) in the presence of a masker broadcasting rhythmic sounds from different directions (for behavioural equivalent see figure 2f) [52]. (d) Objective measure of stream segregation (sensitivity for detecting onset time shift of off-CF tone in ABA tone triplets) relying on the temporal response of European starling primary auditory forebrain neurons (for the behavioural equivalent see figure 2d) [31].

similar effect of  $\Delta f$  between the two tones on the differential representation of A and B tones. Bee & Klump [40] recorded multiunit activity from the awake European starling forebrain in response to ABA-sequences with varying  $\Delta f$ , TRT and TD (figure 3a). They showed that larger  $\Delta f$  and shorter TRT evoked larger a response difference between A and B tones, but showed no significant effect of varying TD. A subsequent study confirmed the hypothesis that the inter-stimulus interval (i.e. the silent interval between successive tones) had the largest effect among various temporal factors (TRT, TD) on the relative response of starling forebrain neurons which also suggests a role of forward masking on streaming [53]. Similar to the monkey and the songbird, Kanwal and co-workers [54] observed the neural responses in the auditory cortex of the Mustached bat (*Pteronotus parnellii*) exhibiting increasing suppression between the two tones in an ABAB stimulus if the TRT was short. The study suggested that this suppression may enhance perception of echolocation sounds. The role of across-frequency suppression has also been demonstrated in an auditory streaming paradigm related to the ABAB paradigm that was termed as the rhythmic masking release paradigm [55] presenting alternating distractor and target tones to neurons in the primary auditory cortex of the crab-

eating macaque. In that paradigm, identical tone frequencies for targets and distractors are used and the response to targets cannot be distinguished from that to the distractors making it impossible for the neurons to reflect the target rhythm. The response to distractors was considerably reduced; however, only if flanking bands were added to the distractor. Although there was a response to the target tone with no flanking bands, the response to the distractor was suppressed, thus making it possible to discern the target tone rhythm from the neural response. Scholes and co-workers [56] recorded spike responses to ABAB stimuli with different  $\Delta f$  and TRT in the auditory cortex of the anaesthetized guinea pig (*Cavia porcellus*) and showed that the pattern of neural responses was consistent with human perceptual stream segregation. Tonotopic patterns of cortical membrane potentials in the anaesthetized guinea pig also reflected the spatially separated representation of the two tone signals in ABAB and ABA-triplet paradigms showing a suppression of the response to the off-BF tones [57]. Noda *et al.* [37] recorded local field potential (LFP) in the primary auditory cortex of anaesthetized rats and suggest that the amplitude and phase of cortical oscillatory activity, especially in the gamma band, carries important information regarding stream segregation. Already

in the cochlear nucleus (CN) of the anaesthetized guinea pig, Pressnitzer *et al.* [58] observed neural responses elicited by ABA-streaming sounds that were similar to those observed in the primary auditory cortex of other mammals and birds. This suggests that the neural basis of perceptual segregation by frequency may already emerge in the auditory periphery as is suggested by the peripheral channelling model.

### (b) Stream segregation by temporal cues

Although peripheral channelling is a primary mechanism for the stream segregation, two sound sequences with overlapping frequency components can also evoke stream segregation if the temporal structure of those sounds is different [4,5]. In primary auditory forebrain neurons of the awake European starling, Itatani & Klump [51] observed a neural correlate of stream segregation based on the response to the temporal envelope of A and B signals in ABA-triplets with A and B being tones with the same carrier frequency and differing sinusoidal amplitude modulation. The carrier frequency of two tones was always set to the characteristic frequency (CF) of the recording site and A and B tone modulation frequencies varied so that the modulation frequency difference was between 0 and 4 octaves. Spike responses reflected the modulation tuning characteristic of each recording site. Although in many cases the spectral components of the modulated tones were processed by the same auditory filter to exclude the effect of peripheral channelling, differential suppression of A and B tone responses was observed (figure 3*b*). The study suggests that, similar to peripheral channelling, there may be channelling at higher levels of the auditory pathway related to acoustic features that are analysed at these levels. Furthermore, both rate and temporal response properties of the neuronal response appeared to be suitable for auditory streaming of 125 ms signals.

