

## Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment

Jagmeet S Kanwal<sup>1,2,4</sup>, Andrei V Medvedev<sup>1</sup> and Christophe Micheyl<sup>3</sup>

<sup>1</sup> Department of Physiology and Biophysics, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington, DC 20057-1460, USA

<sup>2</sup> Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA 22030-4444, USA

<sup>3</sup> Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

E-mail: kanwalj@georgetown.edu

Received 16 October 2002, in final form 13 March 2003

Published 14 May 2003

Online at [stacks.iop.org/Network/14/413](http://stacks.iop.org/Network/14/413)

### Abstract

During navigation and the search phase of foraging, mustached bats emit ~25 ms long echolocation pulses (at 10–40 Hz) that contain multiple harmonics of a constant frequency (CF) component followed by a short (3 ms) downward frequency modulation. In the context of auditory stream segregation, therefore, bats may either perceive a coherent pulse–echo sequence (PEPE...), or segregated pulse and echo streams (P–P–P... and E–E–E...). To identify the neural mechanisms for stream segregation in bats, we developed a simple yet realistic neural network model with seven layers and 420 nodes. Our model required recurrent and lateral inhibition to enable output nodes in the network to ‘latch-on’ to a single tone (corresponding to a CF component in either the pulse or echo), i.e., exhibit differential suppression by the alternating two tones presented at a high rate (>10 Hz). To test the applicability of our model to echolocation, we obtained neurophysiological data from the primary auditory cortex of awake mustached bats. Event-related potentials reliably reproduced the latching behaviour observed at output nodes in the network. Pulse as well as nontarget (clutter) echo CFs facilitated this latching. Individual single unit responses were erratic, but when summed over several recording sites, they also exhibited reliable latching behaviour even at 40 Hz. On the basis of these findings, we propose that a neural correlate of auditory stream segregation is present within localized synaptic activity in the mustached bat's auditory cortex and this mechanism may enhance the perception of echolocation sounds in the natural environment.

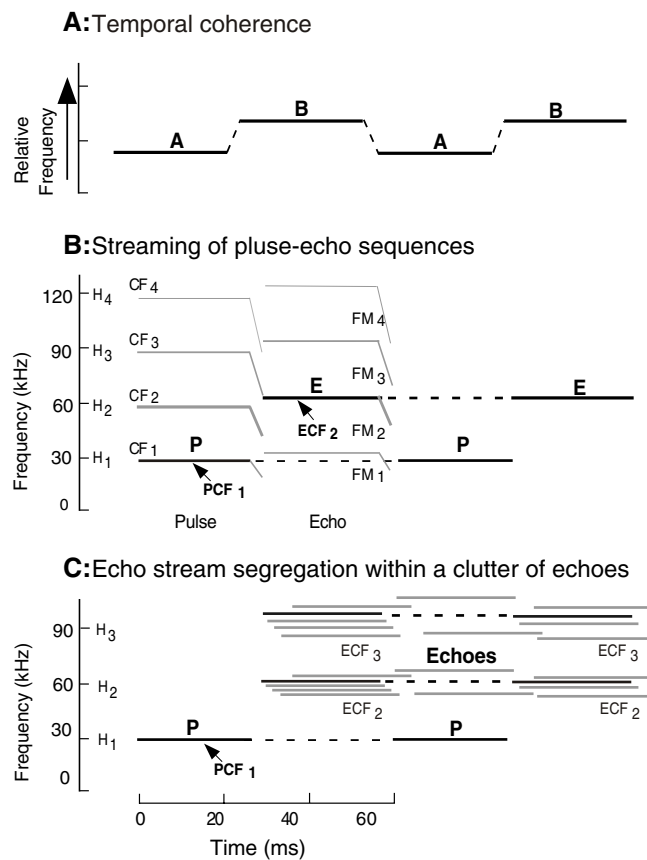
<sup>4</sup> Author to whom any correspondence should be addressed.

## 1. Introduction

Auditory stream segregation occurs in humans when a series of alternating tones of different frequencies (ABAB) are presented sequentially at a high rate. During stream segregation, perception of temporal coherence (see figure 1(A)) in the alternating tones splits into that of two separate and parallel auditory streams (A–A–A–and –B–B–B) and detailed information on the relative timing of the A and B tones is lost (Bregman 1990). This phenomenon has been extensively studied in humans by psychophysicists (Miller and Heise 1950, Bregman and Campbell 1971, for a review see: Bregman 1990) because of its importance as a possible mechanism for tracking speech in a noisy environment and continuous musical pieces within an acoustical or orchestral ‘clutter’. There is also some evidence that stream segregation occurs in monkeys (Izumi 2002), as well as in nonmammalian species (Hulse *et al* 1997, Fay 1998). Only recently attempts have been made to identify the neuronal basis of this phenomenon by recording the stimulus-driven multineuron activity within the cortex of primates (Fishman *et al* 2001).

The present study was inspired by the idea that auditory stream segregation may play a particularly important role in echolocation in at least some species of bats. Mustached bats (*Pteronotus parnellii*) emit brief (30 ms or shorter duration) pulses for acoustic imaging of objects in their environment and prey tracking (Schnitzler 1970). The echolocation signals emitted by these bats consist of four harmonics of a constant frequency (CF) pulse, which terminate in a 3 ms long frequency-modulated (FM) component (figure 1(B)). Much of our understanding of pulse–echo processing by auditory system comes from experiments using single pulse–echo combinations (Suga *et al* 1983, Esser *et al* 1997, Kanwal 1999, Suga 1990). In the natural environment, however, single pulse–echo combinations rarely arrive at the ear in isolation. Rather, mustached bats that hunt for insects in the canopy of rain forests, produce multiple pulses at a high rate during the navigational and search phases of echolocation (Novick and Vaisnys 1964). Under these circumstances, auditory stream segregation is not only likely to take place, but it may also have beneficial consequences. First, by separating the pulse and the echo sequences into two perceptual streams (i.e., forming one ‘pulse stream’ and one ‘echo stream’ for corresponding harmonics, as illustrated in figure 1(B)), each can be perceived largely independently of the other. Second, bats may be able to monitor the information present in successive echoes while suffering little or no interference from the interspersed pulses, or vice versa, bats may also at times need to monitor directly the sequence of pulses that they emit for motor control feedback (the latter would be equivalent to a person trying to talk or sing in a noisy and/or reverberating environment). For example, the animal may need to detect a slight change in Doppler shift in the CF between two consecutive echoes. If the echo stream is not perceptually ‘shielded’ from the pulse stream, this could be masked by the presence of a large frequency difference between the pulse and echo. Alternatively, because complex acoustic reflection and interference patterns often result in the arrival of multiple echoes within a small time and frequency window (Simmons *et al* 1989), the bat may prefer to track over time only echoes that specifically correspond to a target sound source among a clutter of echoes from other objects (figure 1(C)). In this situation, stream segregation may operate *within* the echo clutter itself, and contribute to the isolation and tracking over time of relevant echo components only.

Although sophisticated behavioural investigations are needed to further elucidate the role, if any, of stream segregation in bat echolocation, some progress in understanding this role may be achieved through electrophysiological studies exploring the conditions in which neural mechanisms underlying stream segregation come into play. For example, it would be interesting to study the responses of neurons in the bat’s central auditory system to stimuli



**Figure 1.** (A) Spectrographic representation of the pattern of tones that result in perceptual coherence (ABAB). (B) Stream segregation (AA and BB) is obtained for high rates of stimulus presentation and/or by increasing the frequency difference between the two tones. The mustached bat's echolocation pulse consists of four harmonics ( $H_1$ – $H_4$ ). Each harmonic consists of a CF ( $CF_x$ , where  $x = 1$ – $4$ ) and a frequency modulated ( $FM_x$ , where  $x = 1$ – $4$ ) component. Repeated pulse-echo sequences mimic a streaming type of stimulus, allowing the stream of echoes to be perceived separately from the stream of pulses. (C) Adjacent frequencies as in the  $CF_2$  in multiple echoes can lead to acoustic clutter. Streaming in the echo (BB) channel (shown by dashed lines) will also facilitate its separation from nontarget echoes that do not arrive in a fixed temporal relationship to the pulse.

corresponding to the two possible scenarios described in the previous paragraph, i.e., sequences of pulse and echo CFs as shown in figure 1(B), or of different echo CFs as shown in figure 1(C), succeeding each other and alternating repeatedly. If the electrophysiological responses to such (admittedly artificial and simplified, but well controlled) sequences of stimuli provide evidence for the occurrence of spatial or temporal segregation of responses to successive sounds at the neural level, then it could provide insights into both the conditions for and the mechanisms of auditory stream segregation in bats. In short, streaming may aid in constructing a perceptually consistent and persistent representation or auditory image of the sound source, be it the background or the target for echolocation in the natural environment.

Although the exact neural mechanisms for stream segregation remain uncertain, physiologically plausible computer models of streaming and electrophysiological data that have recently appeared in the scientific literature provide for some progress in this direction.

Specifically, existing models of auditory stream segregation (Beauvois and Meddis 1996, McCabe and Denham 1997) suggest that frequency selectivity and post-stimulus inhibition probably constitute the essential neurophysiological mechanisms for stream segregation. Recent electrophysiological results further indicate that the neural bases of stream segregation may be demonstrated at the level of the primary auditory cortex, and may consist of differential suppression of neural responses to tones one of which is displaced relative to the best frequency (BF) at a considered tonotopic location (Fishman *et al* 2001).

