

# Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain

Mark A. Bee and Georg M. Klump

Animal Physiology and Behaviour Group, Institute for Biology and Environmental Sciences, Carl von Ossietzky University–Oldenburg, D-26111 Oldenburg, Germany

Submitted 9 September 2003; accepted in final form 17 March 2004

**Bee, Mark A. and Georg M. Klump.** Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol* 92: 1088–1104, 2004. First published March 24, 2004; 10.1152/jn.00884.2003. Auditory stream segregation refers to the perceptual grouping of sounds, to form coherent representations of objects in the acoustic scene, and is a fundamental aspect of hearing and speech perception. The perceptual segregation of simple interleaved tone sequences has been studied in humans and European starlings (*Sturnus vulgaris*) using sequences of 2 alternating tones differing in frequency (ABA-ABA-ABA- . . .). The segregation of A and B tones into separate auditory streams is believed to be promoted by preattentive auditory processes that increase the separation of excitation patterns along a tonotopic gradient. We tested the hypothesis that frequency selectivity and forward masking operate as 2 preattentive processes in sequential stream segregation by recording neural responses in the auditory forebrain of awake starlings to repeated ABA- sequences in which we varied the frequency separation ( $\Delta F$ ) between the A and B tones and the tone repetition time (TRT). The A tones were presented at the neurons' characteristic frequency (CF), and B tones differed from the CF over a one-octave range. Larger  $\Delta F$  values and shorter TRTs promote the perceptual segregation of alternating tone sequences in humans and also resulted in larger differences in neural responses to alternating CF (A) and non-CF (B) tones. Our results are consistent with the hypothesis that preattentive auditory processes, such as frequency selectivity and forward masking, contribute to the perceptual segregation of sequential acoustic events having different frequencies into separate auditory streams, but also suggest that additional processes may be required to account for all known perceptual effects related to sequential auditory stream segregation.

## INTRODUCTION

An important goal of neuroscience is to identify the mechanisms by which sensory systems group simultaneous and sequential sensory input to form coherent perceptual representations that correspond to different objects in the environment. In hearing, the processes responsible for grouping auditory input into distinct percepts are commonly referred to as “auditory scene analysis” or “auditory image formation” (reviewed in Bregman 1990; Yost 1991). Based on extensive psychophysical research in humans, Bregman (1990) proposed a distinction between “primitive” and “schema-based” processes in auditory scene analysis. According to Bregman (1990), the former is a data-driven phenomenon that consists of preattentive auditory processes that are automatic and obligatory, and that function to partition the incoming sound wave-

form into smaller elements, and to analyze the acoustic features of these elements. Primitive processes also function in grouping those sound elements that likely come from a common source into a coherent perceptual representation based on their common acoustic properties. Schema-based scene analysis, on the other hand, refers to perceptual grouping processes that require high-level, cognitive input and are influenced by the listener's attention and prior expectations based on previous learning. To date, there have been few neurophysiological studies aimed at identifying the underlying mechanisms of either primitive or schema-based auditory scene analysis (reviewed in Feng and Ratnam 2000).

One form of auditory scene analysis that has been the focus of psychophysical (Bregman 1990; Moore and Gockel 2002), theoretical (Beauvois and Meddis 1991, 1996; McCabe and Denham 1997), and physiological (Fishman et al. 2001; Hung et al. 2001; Sussman et al. 1999) studies, is sequential auditory stream segregation, which involves the segregation of temporally separated sounds from intervening and overlapping sounds, and their integration into separate “auditory streams.” A well-known example of sequential stream segregation is the “streaming effect” (Fig. 1; Bregman 1990). Under some acoustic stimulus conditions, human listeners presented with a repeated 3-tone sequence comprised of 2 alternating tones (ABA-ABA-ABA- . . .) report hearing a galloping rhythm. Under different acoustic stimulus conditions the A and B tones perceptually “split” into separate auditory streams, and listeners report hearing 2 separate tone sequences with different isochronous rhythms corresponding to 2 simultaneous sequences of tones occurring at different rates (A-A-A-A-A-A- . . . and -B---B---B- . . .). Two of the most important stimulus attributes that determine whether a sequence of alternating A and B tones is heard as one coherent stream of alternating tones, or 2 segregated streams of A and B tones, are the *frequency separation* ( $\Delta F$ ) between A and B tones and the *tone repetition time* (TRT) (reviewed in Bregman 1990). As illustrated in Fig. 1, the perceptual streaming effect is more pronounced with a larger  $\Delta F$  and at shorter TRTs.

If the streaming effect can be attributed largely to the operation of preattentive auditory processes, as suggested by Bregman (1990), we should expect to find neural mechanisms that operate in awake and passively listening animals, and for which the response output is influenced by  $\Delta F$  and TRT in ways that parallel their influence on the perceptual streaming effect. Two such auditory processes that have been implicated

Address for reprint requests and other correspondence: M. A. Bee, Carl von Ossietzky Universität Oldenburg, Fakultät V, Institut für Biologie und Umweltwissenschaften, AG Zoophysiologie und Verhalten, Carl von Ossietzky Str. 9-11, 26129 Oldenburg, Germany (E-mail: mark.bee@uni-oldenburg.de).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

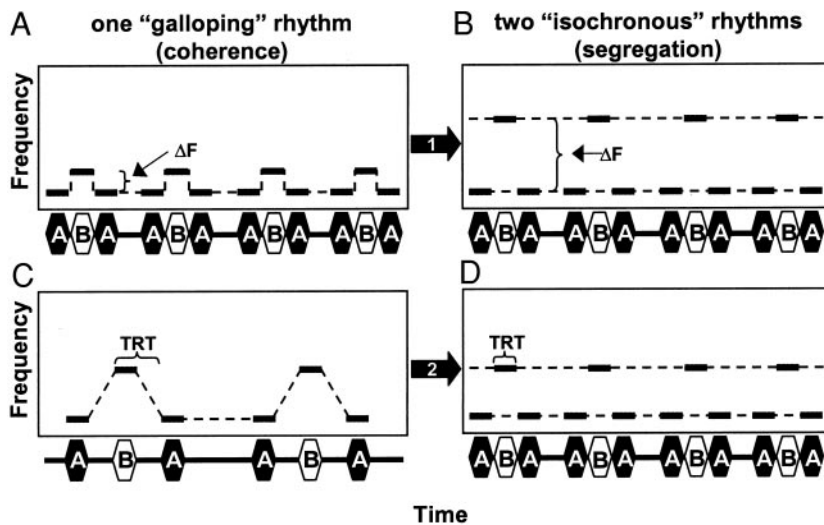


FIG. 1. Schematic diagram illustrating the ABA- stimulus paradigm and the dependency of the streaming effect on (A-B) the frequency separation ( $\Delta F$ ) between A and B tones and (C-D) tone repetition time (TRT). *Left column*: acoustic stimulus conditions under which human listeners report hearing a coherent stream of alternating tones with a galloping rhythm (ABA-ABA-ABA-...). *Right column*: acoustic stimulus conditions under which human listeners report hearing 2 segregated streams with "isochronous" rhythms corresponding to 2 perceptually distinct sequences of tones, one of which occurs at half the rate of the other (A-A-A-A-A-A-... and -B---B---B---B-...). Dashed lines connect adjacent tones that are heard in the same stream. Block arrows (1-2) illustrate that the perceptual streaming effect is promoted (1) at larger  $\Delta F$ s between A and B tones and (2) at shorter TRTs.

in previous theoretical and electrophysiological studies as primitive processes involved in sequential stream segregation are *frequency selectivity* and *physiological forward masking* (Beauvois 1998; Beauvois and Meddis 1991, 1996; Fishman et al. 2001; Hartmann and Johnson 1991; McCabe and Denham 1997). These studies share in common the general hypothesis that the perceptual segregation involved in the streaming effect is promoted when stimulus attributes result in an increase in the spatial separation of neural excitation patterns along tonotopic maps in the auditory system.

Frequency selectivity, which ultimately arises from spectral filtering in the cochlea, is realized in the form of tonotopic maps throughout the ascending auditory system. Tonotopy can promote the spatial separation of excitation by ensuring that alternating tones with different frequencies are encoded by different populations of neurons, with the separation among populations of neurons increasing as a function of  $\Delta F$  (Beauvois 1998; Beauvois and Meddis 1991, 1996; Fishman et al. 2001; Hartmann and Johnson 1991; McCabe and Denham 1997). Physiological forward masking involves the suppression of neural responses to a sound (the signal) after the presentation of a preceding sound (the masker). Although forward masking can be observed in the responses of auditory nerve fibers (Relkin and Turner 1988; Turner et al. 1994), additional processing is thought to contribute to masked neural responses at higher levels of the auditory system (e.g., Brosch and Schreiner 1997; Calford and Semple 1995; Oxenham 2001). In the context of a series of tones alternating in frequency, such as in the ABA- stimulus paradigm, each tone potentially serves as a masker of a subsequent tone, and as a signal tone following a preceding masking tone. Studies of physiological forward masking in cat primary auditory cortex (AI) using pure tones to mask a pure-tone signal presented at the recording site's characteristic frequency (CF) have shown that masking is often more pronounced when the masker and signal are similar in frequency and occur with short masker-signal delays (Brosch and Schreiner 1997; Calford and Semple 1995).

In a recent study using a sequential stream segregation paradigm to study neural ensemble responses in macaque AI, Fishman et al. (2001) provided evidence to suggest that frequency selectivity and forward masking play important roles in

sequential auditory stream segregation. They proposed the hypothesis that the differential suppression by tones presented at best frequency (BF) and non-BF tones resulted from the relatively stronger physiological forward masking of non-BF tones by preceding BF tones. According to this hypothesis, both BF (A) and non-BF (B) tones are able to mask each other to varying degrees, but BF (A) tones are more potent maskers than non-BF (B) tones when these are arranged in an alternating pattern. Fishman et al. (2001) suggested that the time course of suppressive interactions between neural responses to maskers and signals can explain the well-known effects of TRT on the streaming effect.

Here, we report results from a study in an awake songbird, the European starling (*Sturnus vulgaris*), that investigated the potential roles of frequency selectivity and forward masking in sequential stream segregation using the ABA- experimental paradigm. As in humans, starlings and other songbirds rely primarily on acoustic signals for social communication, and numerous parallels exist in the perceptual processing of bird-song and human language (Ball and Hulse 1998; Doupe and Kuhl 1999). Thus the hearing systems of both humans and starlings have likely experienced common evolutionary selection pressures to solve similar detection and perceptual organization tasks (Klump 1996; Klump et al. 2000). Psychophysical and neurophysiological studies of the starling hearing system confirm that starlings and humans share an impressive number of similarities in the spectral and temporal processing of acoustic stimuli (reviewed in Klump et al. 2000). These similarities extend, for example, to frequency selectivity, frequency discrimination, temporal resolution, temporal summation, and duration discrimination. Furthermore, given that vocal communication by songbirds and humans often occurs in large social groups (e.g., dawn choruses for songbirds and cocktail parties for humans), we should expect that songbirds also face the problem of perceptually segregating multiple overlapping and interleaved sequences of sounds into distinct perceptual objects. Current evidence, in fact, suggests that starlings possess capabilities of auditory scene analysis similar to those of humans (reviewed in Hulse 2002). Most relevant to our study is a previous study of sequential auditory stream segregation by MacDougall-Shackleton et al. (1998), which demonstrated that starlings, like humans, experience the

streaming effect using the ABA- stimulus paradigm, and are able to discriminate between coherent and segregated streams based on perceived differences in rhythm (Fig. 1). Given such strong similarities between humans and starlings in perceptual and physiological aspects of hearing, and the fact that starlings are known to experience the perceptual streaming effect, starlings are an excellent animal system for investigating physiological mechanisms underlying auditory scene analysis.

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, coherent or segregated streams, thus minimizing the potential influence of schema-based processes. We recorded neural responses in the tonotopically organized avian auditory forebrain (field L2) to repeated ABA- triplets in which we varied in a factorial design the TRT and the  $\Delta F$  between A tones presented at the site's CF and non-CF (B) tones. The tonotopically organized avian field L2 is the primary target of ascending projections from the auditory dorsal thalamus (reviewed in Carr and Code 2000), and thus represents the avian equivalent of mammalian primary auditory cortex, which is believed to play a role in sequential stream segregation (Fishman and Steinschneider 2003; Fishman et al. 2001; Micheyl et al. 2003).

Our study had 3 primary objectives. First, we investigated the effects of  $\Delta F$  and TRT on neural responses to test the general hypothesis that the degree of overlap in excitation along a tonotopic gradient decreases under stimulus conditions known to promote the perceptual streaming effect. We tested the specific prediction that the *differences* in responses to CF (A) and non-CF (B) tones would increase as an increasing function of  $\Delta F$  and as a decreasing function of TRT. Second, we tested the hypothesis proposed by Fishman et al. (2001) that differential responses to CF and non-CF tones are influenced by the relatively stronger physiological forward masking of non-CF tones by preceding CF tones. We tested the specific prediction that the relatively greater forward suppression of non-CF tones by preceding CF tones would increase as a decreasing function of both  $\Delta F$  and TRT (Brosch and Schreiner 1997; Calford and Semple 1995). Finally, we compare the effects of  $\Delta F$  and TRT on neural responses in the starling forebrain to their well-known influences on the perceptual streaming effect observed in previous psychoacoustic studies in humans and starlings.

