Does common spatial origin promote the auditory grouping of temporally separated signal elements in grey treefrogs?

MARK A. BEE & KASEN K. RIEMERSMA
Department of Ecology, Evolution, and Behavior, University of Minnesota, Twin Cities

(Received 7 August 2007; initial acceptance 3 September 2007; final acceptance 4 January 2008; published online 9 July 2008; MS. number: A10840R2)

‘Sequential integration’ represents a form of auditory grouping in which temporally separated sounds produced by the same source are perceptually bound together over time into a coherent ‘auditory stream’. In humans, sequential integration is important in music and speech perception. In this study of the grey treefrog, *Hyla chrysoscelis*, we took advantage of female selectivity for advertisement calls with conspecific pulse rates to investigate common spatial location as a cue for sequential integration. We presented females with two temporally interleaved pulse sequences with pulse rates of 25 pulses/s, which is half the conspecific pulse rate and more similar to that of *H. versicolor*, a syntopically breeding heterospecific. We tested the hypothesis that common spatial origin between the two pulse sequences would promote their integration into a coherent auditory stream with an attractive conspecific pulse rate. As the spatial separation between the speakers broadcasting the interleaved pulse sequences decreased from 180° to 0°, more females responded and females had shorter response latencies and travelled shorter distances en route to a speaker. However, even in the 180° condition, most females (74%) still responded. Detailed video analyses revealed no evidence that patterns of female phonotaxis resulted from impaired abilities to localize sound sources in the spatially separated conditions. Together, our results suggest that females were fairly permissive of spatial incoherence between the interleaved pulse sequences and that common spatial origin may be only a relatively weak cue for sequential integration in grey treefrogs.

© 2008 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Keywords: acoustic communication; auditory grouping; auditory stream segregation; cocktail party problem; communication; gray treefrog; grey treefrog; *Hyla chrysoscelis*; sequential integration; vocal communication

A fundamental question in sensory biology, and one that is central to understanding the mechanisms of animal communication (Feng & Ratnam 2000; Hulse 2002; Bee & Micheyl, in press), is how do brains parse continuous streams of sensory information into perceptually distinct units? The processes by which the auditory system solves this problem are often referred to as ‘auditory scene analysis’ (Bregman 1990; Hulse 2002; Bee & Micheyl, in press). Analysing acoustic scenes involves integrating or ‘grouping’ sounds that originate from a single source into perceptually coherent ‘auditory streams’ or ‘auditory objects’ that can be attended to and perceptually segregated from other sounds in the environment (Bregman 1990; Carlyon 2004; Bee & Micheyl, in press).

Two forms of perceptual integration are important in human auditory scene analysis (Bregman 1990). ‘Simultaneous integration’ refers to the perceptual grouping of different, simultaneously occurring components of the frequency spectrum (e.g. harmonics, formants) into a representation of a single sound source. ‘Sequential integration’ on the other hand involves the integration of temporally separated sounds from one sound source (e.g. syllables, words, musical notes) into a coherent auditory stream and their segregation from other intervening and overlapping sounds from other sources. Both simultaneous and sequential integration are important in music and speech perception by humans and in allowing human listeners to solve the so-called ‘cocktail party problem,’ which refers to the difficulty we sometimes have understanding speech in noisy social environments (reviewed in Bregman 1990; Bee & Micheyl, in press). Numerous studies have investigated the auditory grouping cues, including common spatial origin and various acoustic
cues, that humans use to perform simultaneous and sequential integration (reviews in Bregman 1990; Darwin & Carlyon 1995; Cusack & Carlyon 2004; Bee & Michely, in press).

While auditory scene analysis constitutes a major paradigm in studies of human hearing and speech communication (Bregman 1990; Carlyon 2004; Cusack & Carlyon 2004), few studies have investigated similar processes in the acoustic communication systems of nonhuman animals (Feng & Ratnam 2000; Hulse 2002; Bee & Michely, in press). Anuran amphibians (frogs and toads) represent one taxonomic group for which aspects of auditory scene analysis may be important for communication (Hulse 2002). Although the cocktail-party-like listening conditions in a frog breeding chorus can impair auditory perception (Gerhardt & Klump 1988; Wollerman 1999; Schwartz et al. 2001; Wollerman & Wiley 2002; Bee 2007, 2008; Bee & Swanson 2007), the advertisement calls of male frogs nevertheless effectively mediate species recognition, source localization, female mate choice, male–male assessment, and even individual recognition under the noisy conditions of a breeding chorus (reviews in Gerhardt & Bee 2006; Wells & Schwartz 2006).