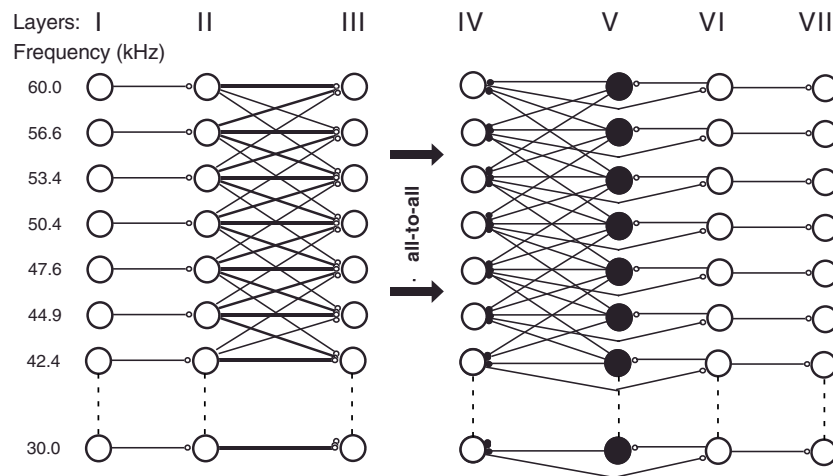
The study presented in this paper is a first attempt to explore the possibility that auditory stream segregation mechanisms contribute to the perceptual aspects of echolocation in a bat. This study combines two approaches. First, we developed and tested a neural network model that implements mechanisms which, based on both our electrophysiological recordings and earlier data in the literature, were likely to reproduce the neurophysiological phenomena that are assumed to be the basis of stream segregation. We incorporated several response properties of neurons in the mustached bat's auditory system in the design of the model (Medvedev *et al* 2002). Second, we obtained electrophysiological recordings of neural responses in the bat's auditory cortex to stimulus sequences for which stream segregation may occur in this animal.

The neural mechanisms incorporated in the model simulate neurophysiological processes known to operate in the central auditory system of most mammals, namely, recurrent and lateral inhibition (Shamma and Symmes 1985, Gao and Suga 1998, Loftus and Sutter 2001). Although the mechanisms and the architecture of the model presented here remain sufficiently general to explain the occurrence of stream segregation in a large variety of species, this model also takes into account specific functional properties of the bat central auditory system. As mentioned above, the mustached bat uses a pulse–echo sequence in order to analyse its environment, and this is reflected in response properties of neurons in the central auditory system. For example, neurons in the central auditory system of the mustached bat have multi-peaked receptive fields that exhibit broad tuning to frequencies of the first harmonic ( $H_1$ ) of the pulse and sharp tuning to the CF component in the Doppler-shifted harmonics in the echo (Suga *et al* 1983, Fitzpatrick *et al* 1993, Kanwal *et al* 1999). By being tuned to different values of the echo  $CF_2$ , 'combination-sensitive' neurons encode or respond to frequencies corresponding to different velocities of the bat relative to the background during the search phase (e.g., in the Doppler shifted CF processing or DSCF area) or the target during the insect pursuit phase of foraging (as in the CF–CF area). The general parallel-hierarchical architecture of the model was inspired from that of the auditory system, and the model output is directly comparable to the actual neural responses recorded in the present and other electrophysiological experiments (Fishman *et al* 2001).

## 2. Materials and methods

### 2.1. Network model

We constructed a realistically structured neural network using commercial software: Neuroimitator<sup>TM</sup> version 4.2 (Neuroma-RD; Cell MicroControls, USA). The model is phenomenological: electrophysiological properties of neurons and synaptic connections, such as resting membrane potential, threshold, absolute and relative refractoriness, synaptic delays, waveforms and durations of the postsynaptic potentials (PSPs) are phenomenological and are described by mathematical functions and parameters imitating the real neural/synaptic properties of classical neurons. The description is limited to the electrophysiological properties of whole neurons and synapses while the underlying molecular mechanisms, such as ionic conductances, metabolic transport, etc, and the real geometry of neurons are not modelled. Further details of this type of network are described elsewhere (Medvedev *et al* 2002). In this

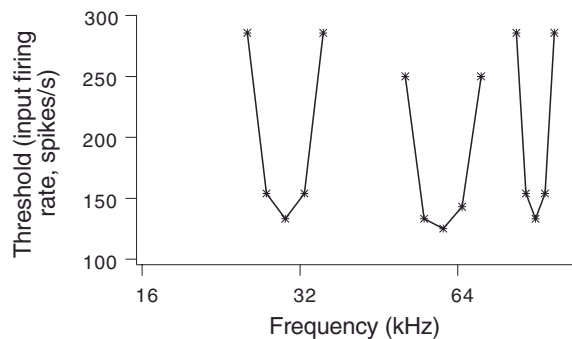


**Figure 2.** Schematic representation of a seven-layered neural network model (with 420 nodes) to simulate auditory stream segregation. The numbers on the left indicate sample frequency inputs on a logarithmic scale to seven representative channels in the network. Lines terminating in small circles indicate axonal projections to the next layer. Large circles indicate nodes. Open circles represent excitatory nodes/terminals and solid circles represent inhibitory nodes/terminals. The line thickness indicates the connection strength and dashed lines indicate the projection of the fundamental (first harmonic) in the pulse that is necessary to produce streaming. Each node in layer III projects to all nodes in layer IV. Lateral and recurrent inhibition is implemented in layer IV through layers V and VI. Layer VII is a recognition layer.

network model for streaming, we refer to the neuron-like processing elements as ‘nodes’ so as to capture the electrophysiological activity (ERPs) of a local population of neurons.

As designed, our network can accept four octaves of frequency inputs with 12 or more channels per octave. To take an example from the frequency scale at which bats emit their echolocation pulses, an octave may be in the frequency range 30–60 kHz. A portion of the network with frequency inputs in this range is presented in figure 2 to highlight the main features of the network. Furthermore, the fourth octave was designed to have 24 frequency channels. This reflects sharper frequency tuning (on a logarithmic scale) of the auditory nerve fibres tuned to higher frequencies. Layer I is an input layer. Layers II and III together generate a pattern of stimulus-driven activity in the auditory nerve. Divergence and convergence between these two layers provide a tuning curve in that each frequency channel (auditory nerve fibre) has a ‘best’ or ‘characteristic’ frequency. Layers IV–VII represent neural processing at the inferior collicular and thalamo-cortical levels.

To simulate multi-frequency tuning of cortical neurons, the neural network was initially trained to associate harmonically related frequencies as described earlier (Medvedev *et al* 2002).  $CF_1/CF_2$  association is implemented through a mechanism of synaptic plasticity incorporated between layers III and IV. This mechanism simulates Hebbian-type plasticity and creates associations between different frequencies. It is implemented uniformly such that all cross-frequency connections from layer III to IV are modifiable through this type of plasticity. Activation of several synapses on the same postsynaptic node (synaptic facilitation, i.e., combination sensitivity) is implemented through increases in the strength of synapses from simultaneously activated inputs to the same node. The parameters for this type of plasticity are chosen to provide a relatively quick (within  $\sim 15$  ms) plastic change of the response in a postsynaptic node. Although combination sensitivity can be a powerful mechanism to increase



**Figure 3.** Multifrequency, excitatory tuning curve for a single node. This is produced as a result of Hebbian learning of harmonic combinations within layer IV of the network and simulates the tuning of single neurons within the mustached bat's auditory system. The response threshold is mapped for input intensity against the frequency axis on a logarithmic scale. Threshold intensity for each frequency is defined as the minimal firing rate at the input required to evoke firing at a node.

the signal-to-noise ratio and have a significant contribution on the phenomenon of streaming at high rates of stimulus presentation, its specific role in our model is not investigated in detail.

A typical tuning curve for a node (at layer IV) tuned to a frequency of 60 kHz is shown in figure 3. After learning a combination of the fundamental frequency of 30 kHz, its first harmonic at 60 kHz and its second harmonic at 90 kHz, this node starts to respond to all three frequencies presented individually to the network. The multiple peaks in this graph are similar to the real V-shaped tuning curves of auditory neurons in both the inferior colliculus and the auditory cortex (Portfors and Wenstrup 2002, Kanwal *et al* 1999, Sutter 2000). In our network,  $CF_2$  frequencies are not over-represented and tuning to these frequencies is not modelled to be sharp (see figure 3) as in the auditory fovea of mustached bats. This feature is not critical to generate streaming (differential response suppression) and was excluded from our network so as not to overspecify the model.

The bats auditory system was simplified and no attempt was made to simulate the enhanced frequency tuning and the presence of a pronounced auditory fovea for a  $CF_2$  frequency (about 59 kHz for this subspecies of mustached bats). Instead, a general wiring scheme, which could be present in any mammalian species, was used. Thus, the streaming behaviour described here does not require any of the special foveal neurophysiological features in this species. Whereas this fovea is important for bat echolocation, 'streaming' is a general phenomenon known to be present in humans and primates and perhaps most mammals. Therefore, even though many auditory neurons in mustached bats are tuned to the resting  $CF_2$  frequencies, the special foveal neurophysiological features in this species are not required to model the neural mechanisms for streaming.

For the initial state of the network, the resting membrane potential for all nodes ranged from  $-60$  to  $-70$  mV, the activation threshold equalled 8–10 mV and the PSP amplitude and duration were chosen to range between 2 and 7 mV and between 5 and 15 ms. Synaptic delays varied from 1 to 5 ms. For the recurrent inhibitory connections at layer IV, synaptic delays were equal to 5 ms and the amplitudes for the IPSP varied from  $-7$  to  $-1$  mV with a relaxation time of 15 ms. Thus, the model parameters, including those of the recurrent inhibitory connections, which are the most important for the phenomenon of streaming, were chosen to correspond to the real neuronal and synaptic parameters known from the neurophysiological data and are common to many mammalian species. We would like to stress that it was not necessary to assign any special values to those parameters in order to realistically simulate the streaming phenomenon.

