Auditory Stream Segregation:
Neurophysiological Perspectives from the Songbird Forebrain

Mark A. Bee[1], Georg M. Klump[2]

Institut für Biologie und Umweltwissenschaften, 26111 Oldenburg, Deutschland
[1]*Email: mark.bee@uni-oldenburg.de*   [2] *Email: georg.klump@uni-oldenburg.de*

Introduction

Auditory stream segregation refers to the processes involved in perceptually segregating temporally separated sounds from other interleaved sounds and integrating them over time into coherent perceptual representations ("auditory streams") that correspond to the sound sources in an acoustic scene [1]. Such processes are a fundamental aspect of human hearing and speech and music perception. Auditory stream segregation, however, is not unique to humans, but instead appears to be taxonomically widespread and may, therefore, represent the operation of auditory processes that arose early in the evolution of vertebrate hearing [2,3,4]. The perceptual segregation of simple interleaved tone sequences has been studied in humans and European starlings (*Sturnus vulgaris*) using sequences of two alternating tones differing in frequency (ABA-ABA-ABA-...) [1,2]. The segregation of A and B tones into separate auditory streams is believed to be promoted by low-level auditory processes that increase the separation of excitation patterns along a tonotopic gradient [5,6]. Here, we report the results of a study [7] in which we tested the hypothesis that frequency selectivity and forward masking function in auditory stream segregation by recording neural responses in the tonotopically-organized auditory forebrain (field L2) of awake starlings to sequences of repeated ABA- triplets.

Methods

Multi-unit recordings were made from electrodes chronically implanted in field L2 of four adult starlings. Surgery and electrode implantation were performed while the birds were under general anesthesia (Isoflurane). Up to four electrodes were attached to a small head-mounted microdrive and could be lowered stepwise into field L2. Data were obtained at a total of 46 recording sites from awake, unrestrained birds using radio telemetry. During an experiment, the birds were located in a test cage inside a semi-anechoic sound chamber with a frequency response that was flat (± 4 dB SPL) over the range of frequencies used in this study. Acoustic stimuli were presented from an overhead speaker. Birds were given *ad libitum* access to food and water. The starlings in our study were passive listeners, in the sense that the birds were not required to perform a learned discrimination task during recordings, and had not been trained to attend to, or to discriminate between, 1-stream (integrated) or 2-stream (segregated) percepts. The care and treatment of the animals were in accordance with the procedures of animal experimentation approved by the Bezirksregierung Weser-Ems. Additional methodological details have been reported elsewhere [7].

Experimental Design: The stimuli consisted of sequences of 30 repeated ABA triplets. In separate stimulus sequences, we varied the *frequency separation* (ΔF) between the A and B tones and the *tone repetition time* (TRT). The frequency of the A tone was always fixed at the neurons' characteristic frequency (CF) [7]. Within a stimulus sequence, the B tone was constant and differed from the CF by either 0, 2, 4, 6, 8, 10, or 12 semitones. The repeated tones within a stimulus sequence were presented at TRTs that where 100, 200, 400, or 800% of the tone duration (TD), which was constant within a stimulus sequence and was either 25, 40, or 100 ms. We included a TRT that was 100% of the TD as the fastest possible tone rate without tone overlap to simulate conditions under which the perceptual streaming effect is known to occur in humans [1] and in starlings [2]. Longer TRTs (e.g., TRT = 800% of TD) were chosen to simulate conditions under which the streaming effect should be weak or absent based on human psychophysical data [1]. At each recording site, we presented stimulus sequences in a different randomized order at 70 dB SPL with a minimum silent interval of 7s between consecutive sequences.

Data Analysis: We used an online window discriminator to reject and then repeat recordings that contained a high number of movement artifacts. Our analyses are based on responses to the first 20 triplets in a sequence of 30 triplets that were free from movement artifacts. For each recording site, and for each stimulus sequence, we determined the mean discharge rate in spikes per second (averaged over artifact-free responses to 20 ABA triplets) in response to the first, second, and third tone presentations in a triplet. Responses were normalized to the responses to a CF(A) tone of the same TD presented at a slow rate.

