Auditory Streaming of Amplitude-Modulated Sounds in the Songbird Forebrain

Naoya Itatani and Georg M. Klump
Animal Physiology and Behaviour Group, Institute for Biology and Environmental Sciences, Carl von Ossietzky University-Oldenburg, Oldenburg, Germany

Submitted 18 December 2008; accepted in final form 31 March 2009

INTRODUCTION

In nature a number of sound sources may be active simultaneously and it is important for the auditory system to segregate temporally overlapping sounds from different sources and integrate consecutive sounds from each source. This task has been called “auditory scene analysis” and the consecutive sounds originating from one source have been described as “auditory streams” (Bregman 1990). The segregation of auditory streams is often referred to as “auditory streaming.”

One of the most common paradigms used in the research on auditory streaming has been proposed by van Noorden (1975) who presented a repeated three-tone sequence that consisted of triplets of low-frequency (A) and high-frequency (B) alternating pure tones and an additional silent interval between each triplet (i.e., ABA-ABA-. . ., where A and B denote the tones and - the silent interval). Depending on the rate at which the tones were presented and on their frequency separation, listeners heard one tone series of both A and B tones with a “galloping” rhythm or separate tone series of A or B tones each with an isochronous rhythm. The latter percept has been viewed as evidence for auditory streaming (for review, see Moore and Gockel 2002). The ABA- stimulus or related stimuli have been used in a number of psychophysical or physiological studies of auditory streaming of pure-tone sequences (e.g., Bee and Klump 2004, 2005; Fishman et al. 2001; Micheyl et al. 2003; Pressnitzer et al. 2008; van Noorden 1975). Beauvois and Meddis (1996) and McCabe and Denham (1997) established a computational model explaining auditory streaming based on peripheral frequency channeling, i.e., that exploited spectral differences of sequential sounds exciting auditory neurons tuned to different pure-tone frequencies. According to these models, one stream is perceived if the frequencies of both A and B tones are represented in one frequency channel and two streams are perceived when A and B tones each excite a different frequency channel. Auditory filters can be viewed as the psychophysical equivalent of a frequency channel in the models and populations of neurons that are defined by their frequency tuning can be viewed as their neurophysiological correlate. Following the line of argument of the peripheral channeling hypothesis, Fishman et al. (2001) proposed a model to explain streaming in tone series in the monkey auditory cortex on the basis of frequency differences of the tones. This model can also explain pure-tone streaming in the forebrain of a songbird (Bee and Klump 2004). The tone presentation rate and tone duration also affected perceptual stream segregation, which suggests that forward suppression of neuronal activity to A and B tones affects auditory streaming (Bee and Klump 2005; Fishman et al. 2004).

Auditory streaming, however, can also be observed for stimuli that do not differ in the frequency range in which they provide excitation (see review by Moore and Gockel 2002). Roberts et al. (2002) demonstrated that stream segregation can be observed between A and B signals that do not differ in their spectral components but only in the phase relationship of those components that result in a different temporal pattern. Stream segregation has also been observed with sinusoidally amplitude modulated (SAM) signals in which the carrier was a broadband noise lacking spectral cues (Grimault et al. 2002). Others have demonstrated psychophysically that stream segregation can occur on the basis of the fundamental (f0) in complex tone stimuli that spectrally overlap (e.g., Singh 1987; Singh and Bregman 1997; Vliegen and Öxenham 1999). A physiological correlate of such auditory streaming by the fundamental in
complex tones was reported by Gutschalk et al. (2007) who studied the cortical activity of humans in response to harmonic complexes with only unresolved harmonic frequency components by using functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). They suggested that processes of forward suppression may also operate with respect to streaming by temporal features of sound.

Although streaming by temporal patterns has been well investigated psychophysically, studies focusing on the neuronal correlate of such streaming are limited to fMRI and MEG experiments (e.g., Gutschalk et al. 2007) that reflect the gross population response, providing indirect evidence as to the mechanisms underlying streaming being based on the global pattern of activation. To elucidate the mechanisms on the cellular level, we studied the responses of cortical neurons in a songbird, the European starling (Sturnus vulgaris), when presented with SAM tones of a constant carrier frequency that differed in the rate of modulation. The choice of the SAM stimuli on the one hand connects to the study by Gutschalk et al. (2007) and on the other hand builds on what is known about modulation coding from observations of single neurons or small clusters of neurons (see review by Joris et al. 2004).

Using pure tones, the auditory streaming effect has been demonstrated in starlings in the ABA- paradigm (MacDougall-Shackleton et al. 1998) and a neural correlate of this percept has been observed in starling cortical neurons (Bee and Klump 2004, 2005). With respect to streaming of pure-tone stimuli, European starlings show many similarities to humans (Bee and Klump 2004). Therefore we think that the European starling can also be a suitable model for studying streaming of SAM tones, especially since songbirds are known to use rapid amplitude modulations (AMs) in their song (see review by Greenewalt 1968). Since neurons in the forebrain of the bird not only are tuned to tone frequency but also can show tuning to the rate of AM (Hose et al. 1987), it is possible that different populations of neurons represent different auditory streams of SAM tones based on the modulation. Thus tuned responses to different features of the sounds, i.e., spectral frequency in the case of pure tones and modulation rate in the case of SAM tones, may underlie auditory streaming. Here we test whether the neural mechanism underlying streaming of SAM tones is similar to the neuronal mechanism proposed for streaming of pure tones, i.e., whether different populations of neurons, each of which shows the tuning to a specific modulation frequency, integrate or segregate successive SAM tones based on their temporal envelope structure. The contribution of forward suppression on stream segregation of SAM tones is also considered in relation with this mechanism, as suggested in the imaging studies by Gutschalk et al. (2007).

METHODS

Surgical and recording procedures

Three wild-caught, adult starlings (one male, two females) were used in the present study. The care and treatment of the animals were in accordance with the procedures of animal experimentation approved by Niedersaechsisches Landesamt fuer Verbraucherschutz und Lebensmittelsicherheit. All procedures were performed in compliance with the American Physiological Society’s Guiding Principles in the Care and Use of Animals.

For extracellular recording from the starling forebrain, implantable microdrives with two types of electrodes were prepared: commercially made tungsten microelectrodes (shank diameter 75 μm; FHC, Bowdoinham, ME) and custom-built Teflon-insulated platinum-iridium wires (wire diameter 25 μm; A-M Systems, Carlsborg, WA). The latter electrodes were sharpened at the tip. The procedure of the sharpening is described in Hofer and Klump (2003). The impedance of electrodes measured in 0.9% NaCl using an isolated differential amplifier (ISO-80; World Precision Instruments, Sarasota, FL) ranged from 4.0 to 7.0 MΩ. An array of four electrodes was fixed to a custom-built small head-mounted microdrive using dental acrylic. The microdrive allowed positioning the electrodes at a depth of ≤5 mm into the forebrain.

Prior to surgery 0.04 ml of atropine solution (B. Braun Melsungen, Melsungen, Germany) was injected subcutaneously. After 2–3 min the animal was anesthetized using 4–5% isoflurane and the concentration of isoflurane was subsequently reduced to 1.5–2.0% for keeping the animal anesthetized during the surgery. The head was fixed using ear bars and the head angle of the animal was adjusted in a stereotaxic apparatus so that the bill of the bird inclined about 45° below the horizontal plane. The electrodes were implanted into the field L complex of the right forebrain hemisphere. Recordings were done primarily from neurons in the input layer field L2, which is the homolog of layer IV of the mammalian primary auditory cortex (Jarvis 2005). These neurons can be identified by a primary-like response pattern (Hofer and Klump 2003; Nieder and Klump 1999). Two custom-built reference electrodes were implanted into the left rostral forebrain hemisphere (stainless steel wire, diameter of 50 μm; A-M Systems). Finally, next to the microdrive and reference electrodes a small socket for attaching a radio transmitter was mounted. After a recovery of between 3 and 7 days after the surgery the recordings started after surgical wounds had healed normally.

