Journal of Comparative Physiology A
https://doi.org/10.1007/s00359-024-01702-9

RESEARCH

Perceptually salient differences in a species recognition cue do not promote auditory streaming in eastern grey treefrogs (*Hyla versicolor*)

Lata Kalra[1] · **Shoshana Altman**[1] · **Mark A. Bee**[1]

Received: 3 December 2023 / Revised: 17 April 2024 / Accepted: 18 April 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Auditory streaming underlies a receiver's ability to organize complex mixtures of auditory input into distinct perceptual "streams" that represent different sound sources in the environment. During auditory streaming, sounds produced by the same source are integrated through time into a single, coherent auditory stream that is perceptually segregated from other concurrent sounds. Based on human psychoacoustic studies, one hypothesis regarding auditory streaming is that any sufficiently salient perceptual difference may lead to stream segregation. Here, we used the eastern grey treefrog, *Hyla versicolor*, to test this hypothesis in the context of vocal communication in a non-human animal. In this system, females choose their mate based on perceiving species-specific features of a male's pulsatile advertisement calls in social environments (choruses) characterized by mixtures of overlapping vocalizations. We employed an experimental paradigm from human psychoacoustics to design interleaved pulsatile sequences (ABAB…) that mimicked key features of the species' advertisement call, and in which alternating pulses differed in pulse rise time, which is a robust species recognition cue in eastern grey treefrogs. Using phonotaxis assays, we found no evidence that perceptually salient differences in pulse rise time promoted the segregation of interleaved pulse sequences into distinct auditory streams. These results do not support the hypothesis that any perceptually salient acoustic difference can be exploited as a cue for stream segregation in all species. We discuss these findings in the context of cues used for species recognition and auditory streaming.

Keywords Acoustic communication · Auditory perception · Auditory stream segregation · Grey treefrog · Species recognition · Vocal communication

Introduction

Acoustic communication, and hearing more generally, frequently requires listeners to perceive relevant sound sequences as distinct from other concurrent sounds (Cherry 1953; McDermott 2009). In humans, for example, following a conversation in noisy social settings (Repp 1988; Remez 2021) or recognizing a melody in an orchestral piece (McDermott and Oxenham 2008; Dowling 2012) involves the ability to hear sound sequences (e.g., words, syllables, musical notes) as distinct from other sounds occurring at the same time. The ability to hear distinct sound sequences amid competing sounds is a non-trivial challenge because sounds from multiple sources sum to form a composite sound wave that impinges on the ears of a listener (Bregman 1990). The composite sound wave must be perceptually organized into distinct "streams," each corresponding to a coherent representation of the sound sequence produced by a given source. This process, called "auditory streaming" (Bregman 1990), involves two complementary processes in which sounds produced by the same source are *integrated* into a coherent auditory stream while sounds produced by different sources are *segregated* into separate streams (Bregman and Campbell 1971; Bregman 1990; Moore and Gockel 2002, 2012).

Psychoacoustic studies in humans have uncovered various cues influencing the integration versus segregation of sounds during auditory streaming. Many of these studies have employed a simple experimental paradigm wherein subjects listen to interleaved sequences of two types of tone

Handling Editor: Andrea Megela Simmons.

✉ Lata Kalra
kalra023@umn.edu

[1] Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA

Published online: 11 May 2024

pulses (A and B) and report their perception of the rhythm or rate of the sequence. The acoustic differences between the A and B pulses are manipulated across trials. Integration versus segregation can be assessed using this ABAB stimulus paradigm to determine whether subjects report hearing, as a function of the acoustic differences between the A and B pulses, a single, integrated sequence (ABAB…) or two segregated sequences (A–A–… and B–B–…), each at half the pulse rate of the actual stimulus sequence (van Noorden 1975). Sufficiently large differences in the spectral content (e.g., fundamental frequency or timbre), temporal patterns (e.g., onset/offset times, amplitude and frequency modulation patterns) or spatial location of A and B sequences promote their segregation, while smaller differences are more likely to result in their integration (reviewed in Bregman 1990; Darwin 1997, 2008; Micheyl and Oxenham 2010). The breadth of acoustic cues that facilitate auditory streaming in humans led Moore and Gockel (2002) to hypothesize that "the extent to which sequential stream segregation occurs is directly related to the degree of perceptual difference between successive sounds. *Any* sufficiently salient perceptual difference may lead to stream segregation" (p. 331, emphasis original).

Many non-human animals communicate using rhythmic sequences of sounds, such as pulsatile calls in frogs and crickets (Prestwich 1994; Gerhardt and Huber 2002), song motifs in songbirds and whales (Winn et al. 1981; Hyland Bruno and Tchernichovski 2019), and echolocation clicks in bats and dolphins (Fenton et al. 2014). Moreover, these signals are perceived in complex acoustic environments consisting of multiple biotic and abiotic sound sources (Gerhardt and Huber 2002; Greenfield 2005; Bee and Micheyl 2008). Auditory streaming is thus essential for accurate recognition, discrimination, and localization of signals across diverse species and behavioral contexts. Even though auditory streaming is a ubiquitous communication challenge, the phenomenon has so far received relatively little attention in studies of non-human animal communication (Hulse 2002; Bee and Micheyl 2008; Dent and Bee 2018). Preliminary investigations using the ABAB paradigm in non-human animals suggest similar auditory streaming cues are used in humans and a diversity of other species. Frequency differences, for example, promote segregation in insects (Schul and Sheridan 2006), frogs (Nityananda and Bee 2011), fish (Fay 1998, 2000), birds (MacDougall-Shackleton et al. 1998; Itatani and Klump 2014; Dent et al. 2016), and mammals (Izumi 2002; Ma et al. 2010; Noda et al. 2013; Christison-Lagay and Cohen 2014). Temporal differences in onset/offset times and amplitude modulation patterns promote segregation in frogs (Gupta and Bee 2020) and birds (Itatani and Klump 2009). Differences in spatial location promote segregation in insects (von Helversen 1984; Weber and Thorson 1988), frogs (Farris et al. 2002, 2005; Bee 2010) and mammals (Middlebrooks and Bremen 2013; Yao et al. 2015). While these studies establish interesting parallels between auditory perception across taxa, it remains to be tested whether perceptual salience per se (sensu Moore and Gockel 2002) is sufficient to promote segregation of sounds in non-human animals.

In this study of the eastern grey treefrog, *Hyla versicolor*, we used the ABAB stimulus paradigm to test the hypothesis that perceptually salient acoustic differences promote auditory streaming. The eastern grey treefrog is a well-studied frog in the context of animal communication that breeds in ponds and wetlands distributed throughout eastern North America (Gerhardt 2001). Males of *H. versicolor* produce pulsatile advertisement calls (Fig. 1a) and breed in choruses. Even in small choruses of only conspecifics, there is a high degree of call overlap among neighboring males (Schwartz et al. 2002). In mixed-species choruses heterospecific males, including those of a morphologically indistinguishable sister species, *Hyla chrysoscelis*, also produce spectrally and temporally overlapping pulsatile advertisement calls (Fig. 1b) (Nityananda and Bee 2011). Auditory streaming is thus crucial for female frogs to perceive the signal of a potential mate amidst other concurrent sounds (Bee 2015). In *H. versicolor*, each advertisement call consists of a sequence of 11 to 25 pulses (Fig. 1a). The amplitude time envelope of each pulse has a slow (approximately 65% of pulse duration) rise from pulse onset to peak amplitude and a fast (approximately 35% of pulse duration) fall from peak amplitude to pulse offset (Fig. 1c) (Gerhardt and Doherty 1988; Ptacek et al. 1994; Gupta et al. 2021). Pulse rise time is a key species recognition cue in *H. versicolor*. Females from a population of *H. versicolor* in Missouri in the central United States prefer pulses shaped with slow rise times typical of conspecific calls (Fig. 1c) over pulses that have faster rise times and an overall shape that more closely resembles the heterospecific pulses of *H. chrysoscelis* (Fig. 1d). Rise time differences as small as 5 ms were perceptually salient and elicited strong behavioral discrimination between signals (Gerhardt and Schul 1999).

Here, we capitalized on the behavioral significance of pulse rise time as a species recognition cue in *H. versicolor* to test the hypothesis that a perceptually salient difference in pulse rise time promotes the segregation of interleaved pulse sequences into separate auditory streams. As a first step in our experimental design, we recorded and analyzed advertisement calls because pulse rise time preferences had not been established previously in our study population in Minnesota. We used the results from our acoustic analyses to determine biologically realistic values of pulse rise time for subsequent behavioral experiments. We exploited a well-known preference for conspecific calls based on pulse-rate selectivity to design experiments based on the ABAB stimulus paradigm. The pulses in a *H. versicolor* call are, on