The key property that allows this network to perform stream segregation is the recurrent connections of layer IV via layer V. Horizontal projections from layer IV go to layer VI and return to layer IV through the inhibitory elements of layer V. Each node at layer IV has inhibitory projections to itself, the adjoining and next-neighbouring frequency channels. This implements recurrent and lateral inhibition that results in an increase in frequency specificity at layer IV. Layer VII is the 'recognition' layer of the network where a specific thresholding of neural inputs from layers IV–VI takes place. This recurrent inhibition parallels thalamocortical loops and corticofugal inhibition, as reported by Gao and Suga (1998). In this version of the model, the activity of layer IV was used as the output of the model. In order to obtain a network analogue of the event-related potential (ERP) at a specific tonotopic location, we averaged over 3–5 trials with the same stimulus presentation, and low-pass filtered (below 70 Hz) the transmembrane potential of the corresponding node at layer IV.

The simulation experiments were designed as follows. The external stimuli to the network consisted of the same (AA... or BB...) or alternating (ABAB...) tone bursts of 25 ms duration each and of different frequencies. Stimulation at a given frequency was simulated by activation of the corresponding frequency channel at the input layer. The stimulus bursts were delivered to the network at presentation rates (PRs) of 5, 10, 20 and 40 Hz. The number of tone bursts within one trial varied from 4 at PR = 5 Hz to 20 at PR = 40 Hz keeping the duration of each trial approximately the same (500–800 ms). In the paradigm which is of particular interest in this paper, frequencies A and B were harmonically related to each other; e.g., A = 60 kHz (CF<sub>2</sub>) and B = 30 kHz (CF<sub>1</sub>, the generalized fundamental frequency). In reference to echolocation, the CF<sub>2</sub> is typically Doppler shifted when a bat is flying.

## 2.2. Neurophysiological experiments

The surgery, acoustic stimulation methods and recording of neural activity for these experiments are similar to those described previously (Kanwal *et al* 1999) and therefore only briefly presented here.

Mustached bats, *Pteronotus parnellii*, were caught in the wild and transported to the Animal Care Facility at Georgetown University. Bats were hand-fed meal worms for 1–2 weeks until they adjusted to their new diet. For surgery, bats were anaesthetized with an isoflurane/air mixture (medical grade, Anaquest, USA). An initial dose of 2.5–3.0% isoflurane supplied for 1–2 min was followed by a continuous stream of approximately 1% isoflurane/air mixture. A skin incision was made at the mid-line on the head and a 2 mm diameter metal post was affixed just behind the intersection of the sagittal and coronal sutures with cyanoacrylic glue (Loctite 411). Bats were allowed to recover for 3 days before the first recording session.

Sound stimuli were presented from two condenser loudspeakers mounted on a vertical hoop and positioned 1 m directly in front of the bat. The loudspeakers were positioned adjacent to each other in the same azimuth to avoid binaural effects. The stimulus generation and delivery system consisted of two analogue channels. Sounds consisted of 25 ms long CF tones. The loudspeakers were calibrated by placing a B&K microphone at the position of the ear; the condenser speakers were flat ( $\pm 6$  dB) between 20 and 100 kHz with a significant roll off at 120 kHz. The maximum amplitude level that could be delivered from speaker A was 98 dB SPL (re 20  $\mu$ Pa, RMS) at 90 kHz and that for speaker B was 95 dB SPL at 85 kHz. The average SPL of the tones in our experiments was approximately  $50 \pm 10$  dB SPL for the high frequencies and  $70 \pm 10$  dB SPL for the low frequencies.

For electrophysiological recordings, the head was restrained by clamping the metal post and the bat's body was suspended in a Styrofoam mould by elastic bands in a heated (31 C), sound proofed and echo-attenuated chamber (IAC 400A). The activity of single neurons

was recorded with sharpened, vinyl-coated tungsten microelectrodes with tip diameters of approximately 10  $\mu\text{m}$  and impedance of about 1  $\text{M}\Omega$ . All surgery and animal experiments were performed with approval of the Georgetown University Animal Care and Use Committee.

In order to test model predictions, we recorded single/few unit (SU) responses as well as ERPs from eight sites in the DSCF area in the primary auditory cortex of four mustached bats. Tuning curve data based on extracellular recordings were obtained from over 200 neurons in the primary auditory cortex of mustached bats (Kanwal *et al* 1999). These data guided our present neurophysiological recordings. ERPs and poststimulus time histograms (PSTHs) from SU activity and summed SU activity were calculated over 200 presentations of acoustic stimuli using a bin width of 5 ms. ERPs were obtained by bandpass filtering between 1 and 300 Hz. SU data was bandpass filtered between 600 Hz and 3 kHz. Neural responses were evoked by sequences consisting of tones at a single frequency (AAA . . . or BBB . . .), or of tones of alternating frequency (ABAB . . .). Stimuli were presented at 10 dB above the threshold of the neuron/s at the recording site.

### 3. Results

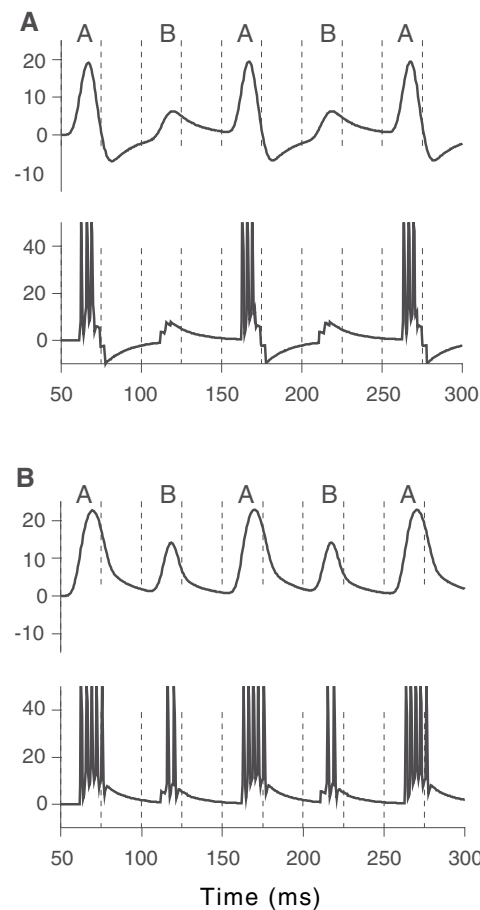
#### 3.1. Neural network for auditory stream segregation

The simulation results obtained using harmonically related A- and B-tone frequencies are presented in figure 4. Within each panel, the simulated 'ERP' is shown at two different tonotopic network locations, which correspond to the presented frequencies: A = 60 kHz, and B = 30 kHz. Those locations are referred to as the '60 kHz node' and '30 kHz node', respectively. Both nodes acquire multi-peaked tuning as a result of associative learning, and are therefore tuned to both frequencies A and B. Nevertheless, a node's initial BF remains to be slightly dominant, i.e., the '60 kHz node' responds better to the sequence AA while the '30 kHz node' responds better to sequence BB. When presented with the alternating sequence ABAB, these nodes respond with a series of ERPs of alternating amplitude (the bottom panels at PR = 5 and 10 Hz). When the PR increases to 40 Hz, the responses of both nodes to their preferred frequencies (AA or BB) actually become smaller. This is because the recurrent inhibition after each response affects the following response.

As a result of the properties modelled, nodes in the output layer respond to either stimulus A or stimulus B when those stimuli are presented at the low PR values of 5 and 10 Hz. When the alternating sequence AB is presented at high repetition rates (20 and 40 Hz), either node responds only to its own BF and the component of the response to another frequency is suppressed. This effect is due to the temporal interaction of responses A and B as follows. For the '60 Hz node', the A-response is slightly greater than the B-response. It is important to note that the recurrent inhibition immediately following each excitatory response (A or B) is proportional to this response. Accordingly, the A-response results in a stronger recurrent inhibition. At the high PRs, the following B-response is affected by this inhibition and becomes partially suppressed at PR = 20 Hz and almost completely suppressed at PR = 40 Hz. In turn, the weak B-response is followed by the proportionally weak inhibition and the following A-response is suppressed to a lesser extent. As a result, the 60 Hz node 'latches on' to its BF at A and fails to respond to a second tone at a non-best frequency. Similarly, the 30 Hz unit 'latches on' to its BF at B and does not respond to frequency A.

The resulting rate of the nodal responses at PR = 20 and 40 Hz is half the PR (10 and 20 Hz, respectively). Note that the nodal responses are out of phase with each other. This means that at high PR, two independent loci of activity are created at layers IV–VII. One locus can be considered as representing sequence A and another as representing sequence B. In this situation, two intermixed streams of tones (A and B) are spatially segregated within the



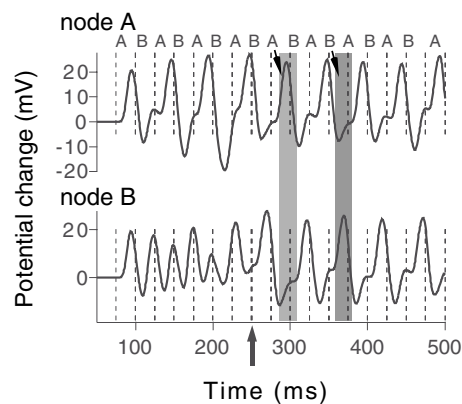


**Figure 5.** Population (nodal) responses (upper graphs) and SU responses (lower graphs) with (A) and without (B) recurrent inhibition. In the presence of recurrent inhibition, the alternating AB stimulus sequence at PR = 20 Hz causes the 'latching-on' effect of the A-node which fires only in response to tones A (A). The same A-node responds to both tones A and B in a disinhibited population (B).