The recording was performed using radio telemetry from the freely behaving birds in a test cage (56 × 36 × 33 cm, L × W × H) located inside a radio-shielded sound chamber (IAC 402A; Industrial Acoustics, Niederku¨chten, Germany, equipped with sound-absorbing foam to reduce echoes; for details see Bee et al. 2007). For radio transmission, a small FM radio transmitter (FHC type 40-71-1; FHC) was used. A dipole antenna was located near the test cage to receive the radio signal. The signals were demodulated by an FM tuner (TX-970; Pioneer), band-pass filtered (500–5,000 Hz), amplified, converted to 16-bit, 44.1-kHz digital signals (Hammerfall DSP Multiface II, RME), and recorded on a Linux workstation for later analysis.

Acoustic stimulation

All stimuli were generated digitally (sampling rate 44.1 kHz, 16-bit resolution) and played back by a Hammerfall DSP (Multiface II, RME) using the same Linux workstation that recorded the neural responses synchronized to the playback. The analog sound output was attenuated (Hewlett-Packard 350D, Böblingen, Germany, and TDT PA4, Tucker-Davis Technologies, Alachua, FL) then amplified (RB-1050; Rotel, Sussex, UK) and presented through a loudspeaker (Type SP3253; KEF Audio, Maidstone, UK) attached on the ceiling of the chamber about 70 cm above the bird in the cage.

For defining the characteristic frequency (CF; see following text) of each recording site, 200-ms tone pips with 10-ms raised cosine shaped rise and fall were used. Stimulus levels were adjusted to take the frequency response of the speaker into account that was generally flat (±3 dB) over the range of frequencies used in this study. First, the CF was estimated audiovisually by presenting a series of pure tones with frequency rising in 0.5-octave steps and observing the neural response. Then 20 repetitions of tone pips with frequencies that varied from 1.5 octaves below to 1.5 octaves above the estimated CF in 0.25-octave steps were presented with a silent interval of 800 ms between the tones and the first 10 repetitions showing no artifacts (i.e., typically high potentials resulting from ongoing movement were about threefold higher than the threshold used to detect spiking
activity of the neurons) were analyzed. The level of the stimuli ranged between 0 and 70 dB SPL and was increased in 10-dB steps.

To characterize the modulation tuning, SAM tones with the CF as the carrier frequency were presented with modulation frequencies (fmod) varying in 0.5-octave steps and ranging from 5 to 320 Hz (in a few exceptional cases ≤640 Hz) to construct a rate modulation transfer function (rMTF). The modulation depth was fixed at 100%. The presentation level was 70 dB SPL and the duration was 600 ms (5-ms raised cosine rise/fall) with a silent interval of 800 ms between the AM tones. The long duration of the signal ensured that a large number of modulation cycles were presented.

The basic stimulation paradigm for investigating auditory streaming of SAM tones was similar to that of the staring study by Bee and Klump (2004) using pure tones. An ABA signal triplet with a silent period after the third tone was repeatedly presented (...ABA-ABA-...) and the responses to those repetitions were recorded. The A signals were SAM tones with a CF carrier and a constant modulation frequency (termed reference modulation frequency [RMF]) of ≤160 Hz that was chosen as described in the following text. The B signals were SAM tones with a CF carrier that had a higher modulation frequency than that of A signals. The modulation frequency of B tones varied between 0.5 and 4.0 octaves above RMF in 0.5-octave steps. By increasing the modulation frequency of the B signals rather than decreasing it, we made sure that the neurons were provided with a sufficient number of modulation cycles, given that the SAM tone duration was limited to 125 ms. In most cases, this also brought the modulation frequencies to a range for which SAM stimuli may evoke a pitch percept rather than a percept of a fluctuating amplitude, thus allowing a better comparison to the study by Gutschalk et al. (2007) in humans that have used stimuli providing a pitch sensation. The ABA signal triplets were repeated 30 times. The duration of each A or B signal and of the silent interval was 125 ms in the case of a repetition period of 100%. SAM tones were ramped with a 5-ms raised cosine (rise/fall). For stimuli with larger repetition periods of 200 or 400% additional silent intervals were introduced to increase the time from the onset of the A signal to the onset of the B signal (and vice versa) to 200 or 400% of the signal duration of 125 ms and adjust the silent interval between the triplets accordingly. Figure 1 shows an example of the ABA-triplet stimulus and the response to a TRT of 200%. As additional controls triplets of all RMF SAM tones (AAA-), B signals of varying modulation frequency surrounded by unmodulated tones of the carrier frequency of the B signal (CBC-), and isolated A and B signals (A-A- and -B–) were presented. The presentation level of all stimuli was 70 dB SPL. The order of the presentation of all triplet stimuli was randomized.

Evaluation of pure-tone responses and tuning characteristics

A frequency tuning curve was constructed based on the rate responses to the 200-ms pure tones varying in frequency and level. Spike rate was calculated by counting spikes over 10 artifact-free responses within a time window incorporating the total length of the tone and considering the response latency. The threshold of the recording site was determined as the minimum stimulus amplitude at which the spike rate at a specific frequency was greater than the spontaneous rate plus 1.8SDs. The tone frequency at which the lowest threshold was found was defined as the CF of the recording site. The bandwidth of the tuning curve was calculated as the frequency difference between the tone frequencies above and below CF, which evoked the threshold spike rate at 70 dB SPL. Recording sites in field L2 of European starlings commonly show primary-like temporal responses and regions of reduced spike activity compared with the spontaneous activity that are often referred to as suppressive sidebands (Nieder and Klump 1999). These suppressive sidebands were determined as regions of the response map at which the spike rate was less than the spontaneous rate minus 1.8SDs. If the response pattern was not primary-like, suggesting a recording site outside the field L2, no further measurement was conducted from that recording site.

Evaluation of rate modulation transfer functions

Spike rates in response to SAM tones with varying fmod were analyzed to construct an rMTF based on the responses to 10 artifact-free repetitions of the stimuli. Spike rates were estimated for a latency-corrected time window of the duration of the stimulus. First, the maximum spike rate and the corresponding modulation frequency (rBMF) were identified. Then it was verified whether the spike rate for modulation frequencies above or below rBMF dropped to ≤75% of the maximum rate. If the spike rate dropped according to this criterion only for fmod above rBMF, the recording site was classified as “low-pass” (LP). If the drop according to this criterion occurred only for fmod below rBMF, the recording site was classified as “high-pass” (HP). If the criterion was met on both sides of rBMF, the recording site was labeled as “band-pass” (BP). In some recording sites, spike rate recovered after the rate first dropped by >25% with increasing fmod. If the recovery was to 12.5% of the maximum rate, these recording sites showing a partial suppression at specific fmod values were classified as “band-reject” (BR). If the rate did not drop by ≥25% of the maximum rate at the different fmod values tested, that recording site was classified as “all-pass” (AP).

The reference modulation frequency (RMF) for the subsequent presentation of streaming stimuli, which was also the modulation frequency of the A signal, was set as follows. In BP units, the fmod evoking the maximum rate was chosen as the RMF. In LP units, RMF was set to the highest fmod that evoked ≥75% of the spike rate compared with the maximum discharge rate observed at rBMF. For BR units, the choice of RMF was similar to that for LP units, i.e., RMF was set to a modulation frequency at which the response started to decline. This fmod generally was below the fmod at which the discharge rate was at a minimum. For recording sites with HP characteristics, a local maximum close to the sloping part of the rate modulation transfer function (in HP the rate first increased with increasing fmod and then varied within the 75–100% range relative to the maximum rate) was chosen as the RMF. In AP recording sites, a local maximum up to fmod of 160 Hz was chosen as the RMF.

Rate analysis of signal triplet response pattern

Rate responses to the first, second, and third signals of a triplet summed up over the duration on the ongoing signal (with a correction
for the recording site response latency) were compared as follows. Absolute spike rates in spikes/s to each signal in the different types of signal triplets and in the single signal type controls were determined for the different \( \Delta f_{\text{mod}} \) values and tone repetition time (TRT) for further analysis. Data from 20 triplet repetitions with artifact-free recordings were analyzed. Normalized responses to all stimuli were also calculated by dividing their absolute spike rates by the absolute rate to the isolated A signal (first A of the A-A- control stimulus) at the largest TRT (400%). The normalized rates were expressed as a function of the difference between the modulation frequencies of A and B signals in octaves (\( \Delta f_{\text{mod}} \)).