Results and Discussion

We assessed the magnitudes of the effects of ΔF, TRT, TD, and position within a triplet (Tone 1, 2, or 3) on responses to the CF (A) and non-CF (B) tones in the ABA triplet sequences by computing a 7 (ΔF) x 4 (TRT) x 3 (TD) x 3 (Tone) rmANOVA. Three important trends were evident in responses to the ABA stimulus (Figure 1A&B). First, there was a significant main effect of ΔF (P < 0.01) and a significant interaction between ΔF and position within a triplet (P < 0.01). Responses to the non-CF (B) tone in the middle triplet position decreased as an increasing function of ΔF, while responses to the CF (A) tones in the first and third triplet positions were largely unaffected by differences in ΔF, at least at the longer TRTs (e.g., 200, 400, and 800%). At the highest levels of ΔF (8-12 semitones), the pattern of responses was dominated by excitatory responses to the A
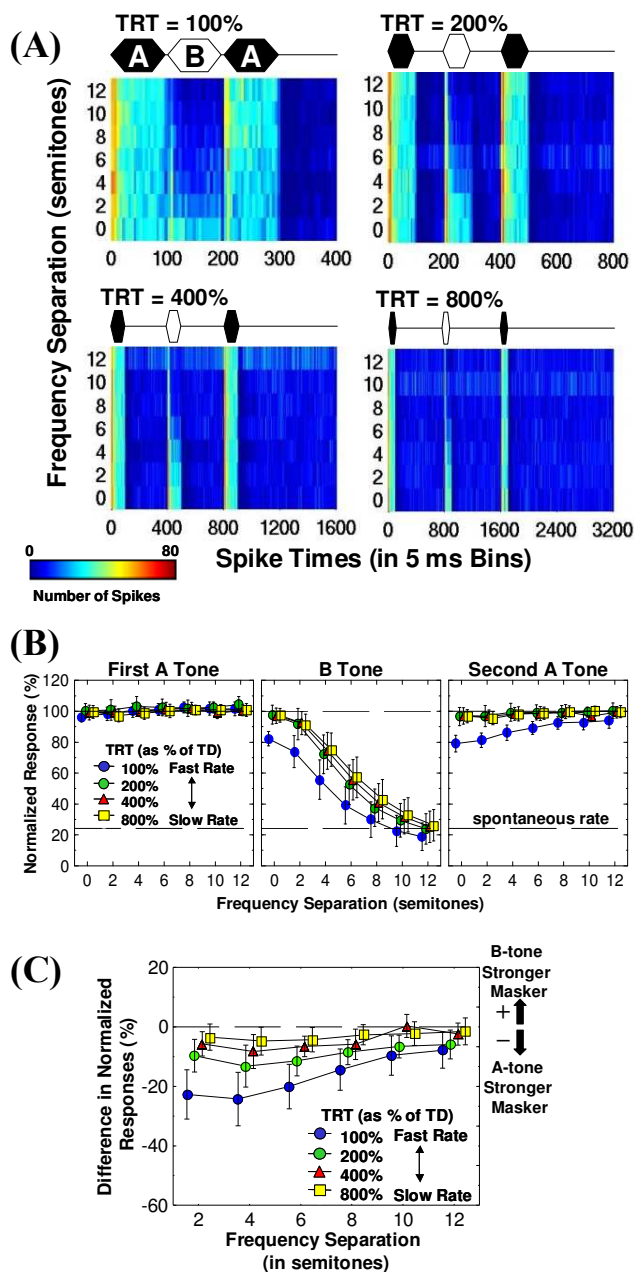
(A)

TRT = 100%   TRT = 200%

A   B   A

TRT = 400%   TRT = 800%

Frequency Separation (semitones)

Spike Times (in 5 ms Bins)