### 3.2. 'Build-up' of the response to tone sequences

An essential feature of the phenomenon of streaming is its time dependence; i.e., the fact that a finite amount of time is generally needed for a listener to start hearing the alternating sequence as two individual streams. In figure 6, at the beginning of a BABA sequence, node B responded both to A and B tones. These tones were taken as BF and off-BF frequencies within one of the two peaks of tuning at the node, e.g., 30 and 32 kHz. Both nodes latched onto their corresponding BFs, either A or B. Node B, however, lags and latched on to the B frequency only ~125 ms after the start of stimulation. This observation suggests that the differential suppression of responses to tones away from the BF may need some time to 'build-up'. The time course of this delayed latching or 'build-up' phase was not always the same.

Latching on of the responses depends on the temporal interaction between the previous and the consequent components of the neurophysiological response to the alternating sequence ABAB. Let us assume that tone 'A' represents the BF of a cortical neuron. Then the response

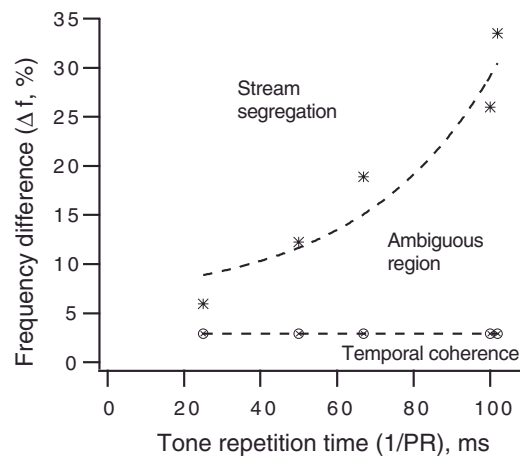


**Figure 6.** Time dependence of the phenomenon of streaming in the model.  $PR = 40$  Hz,  $f_A = 30$  kHz,  $f_B = 32$  kHz ( $\Delta f = 6\%$ ). The bottom and top traces show the activities of nodes A and B, respectively. Stimulation sequence BABA starts 75 ms from initiation of activity in the network. Initially, node B responds both to A and B tones. Two distinct out-of-phase patterns of activity in nodes A and B take 125 ms after the start of stimulation (marked by arrow) to build up.

to 'A' increases during latching on compared to the same response in the 'control' sequence AAAA (figure 6). Such facilitation occurs in parallel with, and is dependent upon, the concurrent decrease of the response to B within the sequence ABAB. The decreased B-response causes the weaker recurrent inhibition and a weaker forward masking effect. Thus, the subsequent response to the next A tone becomes less suppressed if compared with the response to the 'control' stimulus AAAA. In turn, the increased A-response causes a stronger recurrent inhibition and a stronger forward masking effect. As a result, the successive response to tone B becomes suppressed. The latching on represents 'competition' between tones A and B to elicit a response in which the tone eliciting the strongest response wins. This is a time-dependent phenomenon because the very first responses to A and B within the sequence ABAB are 'regular' (there is no facilitation or suppression yet; see the first 150 ms in the B-node response in figure 6). Over a period of about 200 ms there is a gradual development ('build-up') of facilitation/suppression as a result of temporal interactions in the responses in the same neuron to tones A and B. This process may be equated to that of a system achieving an 'attractor', a state of dynamical equilibrium between the responses to tones A and B. The temporal build-up of the latching-on phenomenon appears to be sensitive to the initial conditions of the model because the very first responses to A and B within the sequence ABAB may vary due to the pre-existing excitatory or inhibitory processes on the nodes but this has not been explored in detail and is not presented here.

### 3.3. Model experiments

We examined the dependence of 'streaming' in our model on two parameters:  $\Delta f$ , the frequency separation between tones A and B and the PR. The frequency of tone A was fixed at 30 kHz and the frequency of tone B was adjusted such that the relative value  $\Delta f(\%) = (f_B - f_A)/f_A$  varied within the corresponding peak of the tuning curve for node A from 3 to 33%. For each value of  $\Delta f$  tested, the PR was increased from 10 to 40 Hz, which corresponded to tone repetition times ( $1/PR$ ) of 100 to 25 ms. 'Streaming' was identified within the network when a separation of activities at nodes A and B took place in response to



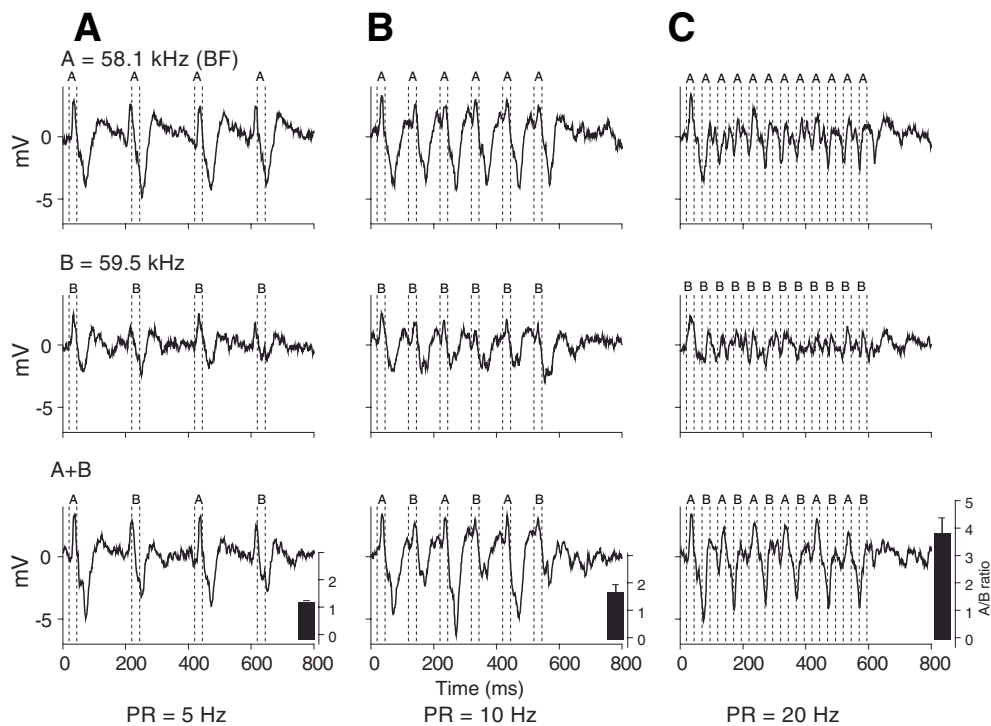
**Figure 7.** Dependence of streaming on frequency separation ( $\Delta f$ ) and PR. Asterisks indicate the higher PRs which were necessary to obtain streaming at progressively smaller values of  $\Delta f$ . The best exponential fit to the asterisks (dashed curve) thus forms a boundary analogous to the 'temporal coherence' boundary in the psychophysical experiments. The horizontal dashed line shows the frequency separation below which it was impossible to obtain streaming at any PR.

the alternating sequence ABAB, i.e., each node responded only to its BF. This was considered a 'temporal coherence boundary' as defined in psychophysical experiments for a similar range of tone repetition times (van Noorden 1975). At lower  $\Delta f$ , streaming occurred at higher PR as shown in figure 7. The best fit of the 'temporal coherence boundary' appeared to be an exponent and it is shown in figure 7 by a dashed curve. In addition, at the minimal frequency separation achievable with our model ( $\frac{1}{24}$ th of an octave,  $\Delta f = 3\%$  at  $f_A = 30$  kHz), streaming did not occur at any PR. Thus, the horizontal line at  $\Delta f = 3\%$  was taken as a 'fission boundary' (van Noorden 1975).

### 3.4. Neurophysiological experiments

BFs of DSCF neurons were obtained from several recording sites to map the boundaries of the DSCF area in each animal. Here we report new temporal integrative properties of DSCF neurons in response to auditory stream stimuli as predicted by our model. ERP activity at recording sites in the DSCF area generally exhibited excitatory frequency tuning to 26 kHz ( $BF_{\text{low}}$ ) and to 58.1 kHz ( $BF_{\text{high}}$ ) tones. These frequencies would correspond to the standardized 30 and 60 kHz frequencies in the model, except that in the subspecies (*Pp rubiginosus*) the mean resting  $CF_2$  is  $58.28 \pm 0.10$  kHz ( $n = 2200$ ) and neurons at the recording site are tuned to a Doppler-shifted frequency of 58.1 kHz. Peak response latencies of ERPs were in the range  $16.3 \pm 2.3$  ms ( $n = 4$ ).