**Forward suppression analysis**

To observe the effect of mutual forward suppression between successive tones, the differences of the responses to different tones (A or B) in different conditions (isolated or surrounded by other tones) were calculated. The analysis method followed that used by Bee and Klump (2004). We further observed the responses to B signals surrounded by nonmodulated pure tones to investigate the effect of modulation on forward suppression. Differences of the normalized response rates for four conditions were calculated: 1) the difference between responses to B signals in ABA- and responses to isolated B signals in -B-, 2) the difference between responses to B signals in CBC- and responses to isolated B signals in -B-, 3) the difference between responses to B signals in ABA- and responses to isolated B signals in -B-, and 4) the difference between responses to A signals in ABA- and responses to A signals in A-A-.

**Temporal analysis of B signal response pattern**

Spike period histograms in which each period was the reciprocal of the modulation frequency of stimulus were constructed. Vector strength (VS; Goldberg and Brown 1969) was then calculated to observe the synchrony of the spiking to the envelope modulation. For the investigation of significance of the synchrony, Rayleigh statistics of VS was used.

**Statistical analysis**

The effects of \( \Delta f_{\text{mod}} \) or \( f_{\text{mod}} \), TRT, triplet type, and response type of recording site on absolute rate, normalized rate, and VS were examined using repeated-measures ANOVA (rmANOVA) using SPSS version 15. Mauchley's sphericity test was used prior to the analysis to inspect whether the sphericity assumption of rmANOVA is violated. For repeated-measures analyses with more than a single factor, the Greenhouse-Geisser (1959) adjusted \( df \) for omnibus tests of within-subjects factors that violated the sphericity assumption of rmANOVA. The unadjusted \( df \) values are shown when reporting statistical results. We also computed for each rmANOVA the partial \( \eta^2 \) as a measure of the effect size for all main effects and interactions. Partial \( \eta^2 \), which can vary from 0 to 1, is the proportion of the combined effect and error variance that is attributable to the effect and thus represents a nonadditive "variance-accounted-for" measure of effect size. The criterion for statistical significance in all tests in the present study was \( \alpha = 0.05 \).

**RESULTS**

**Pure-tone frequency tuning**

In total, data were obtained from 54 recording sites in the auditory forebrain of three European starlings. Figure 2 shows the relationship between CF and bandwidth at 70 dB SPL as defined by the recording sites' pure-tone frequency tuning curve and in relation to recording sites' unit type. The CF ranged from 0.4 to 6 kHz and there were no significant differences between the unit types’ CF values (Kruskal–Wallis H-test). The 70-dB bandwidth increased with increasing CF, the relation being well described by an exponential regression (bandwidth = \( 528e^{(0.0003 \times \text{CF})} \), \( R^2 = 0.63 \)). The response patterns elicited by pure tones at the CF of those 54 recording sites were “primary-like,” exhibiting a strong onset response and a subsequent decay of rate.

**Rate responses to single SAM tones and unit types**

All recording sites were classified into five types by their rate-response patterns to SAM tones with different modulation frequencies (for the classification criteria, see METHODS). AP type units were observed most frequently (15/54), followed by LP (13/54), HP (10/54), BP (8/54), and BR (8/54) types. The modulation frequencies that elicited a strong rate response and that were used as the reference modulation frequencies in the subsequent analysis ranged from 5 to 160 Hz and were mostly <100 Hz (Table 1; also see METHODS for the choice of reference modulation frequency). In a BP-type recording site, the reference modulation frequency is equivalent to the rate best modulation frequency, at which the maximum spike rate is evoked. An example of rate modulation transfer function (rMTF) from a BP recording site together with the VS as a measure of the temporal response pattern is shown in Fig. 3.

**Rate responses to SAM tone triplets**

Responses to SAM tone triplets in each recording site were collected and their spike rates and synchrony were calculated to observe the effect of varying the difference in modulation frequency \( \Delta f_{\text{mod}} \) between the A signals and B signals and the TRT. Figure 4 shows an example of rate responses to ABA-triplets at different \( \Delta f_{\text{mod}} \)s and TRTs. The rMTF is shown in Fig. 3. Since the unit has BP characteristics the fmod of the A signal (i.e., the RMF) for this recording site was chosen as the fmod corresponding to the peak of its rMTF. Similarly to the general pattern (see following text), the example demonstrates a strong response that is phase-locked to the AM of the A signals in the triplets for all TRTs. On the other hand, spike rate in response to B signals and the phase locking to the modulation as expressed by the vector strength decreased as \( \Delta f_{\text{mod}} \).
increased. This was most prominently observed when TRT was short (100%; see top panel of Fig. 4). When responses to ABA-triplets at different TRTs were compared, the trend of the decrease in spike rates as Δfmod increased was similar, although the amount of decrease at the TRTs of 200 and 400% was not as large compared with that at the shortest TRT. Onset responses to the B signal were observed, irrespective of Δfmod, but subsequent spike activities decayed more strongly when fmod of the B signal increased.

Rate responses for the different unit types are summarized in Fig. 5. Mean normalized rates as a function of Δfmod and TRT together with ±2SE are shown. All rate data were normalized to the response elicited by A signals presented alone at the largest TRT. Rate responses to first and second A signals and the interspersed B signal were compared using a rmANOVA (for detail, see METHODS) with sound number (i.e., first, second, or third signal in the triplet), Δfmod, TRT, and unit type as factors. Significant main effects were observed for sound number, Δfmod, and TRT, but not for unit type (see Table 2). The average rate response between all three sounds in the triplet differed (all P < 0.001, t-test). The first A signal in the triplet elicited the highest response, the last A signal in the triplet elicited the second highest response, and the B signal elicited the lowest response. The effect of Δfmod is attributed to the decrease in the B signal response with increasing Δfmod. Since in A signals fmod did not vary and thus the response did not change, the overall effect size remained small. The effect of TRT reflects differential forward suppression that is larger at smaller TRTs (all differences significant with P < 0.001, t-test). Both the B signal and the second A signal in the triplet were suppressed by the preceding signal. Significant two-way interactions were observed between sound number and Δfmod, sound number, and TRT, and sound number and unit type. The interaction between sound number and Δfmod can be explained by the fact that the fmod of A signals is always the RMF, whereas the fmod of B signals changes. The same effect also accounts for the interaction between sound number and unit type since the different unit types respond differently to the varying fmod values of the B signal. The interaction between sound number and TRT reflects the differential suppression on the three signals in the triplet. The significant two-way interaction between unit type and TRT reflects that the change in the rate response in relation to TRT in BR and HP units differed from that in AP, BP, and LP units that all responded similarly to changing TRT. BR units showed higher rate responses at the shortest TRT compared with other unit types. HP units showed the greatest increase in rate by the change of TRT from 100 to 200% and, as a consequence, HP units showed the largest spike rate at a TRT of 200%. The significant three-way interactions are accounted for by the two-way interactions shown here.

Rate responses to B signals in relation to the type of triplet

Now we focus on the rate response to B signals in ABA-triplets and compare these to the responses in two types of controls. In the first control, the first and third signals of the triplet were pure tones instead of SAM tones (triplets termed CBC-), i.e., this control served to demonstrate the effect of AM of preceding sound per se on B signal responses. As an additional control, we also observed responses to B-alone stimuli (−B−), in which forward suppression of the first signal in the triplet on the B signal is absent.

An example of the responses to the three different triplet types at the shortest TRT (100%, which should evoke the maximum effect of preceding sound on B signal) is shown in Fig. 6. The neurons respond to the C signal with a strong onset response that rapidly adapts, whereas the A signal evokes a

---

**TABLE 1. Distribution of reference modulation frequency (RMF) of recording sites (in Hz) applied in the present study and chosen on the basis of the rate response**

<table>
<thead>
<tr>
<th>RMF range (Hz)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10</td>
<td>3</td>
</tr>
<tr>
<td>11–20</td>
<td>6</td>
</tr>
<tr>
<td>21–30</td>
<td>11</td>
</tr>
<tr>
<td>31–40</td>
<td>7</td>
</tr>
<tr>
<td>41–50</td>
<td>9</td>
</tr>
<tr>
<td>51–60</td>
<td>4</td>
</tr>
<tr>
<td>61–70</td>
<td>6</td>
</tr>
<tr>
<td>71–80</td>
<td>0</td>
</tr>
<tr>
<td>81–90</td>
<td>4</td>
</tr>
<tr>
<td>90–160</td>
<td>4*</td>
</tr>
</tbody>
</table>

*In two recording sites an RMF of 120 Hz was used; others were 140 and 160 Hz.