0 — 80
Number of Spikes

(B)

First A Tone   B Tone   Second A Tone

Normalized Response (%)

TRT (as % of TD)
100% Fast Rate
200%
400%
800% Slow Rate

spontaneous rate

Frequency Separation (semitones)

(C)

B-tone
Stronger
Masker
+
−
A-tone
Stronger
Masker

Difference in Normalized Responses (%)

TRT (as % of TD)
100% Fast Rate
200%
400%
800% Slow Rate

Frequency Separation
(in semitones)

Figure 1: **(A)** Color-coded PST histograms showing responses for a typical multi-unit recording site at all combinations of ΔF and TRT for the 100-ms TD, with color representing the number of spikes occurring in 5-ms time bins (see color bar inset for scale). Spike counts in the 5-ms bins were summed over artifact-free responses to 20 sequential repetitions of the ABA- triplet after off-setting the analysis window by the response latency. Oscillograms above each plot depict the timing of the A (filled) and B (open) tones in a repeated triplet. **(B)** Normalized responses to the ABA stimulus sequences showing the effects of ΔF and tone position within a triplet with TRT as the parameter. Points depict the mean ($\pm$2SE) normalized responses, averaged over 20 artifact-free responses to each stimulus sequence and over all three TDs at each of 46 recording sites. The upper dashed lines depict a normalized response that was 100% of the average response to isolated A tones; the lower dashed lines depict the mean normalized spontaneous rate. **(C)** The relative forward masking effects of A and B tones in the ABA stimulus. Points depict mean ($\pm$2SE) differences in the extent to which A and B tones were suppressed (measured relative to single-frequency control stimuli [7].

tones, and responses to the B tone in the middle triplet position were often not significantly different from, or were significantly lower than, the spontaneous rate [7]. These ΔF-dependent changes in responses to B tones reflect the tuning characteristics of the frequency-selective neurons in field L2. Second, there was a significant main effect of TRT ($P <$ 0.01) and a significant interaction between TRT and position within a triplet ($P < 0.01$). Responses to the B tone in the middle triplet position, and the A tone in the third triplet position, were additionally suppressed at the shortest TRT (TRT = 100% of TD) compared to longer TRTs. Third, the magnitude of additional suppression present at the shortest TRT depended on ΔF, being greater when the A and B tones were more similar in frequency (ΔF = 0-6 semitones). This ΔF-dependent suppression of B tones and the second A tone presented at the shortest TRT accounts for the significant ΔF x TRT and ΔF x TRT x Tone interactions ($P < 0.01$). Notably, A tones suppressed responses to B tones to a greater extent than B tones suppressed A tones (Figure 1C). This differential forward masking was most pronounced at small ΔFs and short TRTs.

Larger ΔFs and shorter TRTs promote the perceptual segregation of alternating tone sequences in humans [1] and also resulted in larger differences in neural responses to alternating CF (A) and non-CF (B) tones in the starling forebrain [7]. Importantly, responses to B tones in the ABA stimuli (i) decreased as ΔF increased and (ii) were differentially suppressed by A tones at smaller ΔFs and shorter TRTs. Together, these results indicate that frequency selectivity and forward masking are two low-level processes that can contribute to auditory stream segregation by promoting the spatial separation of neural excitation in tonotopic space that is elicited by CF(A) and non-CF(B) tones when these are presented in a alternating ABA sequence. Similar results have been reported in mammals [6].

Literature

[1] Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, MA, 1990

[2] Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *J Acoust Soc Amer* **103** (1998), 3581-3587

[3] Auditory stream segregation in goldfish (*Carassius auratus*). *Hear Res* **120** (1998), 69-76

[4] Auditory stream segregation in Japanese monkeys. *Cognition* **82** (2002), B113-B122

[5] Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Amer* **99** (1996), 2270-2280

[6] Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* **151** (2001), 167-187

[7] Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *J Neurophys* **92** (2004), 1088-1104