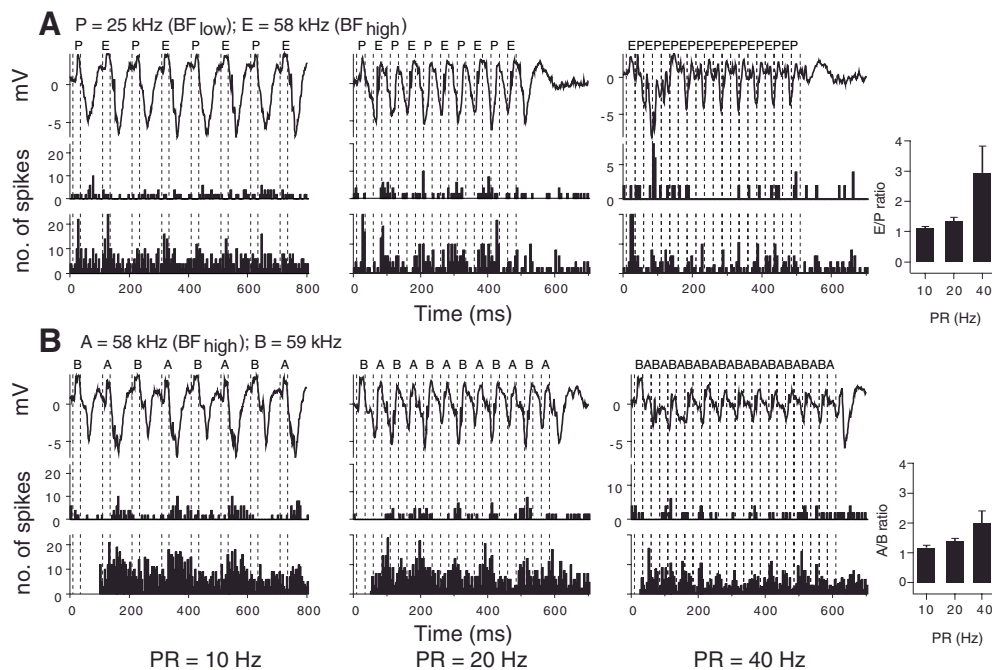
In figure 8, recordings of ERPs from one site in the DSCF area are shown for PRs of 5, 10 and 20 Hz for single tone (AAA and BBB) as two control stimulations and alternating-tone sequences (ABAB). The two tones presented are within the 58.1 kHz excitatory response area. At the slowest PR (5 Hz), ERPs showed equivalent patterns to the three stimulus conditions (A alone, B alone and A + B). At rates of 10 Hz, the waveform is biphasic, for both A and B when presented alone and as alternating tones. At rates of 20 Hz, the response to B alone declined significantly. The response to A alone also declined at high rates, but was restored when presented as a tone sequence alternating with B. The response to B, however, was suppressed in this alternating sequence at a PR = 20 Hz. Thus, even though two excitatory frequencies



**Figure 8.** ERPs from the primary auditory cortex of the mustached bat recorded in response to sequences of pure tones A, B and their alternating combination AB. Stimulus sequences are presented at rates of 5, 10 and 20 Hz. Stimulus B has a frequency of 59.5 kHz, which is slightly different from the BF (off-BF). Bar graphs (with standard deviations) on the right show the magnitude of differential response suppression as a ratio of the peak-to-peak response for A versus B. The time course of the stimulus sequences is shown by the dotted lines.

were present, the local population of neurons at this site responded better to A than to B. This is shown by bar graphs appended at the right of each recording in the ABAB condition. The bar graphs represent the peak-to-peak amplitude for the A versus B response at each rate. The graphs show that with an increase in PR, the peak of the ERP to A improves over that for B. At this recording site, the response magnitude and hence the streaming effect deteriorated at stimulus PRs of 40 Hz. These data represent the condition of a general mammalian paradigm for obtaining auditory stream segregation, except that in the flying bat, the two frequencies may correspond to a target echo and a nontarget echo (clutter) CF.

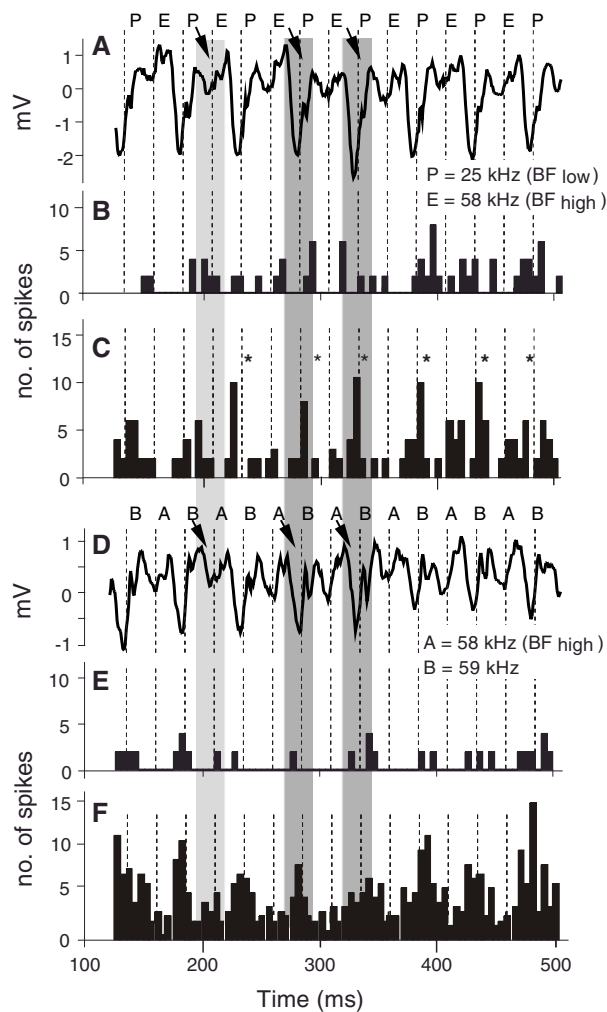
Figure 9(A) shows the effect of different PRs on ERPs, SU responses and on summed SU responses for a sequence of alternating tones. The tonal stimuli were designated as 'P' ( $BF_{low}$ ) and 'E' ( $BF_{high}$ ), respectively, and correspond to the pulse  $H_1$  and echo  $H_2$  frequencies in an echolocation signal. The response of the SU is plotted as a peri-stimulus time histogram (PSTH) that was recorded simultaneously with the same electrode and at the same site and is shown on the same timescale directly below the ERP. Single to few unit responses summed from three different recording sites and plotted as PSTHs are shown at the bottom in each panel. Streaming was observed at PRs of 40 Hz. At a PR of 10 Hz, the response followed each tone regardless of its frequency. At this rate, ERPs showed a clean positive potential (peak) to each tone, whereas the response of the SU showed infrequent increases in response rate to either of the two tones. At PR = 20 Hz, ERPs continue to show a peak for each tone in the sequence,



**Figure 9.** ERPs at a single recording site and PSTHs from the corresponding SU activity in response to (A) a pulse ('P') and echo ('E') CF sequence, and (B) an alternating sequence of two tones, BF ('A') and an off-BF ('B') excitatory frequency presented at different rates. The PSTHs in the lowest portion of each panel show summed SU activity over several recorded units ( $n = 3$ ). Differential response suppression or 'latching-on' is observed in the ERP and summed SU activity at PR = 20 and 40 Hz in A, but is less obvious or absent in B. The observed effect is independent of the order of presentation of the two tones as shown in the 40 Hz stimulus sequence. The bar graphs on the right show the mean response ratio (with standard deviations) to P versus E and A versus B frequencies. The time course of the stimulus sequences is shown by the dotted lines. Bin width of PSTH = 5 ms.

but SUs responded only infrequently, with a slight increase in response rate typically at or around a negative peak in the ERP in response to the echo CF<sub>2</sub>. At a PR = 40 Hz, the ERPs show a 'build-up' in the response pattern over the first 150 ms that closely matches the potential changes at nodes observed in our neural network simulations of streaming. This is followed by a sharp negative peak following the presentation of 'E' (BF at 58 kHz) and coincides with a sharp peak of excitatory response in the single neuron activity represented by the PSTH. The SU response pattern to the pulse and echo sequence was disrupted with transient inhibition following the initial excitatory response whereas the averaged ERPs showed a more consistent response pattern with evidence of streaming as indicated in the bar graph on the right.

For two frequencies corresponding to the echo CF<sub>2</sub> clutter (59 and 58 kHz), differential response suppression was not pronounced. At a PR = 10 Hz, the neuron's response coincided with a negative peak in the ERP immediately following the 'A' tone (figure 9(B)). At rates of 20 Hz, the SU response was inhibited at first and then showed a pattern of excitatory activity to the alternating 'A' tone that corresponded to the negative potential in the ERP as for 'E' (same frequency as 'A') in figure 9(A). At high PRs ( $\geq 20$  Hz), the ERP consisted of a sharp negative peak following tone 'A' and a shallower negative peak following tone 'B'. Once again, SU responses showed a sporadic response pattern to successive CFs. The summed SU activity, in contrast to SU activity, showed a tendency to respond to the 'A' tone (BF<sub>high</sub>) better



**Figure 10.** A 400 ms time slice of a sequence of neurophysiological responses to a pulse  $CF_1$  and echo  $CF_2$  sequence presented at a rate of 40 Hz. (A) ERPs and (B) SU activity in response to the PE stimulus sequence ( $P = 25$  kHz) and echo  $CF_2$  ( $E = 58$  kHz). (D) ERPs, (E) SU activity and summed SU activity in response to two different echo  $CF_2$  (AB) stimulus sequence presented at 40 Hz. Streaming is observed in the ERP at PR = 40 Hz in A, B and C as judged by relative facilitation/inhibition of the responses to the specific components within the stimulus, but is less obvious or absent in D, E and F (see text for details). A ERP consists of a 'positive peak–negative peak' complex. Differential response suppression for ERP in  $A = 3.9 \pm 0.5$ ; and in  $D = 2.0 \pm 0.2$  (mean  $\pm$  s.d.). Examples of facilitated/reduced responses are marked by asterisks for the summed SU activity. The light grey bar indicates an exemplary response for the echolocation pulse and 'B' stimulus frequencies and the dark grey bars indicate two examples for the echo and the 'A' stimulus frequency (58 kHz). Bin width = 5 ms.

than to the B tone. Because of the different starting frequencies in the pulse–echo sequence in different recordings, the first 50 ms of the response are not shown for the summed SU activity. A 'build-up' phase of the response is not as prominent and sharp as in figure 9(A). A large magnitude negative peak, suggestive of a rebound depolarization, is seen at the end of the stimulus sequence and may correspond to a resetting of the state of the network. The peak-to-peak response ratios are quantified and presented as a bar graph to the right.