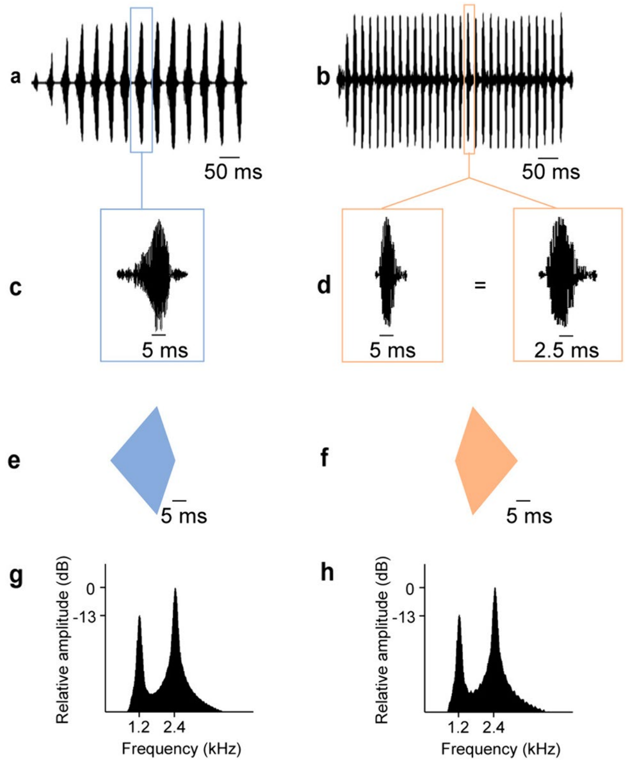


Fig. 1 Natural and synthetic signals of *H. versicolor* and *H. chrysoscelis*. **a** Oscillogram of a natural advertisement call of *H. versicolor*. **b** Oscillogram of a natural advertisement call of *H. chrysoscelis* depicting a faster pulse rate compared to *H. versicolor*. **c** A highlighted natural pulse of *H. versicolor* depicting a slow rise and relatively faster fall in amplitude. **d** left: A highlighted natural pulse of *H. chrysoscelis* (shown in the same time-scale as *H. versicolor* in panel **c**, depicting a fast rise and relatively slow fall in amplitude), and right: The same pulse with a zoomed-in timescale to highlight that the pulse rise and fall times in *H. chrysoscelis* are approximately reversed relative to that of a natural *H. versicolor* pulse. **e** Synthetic "A" pulse (in blue) modelled on the overall duration and the rise and fall-times of a natural *H. versicolor* pulse. **f** Synthetic "B" pulse (in orange), which is a digitally reversed version of "A" pulse, has an overall duration typical of a natural *H. versicolor* and an overall shape typical of a natural *H. chrysoscelis* pulse. **g**, **h** Power spectra of the synthetic pulses shown in **e** and **f**, respectively, showing the two spectral peaks in each pulse [and their relative amplitudes] of 1.2 kHz [-13 dB] and 2.4 kHz [0 dB]

average, about 20 ms long and separated by silent intervals of about 30 ms in duration. This regular rhythm corresponds to a pulse rate of 20 pulses/s (Gerhardt and Doherty 1988; Gupta et al. 2021). Females of *H. versicolor* prefer the pulse rate of conspecific calls (Fig. 1a) over the faster pulse rate of *H. chrysoscelis* calls (Fig. 1b), which is typically about 2 × to 3 × (40 to 65 pulses/s) the pulse rate of conspecific calls (Noble and Hassler 1936; Blair 1958; Gerhardt 1978; Ward et al. 2013). Accurate pulse-rate perception is crucial for species recognition as highlighted by the finding that two interleaved and identical conspecific pulse sequences are perceived by *H. versicolor* females as a single sequence having a fast pulse rate that is less attractive than the conspecific pulse rate (Schwartz and Gerhardt 1995; Schwartz and Marshall 2006; see also Bee and Riemersma 2008). Our use of the ABAB stimulus paradigm was based on a female's pulse-rate selectivity. We broadcast two interleaved sequences of pulses (ABAB…), each having the same conspecific pulse rate (each 20 pulses/s) but differing in their pulse rise time (A and B). The "A" pulses (Fig. 1e) had the pulse duration and rise time typical of conspecific pulses (Fig. 1c) and were expected to be preferred by the females. The "B" pulses (Fig. 1f) were time-reversed versions of the "A" pulses, and therefore, had a pulse duration typical of conspecific pulses but an overall shape typical of heterospecific *H. chrysoscelis* pulses (Fig. 1d). As such, the "B" pulses were expected to be less preferred by the females. We measured stream segregation based on whether, upon hearing the ABAB stimulus, the subjects perceived two separate (A–A–… and B–B–…) sequences (indicating *segregation*), wherein one of the sequences (A–A–) had a preferred pulse rise time and pulse rate, or a single (ABAB…) sequence (indicating *integration*) with a less preferred pulse rate of 40 pulses/s (Fig. 2). According to our hypothesis, we predicted subjects would be attracted to interleaved sequences that could be segregated into separate auditory streams, one of which (A–A–) was attractive, based on a perceptually salient and biologically relevant difference in pulse rise time (Fig. 2).

Methods

Subjects

All acoustic recordings and behavioral tests were conducted using subjects from the Tamarack Nature Center (Ramsey County, MN, USA), which belong to the Midwest clade of *H. versicolor* (Booker et al. 2022). Acoustic recordings of males ($n = 30$) were made at night (between 2100 and 0100 h) in May and June of 2006 and 2021. For behavioral tests, females ($n = 43$) were collected in amplexus at night (between 2100 and 0100 h) in May and June of 2021. Amplexed pairs were returned to the laboratory where they were maintained at approximately 4 °C to delay egg laying and maintain behavioral responsiveness (Gerhardt 1995). Prior to behavioral testing, frogs were placed in an incubator for at least 30 min and allowed to reach a body temperature of 20 °C. Between trials, females were returned to the 20 °C incubator with their mates for a minimum of 5 min to maintain body temperature and preserve responsiveness. Because *H. versicolor* breeds syntopically with *H. chrysoscelis* at our field site, we confirmed the species identity of all subjects in an initial two-alternative choice test in which we broadcast alternating synthetic models of the two species' calls (as

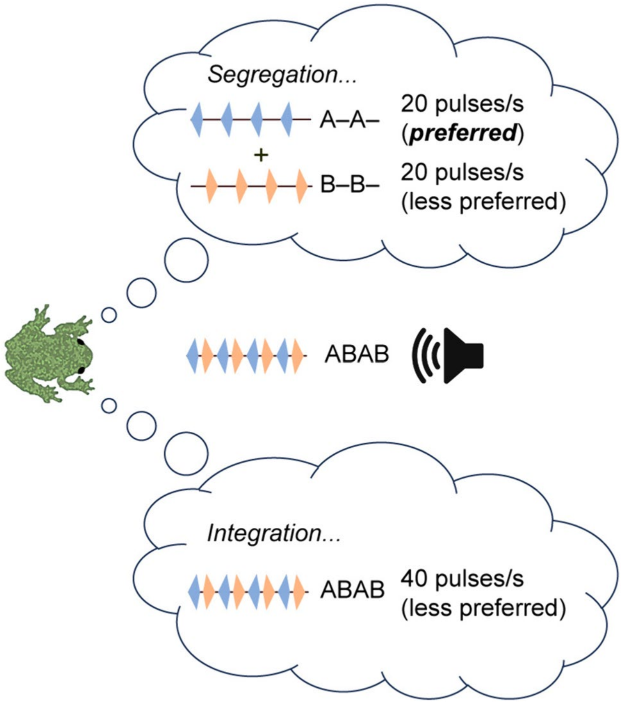


Fig. 2 Protocol for testing auditory streaming. We broadcasted interleaved pulsatile sequences of A (in blue) and B (in orange) pulses (ABAB) to female *H. versicolor*. If pulse rise time differences were sufficient to promote segregation of sounds, we expected the females to perceive two distinct sequences, A–A– and B–B–, both of which had the preferred conspecific pulse rate of 20 pulses/s and one of which (A–A–) also had the preferred conspecific pulse rise time. Consequently, the ABAB stimulus was predicted to be attractive to females. In contrast, if the rise time differences between the A and B pulses were insufficient to promote segregation, we would expect females to perceive a composite ABAB sequence as having a relatively less preferred pulse rate of 40 pulses/s

in Gupta et al. 2021). Only females that approached the *H. versicolor* stimulus were used as subjects in the experiments described below. In some case, females were also used as subjects for other experiments not described here. There is little evidence for "carryover" effects between consecutive phonotaxis tests separated by several minutes (Gerhardt 1981; Akre and Ryan 2010). All frogs were released at their collection site within 48 h of completing behavioral tests.

Acoustic recordings and analysis

Vocalizations were recorded (44.1 kHz sampling rate, 16-bit resolution) using Sennheiser ME66 or ME67 microphones (Sennheiser USA, Old Lyme, CT, USA) connected to Marantz PMD620 or PMD670 recorders (D&M Professional, Itasca, IL, USA). Microphones were held by hand or mounted on a tripod, and the tip of the microphone was positioned approximately 1 m away from the focal male. For each individual male ($n = 30$) we recorded and analyzed a minimum of 5 calls (median = 21 calls/male; range, 5 to 45 calls/male). Since both the acoustic properties of advertisement calls and female preferences for call properties can vary with temperature (Gerhardt 1978), we measured the wet-bulb air temperature and the water temperature at each male's calling site immediately following each recording. We noted the general position from which the male was calling (e.g., in air on emergent vegetation versus floating on the surface of the water) to determine the most appropriate temperature for later use to standardize call properties to a common temperature of 20 °C. We recorded males from two different ponds and from different areas within each pond across nights and years to reduce the chances of recording the same individual multiple times.

Acoustic recordings were analyzed using SoundRuler version 0.9.6.0 (Gridi-Papp 2007), which performs automatic recognition of small repeated acoustic elements and exports an output summary of numerous acoustic properties (Bee 2004). The output summary was further analyzed in R studio (R Core Team, 2020) to derive and analyze specific acoustic properties of interest for which we computed means, standard deviations (SD), and ranges. Our primary focus was on pulse rise time (ms, time from pulse onset to peak amplitude) and also pulse fall time (ms, time from pulse peak amplitude to offset). To place these two measures of pulse amplitude envelope in the overall context of the advertisement call, we also measured other temporal properties, including pulse duration (ms), pulse rate (pulses/s), call duration (pulses/call), call rate (calls/min), and spectral properties including the frequency (Hz) of each pulse's first and second harmonics, which correspond to the fundamental frequency and dominant frequency, respectively. Because the recordings were made at different temperatures, we followed Platz and Forester (1988) to standardize all call properties to 20 °C, which is close to the average temperature observed in our recordings as well as the temperature at which we performed behavioral experiments.

Acoustic stimuli

Synthetic acoustic stimuli (44.1 kHz sampling rate, 16-bit resolution) were generated in MATLAB R2020b (Mathworks, Natick, MA, USA) using parameter values taken from our acoustic analysis of natural signals. Across all experiments, stimuli were designed to stimulate a calling male and were constructed as a 5 min sequence of synthetic calls that repeated at a rate of 10 calls/min. Each call was generated as a sequence of pulses wherein each pulse was 20 ms long and composed of two phase-locked sinusoids (1.2 kHz and 2.4 kHz, corresponding to the fundamental and dominant frequencies, respectively, of the natural signals). Further, based on our acoustic analysis, the amplitude of the 1.2 kHz component was fixed to be 13 dB lower relative to the 2.4 kHz component (Fig. 1g, h). The calls within each stimulus sequence differed in the rise time, rate, and timing of their constituent pulses according to the type of phonotaxis test performed, as described next.