## METHODS

### *Surgery and recordings*

Four wild-caught, adult starlings (2 males, 2 females; 71.2–94.3 g) were used as subjects in this experiment. The care and treatment of the animals were in accordance with the procedures of animal experimentation approved by the Bezirksregierung Weser-Ems. All procedures were performed in compliance with the American Physiological Society's *Guiding Principles in the Care and Use of Animals*.

Detailed descriptions of the manufacturing of electrodes and surgical procedures can be found elsewhere (Hofer and Klump 2003). Briefly, 2 types of extracellular recording electrodes were fashioned from either commercially made tungsten microelectrodes (shank diameter = 75  $\mu\text{m}$ , Frederick Haer and Co., Bowdoinham, ME) or Teflon-insulated platinum-iridium wires (shank diameter = 25  $\mu\text{m}$ , A-M Systems, Carlsborg, WA) that were sharpened at the tip using the procedures described by Hofer and Klump (2003). The impedance

of both types of electrodes measured in 0.9% NaCl ranged from 3.6 to 12.1 M $\Omega$  (1 kHz a/c). An array of 4 electrodes was fixed with dental acrylic to a small head-mounted microdrive that was used to manually lower the electrodes into the brain. The array was fixed so that the recording tips of different electrodes protruded between 0.5 and 1.5 mm from the opening of a 0.8-mm-diameter tube in the microdrive. The absolute distances between electrode tips typically varied between 0.25 and 1.0 mm.

After an initial subcutaneous injection of atropine (0.05 ml) to reduce salivation, surgery was performed under general anesthesia (Isoflurane: 5% for induction, 1.5–2.5% for maintenance). Anesthetized animals were fixed in a stereotaxic holder, with the bill inclined about 45° below the horizontal plane. The caudal bifurcation of the sinus sagittalis served as a reference for making a small hole in the right hemisphere of the skull that was 0.8–0.9 mm lateral and 1.6–1.8 mm rostral of the bifurcation. These coordinates were chosen to reach the input layer of the field L complex, L2 (Nieder and Klump 1999). Electrodes were implanted in the brain through a small incision made in the dura. Two indifferent electrodes (stainless steel wire, diameter = 75  $\mu\text{m}$ , A-M Systems) were implanted through a second small opening in the skull made in the left rostral hemisphere. The microdrive, indifferent electrodes, and a small socket for attaching a radio transmitter were fixed to the exposed skull with dental acrylic. Recordings began 3–9 days after surgery.

Multiunit recordings from a total of 46 recording sites were made from awake and freely behaving birds placed in a test cage (56 × 36 × 33 cm) located inside a radio-shielded sound chamber (IAC 402A, Industrial Acoustics, Norderkuchten, Germany). Neural activity was recorded by radio telemetry using a small FM radio transmitter (FHC type 40-71-1, Frederick Haer and Co). The radio signal was received by a dipole antenna inside the sound chamber and demodulated by an FM tuner (Technics ST-GT 550, Panasonic, Hamburg, Germany) located outside the chamber. The demodulated signals were band-pass filtered (600–4,500 Hz), amplified, digitized (Sound Blaster PCI128, 16-bit, 44.1 kHz), and stored on the hard drive of a Linux workstation (AMD Athlon XP 1900+) for later analysis.

At the beginning of an experimental recording session, the radio transmitter was attached to the head-mounted socket and the bird was temporarily restrained in a cloth jacket to prevent wing and leg movements. The microdrive was lowered stepwise until a site was found at which auditory-evoked activity was elicited in response to a series of test tones. The cells in L2 of the field L complex have an average cell diameter of 5–7  $\mu\text{m}$  (Saini and Leppelsack 1981); therefore we advanced the electrodes a minimum distance of 40  $\mu\text{m}$  into the brain between recordings from the same electrode to ensure that different cells were recorded. Once a suitable recording site was found, the bird was released into the test cage and given unrestricted access to food and water. Because the subjects were completely unrestrained inside the test cage, it was not possible to objectively quantify a bird's arousal level during a recording session. Instead, birds were monitored remotely using a video camera mounted in the chamber and a video monitor located outside the chamber. These video observations revealed that subjects were awake during recording sessions because the birds often took food and water, changed positions by hopping between 2 perches in the cage, and regularly exhibited other behaviors, such as head turning, scratching, ruffling feathers, and bill-wiping.

At the completion of recordings, animals were killed with an overdose of sodium pentobarbital, and their brains were fixed by transcardial perfusion of Zamboni's reagent after an initial flush of the circulatory system with a warm solution containing 0.9% NaCl and 0.5% NaNO<sub>2</sub>. The brain was stored in 30 ml of the fixative containing an additional 30% saccharose for several days before frozen sagittal sections (50  $\mu\text{m}$ ) were sliced and stained with cresyl violet to confirm the position of the electrodes (for additional details see Nieder and Klump 1999). These histological analyses confirmed that recordings were made in field L2.

### Stimulus generation and presentation

Acoustic stimuli were generated at a sampling rate of 44.1 kHz and 16-bit resolution using custom-designed software running on the Linux workstation that allowed for the synchronous playback of acoustic stimuli and recording of neural responses. The analog sound output of the computer soundcard was attenuated (Hewlett-Packard 350D, Böblingen, Germany, and TDT PA4, Tucker-Davis Technologies, Alachua, FL), amplified (Rotel RB-1050, Sussex, UK), and presented through a speaker (Type SP3253, KEF Audio, Maidstone, UK) mounted from the ceiling of the sound chamber about 70 cm above the position of a starling sitting in the test cage. The frequency response inside the test cage was flat ( $\pm 4$  dB) over the range of frequencies used in this study.

Just before presenting the stimulus sequences described below, we generated a frequency tuning curve (Fig. 2) by presenting a series of 20 pure tones (200-ms duration, 10-ms Gaussian rise and fall times, 800-ms intertone interval) at each of 11 frequencies separated by 0.25 octaves within a 2.5 octave range centered around our estimate of the CF based on responses to the series of test tones. An on-line window discriminator automatically rejected responses containing artifacts attributed to the bird's movements and repeated tones until 10 artifact-free responses were obtained at each combination of frequency and level. Presentations began at the lowest level of 0 dB SPL (re 20  $\mu$ Pa) and were increased in 5-dB steps to a level of 70 dB SPL. We determined the recording site's CF as the frequency with the lowest threshold, where threshold was determined as the lowest stimulus amplitude at which the neural response was  $>1.8$  times the spontaneous rate.

The stimuli used in the experiment were based on the well-known ABA- stimulus paradigm commonly used in psychoacoustic studies of sequential auditory stream segregation (Bregman 1990; Moore and Gockel 2002). The "ABA- stimulus" consisted of a repeated triplet of 3 sinusoidal tones that alternated in frequency and consisted of an A

tone, a B tone, and a second A tone, as depicted in Fig. 1 (ABA-ABA-ABA-. . .). In different stimulus sequences, the frequencies of the A and B tones differed along a semitone musical scale (see following text). The third tone in an ABA- triplet (i.e., the second A tone) was followed by a silence (denoted above as "-") that was equal in duration to the time interval between adjacent tones plus the tone repetition time (measured from tone onset to onset) so that the respective periods of the A and B tones were constant across consecutively repeated triplets, and the period of the B tone was twice that of the A tone.

We included 4 types of stimuli as controls. The first control stimulus (the "AAA- stimulus") consisted of a repeated triplet with the same temporal arrangement as the ABA- triplet, but consisted of 3 repetitions of the A tone alone (AAA-AAA-AAA-. . .). Hereafter, this stimulus is usually denoted as an ABA- stimulus having 0 semitones frequency separation between the A and B tones. A second type of control stimulus (the "BBB- stimulus") consisted of a repeated triplet consisting only of the B tone (BBB-BBB-BBB-. . .). The AAA- and BBB- stimulus sequences were designed to allow us to compare the effects of surrounding A and B tones on responses to A and B tones in the middle position of a triplet. A third control stimulus (the "A-A- stimulus") was similar to the ABA- stimulus, except that the B tones were omitted and replaced with silences equivalent to the tone duration (A-A-A-A-A-. . .). A final type of control stimulus (the "-B-- stimulus") consisted of the B tone alone occurring in the same temporal arrangement, relative to the A tones, as it occurred in the ABA- stimuli, with the A tones replaced by silent intervals equivalent to the tone duration (-B---B---B-. . .). The A-A- and -B-- stimulus sequences allowed us to assess responses to isolated, single-frequency tone sequences in relation to triplets containing A tones, B tones, or both. In all 5 types of stimuli (ABA-, AAA-, BBB-, A-A-, and -B--), the "triplet" and the following silent intertriplet interval were repeated 30 times in sequence, and data were collected for artifact-free responses to 20 triplet repetitions.

At each recording site, we presented stimulus sequences in a different randomized order at 70 dB SPL. A silent interval of 7 s separated consecutive sequences to minimize any possible influence of auditory stream biasing between consecutive stimulus presentations (Beauvois and Meddis 1997; Bregman 1978). Spontaneous activity was recorded for 4 s preceding the onset of the first tone in each stimulus sequence. The generation of the frequency tuning curve, and a complete presentation of all stimulus sequences, required about 4 h.

### Experimental design

We examined the effects of  $\Delta F$  between CF and non-CF tones by fixing the frequency of A tones at the recording site's CF and varying the frequency of the non-CF (B) tone away from that of the CF (A) tone over a one-octave range along a semitone scale by a value of 2, 4, 6, 8, 10, or 12 semitones (see Fig. 2). The frequency of the B tone within a given stimulus sequence was constant over all 30 triplet repetitions; therefore there were 6 different ABA-, BBB-, and -B-- stimulus sequences corresponding to the 6 levels of  $\Delta F$  between the CF (A) and non-CF (B) tones. Similar ranges of  $\Delta F$  values have been used in human psychoacoustic studies of the streaming effect (Anstis and Saida 1985; Carlyon et al. 2001; Rogers and Bregman 1993; van Noorden 1975). In starlings (MacDougall-Shackleton 1998), the streaming effect has been demonstrated to occur at a short TRT when the  $\Delta F$  was about 9 semitones or larger, but not when the  $\Delta F$  was less than about 1 semitone. MacDougall-Shackleton (1998) did not test  $\Delta F$ s between 1 and 9 semitones, nor did they test longer TRTs. For recording sites with CFs above about 1 kHz and below about 3 kHz, the direction of  $\Delta F$  imposed on the B tones relative to the CF was determined randomly; for CFs below 1 kHz or above 3 kHz, the frequency of the B tones was increased or decreased, respectively, to ensure that the frequencies of the B tones remained well within the starling's hearing range (Dooling et al. 1986; Klump et al. 2000).

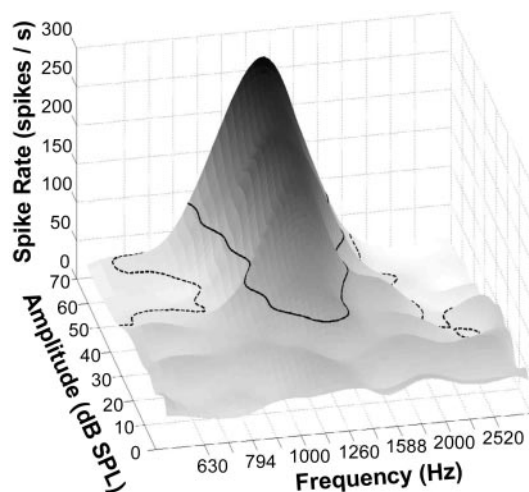


FIG. 2. Frequency tuning curve (FTC) based on multiunit activity recorded in field L2. Characteristic frequency (CF) at this recording site was 1,260 Hz and the response threshold was 18 dB SPL. FTCs from field L2 commonly exhibit a central excitatory region surrounded by one or 2 suppressive sidebands (Nieder and Klump 1999). Solid black line delineates the excitatory region where the discharge rate was above threshold, defined as being 1.8 times greater than the spontaneous rate. Dashed black lines delineate suppressive sidebands in which the discharge rate was less than the spontaneous rate divided by 1.8. Frequency of the A tones in a stimulus sequence was always fixed at the CF determined from an FTC. Gridlines for the frequency axis between 630 and 2,520 Hz are separated by intervals of 2 semitones over a 2-octave range centered on the CF of 1,260 Hz. Thus the appropriate levels of frequency separation ( $\Delta F$ ; either increase or decrease) between CF (A) and non-CF (B) tones for this field L2 site would correspond to the separation of the gridlines along the frequency axis.

To investigate the effects of TRT, the repeated tones within a stimulus sequence were presented at TRTs that were 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 (see following text). Here, we use TRT to refer to the time interval between the onsets of 2 consecutive tones within a sequence of 3 tones, for example, either A to B or B to A in the ABA- triplet. Thus shorter TRTs (expressed as percentages of TD) correspond to faster tone rates, and longer TRTs correspond to slower tone rates. The designated TRTs for the isolated tone sequences (A-A- and -B--) refer to the TRTs of the corresponding ABA- triplets, so that at a given TRT the respective periods of the isolated A and B tones were the same as those in the corresponding ABA- stimulus sequence. 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 (Miller and Heise 1950; van Noorden 1975) and in starlings (MacDougall-Shackleton et al. 1998). 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 (Beauvois and Meddis 1991; van Noorden 1975).