Auditory streaming in human subjects can also be elicited by differences in the phase relation between components of harmonic complex tones (HCTs) resulting in a different temporal structure of the signals while their frequency spectra do not differ [12]. Presenting similar HCT stimuli to auditory forebrain neurons in the awake starling, Itatani & Klump [59] investigated the neural correlate of streaming by component phase. They used ABA-sequences in which A was an HCT with components in cosine phase and B was an HCT with components in cosine, alternating (odd and even harmonics were in cosine and sine phase, respectively), or random phase. The results showed clear differences in neuronal activity elicited by the three types of HCTs which may provide for a representation of the different sounds by separate populations of neurons. Using the HCT ABA-stimuli, Dolležal *et al.* [60] compared the human percept and the neural response patterns in the European starling forebrain for HCTs with a tone duration of 125 ms or 40 ms. Human subjects could segregate HC stimuli differing in phase independent of tone duration. In the responses of starling forebrain neurons, the synchrony measure of the response being related to the temporal structure of the stimulus waveform deteriorated with the reduced tone duration whereas the neuronal rate response measure could represent the different signal types at the 40 ms stimulus duration suggesting that the rate measure provides the better correlate to perception.

### (c) Stream segregation by spatial cues

Human listeners could discriminate two rhythmic patterns if the spatial separation of the sources producing the patterns

was as small as  $8^\circ$  [10] indicating that spatial separation provides potential cues for stream segregation. In the auditory cortex of the anaesthetized cat [52], neuronal activity was recorded in response to ABAB-stimuli in which the alternating noise bursts were spatially separated (figure 3*c*). A majority of auditory cortex neurons was spatially tuned, i.e. synchronized preferentially to sounds from a specific direction. Similar to segregation by frequency difference, the neurons preferentially responded to one of the two spatially separated sound sources and the bias was enhanced by forward suppression. Spatial tuning of neurons became sharper in the presence of competing sounds from a second source compared with stimulation with sounds from a single source in space indicating the enhancement of precision of spatial segregation. Yao *et al.* [61] further investigated the emergence of spatial stream segregation along the auditory pathway by recording neuronal responses elicited by intermingled rhythmic patterns in neurons from the inferior colliculus (IC), the nucleus of the brachium of the IC (BIN), the thalamic medial geniculate body (MGB) and the primary auditory cortex of anaesthetized rats. In response to successive spatially separated sounds, the central nucleus of the IC showed weak neural stream segregation. Spatial sensitivity for stream segregation and forward suppression was increased along the auditory pathway up to the primary auditory cortex providing for separate populations of neurons representing the streams of sounds from the different directions. Application of GABA receptor antagonists to the primary auditory cortex neurons suggested that forward suppression due to depression at the thalamo-cortical synapse is a major factor in auditory stream segregation [61].

### (d) Neural correlate of build-up of stream segregation

Perceiving segregated streams usually takes some time to develop. Psychophysical studies showed that subjects initially tend to perceive two tone sequences (ABAB, or ABA-) as a single stream and only after several seconds of a build-up time period, the two different sounds are heard as distinct streams [4,5]. The neural correlate of this build-up effect has been investigated by comparing spike responses at different time points after onset of stimulation. Micheyl and co-workers [48] compared perceptual build-up of stream segregation in humans and adaptation of neural responses in the Rhesus monkey auditory cortex presented with the same type of ABA-stimuli as were presented to human subjects evaluating the time course of perception. They set the A tone frequency at the neuron's BF and the B tone frequency was 1, 3, 6 or 9 semitones above the A tone frequency. During the presentation of a 10 s series of the stimuli, both neuronal A and B responses decayed over time in all  $\Delta f$  conditions. The magnitude of decay was larger for the B response and it increased with  $\Delta f$ . By using a simple model comparing the response measure for A and B tones and setting a threshold to the response difference, they were able to predict the human build-up. If A and B tone responses differed by less than the threshold, this was counted as a neuronal one-stream response and otherwise it was counted as a two-stream response. With an appropriate threshold value, the neuronal response probabilities matched the human perceptual response probabilities. Strong adaptation that can provide the basis for build-up was also observed in the tonotopic pattern of membrane potentials recorded using voltage sensitive dye imaging in the primary auditory cortex of the anaesthetized guinea pig [57]. Noda *et al.* [37,62] reported