Many frogs produce vocalizations consisting of sequentially produced call types (Ryan 1985; Larson 2004) or temporally repeated pulses (Gerhardt 1991; Castellano & Giacoma 1998; Howard & Young 1998; Friedl 2006). In frogs with pulsed advertisement calls, the rate of temporally repeated pulses is often an important acoustic property for species recognition (reviewed in Gerhardt & Huber 2002), and the number of pulses per call can serve as an honest indicator of male genetic quality (Welch et al. 1998). Although the sequential integration of temporally separated signal elements is potentially important for species recognition and mate choice in frogs, we still know very little about the cues that promote sequential integration (but see Schwartz & Gerhardt 1995; Farris et al. 2002, 2005). In this study, we tested the hypothesis that common spatial origin functions as an auditory grouping cue that promotes the sequential integration of temporally repeated pulses in the context of call recognition by females of Cope’s grey treefrog, Hyla chrysoscelis.

**METHODS**

**Study System and Experimental Rationale**

Cope’s grey treefrog is the diploid member of the cryptic diploid–tetraploid grey treefrog species complex; the eastern grey treefrog, *H. versicolor*, is the tetraploid (Ptacek et al. 1994; Holloway et al. 2006). These two species breed syntopically throughout much of their common range in North America, including Minnesota, where this study was conducted. A great deal is already known about the acoustic properties that mediate species recognition by females of both species (reviewed in Gerhardt 2001). The advertisement calls of male *H. chrysoscelis* comprise a series of pulses produced at a species-typical rate of about 35–50 pulses/s (Gerhardt 2001). The pulsed advertisement calls of *H. versicolor* are spectrally similar but have pulse rates of 19–24 pulses/s, which is about half as fast as those of *H. chrysoscelis* calls (Gerhardt 2001). Females of both species show robust selectivity for conspecific pulse rates (Gerhardt & Doherty 1988; Gerhardt 2001; Bush et al. 2002; Schol & Bush 2002). We took advantage of a female’s pulse rate selectivity to test the hypothesis that spatial coherence between the temporally separated pulses of a male’s calls promotes sequential integration. We presented females with two temporally interleaved pulse sequences broadcast at angular separations of 180°, 90°, 45° or 0°. Each pulse sequence was delivered at a rate that was about half the normal pulse rate of *H. chrysoscelis* calls, and more similar to the unattractive pulse rate of heterospecific *H. versicolor* calls. We predicted that as the spatial separation between the two sequences decreased from 180° to 0°, females would be more likely to integrate the two sequences into a coherent auditory stream with the attractive pulse rate of conspecific calls. Hence, the underlying assumption of our experimental approach was that females would show greater relative phonotaxis under conditions that promoted auditory grouping.

**General Procedures**

Details regarding our field sites, our procedures for collecting, handling and testing females, and the sound chamber and playback system used for conducting experiments have been fully described in Bee (2007). Briefly, we made nightly collections of females in amplexus between 5 May and 29 June 2006 from several local breeding populations. They were returned to the laboratory, used as subjects in several phonotaxis tests, and then returned to their original location of capture, usually within 1–3 days of collection.

We conducted phonotaxis tests in a circular test arena (2-m diameter) located inside a walk-in, hemi-anechoic sound chamber. The sound chamber was temperature controlled and all phonotaxis tests were performed at 20 ± 2°C. The walls of the circular test arena were constructed from hardware cloth (60 cm height) and covered by visually opaque, but acoustically transparent, black cloth. The perimeter of the arena floor was divided into 24 15° bins. Playback speakers (A/D/S L310) were positioned on the floor of the chamber just outside the arena wall, centred in a 15° bin and aimed towards the centre of the arena. Before testing, females were placed in a 20°C incubator and their body temperatures were allowed to reach 20 ± 1°C. At the beginning of a test, we removed the female from the incubator and placed her in an acoustically transparent release cage (9 cm diameter) on the floor in the centre of the test arena. After a 1-min acclimation period, we started broadcasts of the acoustic stimulus, which was broadcast as a continuous loop during the test. After four repetitions of the loop, we remotely released the female. Groups of three to five females were typically tested in sequence, with one female being tested while the other females waited in the incubator during a 5–15 min ‘time-out’ between tests. The absolute positions of the playback speakers around the perimeter of the circular arena were systematically varied between groups of
females to control for any possibility of a directional response bias in the sound chamber. No such bias was detected (M. A. Bee, unpublished data).