We performed the experiment using 3 different TDs (25, 40, and 100 ms) to provide a basis for comparing our results with a broad range of previously published studies. In their study of the streaming effect in macaque AI, Fishman et al. (2001) used 25-ms-duration tones. In one of the most widely cited studies of the streaming effect in humans (van Noorden 1975), a number of important properties of the streaming effect were demonstrated using 40-ms duration tones. A TD of 100 ms is similar to tone durations used in other studies of the streaming effect in humans (Rogers and Bregman 1993; Rose and Moore 1997, 2000; Singh and Bregman 1997), and, more important, it is the TD that was used by MacDougall-Shackleton et al. (1998) to demonstrate that starlings experience the streaming effect. Within a particular stimulus sequence, the duration of all tones was the same, and the amplitude envelope of the individual tones in all stimuli had 5-ms Gaussian rise and fall times. For each of the 3 TDs, a subset of 80 stimulus sequences was created based on the 4 levels of TRT and the 20 possible combinations of  $\Delta F$  and stimulus type (6 ABA-, 1 AAA-, 6 BBB-, 1 A-A-, 6 -B--), for a total set of 240 stimulus sequences.

### Data analysis

For each recording site ( $n = 46$ ), and for each of the 240 stimuli, we determined the mean discharge rate in spikes/s during the first, second, and third tone presentations in a triplet (or their corresponding silences in the A-A- and -B-- stimuli), averaged over artifact-free responses to 20 triplets (for additional details see Nieder and Klump 1999). The timing of analysis windows was adjusted to compensate for the response latency (11–14 ms). Normalized responses are expressed as a percentage of the average response to an isolated CF (A) tone. We normalized discharge rates by dividing the average response to each tone (or corresponding silence) by the average response to the isolated CF (A) tone having the same TD in the A-A- stimulus presented at the slowest rate (TRT = 800% of TD). Thus for the 25, 40, and 100 ms TDs, response rates were normalized to responses to isolated CF (A) tones that were separated by silent intervals of 375, 600, and 1,500 ms, respectively. In a neurophysiological study of forward masking in the starling auditory forebrain, Klump and Nieder (2001) showed that masker-signal delays of just 80 ms lead to an approximately 55-dB reduction in forward masking relative to a 5-ms masker-signal delay. Thus our method of normalization standardized A-tone and B-tone responses to the response to isolated CF (A) tones presented at rates slow enough to minimize any potential effects of forward masking. A normalized spontaneous rate was determined by averaging the discharge rate, first, across 20 artifact-free, 100-ms time windows recorded just before each stimulus presentation and, second,

across all 240 stimulus presentations (480 s total), and then by normalizing this average spontaneous rate to the average stimulus-driven rate in response to the 100-ms CF (A) tone presented at the longest TRT (800%).

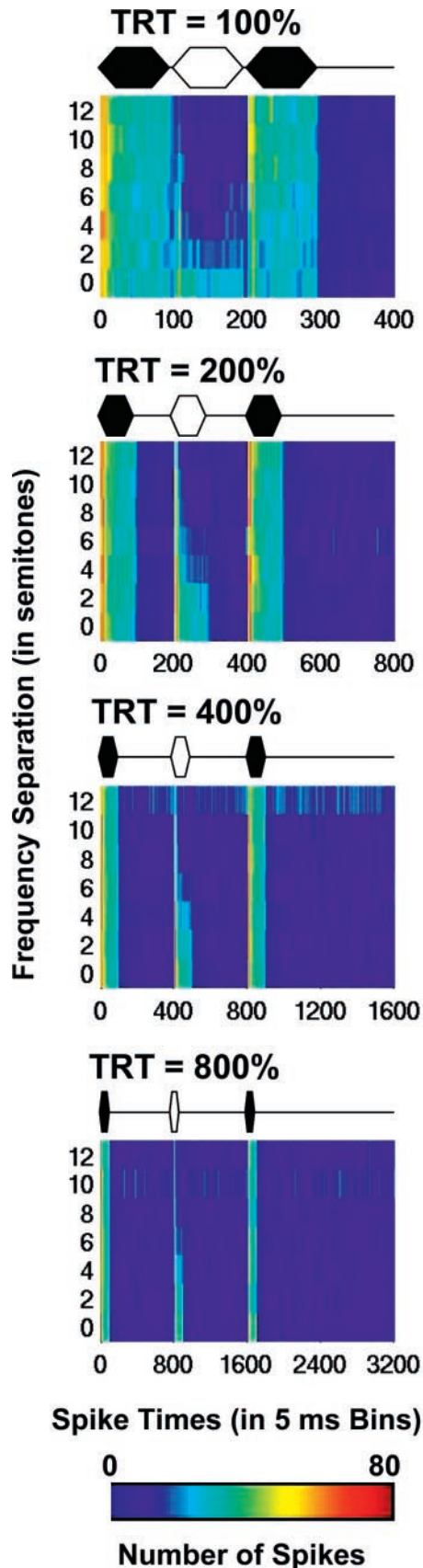
### Statistical analysis

We examined the effects of  $\Delta F$  and TRT using repeated-measures ANOVA (rmANOVA). Tone duration (TD) was also included as a factor in these analyses to partition out any variance that could be attributed to differences in this variable. Planned comparisons were used to test a number of specific predictions (Rosenthal and Rosnow 1985), which we describe below. For repeated-measures analyses with more than a single numerator degree-of-freedom ( $df$ ), we calculated  $P$  values using the Greenhouse and Geisser (1959) adjusted  $df$  for omnibus tests of within-subjects factors that violated the sphericity assumption of rmANOVA (Mauchly's sphericity test). 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, which serves as an estimate of the extent to which the null hypothesis of "no effect" is false. The interpretation of partial  $\eta^2$  values is similar to that of the more familiar coefficient of determination ( $r^2$ ). Although statistical analyses are essential to determine whether there are significant differences among treatments in any experiment, there is the concomitant risk of detecting statistically significant effects of questionable biological importance, especially in studies with high statistical power. In the analyses described below, we pay special attention to the magnitudes of effect sizes in our analyses, and do not judge the influence of a variable solely by the magnitude of the associated  $P$  value. All analyses were performed with Statistica 5.5 or SPSS 11.5, and an experiment-wide criterion of  $\alpha = 0.05$  was used to determine statistical significance.

Preliminary analysis of the normalized responses to the ABA-, BBB-, and -B-- stimuli were conducted to estimate the magnitude of the between-recording-sites effects of the direction of  $\Delta F$  between the A and B tones (increase vs. decrease). These analyses revealed that the between-recording-sites factor of the direction of  $\Delta F$  usually explained <5% of the variation in responses to the

TABLE 1. Results of a 7 ( $\Delta F$ )  $\times$  4 (TRT)  $\times$  3 (TD)  $\times$  3 (Tone) rmANOVA comparing the effects of frequency separation ( $\Delta F$ ), tone repetition time (TRT), tone duration (TD), and triplet position (Tone 1, 2, or 3) on the normalized responses to the A and B tones in the ABA- stimulus

Effect	$df$	$F$	$P$	$\eta^2$
$\Delta F$	6, 270	158.2	<0.001	0.78
TRT	3, 135	21.6	<0.001	0.32
TD	2, 90	0.6	0.571	0.01
Tone	2, 90	304.2	<0.001	0.87
$\Delta F \times$ TRT	18, 810	6.0	<0.001	0.12
$\Delta F \times$ TD	12, 540	0.6	0.809	0.01
$\Delta F \times$ Tone	12, 540	286.0	<0.001	0.86
TRT $\times$ TD	6, 270	1.5	0.197	0.03
TRT $\times$ Tone	6, 270	78.9	<0.001	0.64
TD $\times$ Tone	4, 180	19.4	<0.001	0.30
$\Delta F \times$ TRT $\times$ TD	36, 1620	1.1	0.352	0.02
$\Delta F \times$ TRT $\times$ Tone	36, 1620	6.0	<0.001	0.12
$\Delta F \times$ TD $\times$ Tone	24, 1080	1.7	0.067	0.04
TRT $\times$ TD $\times$ Tone	12, 540	3.5	0.002	0.07
$\Delta F \times$ TRT $\times$ TD $\times$ Tone	72, 3240	0.8	0.676	0.02



stimuli. Therefore this factor was not included in subsequent statistical analyses.

## RESULTS

### Responses to alternating CF and non-CF tone sequences

We assessed the magnitudes of the effects of  $\Delta F$ , 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 by computing a  $7 (\Delta F) \times 4 (\text{TRT}) \times 3 (\text{TD}) \times 3 (\text{Tone})$  rmANOVA (Table 1). Three important trends were evident in responses to the ABA-stimulus (Figs. 3 and 4A). First, there was a significant main effect of  $\Delta F$  and a significant interaction between  $\Delta F$  and position within a triplet (Table 1). Responses to the non-CF (B) tone in the middle triplet position decreased as an increasing function of  $\Delta F$ , whereas responses to the CF (A) tones in the first and third triplet positions were largely unaffected by differences in  $\Delta F$ , at least at the longer TRTs (e.g., 200, 400, and 800%). At the highest levels of  $\Delta F$  (8–12 semitones), the pattern of responses was dominated by excitatory responses to the A 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 (Table 2). Second, there was a significant main effect of TRT and a significant interaction between TRT and position within a triplet (Table 1). 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 with longer TRTs. Third, the magnitude of additional suppression present at the shortest TRT depended on  $\Delta F$ , being greater when the A and B tones were more similar in frequency ( $\Delta F = 0$ –6 semitones). This  $\Delta F$ -dependent suppression of B tones and the second A tone presented at the shortest TRT accounts for the significant  $\Delta F \times \text{TRT}$  and  $\Delta F \times \text{TRT} \times \text{Tone}$  interactions (Table 1).

An additional result of the statistical analyses reported in Table 1 concerns the effects of differences in TD. The main effect of TD was not significant, and 5 of 6 interaction terms containing TD as a factor were nonsignificant and were associated with small effect sizes that explained 4% or less of the variance (Table 1). Therefore because the effects of TD were generally small, we do not specifically consider the effects of differences in TD further, although we include TD as a factor in subsequent analyses to partition out from the error terms any variance attributed to differences in TD.

### Responses to single-frequency CF and non-CF tone sequences

In responses to all 3 tones in the repeated triplets of the BBB-stimulus (Fig. 4B), response magnitudes decreased with increasing  $\Delta F$ . At the shortest TRT there was also

FIG. 3. Color-coded *peristimulus time histograms* showing responses for a typical multiunit recording site at all combinations of  $\Delta F$  and tone repetition time (TRT) for the 100-ms tone duration, 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. Oscilloscope-like insets above each plot depict the timing of the A (filled) and B (open) tones in a repeated triplet.