---

**FIG. 3. An example of a rate-modulation transfer function (rMTF, solid line) and vector strength (VS, dotted line) observed from a recording site determined as band-pass (BP) type. The RMF indicated as the dotted line is used as the modulation frequency of the 1st and 3rd signals of ABA-triplets in the subsequent measurements.**

**FIG. 4. PSTHs as a function of time and the difference in modulation frequency (Δfmod) in response to an ABA-stimulus (data summed over 20 triplet repetitions). Different tone repetition times are represented in the different panels (top: 100%; middle: 200%; bottom: 400%). The response was recorded from the same site as that shown in Fig. 3. Each tone in the triplet was 125 ms with a varying intertone silence period depending on the TRT. Shades of gray represent number of spikes summed over 20 repetitions of the stimulus in 1-ms time bins.**
strong response at each cycle of the modulation. The preceding A signal in ABA- triplets suppressed B signal responses and thus the spike rate of B signal responses in ABA- triplets dropped compared with those observed in B-alone stimuli. Similar responses were observed by CBC- triplets, showing a suppression of spike activity in response to B signals.

To describe the general effects of the types of surround signals on the rate responses to non-RMF B signals, we analyzed the normalized (see preceding text) B signal rate responses using a rmANOVA with triplet type, fmod, TRT, and unit type as factors. Significant main effects of triplet type, fmod, TRT, and unit type were observed (Table 3). Responses to B signals in the different triplet types were all significantly different (P < 0.02, t-test), although the difference of rate responses between ABA- and CBC- triplets was not as large as that between B-alone signals and the other two triplet types. Response rates were significantly different when the modulation frequencies for which rates were compared were largely separated. In particular, B signals at a fmod of 1.0 octave evoked significantly fewer spikes than B signals at a fmod of 1.0 octave (all P < 0.02, t-test). TRTs of 100, 200, and 400% elicited significantly different average spike rates (all P < 0.001, t-test), which can be related to the different magnitudes of forward suppression in each TRT condition as described earlier. The significant interaction between triplet type and TRT can be accounted for by the fact that in the -B– condition no signal precedes the B signal and thus suppression at short TRTs cannot occur that is observed for the other two triplet types. The significant interaction between triplet type and unit type may both reflect the differential susceptibility of the various unit types to suppression since in the -B– condition and for long TRT there is no strong excitation before presentation of the B signal. The relation between relative rate and TRT was

To describe the general effects of the types of surround signals on the rate responses to non-RMF B signals, we analyzed the normalized (see preceding text) B signal rate responses using a rmANOVA with triplet type, fmod, TRT, and unit type as factors. Significant main effects of triplet type, fmod, TRT, and unit type were observed (Table 3). Responses to B signals in the different triplet types were all significantly different (P < 0.02, t-test), although the difference of rate responses between ABA- and CBC- triplets was not as large as that between B-alone signals and the other two triplet types. Response rates were significantly different when the modulation frequencies for which rates were compared were largely separated. In particular, B signals at a fmod of 1.0 octave evoked significantly fewer spikes than B signals at a fmod of 1.0 octave (all P < 0.02, t-test). TRTs of 100, 200, and 400% elicited significantly different average spike rates (all P < 0.001, t-test), which can be related to the different magnitudes of forward suppression in each TRT condition as described earlier. The significant interaction between triplet type and TRT can be accounted for by the fact that in the -B– condition no signal precedes the B signal and thus suppression at short TRTs cannot occur that is observed for the other two triplet types. The significant interaction between triplet type and unit type may both reflect the differential susceptibility of the various unit types to suppression since in the -B– condition and for long TRT there is no strong excitation before presentation of the B signal. The relation between relative rate and TRT was
reflects that the suppression increases with increasing TRT in the BR type but decreases with increasing TRT in the other types. No significant main effect of $\Delta f_{\text{mod}}$ was observed and none of the other interactions involving $\Delta f_{\text{mod}}$ was significant. This indicates that the amount of suppression by preceding A signals on B signals does not differ by the change of the $f_{\text{mod}}$ of the B signals itself.

2) Difference between B in CBC- versus isolated B in -B–. The condition in which the SAM tone preceding the B signal was substituted with a pure tone of the same CF as the carrier of the SAM tone reflects the forward suppression resulting from spectral energy in the frequency range of the B signal (i.e., this signal was in the same spectral range but had no AM). The average difference for the different unit types is shown in the middle column of Fig. 7. A rmANOVA showed a significant main effect of TRT on forward suppression ($P < 0.001$). The shortest TRT condition evoked the largest forward suppression (all $P$ values $< 0.02$, t-test). A significant main effect of $\Delta f_{\text{mod}}$ was also observed ($P = 0.01$, rmANOVA). However, the change of forward suppression with increasing $\Delta f_{\text{mod}}$ was not systematic; e.g., forward suppression at $\Delta f_{\text{mod}}$ of 0.5 octave was significantly different from that at $\Delta f_{\text{mod}}$ of 1.0, 2.0, 3.0, and 4.0 octaves (all $P$ values $< 0.03$, t-test) but not at 1.5, 2.5, and 3.5 octaves. If the eight BR units were excluded from the analysis, the main effect of $\Delta f_{\text{mod}}$ was far from being significant ($P = 0.135$, rmANOVA), indicating that these units may have contributed strongly to the nonsystematic but significant variation. No significant two-way and three-way interactions were observed in this condition.

3) Difference between B in ABA- versus B in CBC-. Responses to B signals in the above-observed two conditions were compared by investigation of the influence of the AM of the masker tones per se on the magnitude of suppression to B signals. We demonstrated earlier (responses to B signals in relation to triplet type) that there is a difference. Here we test whether this difference changes in relation to $f_{\text{mod}}$ and TRT in different unit types (Fig. 7, right column). The differences in the normalized rate observed in different $\Delta f_{\text{mod}}$ conditions ranged between $-0.01$ ($\Delta f_{\text{mod}} = 2.5$ octaves, SE = 0.014)

### Table 3. Results of rmANOVA comparing the effects of triplet type, modulation frequency separation, tone repetition time, and unit types on the normalized rate responses to the B signals in the -B-, CBC-, and ABA- stimuli

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>$F$</th>
<th>$P$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triplet type</td>
<td>2, 98</td>
<td>608,870</td>
<td>$&lt;0.001$</td>
<td>0.926</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$</td>
<td>7, 343</td>
<td>11,524</td>
<td>$&lt;0.001$</td>
<td>0.190</td>
</tr>
<tr>
<td>TRT</td>
<td>2, 98</td>
<td>512,797</td>
<td>$&lt;0.001$</td>
<td>0.913</td>
</tr>
<tr>
<td>Unit type</td>
<td>4, 49</td>
<td>2,764</td>
<td>0.038</td>
<td>0.184</td>
</tr>
<tr>
<td>Triplet type $\times$ $\Delta f_{\text{mod}}$</td>
<td>14, 686</td>
<td>2,029</td>
<td>0.030</td>
<td>0.040</td>
</tr>
<tr>
<td>Triplet type $\times$ TRT</td>
<td>4, 196</td>
<td>28,013</td>
<td>$&lt;0.001$</td>
<td>0.364</td>
</tr>
<tr>
<td>Triplet type $\times$ Unit type</td>
<td>8, 98</td>
<td>2,541</td>
<td>0.026</td>
<td>0.172</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$ $\times$ TRT</td>
<td>14, 686</td>
<td>0.519</td>
<td>0.877</td>
<td>0.010</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$ $\times$ Unit type</td>
<td>28, 343</td>
<td>2,379</td>
<td>$&lt;0.001$</td>
<td>0.163</td>
</tr>
<tr>
<td>TRT $\times$ Unit type</td>
<td>8, 98</td>
<td>3,153</td>
<td>$&lt;0.001$</td>
<td>0.205</td>
</tr>
<tr>
<td>Triplet type $\times$ $\Delta f_{\text{mod}}$ $\times$ TRT</td>
<td>28, 1,372</td>
<td>1,127</td>
<td>0.330</td>
<td>0.022</td>
</tr>
<tr>
<td>Triplet type $\times$ $\Delta f_{\text{mod}}$ $\times$ Unit type</td>
<td>56, 686</td>
<td>1,097</td>
<td>0.320</td>
<td>0.082</td>
</tr>
<tr>
<td>Triplet type $\times$ TRT $\times$ Unit type</td>
<td>16, 196</td>
<td>1,535</td>
<td>0.120</td>
<td>0.111</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$ $\times$ TRT $\times$ Unit type</td>
<td>56, 686</td>
<td>1,186</td>
<td>0.208</td>
<td>0.088</td>
</tr>
<tr>
<td>Triplet type $\times$ $\Delta f_{\text{mod}}$ $\times$ TRT $\times$ Unit type</td>
<td>112, 1,372</td>
<td>1,127</td>
<td>0.251</td>
<td>0.084</td>
</tr>
</tbody>
</table>