Experimental Design

The acoustic stimulus (20 kHz, 16 bit) consisted of two interleaved sequences of identical synthetic pulses (Fig. 1a). One sequence consisted of 23 pulses and started and ended half a pulse period before and after the other sequence of 22 pulses (Fig. 1a). Each sequence was broadcast at a sound pressure level of 75 dB SPL (LCF, re. 20 μPa) at the female release site and repeated with a periodicity of 4.5 s. Sound levels were calibrated by placing the microphone of a Brüel & Kjær Type 2250 sound level meter at the approximate position of a female’s head at the central release site. Each pulse in both sequences was modelled after natural *H. chrysoscelis* pulses and was constructed by adding two phase-locked sinusoids with frequencies (and relative amplitudes) of 1.1 kHz (−6 dB) and 2.2 kHz (0 dB). The duration of each pulse was 10 ms and its amplitude envelope was shaped with a 4.8-ms inverse exponential rise time and a 4.8-ms exponential fall time. The interpulse interval in each sequence was 30 ms so that the resulting pulse period was 40 ms and the pulse rate of each sequence was 25 pulses/s. This pulse rate falls within the range of pulse rates (corrected to 20 °C) recorded in local populations of *H. versicolor* but falls well below the range of variation in local *H. chrysoscelis* pulse rates (M. A. Bee, unpublished data). The pulse rate of the composite interleaved sequence was 50 pulses/s (Fig. 1a), which is close to the average pulse rate of 46 pulses/s (corrected to 20 °C) recorded in local populations of *H. chrysoscelis* (M. A. Bee, unpublished data). The gross temporal envelope of each pulse sequence was shaped so that the composite interleaved pulse sequence had a linear onset of 60 ms.

We performed separate phonotaxis tests in which we simultaneously broadcast the two interleaved sequences either from two speakers separated by an angle (θ) of 180°, 90° or 45° around the circular test arena or from a single speaker (i.e. 0° spatial separation) (see Fig. 1b). The 0° condition served as a control condition in which the stimulus (hereafter ‘standard call’) was an attractive conspecific call with spectral and temporal properties close to the average values recorded in local populations (e.g. 45 pulses, 20 ms pulse period, 50 pulses/s, 890 ms call duration; M. A. Bee, unpublished data). Females were tested in an initial ‘warm-up’ trial in which we presented the standard call to acclimate females to the test situation; data from this condition were discarded. Following this warm-up trial, females were tested in a sequence of five tests that comprised two tests with the standard call and one test each with a spatial separation of 180°, 90° and 45° between the two pulse sequences. This sequence of five tests always began with a test of the spatially coherent standard call (i.e. 0° separation) in an initial control condition (Control 1). In the next three tests in a sequence, we tested females in the three spatially incoherent conditions. The order of these three tests was randomized separately for each female. A test sequence concluded with a test of the standard call in a second control condition (Control 2). Hence, the two spatially coherent tests with the standard call represent temporally replicated control conditions that preceded and followed the three spatially incoherent test conditions.

Hypotheses, Predictions and Data Analyses

Measures of attractiveness: response probability and response latency

According to the hypothesis that common spatial origin promotes sequential integration, we predicted that the ability of females to integrate the two interleaved pulse sequences into a coherent auditory stream with an attractive pulse rate would increase as a function of decreasing angular separation between the two sequences. The probability that females respond to an acoustic stimulus, and the latency with which responses occur, are the two most common measures of signal
attractiveness in phonotaxis studies of frog communication (Gerhardt 1995; Gerhardt & Huber 2002). Responsive females typically initiate phonotaxis within seconds of their release and show directed patterns of zigzag hopping or walking towards a sound source. We gave females up to 5 min to meet a response criterion that required them to touch the wall of the test arena inside the 15° arc centred in front of one of the active playback speakers. (Note that females were allowed to make initial contact with the arena wall somewhere other than directly in front of a speaker. Females that made initial wall contact away from the speaker, however, did not then arrive at the speaker simply by following along the circular wall. Rather, females typically left the wall, moved some distance back towards the centre of the arena, reoriented and then approached the speaker.) As measures of stimulus attractiveness in each condition, we determined the probability of female responses as the proportion of females that met our response criterion and we measured the latency of female responses. We scored a ‘no response’ if a female did not touch the wall in front of one of the speakers within 5 min.

**Measures of source localization behaviour**

One potential problem with probability-based or latency-based measures of phonotaxis behaviour when signal elements are presented from different locations is that some experimental manipulations may not affect stimulus attractiveness, per se, but instead may impair accurate source localization. It was possible, therefore, that females might fail to respond or require more time to respond simply because it was more difficult to determine the position of an active playback speaker in the spatially separated conditions. To assess this possibility, we measured a number of response variables potentially related to sound localization performance in frogs. One of these variables (‘orientation’, see below) was determined in real time for each female that eventually touched the wall of the test arena. The other variables were computed for a subset of 17 females for which we recorded phonotaxis responses as digital videos (sampling rate = 5 samples/s) and used EthoVision to track the females’ movements in offline analyses (Noldus 2005). For these video analyses, we imposed a criterion that required a female to move greater than 1 cm for a movement to be registered. While this criterion was necessary to eliminate the influence of noise in our video analyses, this criterion also precluded analyses of small movements, such as head scanning and reorientation movements, that occurred in a single location.

**Orientation.** As a measure of ‘orientation’ accuracy, we determined the angle at which females first touched the wall of the circular arena. The position of the speaker in the control conditions was designated as 0° around the arena; in the spatially incoherent treatments, the positions of the two speakers were designated as being symmetrical around 0° (i.e. at ±90°; see Fig. 1b).

