Signal Perception in Complex Acoustic Environments: A Study on Treefrog Communication

A DISSERTATION
SUBMITTED TO THE FACULTY OF
THE UNIVERSITY OF MINNESOTA
BY

Lata Kalra

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advised by Dr. Mark A. Bee

August 2024

Acknowledgements

This dissertation would not have been possible without the immense support from multiple people in my professional and personal life. I am beyond grateful to each of these people.

First, I am immensely grateful to my advisor, Mark Bee, for all the time and effort that he spent on mentoring me. He always taught me to think deeply about the rigor and intricacies of my experiments but also to never lose sight of the big picture. His feedback on my research ideas, talks, proposals and manuscripts has been pivotal to my growth as a scientist, communicator and teacher. He always provided me with opportunities that have been immensely helpful for my career goals. I wouldn’t be where I am today without him.

I am also thankful to my committee members, Andrew Oxenham, Marlene Zuk and Trevor Wardill for stimulating discussions on my ideas and feedback on my manuscripts and dissertation. Andrew’s feedback was immensely helpful in designing my experiments and connecting my work to the field of human psychoacoustics. Attending his course ‘Proseminar in perception’ played a significant role in shaping my initial ideas for this dissertation. Marlene’s feedback has always challenged me to think about the big picture and communicate my work with the right mix of detail and concision. My discussions with Trevor have helped me strengthen the neuroethological perspectives of my research. My research collaboration with him introduced me to a new system and the wonderful world of visual perception.

I am also grateful to my collaborator Gerlinde Höbel for spending a generous amount of time helping me set up my first field experiment in Georgia. Her willingness to share her expertise on the study system was instrumental in me bringing a challenging project to realization. A big thanks to Shoshana Altman for her wonderful contributions to one of my chapters and her immensely helpful precision while executing the experiments and data analysis.

I am genuinely grateful to all the lab members of the animal communication lab – Saumya Gupta, Hongyu Li, Katie LaBarbera, Katie Krueger, Satyabhama Weihermann and Megan Freiler – for being a great support system for me and my research projects. I am particularly thankful to Katie Krueger for being the perfect lab sister to me. I appreciate everything she has done for me the past few years. I also am tremendously thankful to all the undergraduates at UMN who spent hours collecting frogs and helping me with behavioral experiments. I am especially indebted to

Miranda Dahl, Madison Deile and Isabelle Hoversten for taking the plunge to drive with me to Georgia and their willingness to mess up their sleep schedules for my field experiments.

This work would not have reached fruition without the financial support from University of Minnesota (UMN) graduate school, College of Biological Sciences at UMN, the Ecology, Evolution, and Behavior department at UMN, the Bell Museum at UMN and the Animal Behavior Society. I am also thankful to neil jahren and Kate Barry for all the time they spent in ensuring that I smoothly navigate the administrative steps for my PhD degree. I also wholeheartedly appreciate the efforts made by all the past and present Directors of Graduate Study and the Heads of Department to ensure a good graduate school experience for EEB graduate students.

I am appreciative of the efforts of the EEB graduate student body towards building a welcoming and connected community for graduate students. I am especially thankful for the EEB international student community for everything that it stands for. A big thanks to my cohort especially Abby, Brie, Janine, Lauren, Max, Megan, Molly, Sean and Taz for being the most welcoming community ever. I will always cherish the memories of all the foundations shenanigans, food parties, game nights, birthday cards and writing retreats with you all.

I feel blessed to have so many non-human animal friends. Thank you Bronzie, Bruno, Lorry, Mang, Mocha, Nia, Rosie and Sidney for coming in my life. I greatly admire my Minnesota community of friends. Thanks to Amod, Chau, Kavy, Laura, Neha, Parikha, Satya, Saumya, Shivona, Swastik and Sutirtha for so many fun-filled times. I am also grateful to my friends and past lab members from India – Anindya, Harshal, Sonam and Sushma – for wonderful philosophical discussions and motivation during some difficult phases of my PhD. A big thanks to Naven, Aarcha and Siddhant for being a great source of support and comfort throughout my PhD journey. I am grateful to have Rimpy and Pragya in my life. I can’t express how much their empathy and unconditional love has shaped me as a human being and a friend. Finding two beautiful best friends in Diksha and Aayush is one of the best things that has happened to me. They inspire me to love myself every day and their confidence in my values and dreams makes me feel that I can achieve anything I want. I might have never overcome some big emotional challenges without them. I am grateful to my partner Khilav, for filling my life with so much love, laughter and deadlifts. He is my home and I am thankful for everything he does for me every single day.