Bold numbers highlight the significant effects.
and 0.036 (Δfmod = 2.0 octaves, SE = 0.014), showing scarcely any change of the amount of forward suppression. The statistical analysis using a rmANOVA showed only a weak but significant (P < 0.02) main effect of TRT on the difference. However, this effect accounted for only a small amount of variance (η² = 0.086). No other main effects or interactions were significant.

4) Difference between A in ABA- versus A in A-A-. We also analyzed the masking effects of responses to B signals on the rate responses to preceding and following A signals. The difference values observed across all Δfmods, TRTs, and unit types for the first A signal were always close to zero, indicating that backward suppression of B signals on the preceding A signals is not an important effect. A rmANOVA with TRT Δfmod and unit type as the factors did not reveal any significant effects. When the response to the A signal following the B signal in ABA- triplets was compared with that to the second A signal in the A-A- stimulus there was a significant main effect of TRT on the change of the magnitude of forward suppression (P < 0.01, rmANOVA with TRT, Δfmod, and unit type as factors). No other significant main effects or interactions were observed.

Temporal responses to ABA- triplets

Vector strength (VS) was calculated from responses with reference to the period of the modulation to obtain a measure of the temporal representation of the modulated signal. The example shown in Fig. 1 demonstrates a strong phase locking to the modulation in both the A signal and the B signal. The phase locking to the modulation of the B signal deteriorated with increasing modulation frequency. Figure 8 shows the mean VS (±2SE) as a function of Δfmod and TRT observed in the different unit types and for different triplet types. VS values of the response to the first and second A signals and the interspersed B signal were compared using a rmANOVA with sound number (i.e., first, second, or third signal in the triplet), Δfmod, TRT, and unit type as factors. Significant main effects were observed for sound number, Δfmod, TRT, and for unit type (see Table 4). Post hoc tests revealed a large significant decrease in VS in the response to the B signals compared with that to the A signals (P < 0.001, t-test). The VS increased significantly (P < 0.02, t-test) from the response to the first A signal to that to the last A signal in the triplet, but this difference was much smaller. The VS of the response decreased with increasing Δfmod; VS for a Δfmod of ≤1.5 octaves was significantly different from the response at larger Δfmod. Post hoc tests revealed that the VS differed significantly between all TRTs tested (P < 0.001, t-test), being largest at the shortest TRT. Of all unit types, AP units had the highest VS value that differed significantly from that in all other unit types and LP units had the lowest VS (P < 0.05). LP units had a significantly lower VS (P = 0.03, t-test) than that of HP units. A number of significant two-factor interactions were observed. The interaction between sound number and Δfmod is trivial since Δfmod was changed in the B signal but...
not in the A signals. The interaction between sound number and unit type reflects that VS exhibits higher values in AP units than in the other unit types. Also the rate of decay of VS with increasing $f_{\text{mod}}$ varied between unit type, which is reflected in the interaction between $f_{\text{mod}}$ and unit type. The interaction between TRT and unit type reflects that VS deteriorates more in AP and HP units with increasing TRT than in the other unit types. The three-way interactions can be deduced from the interactions described so far.

**Temporal responses to B tones in different types of triplets**

Temporal responses to (non-RMF) B signals in different triplet types may vary due to the influence of forward suppression by the surrounding signals. To assess this hypothesis, we compared VS values calculated from the responses to B signals in different types of triplets in which B signals were surrounded either by RMF tones (ABA-) or by pure tones (CBC-). For reference, we also calculated temporal responses to B-alone signals (-B–). VS values of the response to the B signal in the various conditions were compared using a rmANOVA with triplet type, $f_{\text{mod}}$, TRT, and unit type as factors (see Table 5). The main effects of triplet type, $f_{\text{mod}}$, and TRT were significant (although the TRT accounted for changes in VS of only 0.02) but the main effect of unit type was not significant. Post hoc tests revealed that the VS in B-alone stimuli (-B–) was smaller than that in ABA- or CBC- triplets ($P < 0.001$, $t$-test; VS in the latter two conditions was not significantly different). VS decreased with increasing TRT ($P < 0.01$ for all pairwise comparisons, $t$-test). As is typical for modulation transfer functions, VS decreased with increasing $f_{\text{mod}}$. The largest differences were observed in the comparison of $f_{\text{mod}}$ of $\pm 1$ octave and higher $f_{\text{mod}}$ (all $P$ values $< 0.001$, $t$-test). Significant two-way interactions between $f_{\text{mod}}$ and triplet type ($P = 0.001$), $f_{\text{mod}}$ and TRT ($P = 0.001$), and $f_{\text{mod}}$ and unit type ($P = 0.003$) were found.

### Table 4. Results of rmANOVA comparing the effects of triplet position, modulation frequency separation, tone repetition time, and unit types on the VS to the A and B signals in the ABA- triplets

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>$F$</th>
<th>$P$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound-number</td>
<td>2, 94</td>
<td>172.310</td>
<td>$&lt; 0.001$</td>
<td>0.786</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$</td>
<td>8, 376</td>
<td>0.795</td>
<td>$&lt; 0.001$</td>
<td>0.799</td>
</tr>
<tr>
<td>TRT</td>
<td>2, 94</td>
<td>40.109</td>
<td>$&lt; 0.001$</td>
<td>0.460</td>
</tr>
<tr>
<td>Unit type</td>
<td>4, 47</td>
<td>6.244</td>
<td>$&lt; 0.001$</td>
<td>0.347</td>
</tr>
<tr>
<td>Sound-number $\times$ $\Delta f_{\text{mod}}$</td>
<td>16, 752</td>
<td>73.967</td>
<td>$&lt; 0.001$</td>
<td>0.611</td>
</tr>
<tr>
<td>Sound-number $\times$ TRT</td>
<td>4, 188</td>
<td>1.316</td>
<td>0.272</td>
<td>0.027</td>
</tr>
<tr>
<td>Sound-number $\times$ Unit type</td>
<td>8, 94</td>
<td>6.040</td>
<td>$&lt; 0.001$</td>
<td>0.340</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$ $\times$ TRT</td>
<td>16, 752</td>
<td>0.975</td>
<td>0.466</td>
<td>0.020</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$ $\times$ Unit type</td>
<td>32, 376</td>
<td>2.504</td>
<td>$0.002$</td>
<td>0.176</td>
</tr>
<tr>
<td>TRT $\times$ Unit type</td>
<td>8, 94</td>
<td>2.848</td>
<td>$0.019$</td>
<td>0.195</td>
</tr>
<tr>
<td>Sound-number $\times$ $\Delta f_{\text{mod}}$ $\times$ TRT</td>
<td>32, 1,504</td>
<td>0.994</td>
<td>0.461</td>
<td>0.021</td>
</tr>
<tr>
<td>Sound-number $\times$ $\Delta f_{\text{mod}}$ $\times$ Unit type</td>
<td>64, 752</td>
<td>2.950</td>
<td>$&lt; 0.001$</td>
<td>0.201</td>
</tr>
<tr>
<td>Sound-number $\times$ TRT $\times$ Unit type</td>
<td>16, 188</td>
<td>2.527</td>
<td>$0.019$</td>
<td>0.177</td>
</tr>
<tr>
<td>$\Delta f_{\text{mod}}$ $\times$ TRT $\times$ Unit type</td>
<td>64, 752</td>
<td>0.858</td>
<td>0.724</td>
<td>0.068</td>
</tr>
<tr>
<td>Sound-number $\times$ $\Delta f_{\text{mod}}$ $\times$ TRT $\times$ Unit type</td>
<td>128, 1,504</td>
<td>0.919</td>
<td>0.726</td>
<td>0.073</td>
</tr>
</tbody>
</table>

Sound-number, triplet position; $\Delta f_{\text{mod}}$, modulation frequency separation; TRT, tone repetition time; VS, vector strength. Bold numbers highlight the significant effects.