**Turn angles.** Treefrogs show zigzag patterns of walking towards sound sources in phonotaxis tests, and these patterns of movement represent behaviours related to sound localization (Rheinlaender et al. 1979; Rheinlaender & Klump 1988). The angle at which females turn between movements has been used as an indicator of sound localization accuracy. We predicted that females would have greater ‘turn angles’ in the spatially separated conditions if they had difficulty localizing sources in these conditions. To test this prediction, we measured the mean, minimum and maximum turn angles (0°–180°) that females made over the duration of each phonotaxis trial.

**Meander, heading and speed of moving towards.** EthoVision provides estimates of several variables that characterize the directedness of an animal’s movements (Noldus 2005). As one such variable, we determined ‘meander’, which corresponds to the change in direction relative to distance moved and is measured in degrees/cm. High values of meander correspond to situations in which the animal changes direction frequently over short distances, whereas low values of meander represent relatively more directed movements with fewer changes in direction. We predicted that, if females experienced difficulty in source localization in the spatially separated conditions, then they might show higher values for meander. We also measured a female’s angular ‘heading’ during a phonotaxis approach relative to the speaker that was eventually chosen. We calculated relative heading as the mean angle (−180° to 180°) of the vectors formed by two consecutive movements relative to a virtual reference line that connected the release cage with the speaker that the female ultimately chose (0°). Hence, values of heading close to 0° indicated directed movement towards the active speaker that was chosen, whereas higher or lower values of heading indicated less directed movements towards the chosen speaker. We also measured a female’s ‘speed of moving towards’, which functions as a measure of the intensity of approach towards a designated location and was defined as the distance-weighted speed at which the female moved towards the speaker that was ultimately chosen (Noldus 2005). We expected to find lower values for speed of moving towards in conditions in which females had difficulty localizing a source. Because of some noise in our videos of phonotaxis responses, however, we were only able to determine speed of moving towards the chosen speaker for 12 of the 17 females for which responses were recorded and analysed using EthoVision. Note also that four females included in our EthoVision analyses did not actually touch the wall in front of a speaker within 5 min in the 180° condition, although they showed clear phonotaxis behaviour towards the playback speakers; therefore, for these four phonotaxis responses out of the 85 analysed (5 conditions x 17 females), we designated the ‘chosen’ speaker as the speaker that yielded the higher value for speed of moving towards and the value closer to zero for heading.

**Velocity of movement and total distance moved.** If females had more difficulty in source localization in the spatially separated conditions, we expected that they might either spend relatively more time sitting and listening, and
hence have lower average velocities over the duration of time required to reach a speaker, or move longer distances in the test arena in search of the source. We defined a female’s ‘total distance moved’ as the sum of the distances travelled by the female between all temporally consecutive pairs of video samples. A female’s ‘velocity of movement’ was calculated by dividing the total distance moved by the latency required to touch the wall of the test arena in the 15° arc in front of an active speaker or by 5 min for females that did not meet this response criterion.

### Statistical analyses

Twenty five females were collected and tested in this experiment. Because female response motivation can decrease over time (M. A. Bee, personal observation), we included in our final data set only subjects that reached the speaker within 5 min after showing clear phonotaxis behaviours (e.g. head scanning and zigzag patterns of movement; Rheinlaender et al. 1979) in their responses to the attractive standard call in both of the temporally replicated control conditions. Two females did not respond in the final control condition, yielding a final sample size of $N = 23$ females, which included all 17 females for which video analyses were conducted. Such criteria are necessary to ensure that slower responses or ‘no responses’ in the intervening test conditions reflect legitimate disin interest in the stimulus, and not lost or waning response motivation. We and others have used these or similar inclusion criteria to ensure that female grey treefrogs maintain their response motivation over sequences of tests (Bush et al. 2002; Schul & Bush 2002; Beckers & Schul 2004; Bee 2007, 2008; Bee & Swanson 2007; Swanson et al. 2007).

We employed an experiment-wide significance criterion of $\alpha = 0.05$ for all statistical analyses. Response probabilities were compared using Cochran’s Q test ($N = 23$). Response latencies were compared using a repeated measures ANOVA that included only those females that met our response criterion in all five conditions ($N = 14$); that is, we did not arbitrarily assign latencies of 300 s to females that did not meet our 5-min response criterion. We analysed orientation in the control conditions using a $V$ test (Zar 1999) to test the null hypothesis that angles were uniformly distributed against the alternative hypothesis that females oriented directly towards the speaker at 0°. In the spatially incoherent conditions, we had no a priori expectation about an alternative hypothesized angle; therefore, for these conditions, we used Rao’s spacing test (Zar 1999) to test the null hypothesis that orientation angles were uniformly distributed against the alternative hypothesis that responses had some directionality. The sample sizes in analyses of orientation were equal to the number of females that touched the wall in a particular test condition. Response variables determined using EthoVision were compared using repeated measures ANOVAs ($N = 12$ for speed of moving towards; $N = 17$ for all other localization variables). For all ANOVAs, we report partial $\eta^2$ as a measure of effect size and we report $P$ values based on the Greenhouse–Geisser correction (Greenhouse & Geisser 1959). Multiple comparison post hoc tests were performed using Newman–Keuls tests for significant ANOVAs. We used Statistica 7.0 (Tulsa, Ok, U.S.A.) to compute values for Cochran’s Q tests, ANOVAs and post hoc tests; Oriana 2.02 (Kovach Computing, Anglesey, U.K.) was used to compute circular statistics.