that gamma band oscillations of LFP showed adaptation that could be a sign of build-up like responses. Bee and co-workers [47] investigated the effect of  $\Delta f$ , SOA (stimulus onset asynchrony between subsequent signals) and TD on the build-up process in European starling forebrain neurons using a similar modelling approach predicting build-up effects on the basis of the adaptation in neurometric response functions. They reported that SOA had a bigger effect on the dynamic rate and range of adaptation than  $\Delta f$  and TD. These findings support the prediction that with increasing SOA the build-up effect decreases. Similar response decays due to adaptation have also been found at the level of the CN in the guinea pig [58] suggesting that the response properties of neurons in the auditory periphery may already contribute to build-up. In human perceptual studies, it has been observed that after the build-up the two-stream percept can revert back to a one-stream percept, and multiple switching of the percept occurs during a time period of minutes [6]. So far, a corresponding pattern of responses has not been investigated in the neurons.

## 5. Comparison of percept and neuronal response within the same animal model

Usually, neurophysiological animal studies of auditory streaming aim at explaining human psychophysical data assuming that a similar percept exists in animals and humans. Rarely, perception of auditory streams and the neurophysiological correlate have been studied in the same animal model with an identical stimulation paradigm. With such an approach still the assumption is made that the neuronal response patterns observed in the passively listening or anaesthetized animal and the behavioural responses in the actively listening animal can be correlated. However, only simultaneously recording the neuronal activity and the behavioural response in the attentive animal will allow directly relating the neurons' activity to the behaviour. A study of subjective auditory streaming of ABA-tone triplet sequences in the European starling [29] revealed that starlings perceive one-stream triplet sequences as having a galloping rhythm and two-stream sequences as having an isochronous rhythm in a similar way as humans (figure 2c). Analysis of the neuronal responses in the auditory forebrain region of passively listening awake European starlings corresponding to the mammalian auditory cortex demonstrated that the difference in the normalized neuronal response to A and B tones in the triplet correlates with the probability that the birds perceive two streams. Also in the rat (*Rattus rattus*), the probability for perceiving a two-stream percept in ABA-triplets as was inferred from the probability of reporting an isochronous B tone rhythm of probe stimuli is correlated with a neuronal response measure, i.e. the gamma-band response in the primary auditory cortex [37]. However, the high rate of false positives in the behavioural response measure makes it difficult to directly compare the behavioural measure with the neurophysiological measure.

An objective measure of stream segregation obtained in ferrets, i.e. the ability to detect a frequency shift in the stream of tones with the higher frequency in an ABAB tone sequence has been compared with the response to A and B tones in the ferret auditory cortex [32,41]. As found in the starling, ferret cortical neurons in a passively listening animal showed a differential response to A and B tones that increased

with the frequency difference between them. This will result in a representation of the two tones by separate populations of neurons especially at a large frequency difference (e.g. 12 semitones). A behavioural study of frequency discrimination for B tones in the ABAB stimulus by the ferret showed an improvement of the B tone frequency discrimination thresholds with an increasing frequency separation between A and B tones. At a frequency separation of 12 semitones discrimination was superior to that observed for a frequency separation of 6 semitones (figure 2e). This indicates that the objective measure of stream segregation in the ferret can be related to the degree of separation of the neuronal populations representing the A and B signals. As has been concluded from comparing cortical responses in passively listening and attentive ferrets, this separation may be affected by the attentive state of the ferret [41].