Finally, I am where I am today because of the sacrifices my parents, Rajni Kalra and Hari Mohan Kalra, have made to ensure that I study whatever and wherever I wish to. They have always supported me in my academic and non-academic endeavors and respected my ability to choose and decide for myself. They have always taught me the value of hard work as well as the importance of cherishing small moments in life with your loved ones. Big thanks to my brother Shashank for shared love for atrocious movies and songs and for understanding my struggles with adulting. I also want to thank my extended family for always showering me with so much love and teaching me the importance of sharing.

Dedication

To my grandfather who was my first teacher and my first friend

Abstract

How animals organize complex sensory inputs to form behaviorally meaningful percepts of their environments is a fundamental question in animal behavior. For acoustically communicating animals, behavioral decisions rely on their ability to recognize signals of interest in dense social aggregations. In such environments, multiple individuals produce overlapping signals that sum to form a composite sound wave impinging on a receiver's eardrums. Signal recognition requires a receiver to perceptually organize the composite sound wave such that the acoustic components belonging to a signal of interest are integrated into a unified perceptual representation and are also perceptually segregated from other overlapping signals. Although signal recognition has been thoroughly studied, the principles governing perceptual organization have remained relatively unexplored. Drawing parallels from studies on human perception, this dissertation aims to identify the principles of perceptual organization in non-human animals. Specifically, this work builds on the current understanding of vocal communication in treefrogs to investigate how female treefrogs recognize signals of potential mates in dense breeding aggregations. First, I examine how receivers perceptually organize simultaneous spectral components, like harmonics and formants, belonging to overlapping signals. Second, I investigate the perceptual organization of temporally repeated signal components into distinct signal sequences. Finally, I estimate how recognition of temporal patterns within signals is impacted by the presence of multiple concurrent sequences in the environment. Together, this work contributes to the understanding of perceptual mechanisms underlying signal recognition in complex acoustic environments. This work also highlights some key differences and similarities in how animals across diverse taxa perceptually organize complex acoustic inputs.

Table of Contents

List of Tables

List of Figures

Introduction

Behavior requires animals to organize sensory inputs into behaviorally meaningful percepts of their environments. Acoustically communicating animals face this challenge in dense social aggregations, like leks and choruses, in which multiple signalers produce spectrally and temporally overlapping signals (Gerhardt and Huber, 2002; Greenfield, 2005; McGregor, 2005). In such complex acoustic environments, sound waves from multiple sources (i.e., signalers) sum to form a composite sound wave that impinges on the ears of a receiver. Receivers must perceptually organize this composite sound input to form 'auditory objects' or 'auditory streams' corresponding to the unified perceptual representations of relevant signals, like the call of a potential mate, prey item, or predator (Bizley and Cohen, 2013; Bregman, 1990; Dent and Bee, 2018; Griffiths and Warren, 2004). Perceptual organization requires two complementary processes in which the acoustic components – like concurrent harmonics or formants (regions of concentrated acoustic energy) as well as sequential pulses, notes, or syllables – composing a relevant signal are *integrated* into a single auditory object (or auditory stream) that is *segregated* from other concurrent sounds in the environment.