A 400 ms time slice of the recordings is shown in figure 10 to clarify the temporal relationship of the various responses (ERPs and PSTHs of SU and summed SU activity) to successive stimuli at PR of 40 Hz. Here, in contrast to ERPs taken from a single recording site shown in figure 9, ERPs to the two specified stimulus sets (PEPE and ABAB) are averaged across the same three recording sites for which SU responses are also presented as summed PSTHs. Whereas the SU responses are erratic and fail to show clear streaming, the summed SU responses show periodic peaks of activity in the PSTH that roughly correspond to the presentation of only E in the PEPE sequence as highlighted by light (for P) and dark (for E) grey bars (figures 10(B) and (C)). For the two-echo (ABAB) sequence, the peaks of negativity to the 'A' tone (BF) are relatively small (figure 10(D)). SU recordings failed to show a clear and consistent response pattern (figure 10(E)). The summed SU activity is relatively noisy, but does show the emergence of periodic peaks of activity to the 'A' tone (figure 10(F)).

#### 4. Discussion

In psychophysical experiments, the ability of a listener to segregate different tones in an alternating sequence ABAB depends on the frequency separation between tones A and B as well as the PR. In an earlier study (van Noorden 1975), three major zones were identified on the plane defined by these two parameters: (a) an area where streaming always occurs (relatively large frequency separations and high PRs), (b) an area where streaming never occurs and the listener hears a trill, i.e., a connected sequence of alternating frequencies (very low frequency separation, almost no dependence on the PR) and (c) an area between those two regions where the percept is ambiguous. These three features of streaming have been simulated in other modelling schemes that are operational in nature and use a mathematical approach (Beauvois and Meddis 1996, McCabe and Denham 1997). Our network model simulations showed (1) a similar dependence of the stream segregation phenomenon on frequency separation and PR as well as (2) its time dependence.

##### 4.1. Possible mechanisms behind the temporal interactions between neural responses

The fact that we were able to successfully simulate (at least qualitatively) the temporal interactions in neural responses to tones sequences that were observed in our, and other (Fishman *et al* 2001) data, indicates that these neural phenomena may be produced by the mechanisms implemented in our model. Whereas it is difficult to prove with a model that the postulated mechanisms are necessary for (re)producing an observed effect; it is possible to demonstrate that specific mechanisms are sufficient for (re)producing the observed effects. Thus, the results of our formal simulation indicate that recurrent and lateral inhibition, which have been proposed as the neural basis of perceptual stream segregation by Fishman *et al* (2001), are indeed sufficient to reproduce qualitatively the main temporal interaction effects.

Our network model incorporates recurrent inhibition as a neurophysiological mechanism of forward suppression (masking) of rhythmic responses. Forward masking has been explained previously in the context of a limited ability of auditory cortical neurons to follow repetitive stimulation and detect small gaps within it (Eggermont 1995, 2000, 2001, Brosch and Schreiner 2000). In these studies, two mechanisms of forward masking were considered, namely, neuronal membrane after-hyperpolarization (AHP) and recurrent inhibition. Theoretically, both mechanisms may contribute to the phenomenon of forward masking because they have approximately the same time constants (Eggermont 2000). The analytical modelling of the neuronal post-excitatory suppression suggested that it could be best explained if recovery from an AHP was incorporated (Eggermont 2000). However, suppression of a neuron in

the absence of its excitatory response by frequencies outside its tuning curve indicates a dominant role of recurrent inhibition in this process (Brosch and Schreiner 2000). Our data showed a limited ability of SU discharges to consistently follow the ERP time course. The latching-on phenomenon was observed only within the spiking activity summed over several units. This strongly indicates that latching on (and, consequently, streaming) is a population-based synaptic phenomenon. Whereas single neurons fire scarcely and probabilistically in coincidence with specific components of the ERP, their firing is driven by a population response that corresponds well with the ERP. Under such conditions, it is unlikely that the AHP is the only mechanism that can explain the latching-on phenomenon because the AHP is a 'SU-based' mechanism. Contrary to that, recurrent inhibitory projections diverge and converge to many units thus providing a mechanism that is implemented within population behaviour. Clearly, further neurophysiological data on the response patterns of cortical neurons to sequences of tones at different frequencies relative to the BF at the recording site are needed before the detailed mechanisms responsible for the phenomena observed here can be ascertained.

The presence and the role of multifrequency tuning in the mustached bat's auditory system extends our understanding of the neural mechanisms associated with auditory stream segregation as revealed by human psychophysical and primate neurophysiological data. Our modelling and neurophysiological data show that a neural mechanism based on recurrent inhibition is not limited by the critical bandwidth for single frequency tuning. Since neurons in the primary auditory cortex of mustached bats typically have either two or three widely separated excitatory response areas, therefore the same neuronal mechanism may be applicable beyond a single critical band for multifrequency tuned neurons. Furthermore, inhibitory frequency tuning that is common among auditory neurons (Shamma *et al* 1993, Kanwal *et al* 1999, Loftus and Sutter 2001) may create perceptual boundaries for streaming. It is possible that these boundaries may even shift depending on the direction from which sound stimuli are presented since the shape of tuning curves can change depending upon the direction of the sound source (Xu *et al* 1994).

Differential response suppression appears to be a general mechanism, which may come into play each time a neuron is successively excited by sounds to which it responds differently, whether or not the difference in excitation is due to different locations of the sounds on the frequency axis, or some other factor. Consequently, the type of temporal interactions between neural responses described here may underlie stream segregation effects in a larger variety of stimulus conditions than originally suspected. For example, they might explain recently discovered stream segregation effects with sounds that have identical long-term spectra, such as amplitude-modulated broadband noises (Grimault *et al* 2002), or with sounds that induce similar patterns of excitation along the cochlea, like high-pass filtered harmonic complex tones (Vliegen and Oxenham 1999, Vliegen *et al* 1999, Grimault *et al* 2000). As long as there are neurons in the central auditory system that respond differentially to different sound parameters (for example, neurons that are tuned to different rates of amplitude modulation or are otherwise selective to certain temporal characteristics of the sounds), differential suppression may take place. Even in the minimalistic situation where the successive sounds are identical in all respects except for intensity, stream segregation is known to occur at the perceptual level (Bregman 1990), such that responses to the lower-intensity sounds may be suppressed by preceding higher-intensity sounds provided they are sufficiently close in time. This is a classical case of forward masking, which limits the ability of auditory cortical neurons to follow repetitive stimulation and to detect small gaps within it (Brosch and Schreiner 1997, 2000, Brosch *et al* 1999, Eggermont 2001). The ability of our model to reproduce the forward masking phenomenon emphasizes the importance of synaptic inhibitory processes in this population behaviour.

#### 4.2. Correspondence of the neural network model with experimental findings

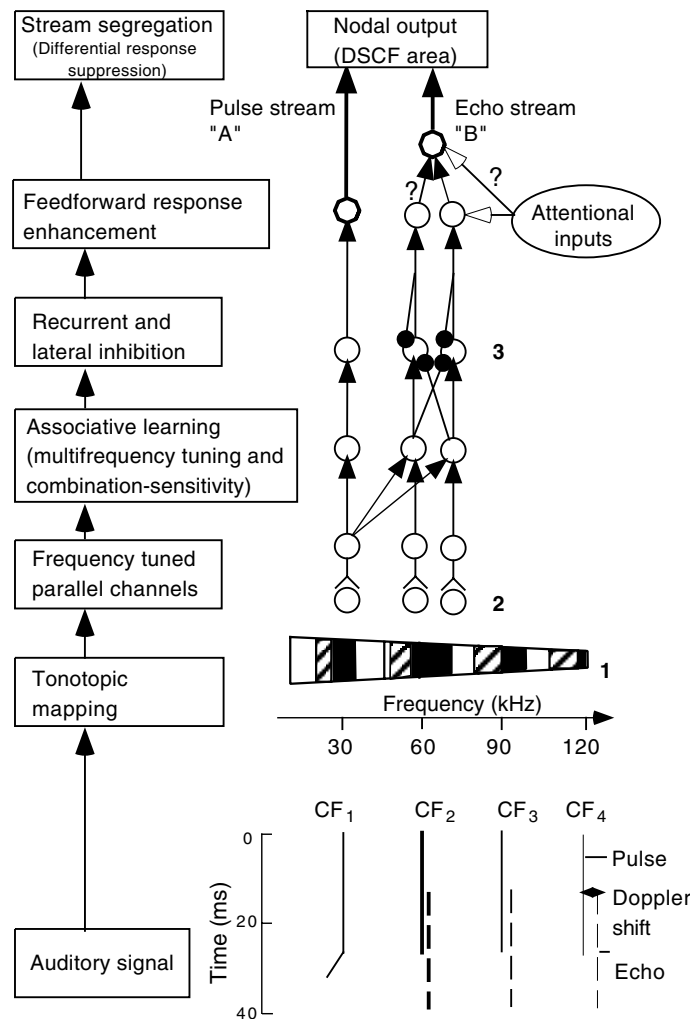
Integrative phenomena such as ‘streaming’ are revealed by the presentation of a sequence of alternating tones. At the neuronal level, this can be based on the implementation of general mechanisms involving two-tone interactions, such as forward masking and may include response properties of specialized combination-sensitive neurons that are well documented in the bat’s auditory system. Since our neural network model is based on the mustached bat’s auditory system, we consider multifrequency tuning, combination sensitivity and recurrent inhibition as playing a central role in the generation of streaming. In the mustached bat, pulse–echo combination-sensitive neurons are present at the level of the inferior colliculus and below (Mittmann and Wenstrup 1995). This means that neurons showing streaming can be present at subcortical levels (layers IV–VI of our model), although presently observed at the cortical level per our recording paradigm.