The most direct comparison between the streaming percept and the neuronal representation can be made if the neurons' responses are recorded while the animal is indicating its percept as can be inferred from an objective measure of stream segregation. This approach allows the most valid comparison between behaviour and the neuronal response pattern because it is well known that attention in a behavioural task may modify the neuronal response to a stimulus [46,63] and also the streaming percept is modified by attention processes [4,5,64]. Furthermore, only this approach allows one to conclude which component of the neuronal response is related to the stimulus features of the streaming stimuli and which to the perception of auditory streams. European starlings were trained to report detecting a time shift of the B tone in ABA-triplets, a task which was also used in a human psychophysical study and that results in higher threshold if two streams are perceived [16]. Both in starlings and humans, the difficulty of detecting the time shift of the B tone depended on the frequency difference  $\Delta f$  between A and B tones. With increasing  $\Delta f$ , the sensitivity for detecting the time shift deteriorates and thresholds increased (figure 2d). Similarly, the sensitivity of the neurons in a primary auditory cortical area in the starling representing the time shift was reduced if  $\Delta f$  was increased (figure 3d) [31]. The neurons' temporal response measure, the van Rossum distance [65] derived from multi-unit activity, already reached an average sensitivity of 40% of the sensitivity observed in the behavioural response being sufficient for explaining the behavioural performance. This primary auditory cortical area, however, did not represent the starling's percept. For stimuli with the same  $\Delta f$  between A and B tones and time shift of the B tone, there was no observed difference in neuronal response during a hit (being more likely during a one-stream percept) or a miss (being more likely during a two-stream percept).

## 6. Concluding remarks

Animal experiments in fish, frogs, birds and mammals ranging from rat to monkey have revealed a number of neural correlates of phenomena of auditory streaming that were observed in psychophysical studies of humans, i.e. have taken an across-species approach. Few studies, however, enable the comparison of neuronal responses with perception in the same species. Such a within-species approach takes into account the representation of stimulus features that provide for stream segregation, whereas in the across-species approach the general assumption is made that in both the human and the animal model stimulus features are represented with similar sensitivity. This is not necessarily

the case, if feature detectors in the auditory system show different tuning. For example, if the peripheral channelling hypothesis is involved, this requires the same auditory filter bandwidth in both humans and the animal species serving as a model. In psychophysics, human subjects are usually attentive when performing a stream segregation task. Many animal studies, however, involve passively listening or even anaesthetized subjects assuming that the critical neuronal response measures are not affected. These state differences between inattentive animal and attentive human subjects make comparisons difficult. Finally, human electrophysiological studies apply non-invasive techniques reflecting the response of large populations of neurons [3] whereas the animal studies rarely look at simultaneously recorded populations of neurons a time [37,57,62], nor have done so at the level of individual neurons. This so far precluded direct observation of anti-correlated neuronal response patterns as are predicted by the temporal coherence hypothesis for the formation of auditory streams [15]. Despite these drawbacks it is promising that a number of correlates of the human streaming percept have been found in animal studies of auditory streaming.

We will reach a new level of understanding of the processes involved in auditory streaming in real-life situations if we aim at applying stimulation paradigms that capture important characteristics of natural scenes. For example, real sound sources usually provide multiple cues. Thus, presenting single cues alone as in most animal studies is quite

unnatural and more studies of cue interactions and weighting of cues would be desirable. So far, most of the animal studies aimed at identifying physical features of sounds that provide for streaming. However, to achieve a new level of understanding of the mechanisms we need more animal studies that investigate the streaming percept in an animal while at the same time recording the neuronal response. Such an approach would reveal the brain areas and neuronal computations underlying the streaming percept itself and its bistability and switching (for a review, see [3]). Furthermore, an approach using behaving animals in a natural stimulus context will allow exposing the processes in which learned templates will affect auditory scene analysis as exemplified in the selective song learning of birds [66] or the way humans perceive speech in relation to the mother tongue [67]. Although such experiments using more natural settings and simultaneous recording of behaviour and neuronal responses are quite demanding, especially if responses from multiple sites are recorded at a time, it will be worth the effort to understand how the brain achieves an astonishingly good performance in the analysis of complex acoustic scenes.