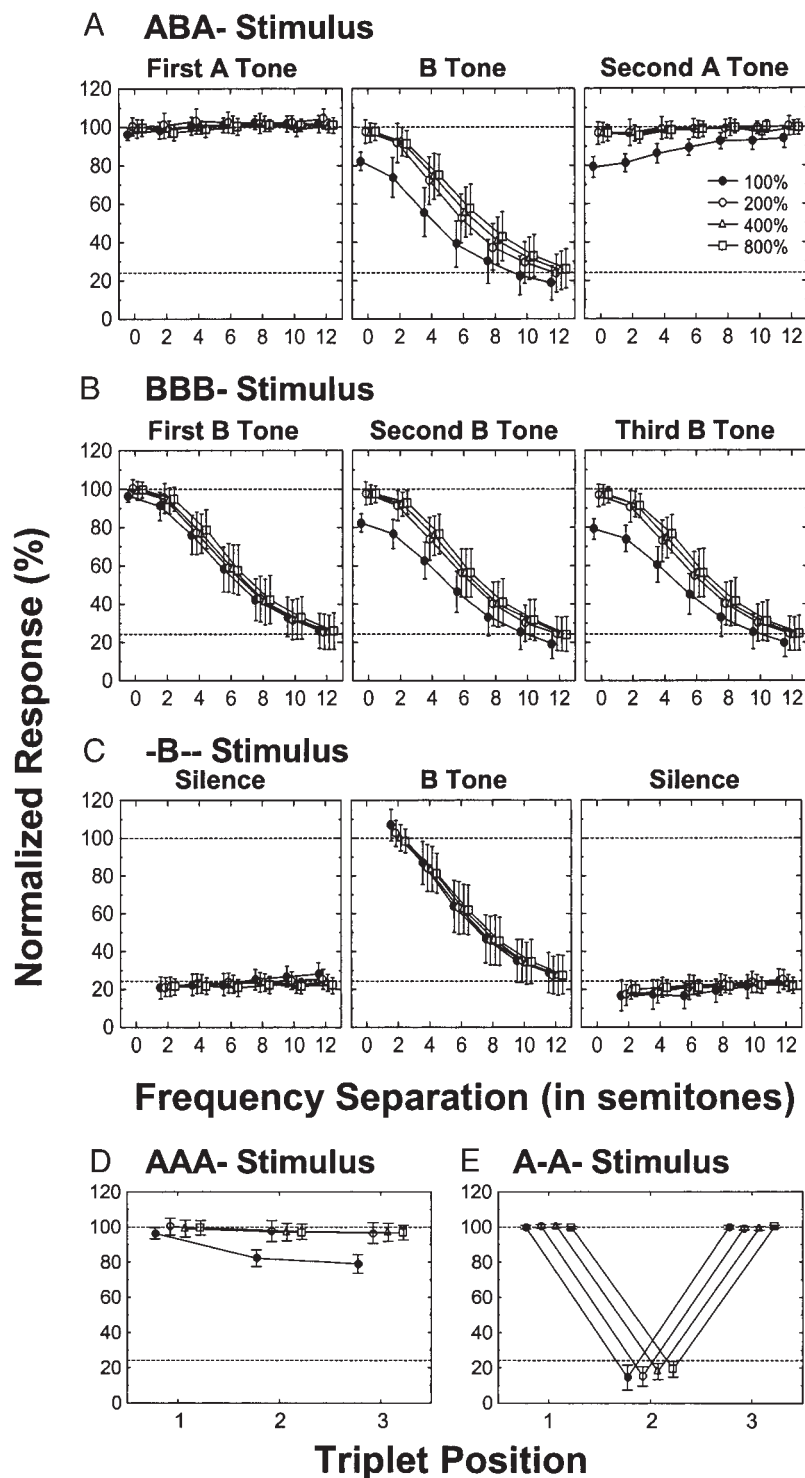


FIG. 4. Normalized responses to the ABA-, BBB-, -B-, AAA-, and A-A- stimulus sequences showing the effects of  $\Delta F$  and TRT, averaged over the 3 tone durations, on responses to the 3 tones in a triplet (or their equivalent silent intervals). Points depict the mean ( $\pm 2$  SE) normalized responses, averaged over 20 artifact-free responses to each stimulus sequence at each of 46 recording sites. Top dashed lines depict a normalized response that was 100% of the average response to isolated A tones; the bottom dashed lines depict the mean normalized spontaneous rate. A–C: responses to the ABA-, BBB-, and -B- stimulus sequences, respectively, showing separately the responses for different levels of  $\Delta F$  with TRT as the parameter (see legend in A). D–E: responses to the AAA- and A-A- stimulus sequences, respectively, for tones (or silent intervals) in triplet positions 1–3 with TRT as the parameter (see legend in A). Responses to the AAA- stimulus shown in D are also depicted in A and B as responses to a  $\Delta F$  of 0 semitones.

additional suppression of responses to the B tones in the second and third triplet positions. This suppression at the shortest TRT was larger at small  $\Delta F$  values, at which the frequency of the B tones was closer to the recording site's CF. Responses to the B tone in the -B- stimulus (Fig. 4C) also decreased with increases in  $\Delta F$ . Unlike responses to the middle B tones in the ABA- and BBB- stimulus sequences, however, responses to isolated B tones in the -B- stimulus presented at the shortest TRT were similar to those

at longer TRTs. Responses to all 3 A tones in the AAA- stimulus were generally similar and near the maximum normalized response rate at TRTs of 200, 400, and 800% (Fig. 4D). However, responses to the second and third A tones in the AAA- stimulus were suppressed at the shortest TRT. The responses to the isolated A tones in the A-A- stimulus (Fig. 4E) were similar at all TRTs, as would be expected for repetitions of isolated and temporally separated CF tones.

TABLE 2. Results from paired-sample *t*-tests in which the normalized responses to the B tone in the ABA- stimulus sequences were either not significantly different from, or were significantly lower than (\*), the normalized spontaneous rate

TD, ms	TRT, %	$\Delta F$ (semitones)	$t_{45}$	<i>P</i>	
25	100	8	2.4	0.020	
		10	1.0	0.342	
		12	-0.6	0.524	
		200	10	1.9	0.064
		12	0.1	0.950	
		12	3.1	0.003	
	400	12	1.6	0.118	
		800	10	3.3	0.002
		12	1.4	0.156	
		100	8	1.5	0.151
			10	-1.1	0.295
			12	-2.9	0.006
200	10		2.1	0.044	
	12		-0.7	0.498	
	400		10	2.6	0.014
	12	0.3	0.766		
	800	10	3.3	0.002	
		12	1.2	0.220	
100		8	1.7	0.089	
		10	-1.9	0.067	
		12	-4.7	<0.001*	
		200	10	1.7	0.088
	12		0.3	0.797	
	400		10	2.1	0.044
12			-1.2	0.248	
800			10	2.7	0.010
			12	-0.5	0.638

Statistical significance is based on a sequential Bonferroni correction for 84 multiple comparisons (Rice 1989). Responses to the B tone were significantly greater than the spontaneous rate in the remaining 57 comparisons ( $3.7 < t_{45} < 38.4$ ,  $P_s < 0.001$ ).

#### Effects of $\Delta F$ and TRT on differential responses to alternating CF and non-CF tones

One goal of the present study was to test the general hypothesis that the degree of overlap in excitation along a tonotopic gradient decreases under stimulus conditions known

TABLE 3. Description of 6 difference scores that were calculated (i) to compare normalized responses to A tones at the characteristic frequency (CF) and non-CF (B) tones in the ABA- stimulus (Difference 1), and (ii) to assess the effects of CF (A) tones on normalized responses to non-CF (B) tones, and the effects of non-CF (B) tones on normalized responses to CF (A) tones, in the ABA-, AAA-, A-A-, and -B- stimulus sequences (Differences 2–6)

Difference Score	Comparison/Effect	Calculation
Difference 1	Responses to middle B tones relative to the average response to surrounding A tones in the ABA- sequence	(B: ABA-) - (average A: ABA-)
Difference 2	Effects of surrounding A tones on responses to B tones in the middle position of the ABA- triplet relative to responses to isolated B tones	(B: ABA-) - (B: -B-)
Difference 3	Effects of B tones on responses to A tones in the first position (i.e., the first A tone) of the ABA- triplet relative to responses to isolated A tones	(A1: ABA-) - (A1: A-A-)
Difference 4	Effects of B tones on responses to A tones in the third position (i.e., the second A tone) of the ABA- triplet relative to responses to isolated A tones	(A2: ABA-) - (A2: A-A-)
Difference 5	Effects of surrounding A tones on middle B tones relative to the effects of middle B tones on the A tone in the first position of the ABA- triplet	Difference 2 - Difference 3
Difference 6	Effects of surrounding A tones on middle B tones relative to the effects of middle B tones on the A tone in the third position of the ABA- triplet	Difference 2 - Difference 4

In the calculations, the tones from the various triplets that entered into the calculations of the difference scores are denoted in parentheses as (tone: corresponding triplet).

to promote the perceptual streaming effect. We predicted that, at any particular site along the tonotopic distribution of CFs in field L2, the differences in responses to CF (A) and non-CF (B) tones in the ABA- stimulus should be larger under acoustic stimulus conditions that promote the perceptual streaming effect, that is, at larger  $\Delta F$  values and shorter TRTs (Fig. 1). To test these predictions, we examined differential responsiveness to the A and B tones in the ABA- stimulus sequence by computing a difference score (Difference 1, Table 3) as the difference between the normalized response to the B tone and the normalized responses to A tones, averaged over responses to the A tones in the first and third triplet positions. Recall that normalized responses are expressed as a percentage of the average response to an isolated CF (A) tone presented at the longest TRT. Therefore values of Difference 1 close to 0% indicate that the normalized responses to the A and B tones were similar, whereas values close to -100% correspond to the situation where there was a large difference between responses to the A and B tones, and the pattern of responses was dominated by excitatory responses to the A tone alone.

The effects of  $\Delta F$  and TRT were assessed in a  $7 (\Delta F) \times 4 (\text{TRT}) \times 3 (\text{TD})$  rmANOVA of the relative responses to A and B tones (Difference 1). The results of this analysis, which are reported in Table 4, revealed significant main effects of  $\Delta F$  ( $\eta^2 = 0.87$ ) and TRT ( $\eta^2 = 0.54$ ), and a significant  $\Delta F \times \text{TRT}$  interaction ( $\eta^2 = 0.10$ ). Although there was a significant main effect of TD ( $\eta^2 = 0.31$ ), the 2-way interactions of  $\Delta F \times \text{TD}$  and TRT  $\times$  TD, and the 3-way interaction of  $\Delta F \times \text{TRT} \times \text{TD}$ , were associated with relatively small effect sizes ( $\eta^2 \leq 0.06$ ; Table 4), indicating that differences in TD had little influence on the effects of  $\Delta F$  and TRT.

We tested the specific prediction that the differences between responses to A and B tones would increase as the  $\Delta F$  became larger by first examining the main effects of  $\Delta F$  separately for each level of TRT, and then, for significant differences, by computing planned linear contrasts comparing the linearly ordered effects of  $\Delta F$  (in semitones:  $0 < 2 < 4 \dots < 12$ ). As the results in Table 5 show, the main effect of  $\Delta F$  was significant at each level of TRT and explained approximately 85% of the variation in responses. Contrasts testing for a

TABLE 4. Results of a  $7 (\Delta F) \times 4 (TRT) \times 3 (TD)$  rmANOVA comparing the effects of frequency separation ( $\Delta F$ ), tone repetition time (TRT), and tone duration (TD) on the difference between normalized responses to the A and B tones in the ABA- stimulus (Difference 1, Table 3)

Effect	df	F	P	Partial $\eta^2$
$\Delta F$	6, 270	298.9	<0.001	0.87
TRT	3, 135	52.1	<0.001	0.54
TD	2, 90	20.0	<0.001	0.31
$\Delta F \times TRT$	18, 810	5.1	<0.001	0.10
$\Delta F \times TD$	12, 540	2.3	0.027	0.05
TRT $\times$ TD	6, 270	2.8	0.022	0.06
$\Delta F \times TRT \times TD$	36, 1620	0.7	0.770	0.02

linearly ordered relationship across  $\Delta F$  at each level of TRT were also significant at each level of TRT and explained more than 90% of the variance in responses. At intermediate  $\Delta F$  values of 4–8 semitones, the discharge rate in responses to the non-CF (B) tones was about 20–60% lower than responses to the surrounding CF (A) tones, whereas at the largest  $\Delta F$  values of 10–12 semitones, the rate responses to B tones was reduced by 70–80% relative to the average responses to the surrounding A tones (Fig. 5). Even at a  $\Delta F$  of 2 semitones, there was a 15% reduction in the rate response to B tones at the shortest TRT. The results for these analyses of the effects of  $\Delta F$  are

TABLE 5. Results of contrasts analyses of the difference between normalized responses to the A and B tones in the ABA- stimulus (Difference 1, Table 3) testing (a) the main effect of  $\Delta F$  separately at each level of TRT, (b) the linearly ordered effects of  $\Delta F$  (in semitones:  $0 < 2 < 4 \dots < 12$ ) at each level of TRT, (c) the main effect of TRT separately at each level of  $\Delta F$ , and (d) the linearly ordered effects of TRT ( $800\% < 400\% < 200\% < 100\%$ ) at each level of  $\Delta F$

Source	TRT, % or $\Delta F$ , semitones	F	P	Partial $\eta^2$
(a) Main effect of $\Delta F$	100%	227.3	<0.001	0.84
	200%	261.4	<0.001	0.85
	400%	262.1	<0.001	0.85
	800%	259.3	<0.001	0.85
(b) Linear contrast for $\Delta F$	100%	654.9	<0.001	0.94
	200%	700.9	<0.001	0.94
	400%	655.5	<0.001	0.94
	800%	570.3	<0.001	0.93
(c) Main effect of TRT	0	14.9	<0.001	0.25
	2	26.6	<0.001	0.37
	4	37.7	<0.001	0.46
	6	42.8	<0.001	0.49
	8	17.3	<0.001	0.28
	10	13.8	<0.001	0.23
(d) Linear contrast for TRT	12	4.7	0.004	0.09
	0	18.8	<0.001	0.29
	2	33.8	<0.001	0.43
	4	58.1	<0.001	0.56
	6	74.2	<0.001	0.62
	8	35.5	<0.001	0.44
	10	19.7	<0.001	0.30
	12	9.5	0.004	0.17

Overall ANOVA results are reported in Table 4. Degrees of freedom: main effect of  $\Delta F$  (6, 270), main effect of TRT (3, 135), linear contrasts (1, 45).

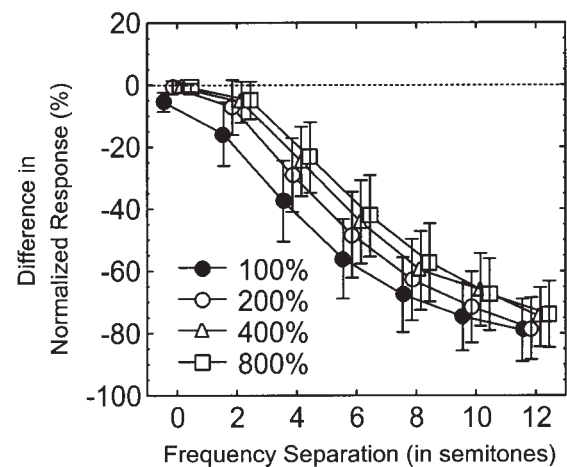


FIG. 5. Differences in normalized responses to the A and B tones in the ABA- stimulus as a function of  $\Delta F$  and TRT. Points depict mean ( $\pm 2$  SE) differences in normalized responses to A and B tones (Difference 1, Table 3), averaged over 20 artifact-free responses to each stimulus at each of 46 recording sites, and averaged over all 3 tone durations.

important because they demonstrate that increases in the  $\Delta F$  between the CF (A) and non-CF (B) tones in an alternating series resulted in greater differences in the rate responses to CF and non-CF tones, as expected as a result of the frequency selectivity of neurons in field L2.