Studies of sensory ecology have extensively investigated how receivers assess multiple signal components to recognize signals across behavioral contexts. However, we have limited understanding of how signal components within complex acoustic mixtures are organized into unified perceptual representations that can be recognized (Bee and Micheyl, 2008; Hulse, 2002; Miller and Bee, 2012). By comparison, perceptual organization is extensively investigated in humans for its role in speech recognition and music perception. Humans exploit various commonalties and differences in the spectral, temporal, and spatial attributes of sounds as 'cues' for perceptual organization. For example, sounds that share a common fundamental frequency, common amplitude modulation patterns, common onset and offset times, and common location of origin are likely to be integrated into a unified perceptual representation. In contrast, sufficiently large differences in these same attributes promote segregation of sounds (Bregman, 1990; Darwin, 1997; Moore and Gockel, 2012).

The broad objective of this dissertation is to identify principles of perceptual organization underlying the recognition of biologically relevant signals in a non-human animal. This dissertation adopts a comparative approach wherein various insights from human auditory perception were used to test hypotheses pertaining to perceptual organization in non-human animals. Treefrogs, the study system for this dissertation, are considered ‘ethological specialists’ for communicating in acoustically complex environments (Bee, 2015; Gerhardt and Huber, 2002). Treefrogs communicate using vocalizations that are characterized by sequences of components like calls, notes or pulses; each of these temporal components may further comprise multiple concurrent spectral components (Gerhardt, 1991; Koehler et al., 2017; Narins and Capranica, 1978; Pettitt et al., 2012; Ryan, 1985; Toledo et al., 2015). Perceiving the species-typical spectral and temporal features is crucial for behavioral decision making across evolutionarily consequential contexts like mate choice and intrasexual selection. Furthermore, treefrogs frequently vocalize in dense breeding aggregations in which males from multiple species concurrently produce sexual advertisement signals (Gerhardt and Huber, 2002; Nityananda and Bee, 2011). Females frequently recognize, assess, localize and discriminate between potential mates in these environments by displaying a behavior, termed ‘phonotaxis,’ that involves movement towards, and ultimately selection of, a preferred mate (Gerhardt, 1982; Gerhardt, 1995; Gerhardt and Bee, 2007). This dissertation builds on the extensive knowledge on vocal communication and mate choice in treefrogs to design perceptual organization tasks in which phonotaxis behavior was used to assess signal recognition. Across its chapters, this dissertation used three different species of treefrogs as a study system. Here, the American green treefrog (*Hyla cinerea*) was used to study perceptual organization of spectral components because individuals of this species recognize conspecific calls based on its species-typical spectral features. Additionally, the eastern gray treefrog (*Hyla versicolor*) and Cope’s gray treefrog (*Hyla chrysoscelis*) were used to investigate perceptual organization of temporal components because species-typical temporal features like pulse rise-time and pulse rate, respectively, are crucial for species recognition and mate choice in these two species.

Chapter 1 investigates how receivers perceptually organize concurrent spectral components belonging to overlapping communication signals. I tested the hypothesis that concurrent spectral components are segregated based on acoustic differences like temporal onset asynchrony ($\Delta T \neq 0$; i.e., concurrent sounds have different onset times) and inharmonicity ($\Delta F \neq 0$; i.e., concurrent sounds do not share a common fundamental frequency). This hypothesis was based on research in humans demonstrating that concurrent formants (spectral peaks) within a vowel are perceptually segregated and, therefore, no longer perceived as a unified vowel category when their temporal onsets are asynchronous and they share an inharmonic relationship (Darwin, 1984). This hypothesis was tested in the context of non-human vocal communication using the green treefrog (*Hyla cinerea*). The advertisement signal of green treefrog consists of a call, comprising multiple formant-like spectral components (Gerhardt, 1974a; Oldham and Gerhardt, 1975). Across its geographic range, green treefrogs breed in mixed species choruses with its sister species, the barking treefrog (*Hyla gratiosa*), which also produces advertisement calls with multiple formants but at relatively lower frequencies than those found in green treefrog calls (Oldham and Gerhardt, 1975). Female green treefrogs were presented with synthetic composite calls that comprised formants typical to both conspecific and heterospecific signals and that were behaviorally attractive *only when* the conspecific and heterospecific formants were perceptually segregated. Across treatments, the relative onset timing and harmonic relationship of the heterospecific formants was shifted in relation to the conspecific formants. Results revealed that asynchronous temporal onset as well as inharmonic relationship between the conspecific and heterospecific formants promoted their perceptual segregation. The findings from this chapter highlight that treefrogs may exploit relative onset timings and harmonic relatedness between components of overlapping signals to perceptually segregate them. These findings are in line with findings in humans demonstrating that relative onset timing and harmonic relatedness act as cues for perceptual organization. As such, this chapter provides evidence for the presence of similar principles mediating perceptual organization in humans and non-human animals.