Experimental design

We performed three different choice tests, described below, using female phonotaxis as a behavioral assay. Each test was replicated twice using stimuli presented at one of two different sound pressure levels (80 dB peak SPL and 100 dB peak SPL). For any given choice test, all alternative stimuli had the same signal level (either 80 dB or 100 dB). We used these signal levels because auditory perception in frogs can be sound-level dependent (Gerhardt 1987, 2005a, 2008) and because these levels encompass much of the natural range of variation in the sound pressure levels of advertisement calls (Gerhardt 1975). Each subject was tested in each of the six tests (3 choice tests × 2 signal levels) in a randomized order. We used two-tailed binomial tests to compare the proportion of females choosing a specific stimulus to the chance expectation if they chose randomly. All data analysis was performed in R studio (R Core Team, 2020).

Perceptual salience test

We performed a two-alternative choice test to determine whether pulse rise time differences are perceptually salient in our study population. This test simulated a choice between two calling males producing calls having attractive pulse rates of 20 pulses/s, with each call comprising 16 pulses that differed only in pulse rise time ("A" versus "B"). "A" pulses (Fig. 1e) had slow, linear rise times (13 ms, 65% of pulse duration) and fast, linear fall times (7 ms, 35% of pulse duration). The "A–A–" stimulus (Fig. 3a) consisted of pulses that simulated the average rise time of conspecific pulses in *H. versicolor* as determined in our acoustic analysis (Noble and Hassler 1936; Blair 1958; Gerhardt 1978). In contrast, the "B–B–" stimulus (Fig. 3b) consisted of time-reversed A pulses that had fast, linear rise times (7 ms, 35% of pulse duration) and slow, linear fall times (13 ms, 65% of pulse duration). These "B" pulses (Fig. 1f) closely resembled the overall shape of pulses in the heterospecific calls of *H. chrysoscelis* (fast pulse rise ~ 35% of pulse duration and slow pulse fall ~ 58% of pulse duration; Ward et al. 2013). By using time-reversed "A" pulses in the "B–B–" stimulus, we ensured both stimuli had pulses of consistent duration and peak sound pressure levels and differed only in pulse rise time and fall time (Diekamp and Gerhardt 1995). Based on the expected biological relevance of these pulse rise-time differences, as demonstrated in previous work by Gerhardt and Schul (1999) in Missouri populations, we predicted a proportion of subjects significantly higher than 0.5 would choose the A–A– stimulus (Fig. 4a). As discussed below, this prediction was supported by the data, thus allowing us to use the perceptually salient and biologically relevant

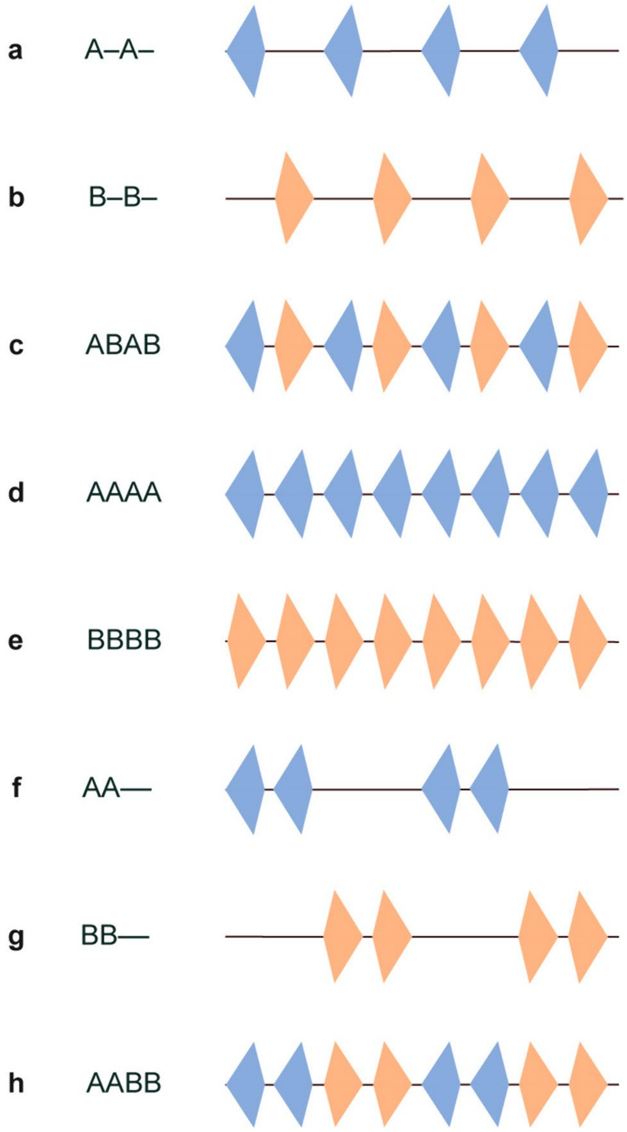
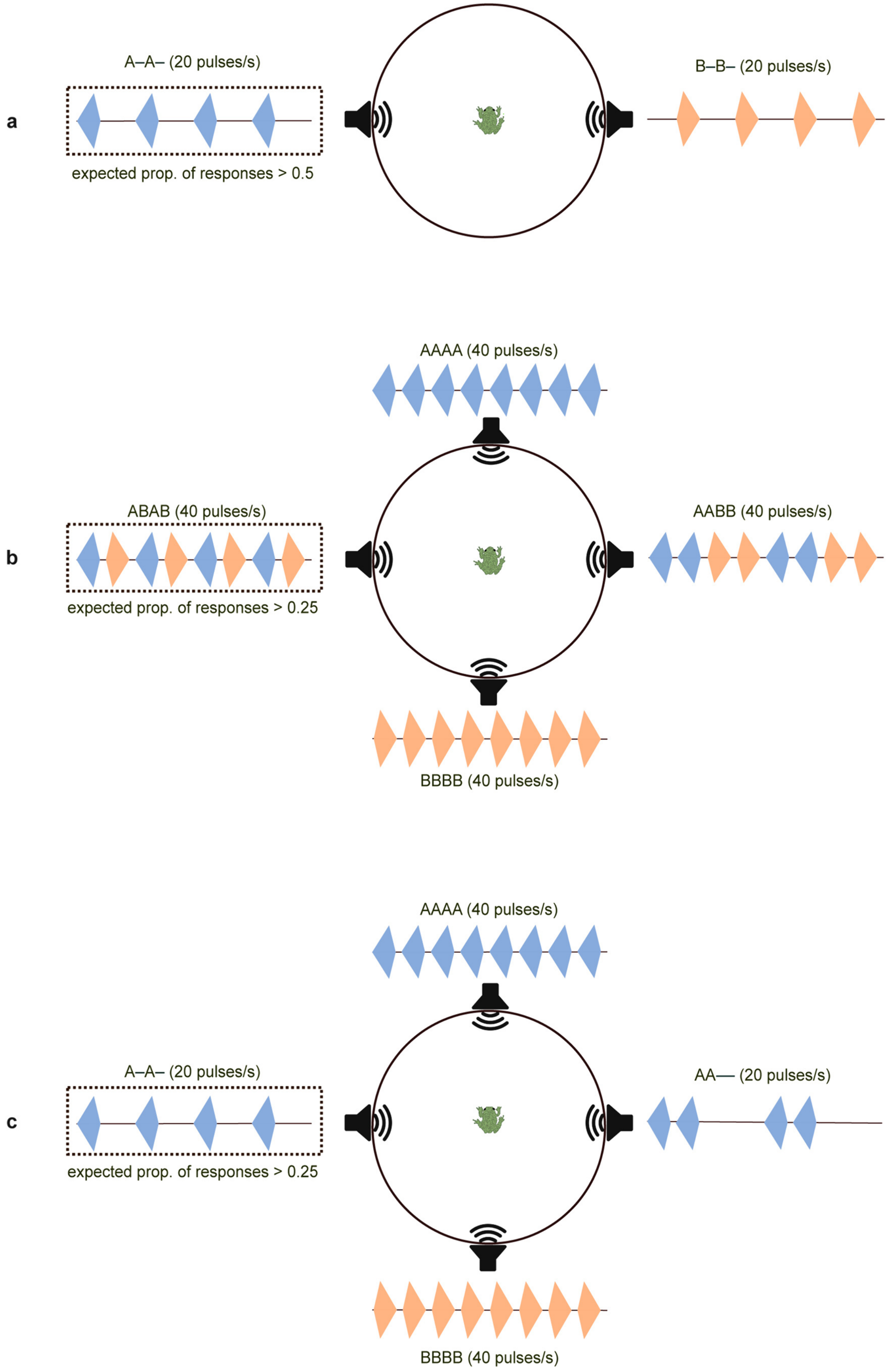


Fig. 3 Schematic segments of the acoustic stimuli for behavioral experiments. **a** "A–A–," and **b** "B–B–" were constructed as sequences of A and B pulses, respectively, repeating at a rate of 20 pulses/s and having a regular inter-pulse interval of 30 ms. **c** "ABAB" was constructed by temporally interleaving the "A–A–" and "B–B–" sequences and had a composite rate of 40 pulses/s. **d** "AAAA," and **e** "BBBB" were constructed as sequences of A and B pulses, respectively, repeating at a rate of 40 pulses/s. **f** "AA—," and **g** "BB––", were constructed as sequences of A and B pulses, respectively, repeating at an average rate of 20 pulses/s and had an irregular inter-pulse interval that shifted between 5 and 55 ms between consecutive pulses. **h** "AABB" was constructed by temporally interleaving the "AA—" and "BB—" sequences and had a composite rate of 40 pulses/s. See text for additional information on the actual numbers of pulses in each stimulus


a
A–A– (20 pulses/s)
expected prop. of responses > 0.5
B–B– (20 pulses/s)
b
AAAA (40 pulses/s)
ABAB (40 pulses/s)
expected prop. of responses > 0.25
AABB (40 pulses/s)
BBBB (40 pulses/s)
c
AAAA (40 pulses/s)
A–A– (20 pulses/s)
expected prop. of responses > 0.25
AA–– (20 pulses/s)
BBBB (40 pulses/s)

◂**Fig. 4** Design and predictions for the behavioral experiments. **a** Perceptual salience test. If the rise time differences between pulses A and B are perceptually salient, subjects were expected to prefer A–A– stimulus more than the chance probability of 0.5. **b** Auditory streaming test. If perceptually salient rise time differences are sufficient to allow auditory streaming, subjects were expected to prefer ABAB stimulus more than the chance probability of 0.25. **c** Pulse-rate and pulse-timing test. If subjects prefer calls with conspecific pulse rates and evenly spaced pulses, they were expected to prefer A–A– stimulus more than the chance probability of 0.25. Schematic segments of the acoustic stimuli are shown; see text for additional information on the actual numbers of pulses in each stimulus

differences between the "A" and "B" pulses to test an auditory streaming hypothesis.