### Results

#### Stimulus Attractiveness

All 23 females (100%) touched the arena wall in front of an active speaker within 5 min in the two spatially coherent (0°) control conditions. These conditions preceded and followed the three spatially incoherent conditions, which were tested in random order. While all 23 females also met this same response criterion in the 45° condition, 2 of 23 females (9%) and 6 of 23 females (26%) failed to do so in the 90° and 180° conditions, respectively. The difference between the proportions of the 23 females that responded across the three spatially incoherent conditions was significant ($Q_2 = 9.3, P < 0.0095$). Although these response proportions were significantly different, large proportions of females (74%, 91% and 100%) still responded to one of the two pulse sequences in the 180°, 90° and 45° conditions, respectively. These differences in response probability also remained significant when we included the 0° control conditions in similar analyses using both the restricted sample of $N = 23$ ($Q_4 = 19.4, P < 0.0006$) and the entire sample of $N = 25$, which did not exclude the two females that failed to respond in the second control condition ($Q_4 = 16.2, P < 0.0027$).

There were significant differences in response latency (Table 1, Fig. 2). The mean latencies in Control 1 (54.0 s; $N = 14$) and Control 2 (46.9 s; $N = 14$) were not significantly different ($P = 0.6706$). The mean response latency in the 180° condition (135.3 s; $N = 14$) was significantly longer than latencies in all other conditions (all $P < 0.0357$). Mean response latencies in the 90° condition (99.3 s; $N = 14$) and the 45° condition (78.0 s; $N = 14$) did not differ significantly ($P = 0.2079$) and were intermediate between latencies in the 180° condition and the control conditions (Fig. 2). While the mean latency in the 90° condition was significantly longer than latencies in both Control 1 ($P = 0.0242$) and Control 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>$F$</th>
<th>$df$</th>
<th>$P$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response latency</td>
<td>9.3</td>
<td>4,52</td>
<td>&lt;0.0001</td>
<td>0.42</td>
</tr>
<tr>
<td>Mean turn angle</td>
<td>0.5</td>
<td>4,64</td>
<td>0.6640</td>
<td>0.03</td>
</tr>
<tr>
<td>Minimum turn angle</td>
<td>1.7</td>
<td>4,64</td>
<td>0.1974</td>
<td>0.10</td>
</tr>
<tr>
<td>Maximum turn angle</td>
<td>1.1</td>
<td>4,64</td>
<td>0.3621</td>
<td>0.06</td>
</tr>
<tr>
<td>Meander</td>
<td>0.5</td>
<td>4,64</td>
<td>0.6316</td>
<td>0.03</td>
</tr>
<tr>
<td>Heading</td>
<td>0.4</td>
<td>4,64</td>
<td>0.5717</td>
<td>0.03</td>
</tr>
<tr>
<td>Speed of moving</td>
<td>0.1</td>
<td>4,44</td>
<td>0.9312</td>
<td>0.01</td>
</tr>
<tr>
<td>Velocity of movement</td>
<td>0.5</td>
<td>4,64</td>
<td>0.6329</td>
<td>0.03</td>
</tr>
<tr>
<td>Total distance</td>
<td>8.1</td>
<td>4,64</td>
<td>0.0035</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Overall, females moved to speakers more rapidly when the stimulus was spatially coherent, and in the 45° condition, females moved significantly faster than in the 90° condition ($Q = 28.7, P < 0.0001$). Females also moved to the 180° stimulus faster than to the 45° stimulus ($Q = 11.4, P < 0.0009$).
Source Localization Behaviours

The orientation angles at which females first touched the wall of the test arena were significantly oriented towards the single active speaker in both control conditions (all Vs > 6.7; all Ps < 0.0001; Fig. 3). The length of the mean vector in the control conditions was greater than r = 0.99, indicating strong orientation when the two interleaved pulse sequences originated from a common location. Females also showed significant orientation in the general direction of the playback speakers in all three of the spatially separated conditions (Rao’s spacing test: all Us > 250, all Ps < 0.01; Fig. 3). Compared to orientation in the control conditions, however, orientation angles were more variable in the spatially incoherent conditions (Fig. 3). When the interleaved pulse sequences were broadcast from opposite sides of the test arena in the 180° condition, females tended to orient towards one of the other speaker (Fig. 3). The distributions of orientation angles in the 90° and 45° conditions fell largely within the arc separating the two speakers (Fig. 3). Some females oriented directly towards one of the two speakers when the interleaved pulse sequences were separated by 90° and 45°, while a few females in these conditions oriented towards a ‘phantom’ location midway between the two speakers (Fig. 3).