**Authors' contributions.** Both N.I. and G.M.K. conceived and wrote the review.

**Competing interests.** We declare we have no competing interests.

**Funding.** N.I. and G.M.K. were funded by grant no. TRR31 from the Deutsche Forschungsgemeinschaft.

## References

- Bee MA, Micheyl C. 2008 The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *J. Comp. Psychol.* **122**, 235–251. (doi:10.1037/0735-7036.122.3.235)
- Bregman AS. 1990 *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: Bradford Books, MIT Press.
- Gutschalk A, Dykstra AR. 2014 Functional imaging of auditory scene analysis. *Hear. Res.* **307**, 98–110. (doi:10.1016/j.heares.2013.08.003)
- Moore BCJ, Gockel H. 2002 Factors influencing sequential stream segregation. *Acta Acust. United Ac.* **88**, 320–333.
- Moore BCJ, Gockel HE. 2012 Properties of auditory stream formation. *Phil. Trans. R. Soc. B* **367**, 919–931. (doi:10.1098/rstb.2011.0355)
- Denham SL, Winkler I. 2006 The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* **100**, 154–170. (doi:10.1016/j.jphysparis.2006.09.012)
- van Noorden LPAS. 1975 Temporal coherence in the perception of tone sequences. PhD thesis, University of Technology, Eindhoven, The Netherlands.
- Rose MM, Moore BCJ. 2000 Effects of frequency and level on auditory stream segregation. *J. Acoust. Soc. Am.* **108**, 1209–1214. (doi:10.1121/1.1287708)
- Hartmann WM, Johnson D. 1991 Stream segregation and peripheral channeling. *Music Percept.* **9**, 155–184. (doi:10.2307/40285527)
- Middlebrooks JC, Onsan ZA. 2012 Stream segregation with high spatial acuity. *J. Acoust. Soc. Am.* **132**, 3896–3911. (doi:10.1121/1.4764879)
- Cusack R, Roberts B. 1999 Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception* **28**, 1281–1289. (doi:10.1068/p2804)
- Roberts B, Glasberg BR, Moore BCJ. 2002 Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am.* **112**, 2074–2085. (doi:10.1121/1.1508784)
- Vliegen J, Oxenham AJ. 1999 Sequential stream segregation in the absence of spectral cues. *J. Acoust. Soc. Am.* **105**, 339–346. (doi:10.1121/1.424503)
- Vliegen J, Moore BCJ, Oxenham AJ. 1999 The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J. Acoust. Soc. Am.* **106**, 938–945. (doi:10.1121/1.427140)
- Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. 2009 Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* **61**, 317–329. (doi:10.1016/j.neuron.2008.12.005)
- Micheyl C, Oxenham AJ. 2010 Objective and subjective psychophysical measures of auditory stream integration and segregation. *JARO* **11**, 709–724. (doi:10.1007/s10162-010-0227-2)
- Dolležal LV, Brechmann A, Klump GM, Deike S. 2014 Evaluating auditory stream segregation of SAM tone sequences by subjective and objective psychoacoustical tasks, and brain activity. *Front. Neurosci.* **8**, 119. (doi:10.3389/fnins.2014.00119)
- Fay RR. 1998 Auditory stream segregation in goldfish (*Carassius auratus*). *Hear. Res.* **120**, 69–76. (doi:10.1016/s0378-595500058-6)
- Bolhuis JJ, Okanoya K, Scharff C. 2010 Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759. (doi:10.1038/nrn2931)
- Tucker MA, Gerhardt HC. 2012 Parallel changes in mate-attracting calls and female preferences in autotriploid tree frogs. *Proc. R. Soc. B* **279**, 1583–1587. (doi:10.1098/rspb.2011.1968)
- Micheyl C, Carlyon RP, Cusack R, Moore BCJ. 2005 Performance measures of auditory organization. In *Auditory signal processing* (eds D Pressnitzer, A de Cheveigné, S McAdams, L Collet), pp. 202–210. New York, NY: Springer.
- Bee MA. 2012 Sound source perception in anuran amphibians. *Curr. Opin. Neurobiol.* **22**, 301–310. (doi:10.1016/j.conb.2011.12.014)
- Bee MA. 2015 Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. *Int. J. Psychophysiol.* **95**, 216–237. (doi:10.1016/j.ijpsycho.2014.01.004)
- Schwartz JJ, Gerhardt HC. 1995 Directionality of the auditory system and call pattern recognition during