Auditory streaming test

We used a four-alternative choice test to evaluate the hypothesis that females of *H. versicolor* can use perceptually salient difference in pulse rise time to segregate temporally overlapping calls into separate auditory streams. The key stimulus was based on the ABAB stimulus paradigm. It was created by temporally interleaving pulses from the A–A– and B–B– stimuli from the perceptual salience test to produce an "ABAB" stimulus (Fig. 3c). This ABAB stimulus had 32 pulses and a composite pulse rate of 40 pulses/s (simulating a less attractive *H. chrysoscelis* call) but was made up of two component pulse sequences (A–A– and B–B–) each having an attractive conspecific pulse rate of 20 pulses/s, only one of which (A–A–) also had the more attractive pulse rise time of conspecific calls. Two other stimuli in this four-alternative choice test also had the less preferred pulse rate of 40 pulses/s and consisted of a sequence of either all A pulses ("AAAA," Fig. 3d) or all B pulses ("BBBB," Fig. 3e). The final stimulus was created by interleaving pairs of A ("AA—," Fig. 3f) and B ("BB—," Fig. 3g) pulses so that, like the ABAB stimulus, this "AABB" (Fig. 3h) stimulus consisted of two pulse sequences having average pulse rates of 20 pulses/s. The main difference between the ABAB and AABB stimuli was that the former comprised two interleaved sequences having "regular" pulse timing (A–A– and B–B–), as determined by their constant 30 ms inter-pulse interval between consecutive pulses (typical of natural advertisement calls), whereas the component sequences in the AABB stimulus had "irregular" pulse timing (AA— and BB—) created by having inter-pulse intervals that alternated between 5 and 55 ms between consecutive pulses but averaged to 30 ms over the duration of each composite stimulus. Among all the pulse sequences used across the four stimuli, the A–A– component sequence in the ABAB stimulus was expected to be the most attractive because it was the only stimulus with the pulse rise times, pulse rate, and pulse timing typical of conspecific calls (Gerhardt and Doherty 1988; Gerhardt and Schul 1999; Gerhardt 2005b). If the perceptually salient difference between the A and B pulse rise times was sufficient to allow auditory streaming, we predicted females would be attracted to the A–A– component of the ABAB stimulus and thus choose the ABAB stimulus over the other three stimuli, which had less preferred pulse rates (AAAA, BBBB, and AABB), pulse rise times (BBBB and AABB), or pulse timing (AABB). Therefore, if auditory streaming of the interleaved A and B pulses in the ABAB stimulus occurred, we predicted that the proportion of subjects choosing the ABAB stimulus would be significantly higher than 0.25 (Fig. 4b).

Pulse-rate and pulse-timing test

We performed a final four-alternative choice test to confirm that females in our population were selective for conspecific pulse rates and regular pulse timing. The A–A– stimulus had a conspecific pulse rate of 20 pulses/s, the conspecific pulse rise time (A), and regular pulse timing. The AAAA and BBBB stimuli both had a faster pulse rate of 40 pulses/s (typical of the heterospecific calls of *H. chrysoscelis*) and regular pulse timing but differed in having either conspecific (A) or heterospecific (B) pulse rise times. Finally, the AA— stimulus had a conspecific pulse rate of 20 pulses/s (on average), the conspecific pulse rise time (A), but irregular pulse timing (alternating 5 ms and 55 ms inter-pulse intervals). We predicted that if females prefer calls with conspecific pulse rates and evenly spaced pulses—two key provisions of our test of auditory streaming—then they would choose the A–A– stimulus at a rate significantly higher than the chance proportion of 0.25 (Fig. 4c).

Testing protocol

Behavioral tests were performed in a 2-m diameter circular phonotaxis arena surrounded by a 60-cm tall wall. The arena wall was constructed from hardware cloth and black fabric to create a visually opaque but acoustically transparent barrier. The arena was set within a hemi-anechoic sound chamber (length × width × height: 2.8 × 2.3 × 2.1 m; Industrial Acoustics Company, IAC, North Aurora, IL, USA). Stimuli were broadcast from an HP ProBook 450 G6 (HP inc., Palo Alto, CA, USA) through a MOTU M4 sound card (MOTU, Inc., Cambridge, MA, USA) using Adobe Audition 3.0 (Adobe Systems Inc. San Jose, CA, USA). The output audio was amplified by a Crown XLS 1000 High-Density Power Amplifier (Crown International, Los Angeles, CA, USA) and played through one of four Orb1 speakers (Orb Audio, Sherman Oaks, CA, USA) located outside the arena wall on the floor of the sound chamber. The four speakers were evenly spaced around the circumference of the circular test arena and positioned to face inward toward the center of the arena. The sound pressure level (SPL, LCpeak, re 20 µPa) of stimuli broadcast through each speaker was measured for calibration using a sound level meter (Larson Davis Model 831, Larson Davis Inc., Depew, NY) attached to a microphone placed at the center of the arena at the same level above the floor as a subject's ears and aimed toward the speaker. In four-alternative choice tests, the alternative stimuli were broadcast through four speakers separated by 90° while in two-alternative choice tests, the alternative stimuli were broadcast through two speakers separated by 180° (Fig. 4). Because all the alternatives had the same call rate of 10 calls/min, the onsets of the subsequent calls within the same alternative stimulus were always separated by a fixed duration of 6 s (Fig. 5). Further, the temporal order of broadcasting different alternatives was such that the onset of each call of a given alternative stimulus was also separated by a fixed duration from the onset of the call of the alternative stimulus (in two-alternative choice tests) or stimuli (in four-alternative choice test) that preceded and followed it. As such, there was no temporal overlap between the calls of the same as well as different alternatives (Fig. 5). We controlled for relative the possible effects of temporal order and relative spatial position among alternative stimuli in the following ways. First, we randomized across subjects the order in which the very first calls of different stimuli were broadcast in a playback. Second, we randomized across subjects the physical speaker around the arena perimeter to which each alternative stimulus was assigned.

At the start of each test a single subject was placed at the center of the circular arena inside an acoustically transparent release cage. After a 60-s acclimation period, we started broadcast of the alternative stimulus sequences and then released the frog once two calls from each stimulus sequence had been broadcast. The frog was remotely released by lifting the lid of the release cage using a pulley system that could be operated from outside the sound chamber. As soon as the frog was released a timer was started. Broadcasts of the stimuli continued until the end of the test. Subjects' responses were monitored and scored in real time using an overhead IR camera mounted directly over the test arena and a video monitor located outside the sound chamber. Subjects were given up to 5 min to respond. A response was recorded if a subject approached to within 10 cm of a speaker and remained there for 30 s. A no-response was recorded if a frog failed to exit the release cage within 3 min after its release or if it failed to meet our response criterion within 5 min.

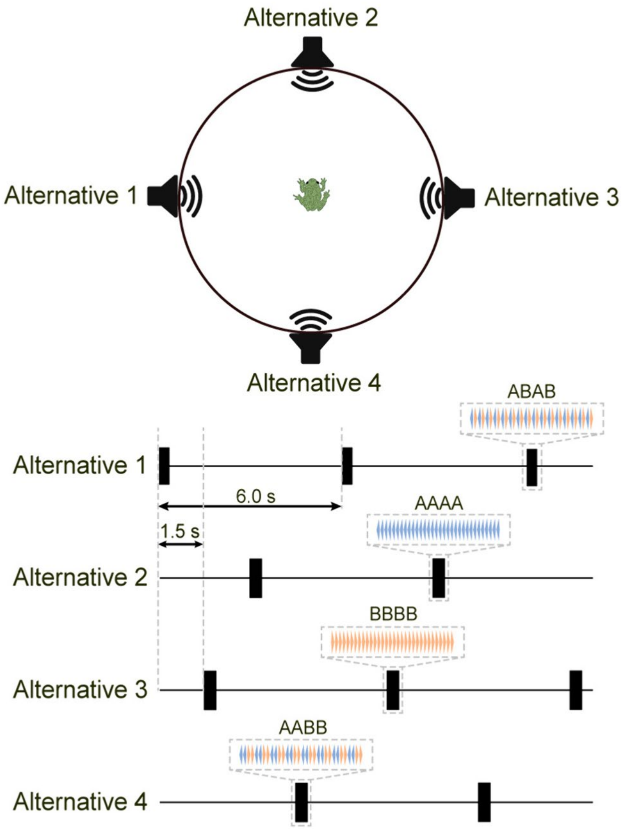


Fig. 5 Setup for broadcasting alternative stimuli in the auditory streaming test. Here, only one example of the random temporal orders and speaker assignments (spatial location) for broadcasting the alternative stimuli is shown. The temporal order and speaker assignments for broadcasting alternative stimuli was randomized across subjects. Across all the tested temporal orders and assigned speaker combinations for the broadcast, the onset of a given call was always separated by a fixed duration of 1.5 s from the onset of the call of a different alternative preceding or succeding it (as shown between the calls of alternative 1 and alternative 3) and also separated by a fixed duration of 6 s from the onset of a subsequent call within the same alternative stimulus (as shown between the two consecutive calls of alternative 1)

Results

Call analyses

The mean (± SD) rise and fall times of *H. versicolor* pulses were 13.0 ms (± 2.6 ms; range: 7.9 to 19.6 ms) and 7.4 ms (± 1.8 ms; range: 4.4 to 14.1 ms), respectively. The mean pulse duration was 20.4 ms (± 3.1 ms; range: 13.8 to 26.8 ms). Thus, on average, the pulse rise and fall times, respectively, were close to 65% and 35% of the call duration. Descriptive statistics for all other acoustic properties are reported in Table 1. Based on these results, we chose the rise and fall times of "A" pulses as 13 ms and 7 ms respectively. Since "B" pulses were digitally reversed versions of "A" pulses, the rise and fall times of "B" pulses were 7 ms and 13 ms, respectively.