**FIG. 8.** Synchrony of the response to the modulation period measured as the VS for ABA- stimulus sequences in relation to $\Delta f_{\text{mod}}$ and TRT. Each row represents the data from a specific unit type. Left, middle, and right columns for each unit type show responses to the 1st, 2nd, and 3rd tones of triplet, respectively. Symbols ($\times$ = TRT 100%, $\circ$ = TRT 200%, $\triangle$ = TRT 400%) represent the mean responses ($\pm 2SE$) averaged over 20 artifact-free responses. Responses to an AAA- stimulus are also shown in the ABA- response panels depicted as the responses to ABA- stimulus sequence with $\Delta f_{\text{mod}}$ of 0.
the -B–, CBC-, and ABA- stimuli

### TABLE 5. Results of rmANOVA comparing the effects of triplet type, modulation frequency separation, tone repetition time, and unit types on the normalized rate responses to the B signals in the -B–, CBC-, and ABA- stimuli

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>F</th>
<th>P</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triplet type</td>
<td>2, 96</td>
<td>85.654</td>
<td>(&lt; 0.001)</td>
<td>0.641</td>
</tr>
<tr>
<td>( \Delta f_{\text{mod}} )</td>
<td>7, 336</td>
<td>47.337</td>
<td>(&lt; 0.001)</td>
<td>0.497</td>
</tr>
<tr>
<td>TRT</td>
<td>2, 96</td>
<td>53.091</td>
<td>(&lt; 0.001)</td>
<td>0.525</td>
</tr>
<tr>
<td>Unit type</td>
<td>4, 48</td>
<td>1.402</td>
<td>0.248</td>
<td>0.105</td>
</tr>
<tr>
<td>Triplet type ( \times \Delta f_{\text{mod}} )</td>
<td>14, 672</td>
<td>3.146</td>
<td>0.011</td>
<td>0.062</td>
</tr>
<tr>
<td>Triplet type ( \times \text{TRT} )</td>
<td>4, 192</td>
<td>9.037</td>
<td>(&lt; 0.001)</td>
<td>0.158</td>
</tr>
<tr>
<td>Triplet type ( \times \text{Unit type} )</td>
<td>8, 96</td>
<td>1.518</td>
<td>0.165</td>
<td>0.112</td>
</tr>
<tr>
<td>( \Delta f_{\text{mod}} \times \text{TRT} )</td>
<td>14, 672</td>
<td>2.112</td>
<td>0.024</td>
<td>0.042</td>
</tr>
<tr>
<td>( \Delta f_{\text{mod}} \times \text{Unit type} )</td>
<td>28, 336</td>
<td>1.917</td>
<td>0.080</td>
<td>0.138</td>
</tr>
<tr>
<td>TRT ( \times \text{Unit type} )</td>
<td>8, 96</td>
<td>0.963</td>
<td>0.462</td>
<td>0.074</td>
</tr>
<tr>
<td>Triplet type ( \times \Delta f_{\text{mod}} \times \text{TRT} )</td>
<td>28, 1,344</td>
<td>0.308</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>Triplet type ( \times \Delta f_{\text{mod}} \times \text{Unit type} )</td>
<td>56, 672</td>
<td>0.328</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>Triplet type ( \times \text{TRT} \times \text{Unit type} )</td>
<td>16, 192</td>
<td>0.099</td>
<td>0.062</td>
<td></td>
</tr>
<tr>
<td>( \Delta f_{\text{mod}} \times \text{TRT} \times \text{Unit type} )</td>
<td>28, 336</td>
<td>0.337</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>Triplet type ( \times \Delta f_{\text{mod}} \times \text{TRT} \times \text{Unit type} )</td>
<td>112, 1,344</td>
<td>0.419</td>
<td>0.079</td>
<td></td>
</tr>
</tbody>
</table>

Bold numbers highlight the significant effects.

0.024), and triplet type and TRT (\( P < 0.001 \)) were also observed. Those interactions may reflect the effect of suppression on VS, which can be related to TRT (shorter TRTs may be providing more suppression) and triplet type (no suppression effects in triplet type -B–).

### Discussion

In this study we tested the hypothesis whether successive modulated signals that differ in modulation frequency and can be segregated into two streams are represented by separate populations of neurons. This separate representation may be similar to the representation of separate streams in the perception of pure-tone sequences by distinct population of neurons (e.g., Bee and Klump 2004; Fishman et al. 2001). Compared with streaming by pure-tone frequency, however, streaming by modulation frequency cannot rely on differences in the location of the main peak in the signal spectrum. Using an approach paralleling that of Bee and Klump (2004), we evaluated the effects of modulation rate differences and temporal proximity of signals on the amount of suppression that may lead to separated neural representations being a possible correlate of stream segregation. The area in the starling auditory forebrain from which we recorded in awake birds is the homolog of the mammalian auditory primary cortex, thus providing an interesting comparison to the imaging study by Gutschalk et al. (2007) focusing on the human auditory cortex.

**Modulation tuning properties and the effect of \( \Delta f_{\text{mod}} \)**

To represent streams of sounds that can be segregated by differences in the rate of modulation, it is necessary to have neurons that show tuning to the modulation. Although the multiunit responses in the current study were obtained from small groups of neurons, the finding that the data provide evidence for modulation tuning suggests that these small groups of neurons do not differ too much in their response properties (as has been found for pure-tone tuning characteristics; see Nieder and Klump 1999). Furthermore, previous studies on masking of neuronal responses in the starling forebrain by temporally structured sounds indicate a well-developed ability of the multiunit activity to represent temporal patterns of sounds (Bee et al. 2007; Nieder and Klump 2001).

We found different types of functions relating the neurons’ rate response to the frequency of sinusoidal AM of a tone. Besides recording sites that were responding best to a certain modulation frequency (i.e., show a band-pass characteristic), a large number of recording sites were also found that showed low-pass or high-pass characteristics regarding the modulation. Neurons showing these three types of response pattern in relation to the modulation could contribute to a segregated representation of sounds with different modulation frequencies. Moreover, the recording sites with a band reject characteristic have the possibility of representing a certain range of modulation frequencies. Thus there is a large population of neurons (>72% of recording sites) in our study in the primary auditory area of the starling forebrain demonstrating tuning characteristics with regard to the modulation that can form the neural substrate for streaming by temporal envelope cues. Recording sites with all pass characteristics should have a limited possibility of contributing to a separate representation of streams of sounds differing in modulation frequency in the rate response. Tuning of the rate response to the modulation frequency has been reported in the field L complex of a close relative of the starling, the mynah bird (Hose et al. 1987). In the study by Hose et al. (1987) the best rate response to the modulation was <200 Hz in >95% of the neurons. Our own data in the starling for recording sites with BP characteristics also show low best modulation frequencies (between 17 and 80 Hz). This compares well with the range of modulation frequencies to which neurons in the mammalian auditory cortex respond (see review by Joris et al. 2004). Since we presented our stimuli with a carrier frequency that corresponded to the recording site’s best frequency when stimulated with pure tones, quadratic distortion products—equivalent to the modulation frequency of presented SAM tone, produced by the spectral components due to the modulation (see McAlpine 2004)—are not likely to have affected the observed response because they generally have a much lower level than that of the main spectral components. Cubic distortion product (2f1-f2, 2f2-f1) otoacoustic emissions in the European starling that have been reported by Kettlembeil et al. (1995) and show a level of \( \leq 10 \) dB SPL at any primary-tone frequencies between 1.9 and 4 kHz when the primary tone level was 70 dB SPL. It is unlikely that this level of the cubic distortion products can lead to a sufficient alteration of the temporal signal structure by an interaction between the spectral components of the signal and the distortion products.