- acoustic interference in the gray treefrog, *Hyla versicolor*. *Audit. Neurosci.* **1**, 195–206.
25. Nityananda V, Bee MA. 2011 Finding your mate at a cocktail party: frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. *PLoS ONE* **6**, e21191. (doi:10.1371/journal.pone.0021191)
  26. Farris HE, Ryan MJ. 2011 Relative comparisons of call parameters enable auditory grouping in frogs. *Nat. Commun.* **2**, 410. (doi:10.1038/ncomms1417)
  27. Fay RR. 2000 Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *JARO* **1**, 120–128. (doi:10.1007/s101620010015)
  28. Hulse SH, MacDougallShackleton SA, Wisniewski AB. 1997 Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *J. Comp. Psychol.* **111**, 3–13. (doi:10.1037/0735-7036.111.1.3)
  29. MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W. 1998 Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *J. Acoust. Soc. Am.* **103**, 3581–3587. (doi:10.1121/1.423063)
  30. Christison-Lagay KL, Cohen YE. 2014 Behavioral correlates of auditory streaming in rhesus macaques. *Hear. Res.* **309**, 17–25. (doi:10.1016/j.heares.2013.11.001)
  31. Itatani N, Klump GM. 2014 Neural correlates of auditory streaming in an objective behavioral task. *Proc. Natl Acad. Sci. USA* **111**, 10 738–10 743. (doi:10.1073/pnas.1321487111)
  32. Ma L, Micheyl C, Yin PB, Oxenham AJ, Shamma SA. 2010 Behavioral measures of auditory streaming in ferrets (*Mustela putorius*). *J. Comp. Psychol.* **124**, 317–330. (doi:10.1037/a0018273)
  33. Javier LK, McGuire EA, Middlebrooks JC. 2016 Spatial stream segregation by cats. *JARO* **17**, 195–207. (doi:10.1007/s10162-016-0561-0)
  34. Wisniewski AB, Hulse SH. 1997 Auditory scene analysis in European starlings (*Sturnus vulgaris*): discrimination of song segments, their segregation from multiple and reversed conspecific songs, and evidence for conspecific song categorization. *J. Comp. Psychol.* **111**, 337–350. (doi:10.1037/0735-7036.111.4.337)
  35. Dent ML, Martin AK, Flaherty MM, Neilans EG. 2016 Cues for auditory stream segregation of birdsong in budgerigars and zebra finches: effects of location, timing, amplitude, and frequency. *J. Acoust. Soc. Am.* **139**, 674–683. (doi:10.1121/1.4941322)
  36. Izumi A. 2002 Auditory stream segregation in Japanese monkeys. *Cognition* **82**, B113–B122. (doi:10.1016/s0010-0277(01)00161-5)
  37. Noda T, Kanzaki R, Takahashi H. 2013 Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. *PLoS ONE* **8**, e0146206. (doi:10.1371/journal.pone.0083544)
  38. Selezneva E, Gorkin A, Mylius J, Noesselt T, Scheich H, Brosch M. 2012 Reaction times reflect subjective auditory perception of tone sequences in macaque monkeys. *Hear. Res.* **294**, 133–142. (doi:10.1016/j.heares.2012.08.014)
  39. Fishman YI, Reser DH, Arezzo JC, Steinschneider M. 2001 Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res.* **151**, 167–187. (doi:10.1016/s0378-5955(00)00224-0)
  40. Bee MA, Klump GM. 2004 Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol.* **92**, 1088–1104. (doi:10.1152/jn.00884.2003)
  41. Shamma S, Elhilali M, Ma L, Micheyl C, Oxenham AJ, Pressnitzer D, Yin P, Xu Y. 2013 Temporal coherence and the streaming of complex sounds. *Adv. Exp. Med. Biol.* **787**, 535–543. (doi:10.1007/978-1-4614-1590-9\_59)
  42. Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. 2005 Neuroimagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* **25**, 5382–5388. (doi:10.1523/jneurosci.0374-05.2005)
  43. Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E. 2007 The role of auditory cortex in the formation of auditory streams. *Hear. Res.* **229**, 116–131. (doi:10.1016/j.heares.2007.01.007)
  44. Kondo HM, Kashino M. 2009 Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* **29**, 12 695–12 701. (doi:10.1523/jneurosci.1549-09.2009)
  45. Bizley JK, Cohen YE. 2013 The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* **14**, 693–707. (doi:10.1038/nrn3565)