Chapter 2 examines principles underlying the perceptual organization of temporally repeated (sequential) components of overlapping signals into distinct signal sequences. I tested a hypothesis from human psychoacoustics that *any* perceptually salient acoustic difference can promote the perceptual segregation of overlapping acoustic sequences (*sensu* Moore and Gockel, 2002). Females of the eastern gray treefrog (*Hyla versicolor*) were used to test this hypothesis in the context of vocal communication in non-human animals. Advertisement signals in this species are characterized by pulsatile sequences that are recognized by a species-typical pulse rate (Bush et al., 2002). This species exploits differences in the pulse rise time (duration from pulse onset to peak amplitude) to differentiate between the signals of conspecific and heterospecific mates (Gerhardt and Doherty, 1988; Gerhardt and Schul, 1999). I tested if perceptually salient differences in this robust species recognition cue (pulse rise time) promote perceptual segregation of overlapping signal sequences. Females were presented with interleaved pulsatile sequences (ABAB…) in which alternate pulses (A and B) differed in the pulse rise time. Integration versus segregation was estimated by whether the subjects perceived a single integrated sequence (ABAB…) or two distinct sequences (A–A–… and B–B–…), each at half the pulse rate of the integrated sequence. I found that while the rise-time difference between A and B pulses was behaviorally salient, there was no evidence that the same difference promoted the segregation of the interleaved pulsatile sequences. These findings suggest that perceptual salience of acoustic differences may not be sufficient for segregation of sequences in all non-human animals. The evidence presented here that perceptually salient differences in a robust species recognition cue did not promote segregation of overlapping sequences did not support a prominent hypothesis based on human psychoacoustic data.

Chapter 3 investigates how recognition of temporal patterns or the “rhythms” within signals is impacted in the presence of multiple concurrent sound sequences in the environment, such as might be encountered in a frog chorus (Nityananda & Bee 2011). Here, I specifically tested the hypothesis that frogs experience the perceptual phenomenon of Rhythmic Masking Release (RMR) (Turgeon et al., 2002a; Turgeon et al., 2005). RMR is described as an improvement in the recognition of a masked signal upon the further addition of concurrent sequences in the environment. Experimental paradigms to investigate RMR typically test the perceptual organization of a pulsatile Target sequence and two other concurrent pulsatile sequences. The distinct temporal pattern (rhythm) of the Target gets masked upon becoming perceptually integrated with the spectrally identical pulses of a Distractor sequence. The masked rhythm of the Target perceptually remerges when a Captor sequence, composed of pulses that are temporally synchronous with the pulses of the Distractor, is added to the Target and Distractor complex. The release from masking of the Target is induced because the Distractor sequence becomes perceptually integrated with the temporally synchronous Captor and, in turn, gets perceptually segregated from the Target. Using females of Cope’s gray treefrog (*Hyla chrysoscelis*), I examined the perceptual organization of a Target sequence, modelled on the species’ advertisement call, in the presence/absence of pulsatile Distractor and Captor sequences. My findings provide evidence that frogs experience Rhythmic Masking Release. Recognition of the Target was impaired in the presence of a spectrally identical Distractor. However, the addition of a Captor induced a release from masking of the Target, as evidenced by an improvement in recognition of the Target. Broadly, these findings highlight that both spectral and temporal commonalities can mediate perceptual organization of multiple concurrent sequences. Although, similar to humans, treefrogs experienced Rhythmic Masking Release, the magnitude of improvement in recognition was much smaller than in humans. Further, these findings highlight that the presence of concurrent sound sequences within a signal may not always impede signal recognition and, in certain circumstances, may even improve signal recognition.