As can be expected from the large proportion of recording sites showing tuning to the modulation, the difference in modulation frequency (\( \Delta f_{\text{mod}} \)) between two successive SAM tones has a significant effect on the spike rate, although in some cases that modulation tuning found for the 125-ms signals used in the triplets was not as good as that observed with the 600-ms signals used to characterize the tuning. This may be due to presenting fewer cycles of modulation. The difference in the response to the A signals and to the B signal in the ABA- triplet is reflected in the significant interaction between signal position in the triplet and the modulation frequency difference \( \Delta f_{\text{mod}} \). The interaction between signal
position and the pattern of change of the rate response with increasing Δfmod varies between different types of recording sites, as indicated by a significant three-way interaction between sound number, Δfmod, and unit type. Although LP or BP recording sites (classified on the basis of the MTF) showed a monotonic decrease in the response to an increasing fmod of the B signals, the rate in response to these signals in HP sites usually first increased with increasing fmod as predicted by the MTF. A similar monotonic increase of spike rate with increasing fmod (i.e., HP characteristic) was commonly observed from recording sites in the primary auditory cortex of the awake squirrel monkey (Bieser and Müller-Preuss 1996), awake marmoset (non-BP; Liang et al. 2002), and cat (Eggermont 1993) as well as other neuronal types such as BP and LP (for a review, see Joris et al. 2004). The similarity of the response pattern in the MTF and in response to the varying modulation frequency of B signals in ABA triplets suggests that the characteristics of modulation tuning in some of the unit types (e.g., LP, BP) may determine the segregation of successive SAM tones in auditory streaming.

By changing the rate of modulation, we changed not only the temporal structure of the envelope, but also at the same time the bandwidth of the signal that increased with larger modulation frequencies. The range of reference modulation frequencies applied in the current study was generally <160 Hz, i.e., strongly overlapped with the range of tuned responses observed in the previous studies described earlier (e.g., Hose et al. 1987). Given the low RMFs, the side bands created by the AM ≤2 octaves above the RMF were generally within the range of frequencies providing excitation (49 of 54 recording sites; only 5 recording sites deviated from the rule; see Fig. 9). Up to four octaves above the RMF, the percentage of units in which the sidebands created by the modulation lay outside the limit of the excitatory tuning curve was increased (between 25 and 69% of the different unit types; see Fig. 9). Thus at modulation frequencies that were ≥2 octaves above the RMF, the drop of spike rates in response to SAM tones may be due to the sidebands created by the modulation falling into suppression areas that were commonly observed in the pure-tone frequency tuning curve in the tested forebrain area. The contribution of suppressive sidebands of the frequency tuning curve on the drop of the spike rate may be important, especially if the RMF of the recording site is high and/or Δfmod is large. In such a condition the effect of the suppression by the sidebands is not clearly distinguishable from the rate drop that is due to the modulation tuning based on temporal cues.

Even if the sidebands created by the modulation were found to lie within the limits of the excitatory tuning curve, it is possible that the decrease of spike rates in response to SAM tones with increasing modulation frequency occurred due to the modulation. Sideband frequencies that are more remote from the carrier frequency may elicit a weaker response than sidebands being close to the carrier frequency. This, however, cannot explain all patterns of spike responses such as HP or AP recording sites, showing no decrease of spike rate with increasing the modulation frequency up to a rate of 320 Hz.

In a previous psychophysical study, Grimault et al. (2002) used SAM noise to investigate whether sounds without spectral cues that differ only in their modulation frequencies could elicit stream segregation. They found that the separation of AM frequencies of ≥1.0 octave could segregate successive sounds paralleling our finding of significant differences in the response to stimuli with a modulation frequency that differs from the RMF by >1 octave. Using stimuli that were similar to those presented to the starlings (ABA- series, 125-ms tone duration, 100% TRT, and carrier frequency of 1 or 4 kHz), Dollezával and Klump (2009 and unpublished data) observed that SAM tones like those presented here could evoke stream segregation in the perception of human subjects. In these experiments, human segregated successive SAM tones with a separation of fmod by >1.5 octaves into two streams, which is a difference that also appears to be effective in starling forebrain neurons. Other studies have suggested the contribution of fundamental frequency of harmonic complex sounds on stream segregation for spectrally unresolved harmonics, which also points to a role of temporal processing in stream segregation (Gutschalk et al. 2007; Vliegen and Oxenham 1999). Those psychophysical studies are consistent with our current investigations, showing modulation-frequency-dependent spike responses to SAM tones with different modulation frequencies.

**Effect of forward suppression**

Forward suppression has been assigned a significant role for the stream segregation of successive sounds. The present study compared the rate responses to B signals in different stimulus sequences (ABA-, CBC- and B–) and showed that the stimulation history (SAM tones, pure tones, or silence) significantly affected the spike-rate responses to B signals. The observation that the B– sequence evoked a significantly larger number of spikes in response to the B signal compared with that in the other two stimulus sequences strongly indicates forward suppression. The result that the change of tone repetition time (TRT) in the present study had a significant effect, as observed in the responses to pure-tone ABA triplets (Bee and Klump 2004), also indicates forward suppression. This effect was observed from all unit types. The finding of the present study that there was a significant monotonic increase of rate for all TRTs of up to the maximum of 400% was in agreement with the responses to pure tones showing a significant increase of spiking activity with increasing TRT ≥800%. The parallels in the effects of TRT indicate that forward suppression is effec-
tive both in the processing of pure-tone sequences and SAM-tone sequences. In contrast to the results from Bee and Klump (2004), who observed that the difference between the response to the first and last A tones in the triplet was almost zero when the TRT was >200%, we still observed a relatively large rate difference between the first and second A signal responses in ABA-triplets at the longest TRT and a general reduction in the A signal responses at TRTs of 100 and 200%. This might also indicate that long-duration suppression is effective in the current paradigm involving SAM signals in which signal energy was always presented at the neurons’ CF. A similar long-lasting suppression between successive SAM tones extending over a few hundred milliseconds was also observed in the study by Bartlett and Wang (2005) in the primate auditory cortex.

Forward suppression (as indicated by Difference 1) was independent of the modulation frequency of B signals, which is in contrast with the previous study of streaming in the starling forebrain using pure-tone stream sequences by Bee and Klump (2004). Bee and Klump found that the amount of forward suppression decreased with increasing frequency difference between A and B tones, whereas it was independent of Δfmod in the present study. Bartlett and Wang (2005), who also observed the response of one SAM tone to the following SAM tone in the auditory cortex of awake marmosets, found a significant masking effect by either SAM tone or noise with modulation frequencies that were dissimilar to those of a following SAM tone/noise by ≤4 octaves or more. This was in agreement with our study. However, they also found a correlation (r = 0.44, P < 0.001) between the weighted average of the modulation frequencies of the first tone providing suppression and the modulation frequency of the second tone that was suppressed, indicating that a similarity in the modulation frequency promoted suppression. However, we observed that an unmodulated first C signal in the triplet could even provide slightly more suppression on the subsequent modulated tone than a modulated first A signal, indicating that similarity in modulation frequency was not that important for the amount of suppression. Our results that show significant main effects of Δfmod and TRT but no significant interaction on the rate response indicate that the effects of suppression, which are mainly mediated by spectral energy in the frequency band occupied by the preceding stimulus and the drop in activity that is related to the modulation tuning, are acting in an additive fashion. Since we increased fmod of only the B sound, we cannot evaluate whether the additive effect also occurs at lowered modulation frequencies.