Perceptual salience test

In the two-alternative choice test comprising the perceptual salience test, approximately, 98% (42 of 43) and

Table 1 Descriptive statistics of acoustic properties of *H. versicolor* advertisement calls ($n = 30$ males) recorded in Minnesota and standardized to a temperature of 20° C

Acoustic property	Mean ± SD	Range
Pulse rise time (ms)	13.0 ± 2.6	7.9–19.6
Pulse fall time (ms)	7.4 ± 1.8	4.4–14.1
Pulse rate (pulses/s)	19.3 ± 3.1	14.9–25.3
Pulse duration (ms)	20.4 ± 3.1	13.8–26.8
Call duration (pulses/call)	15.9 ± 3.3	11.1–25.1
Call rate (calls/min)	14.1 ± 4.0	5.1–22.2
Pulse fundamental frequency (Hz)	1232.5 ± 85.5	1084.7–1501.2
Pulse dominant frequency (Hz)	2465.1 ± 170.9	2169.5–3002.4

The range of temperatures at which males were recorded was 10.2° to 29.0° C

91% (39 of 43) of subjects responded by making a choice at signal levels of 100 dB and 80 dB, respectively. The data were consistent with the prediction that subjects would prefer signals having a conspecific pulse rate and slow pulse rise time (A–A–) over an alternative having a conspecific pulse rate but a fast pulse rise time and overall shape typical of heterospecific *H. chrysoscelis* pulse (B–B–). The proportion of subjects choosing the A–A– stimulus over the B–B– stimulus was significantly higher than expected by chance (0.50) at both signal levels. At 100 dB, approximately 95% of subjects (40 of 42) chose A–A– (two-tailed binomial test: $p < 0.001$), and at 80 dB, approximately 92% of subjects (36 of 39) chose A–A– (two-tailed binomial test: $p < 0.001$) (Fig. 6a). The observed behavioral discrimination based on pulse rise time confirmed that differences in pulse rise time were both perceptually and behaviorally salient.

Auditory streaming test

In the four-alternative choice test (ABAB vs. AAAA vs. BBBB vs. AABB) to investigate auditory streaming, approximately 77% (33 of 43) and 70% (30 of 43) of subjects responded by making a choice at signal levels of 100 dB and 80 dB, respectively. We predicted subjects would be attracted to the A–A– component of the ABAB stimulus if auditory streaming based on pulse rise time differences occurred. The data were not consistent with this prediction. The proportion of subjects choosing ABAB was not significantly higher than expected by chance (0.25) at either signal level. At 100 dB, approximately 15% of subjects (5 of 33) chose ABAB (two-tailed binomial test: $p = 0.231$), and at 80 dB, approximately 37% of subjects (11 of 30) chose ABAB (two-tailed binomial test: $p = 0.143$) (Fig. 6b).

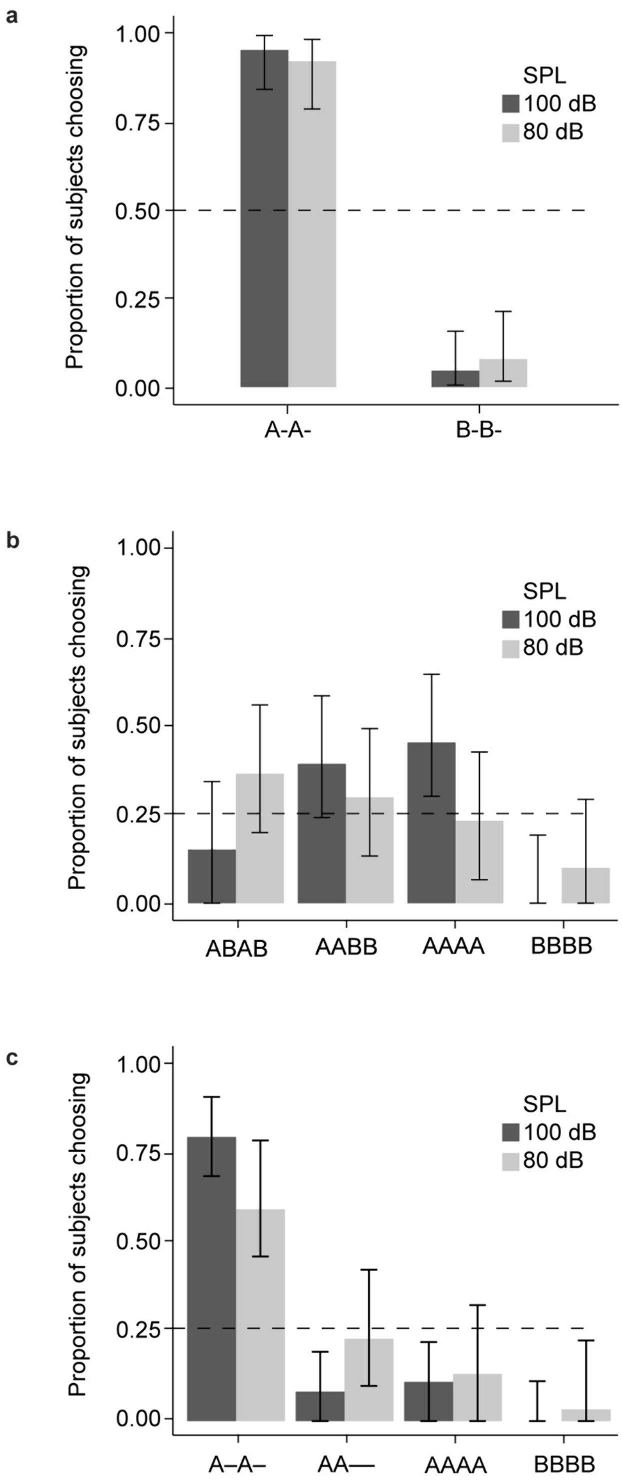


Fig. 6 Results for behavioral experiments. Black and grey bars indicate the proportions of subjects choosing a given stimulus at 100 dB and 80 dB, respectively. **a** Results for the perceptual salience test. Error bars depict exact 95% binomial confidence intervals (CIs). **b** Results for the auditory streaming experiment. Error bars depict 95% multinomial CIs. **c** Results for the pulse-rate and pulse-timing test. Error bars depict 95% multinomial CIs. Horizontal dashed lines depict the chance probability for each experiment.

Pulse-rate and pulse-timing test

In the four-alternative choice test (A–A– vs. AAAA vs. BBBB vs. AA—) to confirm pulse-rate and timing preferences, approximately 84% (36 of 43) and 72% (31 of 43) of subjects responded by making a choice at sound pressure levels of 100 dB and 80 dB, respectively. We predicted subjects would prefer the stimulus with a conspecific pulse rate and regular pulse timing (A–A–) over those with heterospecific pulse rates (AAAA and BBBB) and irregular pulse timing (AA—). The data were consistent with this prediction. The proportion of subjects choosing A–A– (slow rate and regular timing) was significantly higher than expected by chance (0.25) at both signal levels. The percentage of subjects choosing A–A– was approximately 81% (29 of 36) at 100 dB (two-tailed binomial test: $p < 0.001$) and 61% (19 of 31) at 80 dB (two-tailed binomial test: $p < 0.001$) (Fig. 6c).

Discussion

The goal of this study was to test the hypothesis (sensu Moore and Gockel 2002) that perceptual salience per se is sufficient to promote auditory streaming in non-human animals. Our results are inconsistent with this hypothesis. A species-typical difference in pulse rise time was perceptually salient, as evidenced by strong behavioral discrimination based on this acoustic cue in two-alternative choice tests. However, there was no evidence that this salient species recognition cue also promoted the perceptual segregation of two interleaved pulses sequences differing only in pulse rise time. Based on this outcome, we provisionally conclude that the perceptual salience of a biologically relevant acoustic cue was insufficient for auditory streaming in the context of segregating temporal sequences of pulses in overlapping calls in *H. versicolor*.

Our bioacoustic analyses confirmed the presence of species differences in pulse rise time between *H. versicolor* and its sister species, *H. chrysoscelis*, in Minnesota that were similar to differences reported in other populations (Gerhardt and Doherty 1988). Pulses in *H. versicolor* had rise times (range: 7.9 to 19.6 ms; Table 1) that were, on average, about 10 ms slower than those in the calls of *H. chrysoscelis* recorded in the same geographic area (range:1.8 to 4.7 ms; Ward et al. 2013). Moreover, our two-alternative choice test of perceptual salience demonstrated that a rise time difference of just 6 ms was perceptually salient and elicited a robust preference (by 92% to 95% subjects) for slow rise times. This finding corroborates previous work on pulse rise time preferences in female *H. versicolor* from a Missouri population (Gerhardt and Doherty 1988; Gerhardt and Schul 1999). Both the absolute rise times between the A and B pulses used in our study (13 ms versus 7 ms, respectively), and their relative difference (6 ms) were close to those tested by Gerhardt and Schul (1999; e.g., 12.5 ms versus 7.5 ms). Our study used 20-ms pulses with two spectral components (1.2 kHz and 2.4 kHz), whereas Gerhardt and Schul (1999) used 25-ms pulses having just the lower or the higher spectral component alone. In both studies, females of *H. versicolor* rejected a fast rise time more typical of the pulses in calls produced by male *H. chrysoscelis* in favor of a slow rise time typical of the pulses in conspecific calls. Our findings add to the evidence that pulse rise time, along with other fine temporal features like pulse rate, facilitate pre-mating species isolation between *H. versicolor* and *H. chrysoscelis,* which have spectrally similar calls (Gerhardt 2005b). As such, the present findings also contribute to our current understanding of how signal preferences may persist or vary across different populations and geographical lineages of closely-related treefrogs (e.g., Gerhardt et al. 2007; Schrode et al. 2012; Gupta and Bee 2023).