Considering our model and the neurophysiological data available in mustached bats, we can relate some of the processes implemented in the model to the levels of processing within the auditory system (see figure 11). Since the network is mainly process driven, it cannot be directly correlated with the complex ascending and descending organization of the auditory system in general. However, it may be estimated that nodal outputs of the model roughly correspond to hair cell output (layer I), nerve fibre (layer II), brainstem and midbrain (layers III and IV) and thalamocortical levels (layers V–VII) of processing. The ERPs recorded in our experiments appear to correspond well with the output of the nodes in layer IV. These ERPs correspond to synaptic input currents onto probably less than 100 neurons given the small  $\sim 10 \mu\text{m}$  diameter of the recording electrodes. However, at some point in the auditory system, this must be converted into spike activity that influences motor commands. Since the activity of single neurons in the auditory cortex fails to produce streaming-like behaviour, we propose that this perception occurs at levels of the frontal cortex, which receives projections from the auditory cortex and also projects back to it (Fitzpatrick *et al* 1998). The dramatic improvement in the response by summing the activity of only a few units (see figure 10) suggests that the order of this convergence is relatively low and it may take the coordinated activity of tens of neurons to construct the percept of streaming for a particular frequency combination.

#### 4.3. Neurodynamics for streaming: SUs versus ERPs

Extracellular SU recordings in the cortex typically represent the firing of the large pyramidal neurons in layer V of the auditory cortex. However, the properties of cortical neurons do not passively represent the inputs from the thalamic tract and are also shaped by inhibition (Eggermont 1996, Norena and Eggermont 2002). Therefore it is not surprising that ERPs showed a stream preference or segregation effect per the observed differential response suppression, whereas SU responses were erratic and did not show stream segregation at high PRs (40 Hz) of the stimulus set. This property re-emerges, however, when one sums the activity of multiple units recorded from different sites (see figures 9 and 10), suggesting that it is not lost altogether. Our ERP recordings were obtained from relatively superficial to midlevel (approximately  $500 \mu\text{m}$ ) regions of the auditory cortex. We assume that our ERP measurements are based primarily on the activity associated with the thalamo-cortical input fibres since field potentials depend more on synaptic input currents and less on cortical spikes. The ERPs obtained in our data probably represent the summed current flow from a field of about  $50 \mu\text{m}$  spherical area at the tip of the electrode.

The streaming properties in SUs as suggested by our data are at best probabilistic. This is because of the sparseness of the spike response in auditory and especially auditory cortical



**Figure 11.** Schematic representation of the neural mechanisms for stream segregation and their relationship to the auditory system in mustached bats. Rotated spectrogram (time on the vertical axis) at the bottom shows the four harmonics in the CF component of the echolocation pulse (adapted from Suga 1990). For purposes of clarity, the FM sweep is included only in the fundamental of the pulse; CFs in the echo are shown as dashed lines. The basilar membrane (1) produces a tonotopic map separating the pulse–echo CFs in response of the hair cells (2). The mechanisms of multifrequency tuning, associative learning, recurrent and lateral inhibition occur at intermediate stages (3) before the emergence of differential response suppression underlying the perception of two independent pulse and echo streams at the cortical level. Attention probably plays an important role in the perceptual process, but details of this are not known.

neurons. Most likely, SU responses are further constrained to encode additional stimulus properties so that they lose the dynamic characteristics necessary to exhibit robust streaming. This means that streaming is best encoded by the response of multiple neurons, hence our consideration of streaming as a truly integrative property. The output of multiple neurons may of course converge on to single neurons at higher levels, such as the frontal and premotor cortex. Electrophysiological recordings from these locations could yield SU data that would

show more robust responses to streaming than the auditory cortex. In fact, recordings from a frontal auditory field in mustached bats show that responses of single neurons can have latencies of over 100–200 ms which is equivalent to the time over which the ERP builds up (Kanwal *et al* 1999). To obtain robust responses and study this in a more systematic fashion, however, it may be necessary to record SUs from flying or behaving animals.

#### *4.4. Auditory scene analysis in bats: tracking sounds in the natural environment*

Mustached bats live deep inside caves where it is dark and visual stimuli are non-existent. Under these circumstances, auditory scene analysis is not only biologically relevant, but critical for survival. Auditory scene analysis involves processing and integration of different aspects of environmental stimuli. Localization of sound sources is one of the most important aspects of this function of the auditory system and has been extensively studied at both the behavioural and neuronal levels (Pollak and Casseday 1989, Yang and Pollak 1994, Park *et al* 1996). Other aspects of auditory scene analysis include the perceptual organization of simultaneously and sequentially occurring acoustic elements. Neuronal mechanisms underlying this type of analysis are not well understood. Thus, the perception of virtual pitch results from the neuronal processes that either combine the simultaneously present harmonics or track amplitude modulation thereof (Medvedev *et al* 2002).

We know relatively little about the neuronal integration of sequentially occurring acoustic elements, except for combination-sensitivity to FM components in the pulse and echo (O'Neill and Suga 1979). This phenomenon, however, is restricted to response enhancement to the temporal combination of two acoustic components. Tracking of sound stimuli and acoustic images (either background or insects in case of bats), however, involves a stream of sounds. We have used minimal pulse–echo components to demonstrate the neural correlate of streaming in the ERPs and summed SU activity. These components were chosen knowing the response properties of neurons in the DSCF area. A more adequate stimulus should eventually be used that consists of the entire set of pulse and echo harmonics together with the FM components. Different durations of silent intervals can also be introduced in the middle of a pulse–echo (PEPE) or echo–echo (ABAB) sequence to define the robustness and tolerance for the system to trigger differential response suppression. While such experiments would be interesting, they are beyond the scope of this paper. From what we know so far, only the selected pulse–echo frequencies are conveyed to neurons in the DSCF area, so that adding additional components to the stimulus repertoire are unlikely to change the response properties of neurons in this area.

#### *4.5. Do bats segregate pulse and echo streams?*

Dynamic auditory scene analysis is critical for prey tracking and capture in all echolocating bats. The results of experiments in a few species of FM bats demonstrate that they can assemble information about echo delay changes over time (Moss and Surlykke 2001). However, unlike FM bats, which may use a stroboscopic method for insect pursuit and capture (Feng *et al* 1994), CF–FM echolocating bats probably deploy a different strategy, namely that of auditory streaming. Clutter is less of an issue in FM bats that hunt in an unobstructed environment above rather than within the canopy of trees. Repetitive echolocation pulses create an ideal stimulus for the generation of an auditory stream that can facilitate sound source tracking in a cluttered environment despite discontinuities in the stimulus. The tracking of signals in a bat's acoustic environment is perceptually equivalent to the 'cocktail party effect' in humans, where despite brief discontinuities, meaningful modulations in a speaker's voice can be tracked without a significant loss in fidelity. Furthermore, the alternating tones used to demonstrate

the streaming effect may be considered to be somewhat unnatural in humans, but are perfectly natural stimuli for echolocating bats.

It is important to note that auditory stream segregation is relevant only or primarily during the search phase of echolocation when CFs are long and the pulse rate is between 10 and 40 Hz (Novick and Vaisnys 1964). In this phase, repeated pulse–echo stimuli that lead to streaming in the recorded ERPs at specific loci may in fact ‘prime’ single neurons at a locus to respond to the first pulse–echo combination reflected from a target. As shown in figures 9 and 10, single neurons in the DSCF area do not follow pulse–echo sequences as well as the ERPs do; rather they show only a tendency to respond and sometimes may even not respond at all as if waiting for the ‘ideal’ stimulus such as from the wing beats of an insect. The presence of such a signal may trigger the bat to switch to insect pursuit behaviour wherein tracking of different echo parameters becomes more important than tracking the background.

Our data show that the summed synaptic currents act as a driving force whose activity can latch on to an echo stream provided the echo frequency is within the range of frequency tuning at that locus. This mechanism is potentially significant for the perceptual tracking of an echo and may act as a priming mechanism in different regions of the cortex, including the DSCF area. Other echoes, some of whose CF<sub>2</sub> frequencies may also lie within the excitatory frequency tuning of neurons, could also contribute to streaming, although most of them will become ‘invisible’ due to their lack of a fixed temporal relationship with the pulse. Moreover, the frequencies in these extraneous echoes will vary over successive echolocation pulses emitted by the bat since they are not always reflected from the object (background) being tracked.

Stream segregation most likely depends on long term integrative properties of neurons in contrast to response facilitation to single pulse–echo combinations. These properties may be important to understand echolocation at the perceptual level. For example, ‘Doppler-shift compensation’ would always show a lag if the bat had to wait for the echo to arrive before a correction can be made in the frequency of the emitted pulse. Streaming may overcome this problem by providing continuity in the perception of the pulse so that its frequency can be separately monitored and modified to keep the Doppler-shifted frequency reflected from the background within the bat’s auditory fovea (Trappe and Schnitzler 1982). In this context, it is interesting to note that one group of neurons in the DSCF area respond better to pulse frequencies (Kanwal *et al* 1999) and may latch on to the pulse rather than to the echo. Whereas this topic cannot be resolved within the scope of this paper alone, it does open up new lines of thinking about echolocation in bats.

Finally, it is important to keep in mind that much of neural processing occurs in parallel. Therefore, single pulse–echo, FM–FM combinations are still important for computing target distance from pulse–echo delays in the FM–FM area, although one type of neuron (‘tracking neurons’) continue to respond to decreasing pulse–echo delays and increasing echo intensities as the bat closes in on a target (O’Neill and Suga 1979). The FM–FM area is independent of the DSCF area. Furthermore, even within the FM–FM area, neurons show tuning and a facilitative response to combinations of two CF tones (Ohlemiller *et al* 1994). Within the DSCF area, neurons utilize single pulse–echo combinations to compute target characteristics, such as the wing-beat frequency of an insect. These frequencies are represented spatially as frequency maps such that each neuron has a slightly different BF. The identity of the neurons responding to a particular pulse–echo combination identifies a specific wing-beat frequency. Here too, a small percentage of tracking neurons continue to respond to decreasing relative target velocities (lower echo CF<sub>2</sub> frequencies) and increasing echo intensities as the bat slows down to capture its prey (Kanwal *et al* 1999). Thus, it appears that the neural machinery to perform stream segregation does exist in bats, but additional experiments are necessary to unequivocally show that this is indeed used as a perceptual strategy during echolocation.