Tuned AM-selective forward masking was observed in human psychophysics. Using a SAM noise carrier that was modulated for two consecutive time periods, Wojtczak and Viemeister (2005) found that modulation perception during the second period was compromised maximally if the first and second periods were close in modulation frequency. This is in contrast to our finding, which showed that the magnitude of forward suppression was nonsensitive to modulation frequency of B signals. One of the reasons for the discrepancy may lie in the different experimental paradigms. Wojtczak et al. (2007), however, who used the same paradigm as Wojtczak and Viemeister (2005) in the study of IC neurons in the awake rabbit, did not find a correlate of the psychophysical result in the IC neurons’ response pattern. This is similar to our results that indicate that the response to an AM signal is not negatively affected by the modulation of the preceding signal. We even observed slightly higher forward suppression by non-AM maskers (CBC-) on the rate response to the B signal than by SAM tone maskers (ABA-).

Gutschalk et al. (2007) suggested a contribution of modulation-based forward suppression on stream segregation from the investigation of human cortical activity using fMRI and MEG. However, they observed increased cortical activity if the Δfmod of two successive amplitude-modulated sounds was increased, whereas we observed a decrease in the response. This difference between the data obtained with the various methods can be explained by the fact that the MEG and fMRI study by Gutschalk et al. (2007) records the responses to both populations of neurons responding to A and B signals, whereas we look at the population of neurons that is mostly tuned to A signal responses.

Temporal representation of streaming sounds

If stream segregation is based on differences in the temporal patterning of the response rather than differences in the overall rate, the neurons’ ability to reflect the pattern of modulation is important. This ability can be described by the vector strength in response to the period of the modulation of SAM tones investigated in the present study. Different unit types showed different amounts of synchrony to the modulation period of the signals in ABA-triplets and the vector strength in AP recording sites was significantly higher than that in other types of recording sites. In general, vector strength significantly decreased with increasing Δfmod, which indicates a limited ability to synchronize to the modulation at high modulation rates. There was no significant interaction between Δfmod and unit type, indicating similar patterns of decay with increasing Δfmod in all unit types. When VS values from B-tone responses at Δfmod of >1 octave were calculated, scarcely any synchrony of the response with the modulation was observed. Since most RMFs in the present study were <100 Hz, this indicates that the representation of temporal patterns is limited to envelope frequencies of <200 Hz. This limit is in agreement with the results from other studies in birds (Hose et al. 1987; Knipschild et al. 1992) and mammals (e.g., Bieser and Müller-Preuss 1996; Gaese and Ostwald 1995; Liang et al. 2002; Nelson and Carney 2007; for review see Joris et al. 2004). In addition, we observed that the shortest TRT elicited significantly larger VS in the response to B signals in ABA-triplets compared with that at longer TRTs. This may indicate that the reduction of spiking activity due to forward suppression results in a more accurate temporal representation of the modulation. Similarly, VS in response to B signals in ABA- and CBC- was significantly larger than VS in response to -B- stimuli, further indicating that suppression may enhance the temporal representation of modulation frequency. If the temporal representation is enhanced, this may lead to a better segregation of streams since the temporal patterns are better separable. Using SAM noise, Grimault et al. (2002) reported that the increase in modulation depth of sounds results in increased stream segregation in human psychophysics, which is consistent with the hypothesis that a better temporal representation helps segregating sounds from different sources. In summary, although a transformation from a temporal to a rate representation of the modulation occurs at the level of central auditory system (e.g., Schreiner and Lang-
ner 1988; Schulze and Langner 1997; see review by Joris et al. 2004), we cannot exclude that the temporal representation of the modulation by the neurons may be exploited for stream segregation.

Peripheral channeling versus modulation tuning

For stream segregation of pure tones differing in frequency the peripheral channeling hypothesis has suggested that the degree of overlap in excitation patterns at the level of the cochlea is an important factor for stream segregation (Hartmann and Johnson 1991; Moore and Gockel 2002). A neural correlate of stream segregation of tone series can already be observed in neurons of the cochlear nucleus (Pressnitzer et al. 2008). The tonotopically organized peripheral channels are conserved up to the cortical level, but other types of channels representing different features of the sounds (e.g., envelope modulation) are created within the auditory pathway (e.g., Joris et al. 2004; Miller et al. 2001). The tonotopic organization implicates that in the central auditory system streaming on the basis of segregation by spectral frequency can also be observed (e.g., Bee and Klump 2004; Fishman et al. 2001). Beauvois and Meddis (1996) and McCabe and Denham (1997) established computational models to explain the relationship between the excitation patterns evoked in the cochlea by alternating tones and the perception of segregated streams.

Other studies, however, have shown that stream segregation occurs even if two successive complex sounds produce excitation patterns that are identical or very similar (Cusack and Roberts 1999; Dolležal and Klump 2009; Grimault et al. 2002; Gutschalk et al. 2007; Houtsma and Smurzynski 1990; Moore and Gockel 2002; Vliegen and Oxenham 1999). These observations indicate that the temporal structure of sounds may also result in stream segregation if no substantial difference in the representation of the sounds in peripheral frequency channels is provided. Also the results of the present study cannot easily be explained by the peripheral channeling hypothesis. At least for modulation frequencies differing by <2 octaves, the response to the SAM tones that had the same carrier frequency in the triplet differed, although all frequency components of the SAM tones in the triplet fell within the limits of the excitatory tuning curve. The peak of the excitation patterns in the cochlea resulting from A and B signals in the triplet having the same carrier frequency is not different and can be found in the same peripheral channel for both SAM tones that are segregated. At higher levels of the auditory system, modulation filters are found that could support segregation of the SAM tones. Modulation tuning at the central auditory system results in a map that is independent from the tonotopic representation (e.g., see Hose et al. 1987; Joris et al. 2004; Langner 1992). This independent map could lead to a segregated representation of streams differentiated by features other than spectral patterns of excitation. Our observation in the starling forebrain suggests that such a segregated representation of SAM sounds may provide the basis for auditory streaming by temporal modulation patterns. This is a possible mechanism for the implementation of stream segregation that is analogous to the pure-tone stream segregation mechanism (Bee and Klump 2004), suggesting that modulation tuning in general will support the segregation.

Moore and Gockel (2002) pointed out that any difference in the salience of sequential sounds could possibly result in stream segregation. Thus multiple cues are effective and the question is how they interact. Cusack and Roberts (1999) used two-tone stimuli that simultaneously offered different spectral and temporal cues. They found that neither the intermodulation rate of the two-tone stimulus nor the spectral overlap would predict the perception of stream segregation, but both absolute and relative bandwidths of the stimulus were important. Also in the present study the bandwidth of the stimulus was changed and results from the starling forebrain could be explained in terms of not only differentiation by modulation rate but also differentiation by bandwidths. Cusack and Roberts (1999) argued that the negative result regarding the modulation rate might be due to the low pitch strength of their two-tone stimulus. Physiological studies suggest that pitch strength is related to the time locking of action potentials to the temporal waveform of the modulated stimulus (e.g., Cariani and Delgutte 1996) and a well time locked response can be evoked by SAM stimuli in the auditory system. Thus it seems possible that streaming of SAM stimuli can be based on the temporal waveform of the stimulus, although spectral processing of the width of the spectral pattern of excitation that varies with the SAM frequency cannot be excluded.

Concluding remarks

The response pattern observed here in neurons of the avian forebrain area field L can be interpreted as a correlate of the processes occurring during auditory stream segregation by AM that can be observed in an ABA- paradigm. With an increasing difference in modulation frequency between the A and B signals, neurons responding with a high rate to the modulation frequency of A signals will reduce their response to the B signals differing in modulation frequency. Added to this reduction is a forward suppression from A signals on B signals that appears to be related to having spectral energy in a similar frequency range in both signals. Reduction of the response of B signals relative to that of A signals that are attributed to both processes adds up and may result in a more separated representation of A and B signals by two populations of neurons, each responding preferentially to one of the two signal types. This mechanism is consistent with the hypothesis proposed on the basis of fMRI and MEG data by Gutschalk et al. (2007) that auditory streaming by temporal patterns of stimuli evoking a pitch sensation is related to reciprocal feature-specific suppression between the two signal types.

Acknowledgments

We thank the anonymous reviewers for providing useful comments on earlier versions of the manuscript and H. Dierker for help in establishing the experimental setup and for data analysis.

Grants

This research was supported by Deutsche Forschungsgemeinschaft Grant SFB/TRR 31.

References