Despite strong behavioral discrimination between two pulse sequences differing in pulse rise time (i.e. A–A– versus B–B–), there was no evidence that an equivalent difference promoted auditory streaming when the same sequences were temporally interleaved (i.e., ABAB). Females did not prefer the ABAB stimulus when it was presented in a four-alternative choice test with alternatives having less preferred pulse rates, pulse rise times, or pulse timing (AAAA, BBBB, and AABB). This result is contrary to our prediction that the rise time difference would promote segregation of the ABAB sequence into separate streams, one of which corresponded to a pulse sequence (A–A–) having the preferred pulse rate, pulse rise time, and pulse timing typical of conspecific calls. Another way to empirically assess the role of pulse rise time differences in auditory streaming would have been a two-alternative choice test of ABAB versus A–A–. In that experimental design, evidence that females perceived ABAB as two streams (i.e.., A–A– and B–B–) would obtain if females responded equivalently to the ABAB and A–A– stimuli. However, this experimental design has a critical limitation: the streaming hypothesis predicts a null result. By contrast, in the experimental design we used, consisting of a four-alternative choice test, the streaming hypothesis made a directional prediction that departed from a null result. We view the four-alternative choice design to be a more robust empirical test of streaming than a two-alternative choice test that predicts a null result.

One possible explanation for the lack of any preference in our auditory streaming test could be that subjects perceptually segregated the ABAB stimulus into separate A–A– and B–B– streams based on rise time differences but behaviorally avoided the source of the perceived B–B– stream consisting of pulses with fast rise times. This explanation seems unlikely for several reasons based on other work in this species. First, females of *H. versicolor* will approach the calls of a male *H. chrysoscelis* in a no-choice test when it is the only stimulus presented, suggesting stimuli with both fast pulse rates and fast pulse rise times are not inherently aversive (Gerhardt and Doherty 1988). Second, Bush et al. (2002) and Schul and Bush (2002) showed that females responded in no-choice tests to a broad range of stimuli having different pulse rise times, including rise times faster than those of the B pulses in our stimuli. Third, Gerhardt et al. (1994) showed that females of *H. versicolor* did not avoid *H. chrysoscelis* calls while approaching a conspecific call, and Schwartz et al. (2000) showed that females of *H. versicolor* did not preferentially choose a conspecific call by itself over an identical alternative call that was paired with the call of a predator. Consistent with these findings, most females ($\geq 70\%$) chose one of the four stimuli in our auditory streaming test (including the ABAB and AABB stimuli). Results from an additional four-alternative choice test (see Supplementary Information) indicated B pulses can even be attractive in some stimulus contexts. Therefore, it seems highly unlikely that a perceived B–B– stream in the ABAB stimulus was in any way aversive in our test of auditory streaming. Finally, Stratman et al. (2021) demonstrated that females of *H. versicolor* preferentially approach small clusters of calling males over males calling in isolation. Had females perceptually segregated the ABAB stimuli into separate streams, one preferred (A–A–) and one less preferred (B–B–), then we might have expected the perceived presence of two males in close proximity to impart greater behavioral salience to the ABAB stimulus. Based on these previous studies, we interpret the lack of a significant preference for ABAB in our experiment as indicating that the pulse rise time differences did not promote auditory steaming.

Our study is the first investigation of the effects of pulse rise time differences on auditory streaming in a non-human animal. As such, our findings contribute to existing knowledge on the effect of temporal differences on auditory streaming. While our study shows no effects of pulse rise time differences on auditory streaming, it would be worth testing the same hypothesis in other species, such as in some grasshoppers that also use rise time as a behaviorally salient signal trait (Helversen 1993). Besides our study, the only other investigations of the effect of amplitude rise time alone on segregation of sounds have been in humans. Similar to our study, Hartmann and Johnson (1991) tested the segregation of sequential sound elements and found rise time differences to be a weak facilitator of stream segregation. In that study, segregation of short (4 s) interleaved sequences of melodies (A and B) having different rise times was not any better than when melodies A and B had the same rise times. In contrast to our findings and those of Hartmann and Johnson (1991), Bregman et al. (1994a, b) demonstrated that rise time differences can facilitate segregation of sounds that occur simultaneously (as opposed to sequentially). In the studies by Bregman et al. (1994a, b), the discriminability of target tones in a multi-tone complex was better when the target exhibited a sudden rise compared to the other tones in the complex. Bregman et al. (1994a, b) speculated that a sudden onset or change in amplitude of target tones may "reset" the pitch-analysis mechanisms, leading to the segregation of target tones from the complex.

The apparent inability of rise time differences to promote sequential stream segregation in our study and that by Hartmann and Johnson (1991) must be considered in light of a well-known phenomenon in auditory streaming known as the "build up" effect. During segregation of sequential sounds, the percept of two distinct streams does not arise instantaneously but instead builds up over several seconds after stimulus onset (Bregman 1978; Anstis and Saida 1985; Micheyl et al. 2005; Deike et al. 2012). Behavioral measurements in humans (Bregman 1978; Anstis and Saida 1985; Thompson et al. 2011), ferrets (Ma et al. 2010), and budgerigars (Cai et al. 2018) demonstrate that when hearing interleaved tone sequences that differ acoustically, subjects initially perceive a single stream. The probability of perceiving two streams increases as the sequence progresses. This build-up of a two-stream percept over time has been attributed to the long-term adaptation of neural responses, as demonstrated in mammals (Micheyl et al. 2005; Snyder et al. 2006; Pressnitzer et al. 2008) and songbirds (Bee et al. 2010). Importantly, previous studies on the build-up of auditory streaming used long interleaved sequences (> 10 s) and found that the build-up of a two-stream percept took several seconds (5–10 s). In contrast, the study by Hartmann and Johnson (1991), which failed to find strong evidence for sequential stream segregation based on differences in rise time, used an overall stimulus duration that was relatively short at 4 s. While our study involved similar ABAB interleaved sound sequences, our stimulus design was constrained by the requirement to stimulate natural communication signals. Consequently, one limitation of our study is that it only examined auditory streaming over relatively short sequences of pulses within calls that were < 1 s in duration. It is primarily for this reason that our main conclusion, namely that salient pulse rise time differences do not promote stream segregation in grey treefrogs, must remain provisional. The ability of perceptually salient differences in pulse rise time to impact auditory streaming using longer stimulus sequences remains to be investigated in frogs.

Previous investigations of perceptual organization in treefrogs illustrate the importance of considering both stimulus design and the perceptual task. For example, previous studies of *H. chrysoscelis* using short, call-like sequences of pulses similar to those used in the present study have revealed the importance of common onsets/offsets (Gupta and Bee 2020) and common spatial location (Bee 2010) in promoting simultaneous integration of the two harmonics in the pulses of grey treefrogs calls. In contrast to the study by Bee (2010), the effect of spatial separation between consecutive pulses in short, call-like pulse sequences had markedly less impact on promoting sequential segregation in both *H. versicolor* (Schwartz and Gerhardt 1995; Schwartz and Del Monte 2019) and *H. chrysoscelis* (Bee and Riemersma 2008). This discrepancy in the strength of spatial separation as a segregation cue across sequential versus simultaneous segregation tasks parallels the contrast between findings on pulse rise time from the present study of frogs and those of humans by Bregman et al. (1994a, b a, b). One study of sequential segregation in *H. chrysoscelis* found that females could segregate a short, call-like sequence of pulses (A–A–) that was periodically interleaved with the pulses in a long (5 min) and continuous sequence of pulses (B–B–) differing in frequency, provided there was sufficient frequency separation between the A and B pulses (Nityananda and Bee 2011). Whether pulse rise time differences might promote auditory streaming using a similar stimulus paradigm remains to be investigated.

Finally, it is also worth considering the lack of an effect of pulse rise time in the light of complex cue interactions during auditory streaming. In natural auditory scenes, multiple cues, or acoustic differences, are available to a receiver and may be differentially weighed during auditory streaming. For instance, Elhilali et al. (2009) tested auditory streaming in a cue conflict scenario using two sequences (A–A– and B–B–) that exhibited fairly large frequency separation, which promotes segregation, but shared coherent temporal onsets and offsets, which promotes integration. They found that coherent temporal onsets/offsets override frequency separation during auditory streaming as human subjects reported hearing a single stream (indicating integration). In other cases, different cues can also impact auditory streaming in an additive fashion. Micheyl et al. (2013), for example, found that inharmonicity (sounds having different fundamental frequencies) and temporal incoherence additively facilitate the segregation of sounds in humans. Importantly, while Hartmann and Johnson (1991) showed a weak effect of rise times on sound segregation (in the absence of frequency differences), Singh and Bregman (1997) showed an additive effect of rise times and frequency differences on stream segregation in humans. Non-human animals also incorporate cue interactions during auditory streaming, as seen for European starlings, Budgerigars and Zebra finches (Dent et al. 2016; Itatani and Klump 2020). Additionally, there is also evidence for no interaction, as shown by Schwartz and Del Monte (2019) for spectral and spatial cues in *Hyla versicolor*. In the present study, the A and B pulses in the ABAB stimulus had the same carrier frequencies and originated from the same spatial location, but had different pulse rise times. It might be the case that spectral similarity, common spatial location, or both override pulse rise time differences during auditory streaming. In such a case, we would expect that spectral similarity and common spatial location between A and B pulses promote their integration irrespective of the differences in pulse rise times, which is in line with the findings of this study. It remains a possibility that differences in pulse rise time might additively interact with spectral or spatial separation to promote stream segregation. Additional studies that manipulate pulse rise time along with other potential cues will be needed to uncover any interaction effects.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00359-024-01702-9.