Chapter 1

Treefrogs exploit temporal onset synchrony and harmonicity in forming auditory objects of vocal communication signals

Abstract

Animals frequently communicate in dense social aggregations characterized by the presence of overlapping signals from multiple individuals. Receivers have to perceptually organize these overlapping signals into distinct 'auditory objects', each corresponding to the perceptual representation of an individual signal. All the signal components produced by the same individual should be integrated into a unitary auditory object while those produced by distinct individuals should be segregated. The principles of auditory object formation, and their importance in vocal communication, is less understood in non-human animals relative to humans. Here, using American green treefrogs, *Hyla cinerea,* we investigated how receivers exploit the relative timing and harmonic relatedness of multiple spectral components to integrate or segregate sounds during auditory object formation. Using phonotaxis as a behavioral assay, we tested the hypothesis that females exploit temporal onset synchrony and harmonicity as perceptual cues in forming auditory objects of male sexual advertisement signals. We presented females with synthetic 'composite' calls, constituting both conspecific and heterospecific spectral components, that were behaviorally attractive only when the conspecific and heterospecific components were perceptually segregated. Across treatments, we manipulated the relative onset timing and harmonic relatedness of the heterospecific components in relation to the conspecific components. We predicted that asynchronous temporal onset and inharmonic relationship between the conspecific and heterospecific components would promote their perceptual segregation. Our findings are consistent with this prediction. Females segregated the conspecific and heterospecific components within composite calls when these components had asynchronous onset timing, inharmonic relationship or both. We discuss these findings in the light of parallel perceptual processes across animal taxa and neuroethological theories of auditory processing.

Introduction

Animals often communicate in complex acoustic environments like noisy social aggregations, leks or choruses. Such environments are characterized by multiple signalers producing spectrally and temporally overlapping signals (Bee and Micheyl, 2008; Brumm and Slabbekoorn, 2005; Greenfield, 2005; McDermott, 2009). Behavioral decision-making requires a receiver to perceptually organize such overlapping signals into distinct 'auditory objects', corresponding to the perceptual representation of signals produced by distinct individuals (Bizley and Cohen, 2013; Bregman, 1990; Griffiths and Warren, 2004). Auditory object formation requires two complementary processes in which all the signal components (like simultaneous harmonics or formants as well as sequences of words, notes, pulses) produced by the same signaler are perceptually integrated into a unified auditory object while those produced by different signalers are perceptually segregated into separate auditory objects.

Auditory object formation has been extensively studied in the context of speech recognition in noisy social environments. Various acoustic commonalities and differences between sounds can serve as 'cues' for auditory object formation in humans (reviewed in Bregman, 1990; McDermott and Oxenham, 2008; Moore and Gockel, 2012). Sounds tend to be integrated if they possess commonalities in their spectral and temporal features as well as spatial origin. In contrast, sufficiently large differences in these same attributes promote segregation of sounds. Two such cues, temporal synchrony and harmonic relatedness (harmonicity) of sounds, frequently mediate perceptual integration of spectral components within speech signals (Darwin and Gardner, 1987). Speech signals constitute phonetic categories like vowels which, in turn, are composed of multiple formants (i.e., spectral regions with concentrated acoustic energy) that lend a vowel its distinct sound quality (Ladefoged and Disner, 2012). Formants that share temporal synchrony ($\Delta T = 0$; start and end at the same time) or are harmonically related ($\Delta F = 0$; are multiples of a common fundamental frequency) are likely to be perceptually integrated and heard as a single phonetic category. In contrast, sufficiently large deviations from these commonalities ($\Delta T \neq 0$; temporal asynchrony or $\Delta F \neq 0$; inharmonicity) promote segregation of formants from the vowel, thereby causing them to be heard separately from the rest of the vowel (Broadbent and Ladefoged, 1957; Brokx and Nooteboom, 1982; Darwin, 1981; Darwin, 1984; Darwin and Hukin, 1998; Gardner et al., 1989; Hukin and Darwin, 1995).