There were no significant differences in the mean, minimum or maximum turn angles (Table 1, Fig. 4a–c).

We also found no significant differences in mean, minimum or maximum turn angles (Table 1, Fig. 4a–c). Average values of heading ranged between −5° and 14° of the direction of the chosen speaker (0°), but the differences between conditions were not significant (Table 1, Fig. 4e). We also found no significant differences in speed of moving towards the speaker that was chosen (Table 1, Fig. 4f) or in females’ velocity of movement (Table 1, Fig. 4g).

DISCUSSION

Male frogs show a number of behavioural adaptations that function to improve the transmission of their calls to female receivers amid the cacophony of a breeding chorus. Two such adaptations are spacing out within the available breeding habitat (Wilczynski & Brenowitz 1988; Brenowitz 1989; Gerhardt et al. 1989; Dyson & Passmore 1992) and avoiding call overlap with nearby neighbours (Schwartz 1987; Grafe 1996; Schwartz et al. 2002). As chorus density increases, however, intermale distances are necessarily reduced (e.g. Gerhardt et al. 1989), and, even in moderately dense choruses, avoiding overlap with nearby males may become difficult or impossible (e.g. Schwartz et al. 2002). Thus, we might expect females to possess neural mechanisms that allow them to integrate the sequential signal elements produced by one male (e.g. Alder & Rose 1998, 2000) into a coherent auditory stream that can be segregated from the calls of other nearby males (e.g. Schwartz & Gerhardt 1995). The aim of the present study was to exploit the pulse rate selectivity of female grey treefrogs to investigate the influences of common spatial origin on sequential integration.

Decreased Attractiveness or Impaired Source Localization Ability?

Females of both grey treefrog species find calls with conspecific pulse rates more attractive than those with pulse rates similar to the other species (e.g. Gerhardt &
BEE & Riemersma: Sequential Integration in Treefrogs

For example, in no-choice tests with *H. chrysoscelis* similar to those described in the present study, Schul & Bush (2002; see also Bush et al. 2002) reported normalized latencies, or ‘phonotaxis scores’, close to 1.0 (indicating a highly attractive stimulus) for responses to conspecific pulse rates (e.g. 50 pulses/s). In contrast, phonotaxis scores were closer to zero (<0.30; indicating no phonotaxis or very slow phonotaxis) in response to pulse rates closer to those of *H. versicolor* (e.g. 19–28 pulses/s). In that study, when the pulse rate was 19 pulses/s, 4 of 20 females (20%) responded (mean phonotaxis score = 0.01) and 0 of 13 females (0%) responded when the pulse rate was about 23 pulses/s (mean phonotaxis score of 0.0). Ten of 13 females (77%) responded when the pulse rate was about 28 pulses/s (mean phonotaxis score = 0.30), but these responsive females required, on average, more than twice as long to respond compared to their responses to a conspecific pulse rate (J. Schul & S. L. Bush, unpublished data). Although we did not include a condition comprising a single pulse sequence (25 pulses/s) in our study; results from these previous no-choice tests with *H. chrysoscelis* suggest what we would have found in such a condition: response probabilities increase and response latencies decrease when the pulse rate is less similar to that of *H. versicolor* and more similar to that of conspecifics (Bush et al. 2002; Schul & Bush 2002).

In our experiment, response probability increased and response latency decreased as the spatial separation between the two sequences decreased from 180° to 0°. In other words, as the spatial coherence between the two interleaved pulse sequences increased, changes in response probability and latency were consistent with females behaving as if the pulse rate had increased from that of an unattractive heterospecific call to one more typical of an attractive conspecific call (e.g. Gerhardt & Doherty 1988; Bush et al. 2002; Schul & Bush 2002). Moreover, stimulus attractiveness was inversely related to the spatial separation between the two interleaved pulse sequences. Based on more detailed assessments of phonotaxis behaviours, however, we believe this increase in attractiveness with decreasing spatial incoherence might be more apparent than real (see below).