  46. Gutschalk A, Rupp A, Dykstra AR. 2015 Interaction of streaming and attention in human auditory cortex. *PLoS ONE* **10**, e0118962. (doi:10.1371/journal.pone.0118962)
  47. Bee MA, Micheyl C, Oxenham AJ, Klump GM. 2010 Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. *J. Comp. Physiol. A* **196**, 543–557. (doi:10.1007/s00359-010-0542-4)
  48. Micheyl C, Tian B, Carlyon RP, Rauschecker JP. 2005 Perceptual organization of tone sequences in the auditory cortex of awake Macaques. *Neuron* **48**, 139–148. (doi:10.1016/j.neuron.2005.08.039)
  49. Beauvois MW, Meddis R. 1996 Computer simulation of auditory stream segregation in alternating-tone sequences. *J. Acoust. Soc. Am.* **99**, 2270–2280. (doi:10.1121/1.415414)
  50. Fishman YI, Arezzo JC, Steinschneider M. 2004 Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am.* **116**, 1656–1670. (doi:10.1121/1.1778903)
  51. Itatani N, Klump GM. 2009 Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J. Neurophysiol.* **101**, 3212–3225. (doi:10.1152/jn.91333.2008)
  52. Middlebrooks JC, Bremen P. 2013 Spatial stream segregation by auditory cortical neurons. *J. Neurosci.* **33**, 10 986–11 001. (doi:10.1523/jneurosci.1065-13.2013)
  53. Bee MA, Klump GM. 2005 Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav. Evol.* **66**, 197–214. (doi:10.1159/000087854)
  54. Kanwal JS, Medvedev AV, Micheyl C. 2003 Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Netw. Comp. Neural Syst.* **14**, 413–435. (doi:10.1088/0954-898x/14/3/303)
  55. Fishman YI, Micheyl C, Steinschneider M. 2012 Neural mechanisms of rhythmic masking release in monkey primary auditory cortex: implications for models of auditory scene analysis. *J. Neurophysiol.* **107**, 2366–2382. (doi:10.1152/jn.01010.2011)
  56. Scholes C, Palmer AR, Sumner CJ. 2015 Stream segregation in the anaesthetised auditory cortex. *Hear. Res.* **328**, 48–58. (doi:10.1016/j.heares.2015.07.004)
  57. Farley BJ, Noreña AJ. 2015 Membrane potential dynamics of populations of cortical neurons during auditory streaming. *J. Neurophysiol.* **114**, 2418–2430. (doi:10.1152/jn.00545.2015)
  58. Pressnitzer D, Sayles M, Micheyl C, Winter IM. 2008 Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* **18**, 1124–1128. (doi:10.1016/j.cub.2008.06.053)
  59. Itatani N, Klump GM. 2011 Neural correlates of auditory streaming of harmonic complex sounds with different phase relations in the songbird forebrain. *J. Neurophysiol.* **105**, 188–199. (doi:10.1152/jn.00496.2010)
  60. Dolležal LV, Itatani N, Gunther S, Klump GM. 2012 Auditory streaming by phase relations between components of harmonic complexes: a comparative study of human subjects and bird forebrain neurons. *Behav. Neurosci.* **126**, 797–808. (doi:10.1037/a0030249)
  61. Yao JD, Bremen P, Middlebrooks JC. 2015 Emergence of spatial stream segregation in the ascending auditory pathway. *J. Neurosci.* **35**, 16 199–16 212. (doi:10.1523/jneurosci.3116-15.2015)
  62. Noda T, Kanzaki R, Takahashi H. 2014 Amplitude and phase-locking adaptation of neural oscillation in the rat auditory cortex in response to tone sequence. *Neurosci. Res.* **79**, 52–60. (doi:10.1016/j.neures.2013.11.002)
  63. Fritz JB, David S, Shamma S. 2013 Attention and dynamic, task-related receptive field plasticity in adult auditory cortex. In *Neural correlates of auditory cognition* (eds YE Cohen, AN Popper, RR Fay), pp. 251–291. New York, NY: Springer.
  64. Caporello Bluvas E, Gentner TQ. 2013 Attention to natural auditory signals. *Hear. Res.* **305**, 10–18. (doi:10.1016/j.heares.2013.08.007)
  65. van Rossum MCW. 2001 A novel spike distance. *Neural Comput.* **13**, 751–763. (doi:10.1162/089976601300014321)
  66. Fehér O, Wang H, Saar S, Mitra PP, Tchernichovskii O. 2009 De novo establishment of wild-type song culture in the zebra finch. *Nature* **459**, U564–U594. (doi:10.1038/nature07994)
  67. Kuhl PK. 2004 Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831–843. (doi:10.1038/nrn1533)