To examine the effects of TRT more closely, we tested the specific prediction that the differences between responses to A and B tones would increase as the TRT became shorter, using contrast analyses parallel to those described above for the analysis of  $\Delta F$ . The main effect of TRT was significant at each level of  $\Delta F$  (Table 5). The effects of differences in TRT were greatest at  $\Delta F$  values of 2–6 semitones, at which 37–49% of the variance was explained, compared with 23–28% of the variance explained at  $\Delta F$  values of 0, 8, and 10 semitones, and compared with 9% explained at 12 semitones (Table 5). Linear contrasts comparing the ordered relationships among the levels of TRT ( $800\% < 400\% < 200\% < 100\%$ ) at each level of  $\Delta F$  revealed the predicted pattern of significantly larger differences in responses at smaller TRTs (Table 5). For  $\Delta F$  values between 2 and 8 semitones, the linear contrasts explained 43–62% of the variance in responses. At these levels of  $\Delta F$ , there was an additional 10–15% reduction in the rate response to non-CF (B) tones, relative to CF-tone responses, in ABA- triplets that occurred at the shortest TRT (100%) compared with the longest TRT (800%) (Fig. 5). At  $\Delta F$  values of 10–12 semitones, the linear contrasts of TRT explained 30% or less of the variance, and the corresponding differences in the rate responses to A and B tones between the shortest and longest TRTs were smaller (5–7%). The important implications of these results are the following. First, shorter TRTs resulted in larger differences in responses to CF (A) and non-CF (B) tones compared with longer TRTs. Second, the effects of TRT on the relative responses to CF and non-CF tones depended on the level of  $\Delta F$ ; the effects of TRT were generally greater when  $\Delta F$  was between 2 and 8 semitones compared with larger  $\Delta F$  values.

#### Effects of $\Delta F$ and TRT on forward masking by alternating CF and non-CF tones

The second major goal of this study was to test the hypothesis of Fishman et al. (2001), which proposes that physiolog-

ical forward masking results in the *differential* suppression of responses to CF and non-CF tones when these are arranged in an alternating tone sequence. If this hypothesis is true, then we should expect relatively greater forward suppression by CF tones, and this suppression should be more pronounced at shorter TRTs, especially when the A and B tones are similar in frequency (Brosch and Schreiner 1997; Calford and Semple 1995). We tested this prediction by determining the *relative* masking effects of CF (A) tones and non-CF (B) tones when these were arranged in the alternating ABA- stimulus sequence as functions of  $\Delta F$  and TRT. We assessed these effects by computing a number of additional difference scores (see Table 3) between responses to A and B tones in various stimulus sequences.

First, we asked, what were the masking effects of surrounding CF (A) tones on the non-CF (B) tone in the middle position of the ABA- triplet relative to an isolated non-CF (B) tone condition, in which the surrounding CF tones were absent and no masking was expected? To address this question, we computed the difference between responses to isolated B tones in the -B- stimulus and responses to middle B tones in the ABA- stimulus (*Difference 2*, Table 3, Fig. 6A). Because the  $\Delta F$ -dependent changes in responses to the -B- stimulus result from the neuron's tuning characteristics alone, any differences between responses to B tones in the ABA- and -B- stimuli thus

reflect the effects of CF (A) tones on non-CF (B) tones in the context of the ABA- stimulus. As Fig. 6A shows, responses to the non-CF (B) tones in the ABA- triplet were suppressed relative to the condition in which the surrounding CF (A) tones were absent. This suppression was most pronounced at the shortest TRT (100% of TD), at which the degree of suppression was also related to the magnitude of  $\Delta F$ . At the shortest TRT and a  $\Delta F$  of 2 semitones, the normalized discharge rate of responses to B tones in the ABA- triplet was 33% lower when compared with isolated B-tone responses. The greater suppression of B tones in the ABA- triplet decreased as  $\Delta F$  increased toward 12 semitones, at which there was a difference of only 10% in the normalized rates. In contrast to these suppressive effects at the shortest TRT, suppression of responses to a non-CF (B) tone surrounded by CF (A) tones was much less pronounced at the longer TRTs, although there was a slight trend toward some suppressive effect at longer TRTs and intermediate  $\Delta F$  values (2–6 semitones). These results for *Difference 2* confirm that non-CF (B) tones in the middle position of an ABA- triplet were additionally suppressed when presented in the context of surrounding CF (A) tones relative to a condition in which surrounding CF tones were absent, especially at the shortest TRT and when the CF and non-CF tones were more similar in frequency.

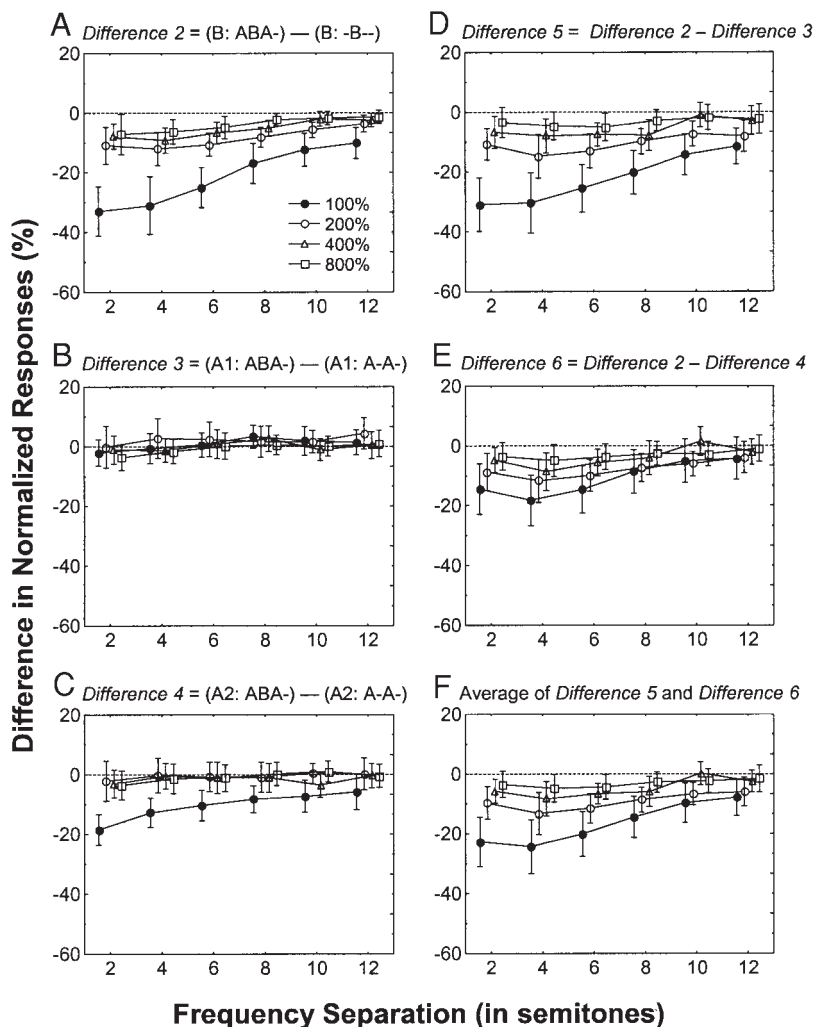


FIG. 6. Relative forward masking effects of A tones at the CF and of non-CF (B) tones as a function of  $\Delta F$  and TRT. In A–F, points depict mean ( $\pm 2$  SE) difference scores showing the differences in normalized responses to the CF tone and non-CF tones between various stimulus types as functions of  $\Delta F$  and TRT (see legend in A), averaged over 20 artifact-free responses to each stimulus at each of 46 recording sites, and averaged over all 3 tone durations. See Table 3 for a full description of the calculation of difference scores. Note, in D–F, that negative values indicate that the effects of CF tones on non-CF tones were greater than the effects of non-CF tones on CF tones.

We next asked, what were the effects on CF (A) tones when these were presented in the context of non-CF (B) tones (ABA-) compared with an isolated CF tone condition (A-A-)? To assess these effects required that we compare separately responses to A tones in the first and third positions of the ABA-triplet to the equivalent A tones in the A-A- control sequence (*Difference 3* and *Difference 4*, respectively; Table 3, Fig. 6, *B* and *C*). At TRTs of 200, 400, and 800%, responses to the CF (A) tones in both the first and third positions of the ABA-triplet were largely unaffected by the presence of non-CF (B) tones at all levels of  $\Delta F$  (Figs. 6*B,C*). This is in contrast to the results for the shortest TRT. At the TRT of 100%, the presence of non-CF (B) tones had a  $\Delta F$ -dependent suppressive effect on responses to CF (A) tones in the third triplet position (Fig. 6*C*), but not on tones in the first triplet position (Fig. 6*B*). Suppression of responses to the A tone in the third triplet positions was greatest when the frequency of the non-CF (B) tone was similar to the CF, and decreased as  $\Delta F$  increased (Fig. 6*C*).

Finally, we asked, what were the *relative* masking effects of CF tones and non-CF tones when these were arranged in the alternating ABA- stimulus sequence? The relative suppressive effects of CF (A) tones and non-CF (B) tones in the ABA-stimulus correspond to the differences between *Difference 2* (effects of A tones on B tones) and *Differences 3* and *4* (effects of B tones on A tones). To compare the relative suppressive effects of A and B tones when these occurred in the ABA-stimulus arrangement, we therefore computed 2 additional difference scores: *Difference 5* was the difference between *Difference 2* and *Difference 3*, and *Difference 6* was the difference between *Difference 2* and *Difference 4* (Table 3). Note that negative values of *Difference 5* and *Difference 6* indicate that the suppressive effects of CF (A) tones on non-CF (B) tones were relatively greater than the effects of non-CF (B) tones on CF (A) tones, whereas positive values indicate the opposite. The values for *Difference 5* and *Difference 6* are depicted in Fig. 6, *D* and *E*, respectively. Figure 6*F* depicts the mean difference scores averaged over the values of *Difference 5* and *Difference 6*.

Two important trends are evident from an examination of the average of *Difference 5* and *Difference 6* (Fig. 6*F*). First, the values for the difference scores are generally either close to zero or negative. Two-tailed, paired-sample *t*-tests, followed by a sequential Bonferroni correction (Rice 1989) for 72 multiple comparisons ( $6 \Delta F \times 4 \text{ TRT} \times 3 \text{ TD}$ ), revealed that in 26 of the 72 possible combinations of  $\Delta F$ , TRT, and TD, the negative values for the average of *Difference 5* and *Difference 6* were significantly below the null expectation of zero difference ( $3.6 < t_{S_{45}} < 8.5$ ,  $P_s < 0.001$ ). In 23 of the 26 comparisons that were significantly different, the  $\Delta F$  was 8 semitones or less, and the TRT was 100 or 200% of the TD. In no instances were the average values of *Differences 5* and *6* significantly above the null expectation of zero difference. These results confirm that under some stimulus conditions the relative suppressive effects of CF (A) tones were significantly greater than the effects of non-CF (B) tones. Second, the relatively greater suppression of non-CF (B) tones by CF (A) tones varied as a function of  $\Delta F$  and TRT. To statistically examine the effects of  $\Delta F$  and TRT, we analyzed the average of *Difference 5* and *Difference 6* in a  $6 (\Delta F) \times 4 (\text{TRT}) \times 3 (\text{TD})$  rmANOVA (Table 6). There were significant main effects of  $\Delta F$  ( $\eta^2 = 0.35$ ) and TRT ( $\eta^2 = 0.40$ ), and the  $\Delta F \times$

TABLE 6. Results of a  $6 (\Delta F) \times 4 (\text{TRT}) \times 3 (\text{TD})$  rmANOVA comparing the effects of frequency separation ( $\Delta F$ ), tone repetition time (TRT), and tone duration (TD) on the relative suppressive effects of CF (A) tones and non-CF (B) tones (average of *Difference 5* and *Difference 6*, Table 3)

Effect	df	F	P	Partial $\eta^2$
$\Delta F$	5, 225	24.3	<0.001	0.35
TRT = 100%	1, 45	36.4	<0.001	0.45
TRT = 200%	1, 45	13.3	0.001	0.23
TRT = 400%	1, 45	25.0	<0.001	0.36
TRT = 800%	1, 45	4.4	0.041	0.09
TRT	3, 135	29.5	<0.001	0.40
$\Delta F = 2$	1, 45	59.0	<0.001	0.57
$\Delta F = 4$	1, 45	57.9	<0.001	0.56
$\Delta F = 6$	1, 45	48.2	<0.001	0.52
$\Delta F = 8$	1, 45	26.7	<0.001	0.37
$\Delta F = 10$	1, 45	15.4	<0.001	0.26
$\Delta F = 12$	1, 45	9.9	0.003	0.18
TD	2, 90	1.9	0.154	0.04
$\Delta F \times \text{TRT}$	15, 675	5.2	<0.001	0.10
$\Delta F \times \text{TD}$	10, 450	1.7	0.119	0.04
TRT $\times$ TD	6, 270	1.0	0.436	0.02
$\Delta F \times \text{TRT} \times \text{TD}$	30, 1350	1.0	0.475	0.02

Under the main effects of  $\Delta F$  and TRT are shown the results of linear contrasts of each main effect computed separately at each level of the other main effect.