Non-human animals face the analogous challenge of communicating in complex acoustic environments using rhythmic signals frequently composed of multiple concurrent spectral components (Fenton et al., 2014; Gerhardt and Huber, 2002; Hyland Bruno and Tchernichovski, 2019; Kershenbaum et al., 2016; Prestwich, 1994; Winn et al., 1981). Therefore, auditory object formation is crucial for accurate signal recognition, discrimination and localization in evolutionarily consequential contexts like survival and reproduction. Although auditory object formation is a ubiquitous communication challenge, our understanding of this process in non-human animal communication remains limited (Bee and Micheyl, 2008; Dent and Bee, 2018; Hulse, 2002). Non-human animals can employ similar auditory object formation cues as humans (Bee, 2010; Cai et al., 2018; Farris et al., 2005; Fay, 1998; Gupta and Bee, 2020; Klinge and Klump, 2009; Ma et al., 2010; Nityananda and Bee, 2011; Schul and Sheridan, 2006 *but also see* Bee and Riemersma, 2008; Kalra et al., 2024; Schwartz and Gerhardt, 1995; Schwartz and Serratto Del Monte, 2019). However, few animal studies have investigated multiple such cues in the same species. And as in humans (Elhilali et al., 2009; Micheyl et al., 2013b; Singh and Bregman, 1997), preliminary investigations of non-human animals suggest that multiple cues may be differentially weighed (Dent et al., 2016; Schwartz and Serratto Del Monte, 2019) or have additive effects (Itatani and Klump, 2020) on auditory object formation.

In this study, we investigated the role of temporal synchrony and harmonicity, two cues pivotal for speech perception in humans, in mediating auditory object formation in the American green treefrog (*Hyla cinerea*). Green treefrogs have been used extensively to investigate evolutionary (Ehret and Gerhardt, 1980; Gerhardt, 1987; Höbel and Gerhardt, 2003; Jones et al., 2014) and mechanistic (Fuzessery and Feng, 1983; Gerhardt et al., 1990; Lee et al., 2021; Schwartz and Gerhardt, 1989) aspects of vocal communication in complex acoustic environments. During their breeding season, male *H. cinerea* aggregate in dense choruses where they attract females by producing short (100-300 ms) advertisement calls that are repeated every 1-3 s during bouts of calling (Gerhardt, 1974a; Oldham and Gerhardt, 1975). Females select a mate by exhibiting phonotaxis toward a calling male and initiating physical contact that results in amplexus. The frequency spectrum of the advertisement call has a formant-like structure with acoustic energy concentrated in separate low-frequency (0.7–1.2 kHz) and high-frequency (2.4–3.6 kHz) bands. Most of the acoustic energy in the lower formant is concentrated in a single spectral component, whereas the higher formant typically emphasizes two to four spectral components (Gerhardt, 1974a,b; Gerhardt, 1976; Gerhardt, 2001; Oldham and Gerhardt, 1975). There is little acoustic energy between the two formants (1.2–2.4 kHz) (Figure 1-1). Across much of its geographic range, *H. cinerea* frequently breeds in mixed-species choruses with the similar but somewhat larger barking treefrog (*Hyla gratiosa*). Green and barking treefrogs are closely related sister species that produce acoustically similar advertisement calls (Hua et al., 2009; Mecham, 1965; Robillard et al., 2006). The frequency spectrum of the *H. gratiosa* advertisement call also has a formant-like structure consisting of a lower formant and a higher formant, both of which are shifted to lower frequencies relative to the same formants in the *H. cinerea* call (Figure 1-1). Females of *H. cinerea* do not avoid heterospecific (*H. gratiosa*) calls and will even approach them when given no other choice (Höbel, 2015). Importantly, however, they strongly prefer conspecific calls over heterospecific (*H. gratiosa*) calls (Oldham and Gerhardt, 1975). Moreover, the addition of spectral components from a heterospecific (*H. gratiosa*) call to a conspecific call starkly reduces the attractiveness of the signal relative to an unaltered conspecific call (Gerhardt, 1974b; Gerhardt and Höbel, 2005). The ability of simultaneous spectral components from heterospecific calls to render conspecific calls less attractive is biologically relevant for this study species and also key to the experimental design of the present study.