In summary, we have shown that with the implementation of a few basic neural mechanisms, a relatively simple neural network can effectively simulate a neurophysiological output that is considered to be the basis of auditory stream segregation (Fishman *et al* 2001). Furthermore, we provide neurophysiological recordings from the bat's auditory cortex that match the network output and validate the realistic nature of our model. Our network model and neurophysiological findings emphasize the importance of testing streaming in bats at the behavioural level so that we begin to appreciate the dynamic aspects of neural processing in the natural environment.

### Acknowledgments

We thank Drs Rauschecker and Tian for discussions on streaming that precipitated some of the ideas presented here. The Ministry of Agriculture, Land and Marine Resources in Trinidad kindly permitted us to export mustached bats, and F Muradali assisted with collection and exportation procedures. This research was supported in part by the NIH research grant DC02054 from the National Institute for Deafness and other Communication Disorders to JK. CM received partial funding from an Engineering and Physical Sciences Research Council Life Sciences Network grant GR/M90146 to Professor C J Darwin, the University of Sussex, England.

### References

- Beauvois M W and Meddis R 1996 Computer simulation of auditory stream segregation in alternating-tone sequences *J. Acoust. Soc. Am.* **99** 2270–80
- Bregman A S 1990 *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press)
- Bregman A S and Campbell J 1971 Primary auditory stream segregation and perception of order in rapid sequences of tones *J. Exp. Psychol.* **89** 244–9
- Brosch M and Schreiner C E 1997 Time course of forward masking tuning curves in cat primary auditory cortex *J. Neurophysiol.* **77** 923–43
- Brosch M and Schreiner C E 2000 Sequence sensitivity of neurons in cat primary auditory cortex *Cereb. Cortex* **10** 1155–67
- Brosch M, Schulz A and Scheich H 1999 Processing of sound sequences in macaque auditory cortex: response enhancement *J. Neurophysiol.* **82** 1542–59
- Eggermont J J 1995 Neural correlates of gap detection and auditory fusion in cat auditory cortex *Neuroreport* **6** 1645–8
- Eggermont J J 1996 How homogenous is cat primary auditory cortex? Evidence from simultaneous SU recordings *Auditory Neurosci.* **2** 79–96
- Eggermont J J 2000 Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap-detection thresholds *J. Neurophysiol.* **84** 1453–63
- Eggermont J J 2001 Between sound and perception: reviewing the search for a neural code *Hear. Res.* **157** 1–42
- Esser K H, Condon C J, Suga N and Kanwal J S 1997 Syntax processing by auditory cortical neurons in the FM–FM area of the mustached bat *Pteronotus parnellii Proc. Natl Acad. Sci., USA* **94** 14019–24
- Fay R R 1998 Auditory stream segregation in goldfish, *Carrassius auratus Hear. Res.* **120** 69–76
- Feng A S, Condon C J and White K R 1994 Stroboscopic hearing as a mechanism for prey discrimination in frequency-modulated bats? *J. Acoust. Soc. Am.* **95** 2736–44
- Fishman Y I, Reser D H, Arezo J C and SteinSchneider M 2001 Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey *Hear. Res.* **151** 167–87
- Fitzpatrick D C, Kanwal J S, Butman J A and Suga N 1993 Combination-sensitive neurons in the primary auditory cortex of the mustached bat *J. Neurosci.* **13** 931–40
- Fitzpatrick D C, Olsen J F and Suga N 1998 Connections among functional areas in the mustached bat auditory cortex *J. Comput. Neurol.* **391** 366–96
- Gao E and Suga N 1998 Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system *Proc. Natl Acad. Sci., USA* **95** 12663–70
- Grimault N, Bacon S P and Michéyl C 2002 Auditory stream segregation on the basis of amplitude-modulation rate *J. Acoust. Soc. Am.* **111** 1340–8

- Grimault N, Micheyl C, Carlyon R P, Arthaud P and Collet L 2000 Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency *J. Acoust. Soc. Am.* **108** 263–71
- Hulse S H, MacDougall-Shackleton S A and Wisniewski A B 1997 Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*) *J. Comput. Psychol.* **111** 3–13
- Izumi A 2002 Auditory stream segregation in Japanese monkeys *Cognition* **82** 113–22
- Kanwal J S 1999 Processing species-specific calls by combination-sensitive neurons in an echolocating bat *The Design of Animal Communication* ed M D Hauser and M Konishi (Cambridge, MA: MIT Press) pp 134–57
- Kanwal J S, Fitzpatrick D C and Suga N 1999 Facilitatory and inhibitory frequency tuning of combination-sensitive neurons in the primary auditory cortex of mustached bats *J. Neurophysiol.* **82** 2327–345
- Lofthus W C and Sutter M L 2001 Spectrotemporal organization of excitatory and inhibitory receptive fields of cat posterior auditory field neurons *J. Neurophysiol.* **86** 475–91
- McCabe S L and Denham M J 1997 A model of auditory streaming *J. Acoust. Soc. Am.* **101** 1611–21
- Medvedev A V, Chiao F and Kanwal J S 2002 Modeling complex sound perception: grouping harmonics with combination-sensitive neurons *Biol. Cybern.* **86** 497–505
- Miller G A and Heise G A 1950 The trill threshold *J. Acoust. Soc. Am.* **22** 637–8
- Mittmann D H and Wenstrup J J 1995 Combination-sensitive neurons in the inferior colliculus *Hear. Res.* **90** 185–91
- Moss C F and Surlykke A 2001 Auditory scene analysis by echolocation in bats *J. Acoust. Soc. Am.* **110** 2207–26
- Norena A and Eggermont J J 2002 Comparison between local field potentials and unit cluster activity in primary auditory cortex and anterior auditory field in the cat *Hear. Res.* **166** 202–13
- Novick A and Vaisnys J R 1964 Echolocation of flying insects by the bat, *Chilonycteris parnellii* *Biol. Bull.* **127** 478–88
- Ohlemiller K K, Kanwal J S, Butman J and Suga N 1994 Stimulus design for auditory neuroethology: synthesis and manipulation of complex communication sounds *Auditory Neurosci.* **1** 19–37
- O'Neill W E and Suga N 1979 Target range-sensitive neurons in the auditory cortex of the mustache bat *Science* **203** 69–73
- Park T J, Grothe B, Pollak G D, Schuller G and Koch U 1996 Neural delays shape selectivity to interaural intensity differences in the lateral superior olive *J. Neurosci.* **16** 6554–66
- Pollak G D and Casseday J H 1989 *The Neural Basis of Echolocation in Bats* 1st edn (Berlin: Springer) pp 77–110
- Portfors C V and Wenstrup J J 2002 Excitatory and facilitatory frequency response areas in the inferior colliculus of the mustached bat *Hear. Res.* **168** 131–8
- Schnitzler H-U 1970 Echoortung bei der Fledermaus *Chilonycteris rubiginosa* *Z. Vergl. Physiol.* **68** 25–38
- Shamma S A, Fleshman J W, Wiser P R and Versnel H 1993 Organization of response areas in ferret primary auditory cortex *J. Neurophysiol.* **69** 367–83
- Shamma S A and Symmes D 1985 Patterns of inhibition in the responses of auditory cortical neurons in the squirrel monkey *Hear. Res.* **19** 1–13
- Simmons J A, Freedman E G, Stevenson S B, Chen L and Wohlgenant T 1989 Clutter interference and the integration time of echoes in the echolocating bat, *Eptesicus fuscus* *J. Acoust. Soc. Am.* **86** 1318–32
- Suga N 1990 Biosonar and neural computation in bats *Sci. Am.* **262** 60–8
- Suga N, Niwa H and Taniguchi I 1983 Representation of biosonar information in the auditory cortex of the mustached bat, with emphasis on representation of target velocity information *Advances in Vertebrate Neuroethology* ed J P Ewert (New York: Plenum) pp 829–67
- Sutter M L 2000 Shapes and level tolerances of frequency tuning curves in primary auditory cortex: quantitative measures and population codes *J. Neurophysiol.* **84** 1012–1025
- Trappe M and Schnitzler H-U 1982 Doppler-shift compensation in insect-catching horseshoe bats *Naturwissenschaften* **69** 193–4
- van Noorden L P A S 1975 Temporal coherence in the perception of tone sequences *Doctoral Dissertation* Institute for Perception Research, PO Box 513 Eindhoven, NL
- Vliegen J, Moore B C and Oxenham A J 1999 The role of spectral and periodicity cues in auditory stream segregation measured using a temporal discrimination task *J. Acoust. Soc. Am.* **106** 938–45
- Vliegen J and Oxenham A J 1999 Sequential stream segregation in the absence of spectral cues *J. Acoust. Soc. Am.* **105** 339–46
- Xu J, Gooler D M and Feng A S 1994 Neurons in the frog inferior colliculus exhibit direction-dependent frequency selectivity to isointensity tone bursts *J. Acoust. Soc. Am.* **95** 2160–70
- Yang L and Pollak G D 1994 Binaural inhibition in the dorsal nucleus of the lateral lemniscus of the mustache bat affects responses to multiple sounds *Auditory Neurosci.* **1** 1–17