TRT interaction ( $\eta^2 = 0.10$ ) was significant. No other effects in the main ANOVA model were significant ( $\eta^2 \leq 0.05$ ).

Recall that our specific predictions were that forward masking effects should be more pronounced at shorter TRTs compared with longer TRTs, at smaller  $\Delta F$  levels compared with larger  $\Delta F$  levels, and that the effects of TRT should be more pronounced at smaller  $\Delta F$  values compared with larger  $\Delta F$  values (Brosch and Schreiner 1997; Calford and Semple 1995). To test these specific predictions, we computed 2 sets of planned linear contrasts (Table 6). The first set of contrasts compared separately at each level of TRT the linearly ordered effects of  $\Delta F$  (in semitones:  $2 > 4 > \dots > 12$ ), and the second set compared separately at each level of  $\Delta F$  the linearly ordered effects of TRT (100% > 200% > 400% > 800%). Results from the first set of contrasts indicated that the linearly ordered effects of  $\Delta F$  on the greater relative masking by CF (A) tones were significant at all levels of TRT. The linear effects of  $\Delta F$  were most pronounced at the shortest TRT ( $\eta^2 = 0.45$ ), and least pronounced at the longest TRT ( $\eta^2 = 0.09$ ). The effects at TRTs of 200 and 400% were intermediate. The second set of contrasts, which examined the linearly ordered effects of TRT, revealed significant effects at all levels of  $\Delta F$ . The magnitude of the linearly ordered effects among the levels of TRT decreased as  $\Delta F$  increased from 2 to 12 semitones. At a  $\Delta F$  of 2 semitones, the linearly ordered effects of TRT explained 57% of the variation in the relatively greater suppressive effects of CF (A) tones on non-CF (B) tones, whereas at  $\Delta F$  levels of 10–12 semitones, 26% or less of the variation was explained by the linearly ordered effects of TRT.

The data depicted in Fig. 6*F*, and the statistical results reported in Table 6, can be summarized as follows. The relatively greater forward suppression caused by CF (A) tones, compared with non-CF (B) tones, was most pronounced at the stimulus combinations that included short TRTs (e.g., 100%)

and small  $\Delta F$  values (e.g., <6–8 semitones). At the shortest TRT and  $\Delta F$  values of 2–6 semitones, the relatively greater forward suppression of non-CF (B) tones caused by surrounding CF (A) tones was equivalent to a 20–25% greater reduction in the discharge rate. These results are important because they conform to the pattern of responses expected if forward masking played a role in determining the relative magnitude of responses to the A and B tones when these were embedded in the context of an alternating ABA- sequence.

## DISCUSSION

Hearing involves grouping the sounds that emanate from a common source, and segregating these from the sounds arising from other sources, to form a perceptual representation of the acoustic scene (Bregman 1990; Yost 1991). An ability to perceptually segregate overlapping and intervening sequences of sounds from multiple sources is fundamental in the perception of speech and music by humans (Bregman 1990), as well as in the perception of acoustic communication signals by other animals, like songbirds (Feng and Ratnam 2000; Hulse 2002). Sequential auditory stream segregation has been studied at a perceptual level in humans (reviewed in Bregman 1990; Moore and Gockel 2002), Japanese macaques (Izumi 2001, 2002), starlings (MacDougall-Shackleton et al. 1998), and goldfish (Fay 1998, 2000) using repeated series of interleaved tones differing in frequency. According to Bregman (1990), sequential auditory stream segregation can be attributed largely to the operation of low-level, preattentive auditory processes. Here, we examined neural responses in the tonotopically organized auditory forebrain of the starling to investigate the potential roles of frequency selectivity and forward masking in primitive sequential stream segregation using the same stimulus paradigm that has been used in previous psychophysical studies of humans and starlings.

### *Effects of $\Delta F$ and TRT*

Frequency differences ( $\Delta F$ ) between tones and the tone repetition time (TRT) have strong influences on the perceptual segregation of interleaved tones into separate streams, with larger  $\Delta F$  values and shorter TRTs promoting the streaming effect (Bregman 1990; Fig. 1). Frequency selectivity and forward masking are believed to play roles in the perceptual segregation of sequences of alternating tones by promoting the spatial separation of neural responses in tonotopic space (Beauvois and Meddis 1991, 1996; Fishman et al. 2001; Hartmann and Johnson 1991; McCabe and Denham 1997). Our goals were to determine whether larger  $\Delta F$  values and shorter TRTs result in larger differential responses to CF (A) and non-CF (B) tones, and whether differential forward masking between CF and non-CF tones could account for the effects of TRT on neural responses.

In starlings, the 10-dB bandwidths of excitatory tuning curves in the auditory periphery and in the tonotopically organized field L2 are similar, the latter depending somewhat less on level (Klump et al. 2000; Nieder and Klump 1999). We therefore predicted that the overlap in neural excitation in field L2 elicited by CF (A) and non-CF (B) tones in an ABA-stimulus would decrease as  $\Delta F$  increased as a result of frequency selectivity. Over the range of stimulus properties used

in this study,  $\Delta F$  consistently had large effects on the differential responses to CF (A) and non-CF (B) tones. A difference in responses to CF and non-CF tones was observed at the smallest  $\Delta F$  used in this study (2 semitones), and the differences in responses increased as the  $\Delta F$  between A and B tones increased up to the largest  $\Delta F$  values (10–12 semitones), at which responses were dominated by excitatory responses to the CF (A) tone alone. In recordings of neural ensembles from macaque AI, Fishman et al. (2001) found a similar effect of  $\Delta F$  on responses to A and B tones presented in an alternating ABAB sequence when the A tones were presented at the site's BF and B tones were situated at  $\Delta F$  levels that were 10–50% (1.65–7.0 semitones) away from the BF. Our results, and those from macaque AI, are thus consistent with a major prediction of hypotheses proposing that frequency selectivity in the auditory system plays a role as a low-level, preattentive process in sequential stream segregation, at least for sequences of pure tones (Beauvois and Meddis 1991, 1996; Fishman et al. 2001; Hartmann and Johnson 1991; McCabe and Denham 1997).

We also predicted that if physiological forward masking plays a role in determining the responses to alternating CF and non-CF tones, as proposed by Fishman et al. (2001), then the differences in responses of starling forebrain neurons to CF (A) and non-CF (B) tones would increase as the TRT decreased, and that the effects of TRT would be larger at smaller  $\Delta F$  values (Brosch and Schreiner 1997; Calford and Semple 1995). Our results show that differential neural responses to A and B tones increased as the TRT decreased, and that the magnitude of the effects of TRT were larger at  $\Delta F$  values between 2 and 8 semitones compared with larger  $\Delta F$  values. The smaller effects of TRT at  $\Delta F$  values of 10–12 semitones may be explained by the fact that B tone responses were already near the spontaneous rate (Table 2). Fishman et al. (2001) reported somewhat similar results for neural ensemble responses in macaque AI to ABAB tone sequences. Using 25-ms tones, they showed that the ratio of non-BF-tone responses to BF-tone responses was near 1.0 at the 2 slowest tone rates used (5 and 10 Hz, or TRT = 800 and 400% of TD, respectively), and decreased as the tone rate was increased to 20 and 40 Hz (TRT = 200 and 100% of TD, respectively). Fishman et al. (2001) attributed the effects of faster tone rates to differential physiological forward masking between BF and non-BF tones. However, Fishman et al. (2001) analyzed the effects of TRT only after collapsing over the different  $\Delta F$  values used in their study, thereby making it impossible to assess whether the effects of TRT depended on the magnitude of  $\Delta F$ , as one would predict for physiological forward masking (Brosch and Schreiner 1997; Calford and Semple 1995). Our study thus extends that by Fishman et al. (2001) by showing that the effects of differences in TRT varied reliably as a function of  $\Delta F$ .

The  $\Delta F$ -dependent effects of TRT observed in the present study can be explained by the greater forward masking of non-CF (B) tones by CF (A) tones in the ABA- stimulus sequence. Both psychophysical and physiological studies of forward masking demonstrate that masking is often more pronounced when the masker and signal are similar in frequency and when the signal occurs at shorter masker-signal delays (Brosch and Schreiner 1997; Calford and Semple 1995; Fastl 1979; Jesteadt et al. 1982; Moore and Glasberg 1983). In our study, the relatively greater suppression of non-CF (B) tones by CF (A) tones was more pronounced when the A and

B tones were more similar in frequency, and at shorter TRTs, at which the intervals between consecutive tones was reduced. These findings are consistent with the hypothesis proposed by Fishman et al. (2001) that differential physiological forward masking plays a role in determining neural responses at a particular location along a tonotopic map to CF (A) and non-CF (B) tones in an interleaved sequence of alternating tones. Thus in addition to frequency selectivity, forward masking is likely an additional low-level, preattentive auditory process that contributes to sequential stream segregation.

#### *Relating multiunit responses to excitation in tonotopic space*

The differential responses to A and B tones reported here for a particular tonotopic site reflect a narrowing of the neurons' receptive fields that would result in a larger spatial separation of excitation within the population of neurons constituting the tonotopic gradient in field L2. In their study of sequential stream segregation in macaque AI, Fishman et al. (2001) developed a conceptual model to illustrate the decrease in the overlap in excitation along a tonotopic gradient that results from increasing  $\Delta F$  and decreasing TRT. As we show in Fig. 7, our results from the tonotopically organized field L2 in starlings are consistent with their model. The model is based on the level of excitation elicited by repeated sequences of A and B tones in 3 hypothetical auditory filters, one centered around the frequency of the A tone, one centered around the frequency of the B tone, and one with a center frequency intermediate between that of the A and B tones (Fishman et al. 2001). In Fig. 7A, we show a schematic diagram illustrating the tonotopic organization of field L2 in which we highlight 3 such hypothetical auditory filters. In Fig. 7B, we depict a subset of our results for the 100-ms TD in a format based on the model of Fishman et al. (2001). We show the magnitude of normalized responses to the A and B tones in an ABA- stimulus at all 3 tonotopic positions ("A", "B", and "X") based on the normalized responses to the ABA- and BBB- stimuli (see Fig. 7 legend).

The data presented in Fig. 7B illustrate 2 important points predicted by the model of Fishman et al. (2001). First, for both the slow and the fast tone rates, the spatial separation in the level of excitation elicited by the A and B tones in a repeated ABA- triplet increases as  $\Delta F$  increases. This is illustrated for  $\Delta F$  values of 2, 6, and 12 semitones by the *differences* that occur as a function of  $\Delta F$  between responses to CF and non-CF tones at the "A" and "B" sites, and by the decrease in response at a tonotopic site ("X") midway between the "A" and "B" sites (black vs. striped bars). The solid black arrows and the open arrows in Fig. 7B depict these changes for the slower and faster tone rates, respectively. This result illustrates how frequency selectivity can promote the spatial separation of neural responses in tonotopic space. Second, at each level of  $\Delta F$  depicted in Fig. 7B, the spatial separation in the level of excitation increases as the tone rate increases because of the greater forward masking of non-CF tones by CF tones at faster tone rates. In Fig. 7B, the differences between the fast and slow rates in the positions of the solid black arrows and the open arrows, relative to responses to a CF tone at the same rate, depict the additional suppression of responses to non-CF tones that results from increasing the tone rate. Notice that the magnitude of the additional increase in the separation of

excitation that results from increasing the tone rate is larger at  $\Delta F$  values of 2 and 6 semitones compared with a  $\Delta F$  of 12 semitones. This again illustrates the  $\Delta F$ -dependent effects of TRT discussed earlier. Thus the data presented in Fig. 7B provide physiological evidence for the conceptual model proposed by Fishman et al. (2001), and illustrate that increases in  $\Delta F$  and decreases in TRT should lead to increased spatial separation in the positions along a tonotopic gradient that encode the A and B tones in an ABA- stimulus. Our data confirm that frequency selectivity and forward masking are 2 preattentive processes that promote the spatial separation of neural responses in tonotopic space to sequences of alternating pure tones depending on  $\Delta F$  and TRT. Therefore we concur with Fishman et al. (2001) in that frequency selectivity and forward masking can play a role in the perceptual segregation of interleaved tone sequences, and thus are likely mechanisms involved in sequential auditory stream segregation.