Acknowledgements We thank Saumya Gupta, Katie Krueger, Sophie Barno, Kaitlyn Bonnema, Sarah Bonnema, Olivia Groth, Rishi Gulati, Collin Meyer and Annika Ruppert for their help collecting and testing frogs, and Mike Goodnature from Ramsey County Parks and Recreation for generous access to collection sites. We are also grateful to Marlene Zuk, Andrew Oxenham, Katie Krueger, Satyabhama de Oliveira, Andrea Simmons, and two anonymous referees for feedback on this manuscript.

Author contributions LK and MAB: designed the study. LK and SA: conducted experiments. LK and SA: analyzed data. LK, SA and MAB: prepared the manuscript.

Funding This research was funded in part by grants to MAB from the National Science Foundation (IOS-1452831 and IOS-2022253) and by grants and fellowships to LK from the University of Minnesota Graduate Program in Ecology, Evolution, and Behavior, the Bell Museum of Natural History, the University of Minnesota Graduate School and the Animal Behavior Society.

Data availability The experimental data collected and analyzed for this study are available in the Digital Repository of the University of Minnesota (https://doi.org/10.13020/jw2s-pn84).

Declarations

Conflict of interest The authors declare no competing or financial interests.

Ethical approval This research was approved by the University of Minnesota Institutional Animal Care and Use Committee (#0602A-81890 and #2001-37746A).

References

Akre KL, Ryan MJ (2010) Proximity-dependent response to variably complex mating signals in túngara frogs (*Physalaemus pustulosus*). Ethology 116:1138–1145. https://doi.org/10.1111/j.1439-0310.2010.01825.x

Anstis S, Saida S (1985) Adaptation to auditory streaming of frequency-modulated tones. J Exp Psychol Hum Percept Perform 11:257–271. https://doi.org/10.1037/0096-1523.11.3.257

Bee MA (2004) Equipment review: Sound Ruler acoustical analysis: a free, open code, multi-platform sound analysis and graphing package. Bioacoustics 14:171–178. https://doi.org/10.1080/09524622.2004.9753520

Bee MA (2010) Spectral preferences and the role of spatial coherence in simultaneous integration in gray treefrogs (*Hyla chrysoscelis*). J Comp Psychol 124:412–424. https://doi.org/10.1037/a0020307

Bee MA (2015) Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. Int J Psychophysiol 95:216–237. https://doi.org/10.1016/j.ijpsycho.2014.01.004

Bee MA, Micheyl C (2008) The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? J Comp Psychol 122:235–251. https://doi.org/10.1037/0735-7036.122.3.235

Bee MA, Riemersma KK (2008) Does common spatial origin promote the auditory grouping of temporally separated signal elements in grey treefrogs? Anim Behav 76:831–843. https://doi.org/10.1016/j.anbehav.2008.01.026

Bee MA, Micheyl C, Oxenham AJ, Klump GM (2010) Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. J Comp Physiol A 196:543–557. https://doi.org/10.1007/s00359-010-0542-4

Blair WF (1958) Call difference as an isolation mechanism in Florida species of hylid frogs. Q J Fla Acad Sci 21:32–48

Booker WW, Gerhardt HC, Lemmon AR et al (2022) The complex history of genome duplication and hybridization in North American gray treefrogs. Mol Biol Evol 39:msab316. https://doi.org/10.1093/molbev/msab316

Bregman AS (1978) Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform 4:380–387. https://doi.org/10.1037/0096-1523.4.3.380

Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. MIT press, Cambridge

Bregman AS, Campbell J (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol 89:244–249. https://doi.org/10.1037/h0031163

Bregman AS, Ahad PA, Kim J, Melnerich L (1994a) Resetting the pitch-analysis system: 1. effects of rise times of tones in noise backgrounds or of harmonics in a complex tone. Percept Psychophys 56:155–162. https://doi.org/10.3758/BF03213894

Bregman AS, Ahad PA, Kim J (1994b) Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. J Acoust Soc Am 96:2694–2703. https://doi.org/10.1121/1.411277

Bush SL, Gerhardt HC, Schul J (2002) Pattern recognition and call preferences in treefrogs (Anura: Hylidae): a quantitative analysis using a no-choice paradigm. Anim Behav 63:7–14. https://doi.org/10.1006/anbe.2001.1880

Cai H, Screven LA, Dent ML (2018) Behavioral measurements of auditory streaming and build-up by budgerigars (*Melopsittacus undulatus*). J Acoust Soc Am 144:1508–1516. https://doi.org/10.1121/1.5054297

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979. https://doi.org/10.1121/1.1907229

Christison-Lagay KL, Cohen YE (2014) Behavioral correlates of auditory streaming in rhesus macaques. Hear Res 309:17–25. https://doi.org/10.1016/j.heares.2013.11.001

Darwin CJ (1997) Auditory grouping. Trends Cogn Sci 1:327–333. https://doi.org/10.1016/S1364-6613(97)01097-8

Darwin CJ (2008) Spatial hearing and perceiving sources. In: Yost WA, Popper AN, Fay RR (eds) Auditory perception of sound sources. Springer, Boston, pp 215–232. https://doi.org/10.1007/978-0-387-71305-2_8

Deike S, Heil P, Böckmann-Barthel M, Brechmann A (2012) The build-up of auditory stream segregation: a different perspective. Front Psychol 3:461. https://doi.org/10.3389/fpsyg.2012.00461

Dent ML, Bee MA (2018) Principles of auditory object formation by nonhuman animals. In: Slabbekoorn H, Dooling R, Popper A, Fay R (eds) Effects of anthropogenic noise on animals. Springer, New York, pp 47–82. https://doi.org/10.1007/978-1-4939-8574-6_3

Dent ML, Martin AK, Flaherty MM, Neilans EG (2016) Cues for auditory stream segregation of birdsong in budgerigars and zebra finches: effects of location, timing, amplitude, and frequency. J Acoust Soc Am 139:674–683. https://doi.org/10.1121/1.4941322

Diekamp B, Gerhardt HC (1995) Selective phonotaxis to advertisement calls in the gray treefrog *Hyla versicolor*: behavioral experiments and neurophysiological correlates. J Comp Physiol A 177:173–190. https://doi.org/10.1007/BF00225097

Dowling WJ (2012) Music perception. In: Plack CJ (ed) Oxford handbook of auditory science: hearing. Oxford library of psychology, Oxford, pp 231–248

Elhilali M, Ma L, Micheyl C et al (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. Neuron 61:317–329. https://doi.org/10.1016/j.neuron.2008.12.005

Farris HE, Rand AS, Ryan MJ (2002) The effects of spatially separated call components on phonotaxis in túngara frogs: evidence for auditory grouping. Brain Behav Evol 60:181–188. https://doi.org/10.1159/000065937

Farris HE, Rand AS, Ryan MJ (2005) The effects of time, space and spectrum on auditory grouping in túngara frogs. J Comp Physiol A 191:1173–1183. https://doi.org/10.1007/s00359-005-0041-1

Fay RR (1998) Auditory stream segregation in goldfish (*Carassius auratus*). Hear Res 120:69–76. https://doi.org/10.1016/S0378-5955(98)00058-6

Fay RR (2000) Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). J Assoc Res Otolaryngol 1:120–128. https://doi.org/10.1007/s101620010015

Fenton B, Jensen FH, Kalko EK, Tyack PL (2014) Sonar signals of bats and toothed whales. In: Surlykke A, Nachtigall PE, Fay RR, Popper AN (eds) Biosonar. Springer, New York, pp 11–59. https://doi.org/10.1007/978-1-4614-9146-0_2

Gerhardt HC (1975) Sound pressure levels and radiation patterns of the vocalizations of some North American frogs and toads. J Comp Physiol A 102:1–12. https://doi.org/10.1007/BF00657481

Gerhardt HC (1978) Temperature coupling in the vocal communication system of the gray tree frog, *Hyla versicolor*. Science 199:992–994. https://doi.org/10.1126/science.199.4332.992

Gerhardt HC (1981) Mating call recognition in the green treefrog (*Hyla cinerea*): importance of two frequency bands as a function of sound pressure level. J Comp Physiol A 144:9–16. https://doi.org/10.1007/BF00612792

Gerhardt HC (1987) Evolutionary and neurobiological implications of selective phonotaxis in the green treefrog, *Hyla cinerea*. Anim Behav 35:1479–1489. https://doi.org/10.1016/S0003-3472(87)80020-9

Gerhardt HC (1995) Phonotaxis in female frogs and toads: execution and design of experiments. In: Klump GM, Dooling RJ, Fay RR, Stebbins WC (eds) Methods in comparative psychoacoustics. Basel, Birkhäuser, pp 209–220. https://doi.org/10.1007/978-3-0348-7463-2_18

Gerhardt HC (2001) Acoustic communication in two groups of closely related treefrogs. Adv Study Behav 30:99–167. https://doi.org/10.1016/S0065-3454(01)80006-1

Gerhardt HC (2005a) Acoustic spectral preferences in two cryptic species of grey treefrogs: implications for mate choice and sensory mechanisms. Anim Behav 70:39–48. https://doi.org/10.1016/j.anbehav.2004.09.021

Gerhardt HC (2005b) Advertisement-call preferences in diploid-tetraploid treefrogs (*Hyla chrysoscelis* and *Hyla versicolor*): implications for mate choice and the evolution of communication systems. Evolution 59:395–408. https://doi.org/10.1111/j.0014-3820.2005.tb00998.x

Gerhardt HC (2008) Phonotactic selectivity in two cryptic species of gray treefrogs: effects of differences in pulse rate, carrier frequency and playback level. J Exp Biol 211:2609–2616. https://doi.org/10.1242/jeb.019612

Gerhardt HC, Doherty JA (1988) Acoustic communication in the gray treefrog, *Hyla versicolor*: evolutionary and neurobiological implications. J Comp Physiol A 162:261–278. https://doi.org/10.1007/BF00606090

Gerhardt HC, Huber F (2002) Acoustic communication in insects and anurans: common problems and diverse solutions. University of Chicago Press, Chicago

Gerhardt HC, Schul J (1999) A quantitative analysis of behavioral selectivity for pulse rise-time in the gray treefrog, *Hyla versicolor*. J Comp Physiol A 185:33–40. https://doi.org/10.1007/s003590050363

Gerhardt HC, Dyson ML, Tanner SD, Murphy CG (1994) Female treefrogs do not avoid heterospecific calls as they approach conspecific calls: implications for mechanisms of mate choice. Anim Behav 47:1323–1332. https://doi.org/10.1006/anbe.1994.1180

Gerhardt HC, Martínez-Rivera CC, Schwartz JJ et al (2007) Preferences based on spectral differences in acoustic signals in four species of treefrogs (Anura: Hylidae). J Exp Biol 210:2990–2998. https://doi.org/10.1242/jeb.006312

Greenfield MD (2005) Mechanisms and evolution of communal sexual displays in arthropods and anurans. Adv Study Behav 35:1–62. https://doi.org/10.1016/S0065-3454(05)35001-7

Gridi-Papp M (2007) Sound ruler. Acoustic Analysis. Version 0.9.6.0.