There was little evidence to suggest that spatial incoherence impaired source localization. Compared to its effect on response latencies ($\eta^2 = 0.42$; Table 1), the manipulation of spatial coherence had small or negligible effects ($0.01 \leq \eta^2 \leq 0.10$; Table 1) on differences in turn angles, velocity of movement and the directedness of a female’s approach towards a sound source (e.g. meander, heading and speed of moving towards). Results for orientation, heading and speed of moving towards confirm that female phonotaxis was generally directed towards the location of a sound source. The manipulation of spatial

**Figure 3.** Circular histograms showing the distributions of angles at which females first touched the wall of the arena in response to the standard call (Control 1 and Control 2) and the three spatially incoherent conditions (180°, 90° and 45°). The mean (SD) angle ($\mu$), the vector length ($r$) and the sample size ($N$) are provided to the left of each histogram.
Figure 4. Values for behavioural variables potentially related to sound localization in response to the spatially coherent standard call in two control conditions (Con 1 and Con 2) and the three spatially incoherent conditions (180°, 90° and 45°): (a) mean turn angle, (b) minimum turn angle, (c) maximum turn angle, (d) meander, (e) heading, (f) speed of moving towards, (g) velocity and (h) total distance moved (see text for detailed descriptions of each variable). Points depict mean values, boxes depict ±1 SE and whiskers depict ±1 SD (N = 12 for speed of moving towards; N = 17 for all other variables).
coherence did affect the total distance moved during a phonotaxis trial. The pattern of results for total distance moved parallel those for response latency (cf. Figs 2, 4h), suggesting that latencies were longer in the spatially incoherent conditions, at least in part, because females spent more time moving about in the test arena.

The causal link between total distance moved and response latency, as well as the larger variability in mean heading in the spatially incoherent conditions (Fig. 4e), resulted because females often altered the direction of their movements between the two sound sources in the spatially incoherent conditions. Figure 5 illustrates this point. For example, in responses to the standard call in the control conditions, females tended to move directly towards the single sound source (Fig. 5). In contrast, responses in the spatially incoherent conditions were still generally directed towards a sound source (as indicated by results for orientation, heading and speed of moving towards), but females did not always immediately approach the speaker they eventually chose (Fig. 5). Rather, some females crossed back and forth between the two speakers, often several times, before finally meeting our
response criterion at one of the two speakers (Fig. 5). This pattern of behaviours was most pronounced in the 180° condition, but similar behavioural patterns were also evident in the responses of some females in the 90° and 45° conditions (Fig. 5).

Does Common Spatial Origin Promote Auditory Grouping in Grey Treefrogs?

Taken together, we believe our results provide only marginal support for the hypothesis that common spatial origin functions as an auditory grouping cue in grey treefrogs. For example, despite changes in response probability and response latency, it is important to bear in mind that 74%, 91% and 100% of females still chose one of the two pulse sequences within 5 min in the 180°, 90° and 45° conditions, respectively. These response rates are higher than what might be expected based on no-choice tests of grey treefrogs using stimuli with pulse rates similar to heterospecific calls (Gerhardt & Doherty 1988; Bush et al. 2002; Schul & Bush 2002). Thus, in the present study, the relatively high response rates of females in the spatially incoherent conditions suggest that females were able to group the two pulse sequences together over fairly large angles of separation. While common spatial origin may promote sequential integration in grey treefrogs to some extent, it is also clear that spatial incoherence was unable to disrupt completely the tendency for most females to group the two interleaved pulse sequences together under our experimental conditions.

We believe the most parsimonious explanation for our results is that the manipulation of spatial coherence, combined with the ability of females to change their location relative to the fixed locations of the speakers, altered the spatial location that females assigned to the attractive (i.e. grouped) stimulus. Consider the 180° condition (Fig. 6). As females approached a speaker at one side of the arena, the amplitude of pulses emitted by that speaker at the female’s position increased during her approach, while those emitted by the speaker on the opposite side of the arena correspondingly decreased (Fig. 6). Hence, the manipulation of spatial coherence had the effect of creating differences in the perceived relative amplitudes of the two pulse sequences as females approached closer to one sound source (Schwartz & Gerhardt 1995). As females moved closer to one sound source, the dominant pulse rate emitted by that source became that rate = 50 Hz dominant' pulse rate = 50 Hz and the closer speaker, since after approaching one speaker they often reoriented towards and approached the other speaker (Fig. 5). Thus, we believe that females were fairly permissive in grouping the two spatially separated pulse sequences together.

Common Spatial Origin as a Weak Auditory Grouping Cue

A number of studies of humans suggest that acoustic cues can override cues to auditory grouping provided by spatial origin (Bregman 1990; Darwin & Carlyon 1995). These acoustic cues include harmonicity (i.e. common fundamental frequency), common timbre, common onsets and offsets, and common amplitude modulation (reviewed in: Bregman 1990; Darwin & Carlyon 1995; Cusack & Carlyon 2004; Bee & Micheyl, in press). Results from such studies have led to a general view that common spatial origin may be a relatively weak auditory grouping cue in humans, especially when conflicting acoustic cues (e.g. harmonicity) provide evidence that two spatially separated sounds should be grouped together. Note that in our study, the two interleaved pulses sequences showed harmonicity (i.e. they had a common fundamental frequency of 1.1 kHz), common timbre (i.e. the same relative amplitudes between the two spectral peaks) and common patterns of amplitude modulation (i.e. similar pulse shapes). These commonalities may have overridden cues for segregating the two pulse sequences into separate auditory streams based on differences in their spatial origin. Our results suggest that common spatial origin might also be a relatively weak cue for auditory grouping in frogs. This view is consistent with previous work with other frog species.