#### *Comparisons between physiology and perception*

Earlier work by van Noorden (1975) based on the ABA-stimulus paradigm showed that the effects of  $\Delta F$  and TRT on the perception of a repeating ABA- sequence as either one coherent stream or 2 segregated streams depended on the instructions given to the listener. Van Noorden (1975) defined 2 perceptual boundaries—the "fission boundary" and the "temporal coherence boundary"—that separate 3 perceptual fields along combinations of  $\Delta F$  (measured in semitones) and TRT (measured in milliseconds). The fission boundary represents the combinations of  $\Delta F$  and TRT below which one coherent stream of alternating tones was always heard, even when listeners were instructed to try to hear 2 segregated streams (Fig. 8, white circles). Thus the fission boundary represents the combinations of  $\Delta F$  and TRT at which the segregation of A and B tones into separate streams first *becomes possible*. The  $\Delta F$  required to cross the fission boundary varies only slightly with increasing TRTs. It has been suggested that the fission boundary is directly related to the bandwidth of auditory filters (Beauvois and Meddis 1991, 1996; Rose and Moore 1997). The temporal coherence boundary, on the other hand, represents the combinations of  $\Delta F$  and TRT above which only 2 segregated streams could be heard, even when listeners were instructed to try to hear one temporally coherent stream of alternating tones (Fig. 8, black circles). Thus the temporal coherence boundary reflects those combinations of  $\Delta F$  and TRT at which stream segregation *becomes obligatory*. In contrast to the fission boundary, an interaction between  $\Delta F$  and TRT has a pronounced effect on the temporal coherence boundary: the level of  $\Delta F$  required to cross the temporal coherence boundary lies only slightly above the fission boundary at short TRTs, but increases rather sharply with increases in TRT. The area between the fission boundary and the temporal coherence boundary represents an ambiguous area where listeners report hearing either 2 isochronous rhythms (segregation) or a single integrated sequence of galloping tones (coherence), and are able to alternate between the 2 percepts.

One important question that can be addressed with our data is whether, under passive listening conditions, the observed differences in A-tone and B-tone responses ascribed to the effects of  $\Delta F$  and TRT represent potential physiological correlates of either the fission boundary or the temporal coherence

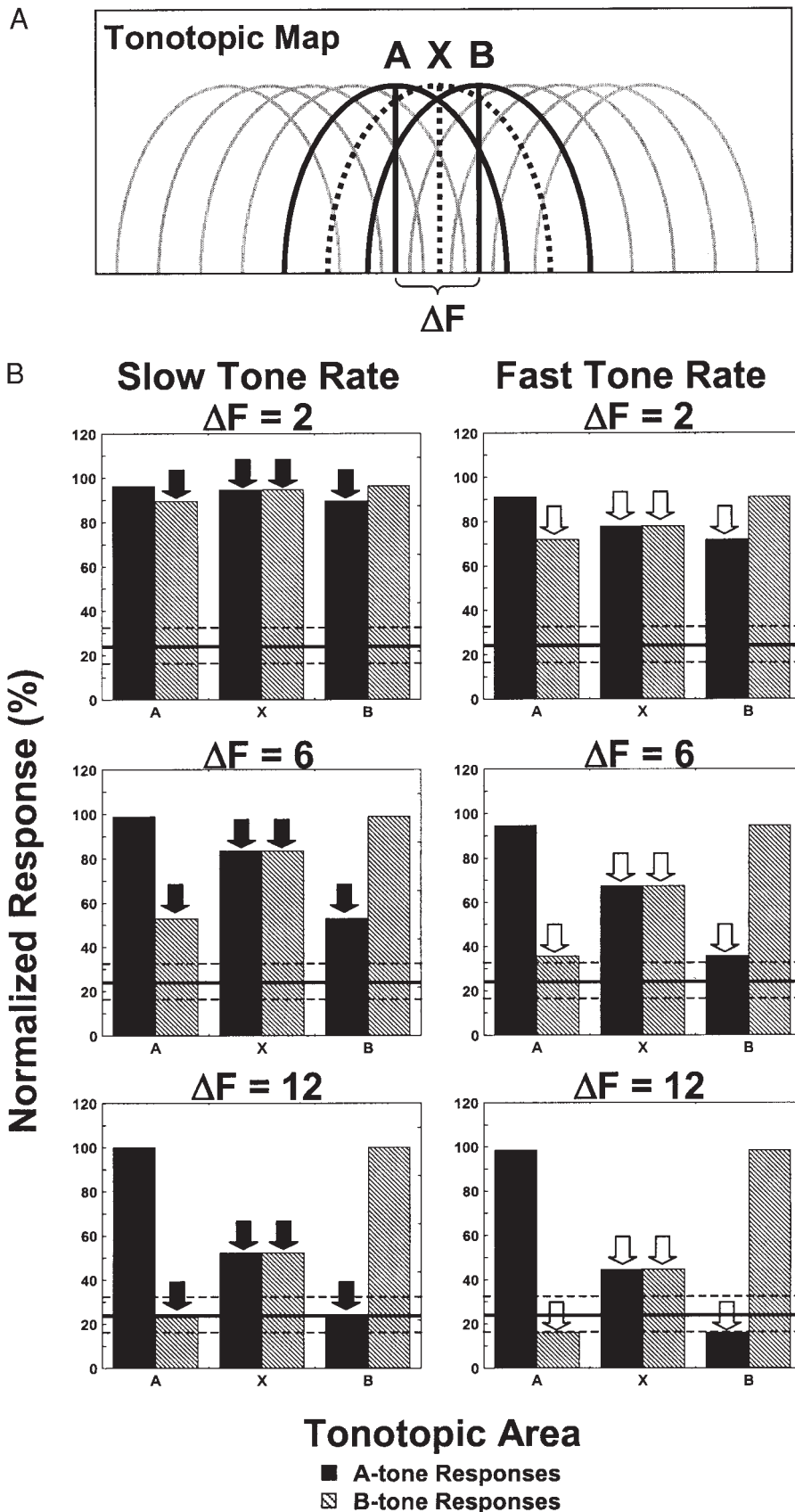


FIG. 7. Relationship between multiunit responses and excitation in tonotopic space based on the conceptual model proposed by Fishman et al. (2001). *A*: schematic diagram illustrating several auditory filters along a tonotopic map. Three hypothetical auditory filters—"A", "B", and "X"—are highlighted. Neurons at the "A" and "B" sites have CFs that correspond to the frequencies of the A and B tones, respectively, in an ABA- stimulus, and the CFs at these sites are separated by a level of  $\Delta F$  corresponding to the  $\Delta F$  in an ABA- stimulus. CFs of neurons at tonotopic position "X" are midway between those of sites "A" and "B". *B*: bar graphs depicting the responses to CF and non-CF tones at positions "A", "B", and "X" along a tonotopic map at 2 levels of tone repetition rate (slow: TRT = 800%; fast: TRT = 100%) and 3 levels of  $\Delta F$ . Responses at areas "A" and "B" are based on the normalized responses to CF and non-CF tones at the specified level of  $\Delta F$  and TRT for the 100-ms tone duration. Responses to non-CF tones at position "X" are based on the normalized responses to 100-ms non-CF tones in the BBB- stimulus. We interpolated from the BBB- response data to determine responses at  $\Delta F$  values of 1 and 3 semitones, which are intermediate between  $\Delta F$  values of 2 and 6 semitones, respectively. Solid and dashed horizontal lines represent the mean  $\pm$  SD of the normalized spontaneous rate. For the purpose of illustrating the differences in responses to A and B tones across a population of forebrain neurons, the following simplifying assumptions were made: 1) the bandwidths of the auditory filters in field L2 are symmetrical on a semitone scale around the CF and have a relatively constant bandwidth; 2) responses to a CF tone at both the "A" and the "B" tonotopic sites are equal in magnitude to the normalized response to CF (A) tones in the ABA- stimulus, averaged over CF tones in the first and third triplet positions when the non-CF tone in the middle triplet position was at the specified level of  $\Delta F$ ; 3) responses to a non-CF tone at both the "A" and the "B" tonotopic sites are equal in magnitude to the normalized response to a non-CF (B) tone in the ABA- stimulus at the specified level of  $\Delta F$ ; 4) responses at the "X" tonotopic site are equal in magnitude to the average normalized responses to non-CF (B) tones in the BBB- stimulus at one half the specified level of  $\Delta F$ ; 5) responses to a triplet of non-CF tones with the same frequency, as presented in the BBB- stimulus, would be the same as responses to a BBB- triplet in which the frequency of the non-CF tone in the middle position and those in the first and third positions were different, but were equally spaced (in semitones) away from the site's CF.

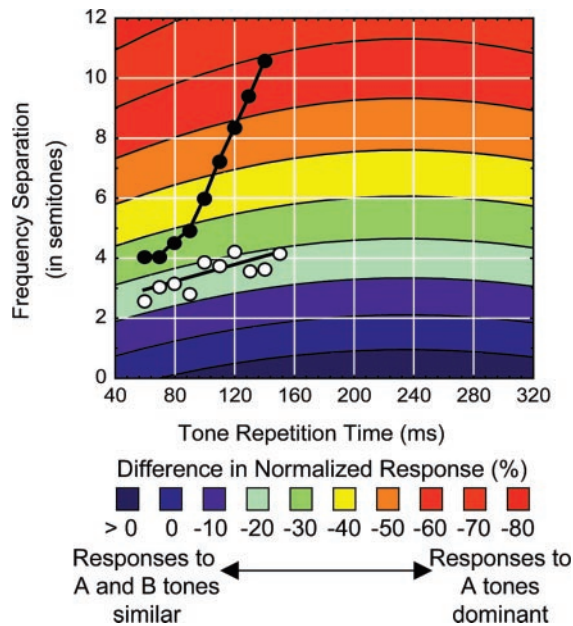


FIG. 8. Comparisons between starling physiology and human perception. Relationships between differences in neural responses to A and B tones as functions of  $\Delta F$  and TRT and the perceptual fission and temporal coherence boundaries in humans. Isoresponse contours for the 40-ms tone duration (TD) are plotted showing the patterns of changes in the differences between responses to A and B tones (Difference 1, Table 3) as a function of both  $\Delta F$  and TRT. Contour lines delineate the area on the  $\Delta F \times$  TRT surface in which the differences in responses to A and B tones were the same (see legend color bar). Data from van Noorden's (1975; Fig. 2.7) experiment using the ABA-stimulus paradigm are plotted showing the fission boundary (white circles) and the temporal coherence boundary (black circles). Tone duration in van Noorden's study was 40 ms and the TRT values were chosen in 10-ms intervals between 60 and 150 ms.

boundary. In Fig. 8 we plot isoresponse contours of the differences in responses to A and B tones (Difference 1, Table 3) for the 40-ms TD in a 2-dimensional space bounded by the values of  $\Delta F$  and TRT used in the present study. These contours delineate the areas in this 2-dimensional space in which the differences in responses to the A and B tones were similar. As is clear in Fig. 8, the slopes of the isoresponse contour lines more closely resemble the pattern of changes in responses that might be expected to occur if differential responses to the A and B tones represented a physiological correlate of the perceptual fission boundary. This finding suggests that the level of spatial separation between excitation patterns could be relatively constant at the levels of  $\Delta F$  and TRT at which stream segregation first becomes possible. If changes in the relative neural responses to A and B tones represented a physiological correlate of the perceptual temporal coherence boundary, we would have expected much steeper rises in isoresponse contours with increases in TRT, similar to the increase of the perceptual temporal coherence boundary shown in Fig. 8 (black circles).

Previous work has shown that, at a given TRT and TD, increases in  $\Delta F$  above the fission boundary result in a decrease in the proportion of listener responses, indicating the perception of a single, coherent stream of alternating tones. The effects of increasing  $\Delta F$  on the relative neural responses to A and B tones in the starling forebrain parallel the effects of increasing  $\Delta F$  on the ability of both starlings and humans to perceive an alternating sequence of A and B tones as a single,

coherent stream. This point is illustrated in Fig. 9, in which we compare differences in neural responses to A and B tones as a function of  $\Delta F$  with behavioral data from starlings (MacDougall-Shackleton et al. 1998) and humans (Anstis and Saida 1985). The probability of starlings and humans hearing one coherent stream decreased as  $\Delta F$  increased, and this decrease parallels the changes in the relative neural responses to A and B tones found in the present study. Although this comparison reveals only correlated effects in physiology and perception, the data in Fig. 9 are consistent with the hypothesis that frequency selectivity is one potential mechanism underlying the effects of  $\Delta F$  on the perception of temporally coherent or segregated streams of pure tones at a given TRT and TD.

In conclusion, our results from the starling forebrain, and those from macaque AI (Fishman and Steinschneider 2003; Fishman et al. 2001; Michey et al. 2003) provide some support for the data-driven view of primitive auditory stream segregation proposed by Bregman (1990). This view suggests that low-level, preattentive processing in the auditory system operates in the segregation of interleaved sequences of sounds into separate auditory streams in the absence of higher-level attentional and cognitive inputs. Data from starlings and macaques indicate that frequency selectivity and forward masking are likely to be 2 such low-level processes involved in sequential auditory stream segregation.