Gupta S, Bee MA (2020) Treefrogs exploit temporal coherence to form perceptual objects of communication signals. Biol Lett 16:20200573. https://doi.org/10.1098/rsbl.2020.0573

Gupta S, Bee MA (2023) Female preferences for the spectral content of advertisement calls in Cope's gray treefrog (*Hyla chrysoscelis*). J Comp Physiol A 209:31–45. https://doi.org/10.1007/s00359-022-01575-w

Gupta S, Alluri RK, Rose GJ, Bee MA (2021) Neural basis of acoustic species recognition in a cryptic species complex. J Exp Biol 224:jeb243405. https://doi.org/10.1242/jeb.243405

Hartmann WM, Johnson D (1991) Stream segregation and peripheral channeling. Music Percept 9:155–183. https://doi.org/10.2307/40285527

Hulse SH (2002) Auditory scene analysis in animal communication. Adv Study Behav 31:163–200. https://doi.org/10.1016/s0065-3454(02)80008-0

Hyland Bruno J, Tchernichovski O (2019) Regularities in zebra finch song beyond the repeated motif. Behav Processes 163:53–59. https://doi.org/10.1016/j.beproc.2017.11.001

Itatani N, Klump GM (2009) Auditory streaming of amplitude-modulated sounds in the songbird forebrain. J Neurophysiol 101:3212–3225. https://doi.org/10.1152/jn.91333.2008

Itatani N, Klump GM (2014) Neural correlates of auditory streaming in an objective behavioral task. Proc Natl Acad Sci 111:10738–10743. https://doi.org/10.1073/pnas.1321487111

Itatani N, Klump GM (2020) Interaction of spatial and non-spatial cues in auditory stream segregation in the European starling. Eur J Neurosci 51:1191–1200. https://doi.org/10.1111/ejn.13716

Izumi A (2002) Auditory stream segregation in Japanese monkeys. Cognition 82:B113–B122. https://doi.org/10.1016/S0010-0277(01)00161-5

Ma L, Micheyl C, Yin P et al (2010) Behavioral measures of auditory streaming in ferrets (*Mustela putorius*). J Comp Psychol 124:317–330. https://doi.org/10.1037/a0018273

MacDougall-Shackleton SA, Hulse SH, Ball GF (1998) Neural bases of song preferences in female zebra finches (*Taeniopygia guttata*). NeuroReport 9:3047–3052. https://doi.org/10.1097/00001756-199809140-00024

McDermott JH (2009) The cocktail party problem. Curr Biol 19:R1024–R1027. https://doi.org/10.1016/j.cub.2009.09.005

McDermott JH, Oxenham AJ (2008) Music perception, pitch, and the auditory system. Curr Opin Neurobiol 18:452–463. https://doi.org/10.1016/j.conb.2008.09.005

Micheyl C, Oxenham AJ (2010) Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. Hear Res 266:36–51. https://doi.org/10.1016/j.heares.2009.09.012

Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48:139–148. https://doi.org/10.1016/j.neuron.2005.08.039

Micheyl C, Kreft H, Shamma S, Oxenham AJ (2013) Temporal coherence versus harmonicity in auditory stream formation. J Acoust Soc Am 133:EL188–EL194. https://doi.org/10.1121/1.4789866

Middlebrooks JC, Bremen P (2013) Spatial stream segregation by auditory cortical neurons. J Neurosci 33:10986–11001. https://doi.org/10.1523/JNEUROSCI.1065-13.2013

Moore BCJ, Gockel H (2002) Factors influencing sequential stream segregation. Acta Acust United with Acust 88:320–333

Moore BCJ, Gockel H (2012) Properties of auditory stream formation. Philos Trans R Soc B Biol Sci 367:919–931. https://doi.org/10.1098/rstb.2011.0355

Nityananda V, Bee MA (2011) Finding your mate at a cocktail party: frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. PLoS ONE 6:e21191. https://doi.org/10.1371/journal.pone.0021191

Noble GK, Hassler WG (1936) Three Salientia of geographic interest from southern Maryland. Copeia 1936:63. https://doi.org/10.2307/1436381

Noda T, Kanzaki R, Takahashi H (2013) Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. PLoS ONE 8:e83544. https://doi.org/10.1371/journal.pone.0083544

Platz JE, Forester DC (1988) Geographic variation in mating call among the four subspecies of the chorus frog: *Pseudacris triseriata* (Wied). Copeia 1988:1062–1066. https://doi.org/10.2307/1445734

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128. https://doi.org/10.1016/j.cub.2008.06.053

Prestwich KN (1994) The energetics of acoustic signaling in anurans and insects. Integr Comp Biol 34:625–643. https://doi.org/10.1093/icb/34.6.625

Ptacek MB, Gerhardt HC, Sage RD (1994) Speciation by polyploidy in treefrogs: multiple origins of the tetraploid, *Hyla versicolor*. Evolution 48:898–908. https://doi.org/10.2307/2410495

Remez RE (2021) Perceptual organization of speech. Handb Speech Percept 18:1–27. https://doi.org/10.1002/9781119184096.ch1

Repp BH (1988) Integration and segregation in speech perception. Lang Speech 31:239–271. https://doi.org/10.1177/002383098803100302

Schrode KM, Ward JL, Vélez A, Bee MA (2012) Female preferences for spectral call properties in the western genetic lineage of Cope's gray treefrog (*Hyla chrysoscelis*). Behav Ecol Sociobiol 66:1595–1606. https://doi.org/10.1007/s00265-012-1413-5

Schul J, Bush SL (2002) Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. Proc R Soc B Biol Sci 269:1847–1852. https://doi.org/10.1098/rspb.2002.2092

Schul J, Sheridan RA (2006) Auditory stream segregation in an insect. Neuroscience 138:1–4. https://doi.org/10.1016/j.neuroscience.2005.11.023

Schwartz JJ, Gerhardt HC (1995) Directionality of the auditory system and call pattern recognition during acoustic interference in the gray treefrog, *Hyla versicolor*. Audit Neurosci 1:195–206

Schwartz JJ, Marshall VT (2006) Forms of call overlap and their impact on advertisement call attractiveness to females of the gray treefrog, *Hyla versicolor*. Bioacoustics 16:39–56. https://doi.org/10.1080/09524622.2006.9753563

Schwartz JJ, Serratto Del Monte ME (2019) Spatially-mediated call pattern recognition and the cocktail party problem in treefrog choruses: can call frequency differences help during signal overlap? Bioacoustics 28:312–328. https://doi.org/10.1080/09524622.2018.1443836

Schwartz JJ, Bee MA, Tanner SD (2000) A behavioral and neurobiological study of the responses of gray treefrogs, *Hyla versicolor*, to the calls of a predator, *Rana catesbeiana*. Herpetologica 56:27–37

Schwartz JJ, Buchanan BW, Gerhardt HC (2002) Acoustic interactions among male gray treefrogs, *Hyla versicolor*, in a chorus setting. Behav Ecol Sociobiol 53:9–19. https://doi.org/10.1007/s00265-002-0542-7

Singh PG, Bregman AS (1997) The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. J Acoust Soc Am 102:1943–1952. https://doi.org/10.1121/1.419688

Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18:1–13. https://doi.org/10.1162/089892906775250021

Stratman KD, Oldehoeft EA, Höbel G (2021) Woe is the loner: female treefrogs prefer clusters of displaying males over single "hotshot" males. Evolution 75:3026–3036. https://doi.org/10.1111/evo.14376

Thompson SK, Carlyon RP, Cusack R (2011) An objective measurement of the build-up of auditory streaming and of its modulation by attention. J Exp Psychol Hum Percept Perform 37:1253–1262. https://doi.org/10.1037/a0021925

van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. Dissertation, University of Technology, Eindhoven

von Helversen D (1984) Parallel processing in auditory pattern recognition and directional analysis by the grasshopper *Chorthippus biguttulus L.* (Acrididae). J Comp Physiol A 154:837–846. https://doi.org/10.1007/BF00610684

von Helversen D (1993) "Absolute steepness" of ramps as an essential cue for auditory pattern recognition by a grasshopper (Orthoptera; Acrididae; *Chorthippus biguttulus L.*). J Comp Physiol A 172:633–639. https://doi.org/10.1007/BF00213685

Ward JL, Love EK, Vélez A et al (2013) Multitasking males and multiplicative females: dynamic signalling and receiver preferences in Cope's grey treefrog. Anim Behav 86:231–243. https://doi.org/10.1016/j.anbehav.2013.05.016

Weber T, Thorson J (1988) Auditory behavior of the cricket—IV. Interaction of direction of tracking with perceived temporal pattern in split-song paradigms. J Comp Physiol A 163:13–22. https://doi.org/10.1007/BF00611992

Winn HE, Thompson TJ, Cummings WC et al (1981) Song of the humpback whale - population comparisons. Behav Ecol Sociobiol 8:41–46. https://doi.org/10.1007/BF00302842

Yao JD, Bremen P, Middlebrooks JC (2015) Emergence of spatial stream segregation in the ascending auditory pathway. J Neurosci 35:16199–16212. https://doi.org/10.1523/JNEUROSCI.3116-15.2015

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.