Using two-choice discrimination tests, Schwartz & Gerhardt (1995) presented females of H. versicolor with a choice between two alternating pairs of interleaved pulse sequences. Each sequence within a pair was presented at the conspecific rate of 20 pulses/s, so that the composite pulse rate of 40 pulses/s was more similar to that of H. chrysoscelis calls. In two experiments, one pair of interleaved sequences was separated by either 120° or 45°;
the alternate pair of interleaved pulse sequences was separated by 5°. Females preferentially chose one of the interleaved pulse sequences in the alternate pair with greater spatial separation when this separation was 120°, but not when it was 45°. Thus, an increase in spatial separation from 5° to 45° failed to improve perceptual segregation of the two pulse sequences. These results suggest that the two pulse sequences could be integrated over fairly wide angles of separation.

In a field playback experiment with the Australian quacking frog, *Crinia georgiana*, Farris et al. (2000) showed that males attempt to match the number of call notes in a nearby neighbour's calls. In tests with an eight-note stimulus in which the first four notes and the last four notes were broadcast from different speakers separated by 180° around the subject, males still integrated the spatially separated notes. There was also evidence for integration when the second four notes were attenuated by 6 dB relative to the first four notes, simulating two neighbours calling at different distances. Thus, under natural conditions, these frogs were fairly permissive of spatial separation between sequentially produced signal elements.

Two recent studies of the túngara frog, *Physalaemus pustulosus*, also suggest that auditory grouping of sequentially produced signal elements can occur over relatively large spatial separations. Male túngara frogs produce either a 'simple call' consisting of a downward frequency-modulated sweep (the 'whine') or 'complex calls' consisting of a whine followed by one to seven short, harmonically structured 'chucks' (Ryan 1985). Based on analyses of orientation angles in phonotaxis trials, Farris et al. (2002, 2005) found that females group the whine together with a chuck even when these two signal elements are separated by angles of up to 135°.

Conclusions

Our study of *H. chrysoscelis*, and those of *H. versicolor* (Schwartz & Gerhardt 1995), *C. georgiana* (Gerhardt et al. 2000), and *P. pustulosus* (Farris et al. 2002, 2005), suggest that common spatial origin is at best only a weak auditory grouping cue for sequential integration in frogs. The extent to which the salient acoustic (i.e. nonspatial) cues for auditory grouping identified for humans might also function in frogs and other nonhuman animals remains largely an open question. Given that animal acoustic signals often comprise sequences of harmonically rich elements, a fundamental issue for understanding the mechanisms that receivers use to perceive such signals concerns identifying the cues that promote simultaneous and sequential integration. Many of the low-level mechanisms underlying auditory scene analysis may have arisen early in the evolution of vertebrate hearing and, therefore, may be shared between humans and other vertebrates (Fay & Popper 2000; Feng & Ratnam 2000; Hulse 2002; Lewis & Fay 2004). Available data suggest that the principles of auditory scene analysis play some role in song perception in songbirds (reviewed in Hulse 2002), echolocation in bats (Moss & Surlykke 2001; Barber et al. 2003; Kanwal et al. 2003) and call perception in frogs (Schwartz & Gerhardt 1995; Simmons & Bean 2000; Farris et al. 2002, 2005; Bee 2007, 2008), penguins (Aubin & Jouventin 1998), monkeys (Miller et al. 2001; Petkov et al. 2003) and mice (Geissler & Ehret 2002, 2004). Invertebrates also appear equipped to analyse biologically relevant scenes for acoustic communication (Schul & Sheridan 2006). It will be important for future studies of animal acoustic communication to adopt auditory scene analysis as an experimental paradigm so that we may continue to understand better the evolution of communication systems and their underlying perceptual and neurophysiological mechanisms (Bee & Micheyl, in press).

Acknowledgments

We thank John Moriarty and the Ramsey County Parks Department for access to the Tamarack Nature Center, Madeleine Linck and the Three Rivers Park District for access to the Carver Park Reserve, and Ed Quinn, Mark Crawford and the Minnesota Department of Natural Resources for access to the Lake Maria State Park. Laura Cremin, Katherine Grillaert, Melanie Harrington, Eli Swanson and Sandra Tekmen provided helpful assistance with collecting and testing frogs. Vince Marshall, Beth Pettit, Johannes Schul, Josh Schwartz and Alejandro Velez provided helpful feedback on earlier versions of the manuscript. This research adhered to the legal requirements of the United States of America, all institutional guidelines, and was approved by the University of Minnesota's IACUC (no. 0510A76966) on 15 December 2005. Animal collections were made under Special Permit 13313 from the Minnesota Department of Natural Resources. This work was supported by a Grant-in-Aid from the Dean of the Graduate School of the University of Minnesota and NIDCD 1R03DC008396-01.

References