However, it is also clear from more recent human studies that frequency selectivity and forward masking by themselves are not sufficient to account for the range of perceptual phe-

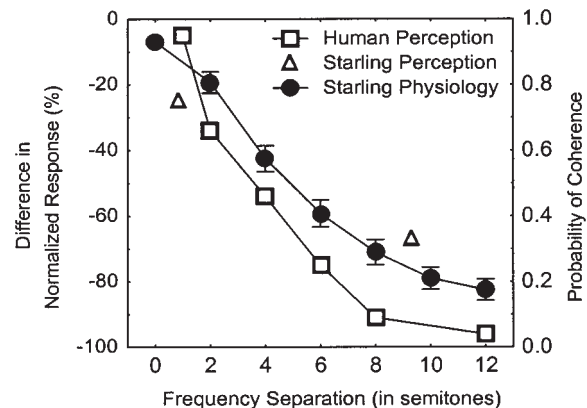


FIG. 9. Comparisons between starling physiology and human and starling perception. Differential neural responses of starling auditory forebrain neurons to A and B tones in the ABA-stimulus (filled circles) as functions of  $\Delta F$  between A and B tones are compared with the probability that starlings (open triangles; MacDougall-Shackleton et al. 1998) and humans (open squares, Anstis and Saida 1985) perceived one temporally coherent stream consisting of alternating A and B tones as functions of  $\Delta F$  in a psychoacoustic task. Physiological data represent the mean ( $\pm 2$  SE) differences in responses to A and B tones (Difference 1, Table 3) when the TD was 100 ms and the tone repetition time was 100% of the TD (TRT = 100 ms). Starling behavioral data represent the mean probability that birds reported hearing a galloping rhythm, which corresponds to the probability of temporal coherence, as a function of  $\Delta F$  in a 2-alternative, forced-choice procedure in which they were trained to discriminate between galloping and isochronous rhythms. TD was 100 ms and the TRT was 100% of the TD (TRT = 100 ms); the frequency of the B tone was fixed at 1,000 Hz, and the frequencies of the A tones in different probe sequences were 1,050 Hz ( $\Delta F = 0.9$  semitones) or 1,710 Hz ( $\Delta F = 9.3$  semitones). Human behavioral data represent the mean probability that listeners reported hearing a single, coherent stream of alternating tones while listening to a 30-s ABAB sequence as a function of  $\Delta F$ . Frequency of the A and B tones was centered around 1,000 Hz, the TD was 125 ms, and the TRT was 100% of the TD (TRT = 125 ms).

nomina experienced by human listeners in sequential stream segregation tasks. For example, a number of recent human psychoacoustic studies (reviewed in Moore and Gockel 2002) demonstrate that sequential stream segregation is possible under conditions in which there is complete or nearly complete overlap of excitation patterns in the cochlea. In addition, Carlyon et al. (2001) recently suggested that high-level attentional processes are required in the segregation of even acoustically simple interleaved sequences of pure tones. These human studies clearly implicate the operation of processes separate from any low-level processes responsible for decreasing the overlap of excitation along a tonotopic gradient, such as frequency selectivity and forward masking.

Some results from the present study are also consistent with the view that frequency selectivity and forward masking are insufficient to account for the range of perceptual experiences of human listeners. For example, at large  $\Delta F$  values (e.g., 10–12 semitones), the perceptual segregation by humans of interleaved A and B tones is automatic and obligatory at fast tone rates, but not at slower tone rates (Bregman 1990; van Noorden 1975). However, we found little evidence that the pattern of differences in the responses of starling auditory forebrain neurons to CF and non-CF tones varied markedly as a function of TRT at  $\Delta F$  values of 10–12 semitones. This may be attributable in part to a “floor effect,” given that responses to B tones at  $\Delta F$  values of 10–12 semitones were already close to the spontaneous rate. Thus shorter TRTs (i.e., faster tone rates) led to relatively small additional decreases in the overlap of excitation when the  $\Delta F$  was large. Thus our results do not provide an explanation for the distinct perceptual effect of decreasing TRT at large  $\Delta F$  values present in humans. Related to this point is the fact that we found no evidence for a physiological correlate of the temporal coherence boundary, suggesting that processes in addition to frequency selectivity and forward masking may be required to account for the perceptual experience of human listeners at the temporal coherence boundary. At present, we cannot provide explanations for these differences between starling physiology and human perception in the effects of  $\Delta F$  and TRT. We cannot rule out the possibility that species differences in perception could explain these differences because the effects of differences in TRT on starling perception have not been investigated. To our knowledge, the effects of TRT at large  $\Delta F$  values (e.g., 10–12 semitones) on neural responses to sequences of alternating tones has not been examined in a similar study in a mammalian auditory system.

We believe a better understanding of the mechanisms of sequential stream segregation might be achieved in 2 ways. First, psychophysical experiments are required in which auditory stream segregation is measured in ways that are not dependent on subjective reports, and could thus be performed in parallel with human and animal subjects. Micheyl et al. (2004), for example, recently described such experimental designs that are based on performance measures in psychoacoustic detection tasks that can be made either more or less difficult depending on whether subjects experience a single coherent stream of interleaved tones, or 2 segregated streams. The use of such designs in future animal studies is clearly desirable. Second, we believe further insights into the mechanisms of auditory stream segregation can be gained by recording neural activity while at the same time precisely manipu-

lating the perceptual experience of the subject. This could be accomplished, for example, by presenting stimuli inside the ambiguous region between the fission and temporal coherence boundaries, which can be perceived as either segregated or coherent streams, and directing the subject’s attention to one or the other percept. The results of such an experiment could then be compared with those obtained in tests with passively listening animals, such as those used in the present study. In such a comparison, one would predict that the same stimulus sequence should result in different patterns of neural responses depending on the listener’s perception, and depending on whether the subject was a passive or engaged listener, thus providing further insights into the neural mechanisms that segregate acoustic events in different frequency ranges into separate auditory streams. Such additional studies will be required to determine the extent to which differences in neural rate responses are able to predict performance in psychoacoustic tasks.

#### ACKNOWLEDGMENTS

We thank Dr. Michael Brosch and two anonymous reviewers for helpful comments on earlier versions of the manuscript.

#### GRANTS

This research was supported by the Deutsche Forschungsgemeinschaft (FOR 306) and by postdoctoral fellowships to M. A. Bee from the National Science Foundation (INT-0107304) and the International Graduate School for Neurosensory Science and Systems.

#### REFERENCES

- Anstis S and Saida S. Adaptation to auditory streaming of frequency-modulated tones. *J Exp Psych Hum Percept Perform* 11: 257–271, 1985.
- Ball GF and Hulse SH. Birdsong. *Am Psychol* 53: 37–58, 1998.
- Beauvois MW. The effect of tone duration on auditory stream formation. *Percept Psychophys* 60: 852–886, 1998.
- Beauvois MW and Meddis R. A computer-model of auditory stream segregation. *Q J Exp Psychol A* 43: 517–541, 1991.
- Beauvois MW and Meddis R. Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Am* 99: 2270–2280, 1996.
- Beauvois MW and Meddis R. Time decay of auditory stream biasing. *Percept Psychophys* 59: 81–86, 1997.
- Bregman AS. Auditory streaming is cumulative. *J Exp Psych Hum Percept Perform* 4: 380–387, 1978.
- Bregman AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.
- Brosch M and Schreiner CE. Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* 77: 923–943, 1997.
- Calford MB and Semple MN. Monaural inhibition in cat auditory cortex. *J Neurophysiol* 73: 1876–1891, 1995.
- Carlyon RP, Cusack R, Foxton JM, and Robertson IH. Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psych Hum Percept Perform* 27: 115–127, 2001.
- Carr CE and Code RA. The central auditory system of reptiles and birds. In: *Comparative Hearing: Birds and Reptiles*, edited by Dooling RJ, Fay RR, and Popper AN. New York: Springer-Verlag, p. 197–248, 2000.
- Dooling RJ, Okanoya K, Downing J, and Hulse S. Hearing in the starling (*Sturnus vulgaris*): absolute thresholds and critical ratios. *Bull Psych Soc* 24: 462–464, 1986.
- Doupe AJ and Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22: 567–531, 1999.
- Fastl H. Temporal masking effects. III. Pure tone masker. *Acoustica* 43: 282–294, 1979.
- Fay RR. Auditory stream segregation in goldfish (*Carassius auratus*). *Hear Res* 120: 69–76, 1998.
- Fay RR. Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *JARO (J Assoc Res Otolaryngol)* 1: 120–128, 2000.
- Feng AS and Ratnam R. Neural basis of hearing in real-world situations. *Annu Rev Psych* 51: 699–725, 2000.

- Fishman YI, Reser DH, Arezzo JC, and Steinschneider M.** Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 151: 167–187, 2001.
- Fishman YI and Steinschneider M.** Neural correlates of auditory stream segregation in primary auditory cortex (A1) of awake monkeys: effects of frequency-separation, presentation rate, and tone duration. *Assoc Res Otolaryngol Abstr* 26: 920, 2003.
- Greenhouse SW and Geisser S.** On methods in the analysis of profile data. *Psychometrika* 24: 95–112, 1959.
- Hartmann WM and Johnson D.** Stream segregation and peripheral channeling. *Music Percept* 9: 155–184, 1991.
- Hofer S and Klump GM.** Within- and across-channel processing in auditory masking: a physiological study in the songbird forebrain. *J Neurosci* 23: 5735–5739, 2003.
- Hulse SH.** Auditory scene analysis in animal communication. *Adv Study Behav* 31: 163–200, 2002.
- Hung J, Jones SJ, and Pato MV.** Scalp potentials to pitch change in rapid tone sequences: a correlate of sequential stream segregation. *Exp Brain Res* 140: 56–65, 2001.
- Izumi A.** Auditory sequence discrimination in Japanese monkeys: effect of frequency proximity on perceiving auditory stream. *Psychologia* 44: 17–23, 2001.
- Izumi A.** Auditory stream segregation in Japanese monkeys. *Cognition* 82: B113–B122, 2002.
- Jesteadt W, Bacon SP, and Lehmen JR.** Forward masking as a function of frequency, masker level, and signal delay. *J Acoust Soc Am* 71: 950–962, 1982.
- Klump GM.** Bird communication in the noisy world. In: *Ecology and Evolution of Acoustic Communication in Birds*, edited by Kroodsmas DE and Miller EH. Ithaca, NY: Cornell University Press, 1996.
- Klump GM, Langemann U, and Gleich O.** The European starling as a model for understanding perceptual mechanisms. In: *Auditory Worlds: Sensory Analysis and Perception in Animals and Man*, edited by Manley GA, Fastl H, Kössl M, Oeckinghaus H, and Klump GM. Weinheim, Germany: Wiley–VCH, 2000.
- Klump GM and Nieder A.** Forward masking in the European starling's primary auditory forebrain. *Assoc Res Otolaryngol Abstr* 24: 1002, 2001.
- MacDougall-Shackleton SA, Hulse SH, Gentner TQ, and White W.** Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences. *J Acoust Soc Am* 103: 3581–3587, 1998.
- McCabe SL and Denham MJ.** A model of auditory streaming. *J Acoust Soc Am* 101: 1611–1621, 1997.
- Micheyl C, Carlyon RP, Cusack R, and Moore BCJ.** Performance measures of auditory organization. In: *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by Pressnitzer D, Cheveigne A, McAdams S, and Collet L. New York: Springer-Verlag, p. 167–174, 2004.
- Micheyl C, Tian B, Carlyon RP, and Rauschecker JP.** The neural basis of stream segregation in the auditory cortex. *Assoc Res Otolaryngol Abstr* 26: 919, 2003.
- Miller GA and Heise GA.** The trill threshold. *J Acoust Soc Am* 22: 637–638, 1950.
- Moore BCJ and Glasberg BR.** Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise. *J Acoust Soc Am* 73: 1249–1259, 1983.
- Moore BCJ and Gockel H.** Factors influencing sequential stream segregation. *Acta Acustica United with Acustica* 88: 320–333, 2002.
- Nieder A and Klump GM.** Adjustable frequency selectivity of auditory forebrain neurons recorded in a freely moving songbird via radiotelemetry. *Hear Res* 127: 41–54, 1999.
- Oxenham AJ.** Forward masking: adaptation or integration? *J Acoust Soc Am* 109: 732–741, 2001.
- Relkin EM and Turner CW.** A reexamination of forward masking in the auditory nerve. *J Acoust Soc Am* 84: 584–591, 1988.
- Rice WR.** Analyzing tables of statistical tests. *Evolution* 43: 223–225, 1989.
- Rogers WL and Bregman AS.** An experimental evaluation of three theories of auditory stream segregation. *Percept Psychophys* 53: 179–189, 1993.
- Rose MM and Moore BCJ.** Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. *J Acoust Soc Am* 102: 1768–1778, 1997.
- Rose MM and Moore BCJ.** Effects of frequency and level on auditory stream segregation. *J Acoust Soc Am* 108: 1209–1214, 2000.
- Rosenthal R and Rosnow RL.** *Contrast Analysis: Focused Comparisons in the Analysis of Variance*. Cambridge, UK: Cambridge Univ. Press, 1985.
- Saini KD and Leppelsack HJ.** Cell-types of the auditory caudomedial neostriatum of the starling (*Sturnus vulgaris*). *J Comp Neurol* 198: 209–229, 1981.
- Singh PG and Bregman AS.** The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *J Acoust Soc Am* 102: 1943–1952, 1997.
- Sussman E, Ritter W, and Vaughan HG.** An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36: 22–34, 1999.
- Turner CW, Relkin EM, and Doucet J.** Psychophysical and physiological forward masking studies: probe duration and rise-time effects. *J Acoust Soc Am* 96: 795–800, 1994.
- van Noorden LPAS.** *Temporal Coherence in the Perception of Tone Sequences* (PhD dissertation). Eindhoven, The Netherlands: Eindhoven University of Technology, 1975.
- Yost WA.** Auditory image perception and analysis—the basis for hearing. *Hear Res* 56: 8–18, 1991.