Using synthetic advertisement calls, we exploited the robust behavioral discrimination of female *H. cinerea* against heterospecific (*H. gratiosa*) spectral components to test the hypothesis that both temporal synchrony ($\Delta T = 0$) and harmonicity ($\Delta F = 0$) promote the perceptual integration of concurrent spectral components. In a series of two-alternative phonotaxis tests, we gave females a choice between a standard *H. cinerea* call with three simultaneous spectral components (900, 2700, and 3000 Hz) versus an alternative composite call consisting of the same three *H. cinerea* spectral components (900, 2700, and 3000 Hz) combined with three additional spectral components representative of *H. gratiosa* calls (450, 1800, and 2250 Hz) (Figure 1-1). The three conspecific components and the three heterospecific components each constituted a three-harmonic complex tone with fundamental frequencies (f_0) of 300 Hz and 450 Hz, respectively. We capitalized on the coincidental property that, when combined, all six spectral components (450, 900, 1800, 2250, 2700, and 3000 Hz) were harmonics of a common fundamental frequency of 150 Hz. In a fully factorial design, we manipulated the degree of temporal synchrony and harmonicity between the conspecific and heterospecific spectral components. These manipulations were done by shifting the onset timing and frequency of the heterospecific components relative to the conspecific components. We used the relative attractiveness of the standard and composite calls, as measured by selective phonotaxis, to assess the extent to which temporal synchrony and harmonicity promote perceptual integration (Figure 1-2a). Based on earlier work in frogs (Gerhardt 1974; Gerhardt and Höbel, 2005), we predicted females would prefer the standard call under conditions ($\Delta T = 0$ ms; $\Delta F = 0\%$) expected to promote integration of the conspecific and heterospecific spectral components of the alternative composite call into a single (and less attractive) auditory object

(Figure 1-2a, b). Conversely, we predicted that introducing temporal onset asynchrony ($\Delta T \neq 0$ ms) and inharmonicity ($\Delta F \neq 0\%$) between the conspecific and heterospecific components of the composite call would promote their perceptual segregation; in turn, we expected perceptual segregation to render the composite and standard calls equally attractive based on their identical conspecific spectral components (Figure 1-2a, c). Our factorial design allowed us to further explore potential interactions between temporal onset asynchrony and inharmonicity in promoting segregation.

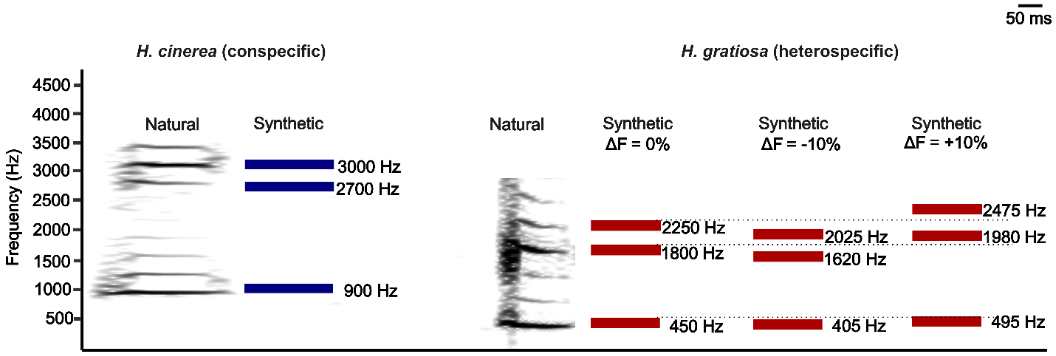

Figure 1-1. Natural and synthetic advertisement calls showing the formant-like spectra for *H. cinerea* (conspecific) and *H. gratiosa* (heterospecific). Sonograms depict the spectral components within a natural conspecific (left) and heterospecific (right) call. Here, greater intensity of the black shade depicts higher acoustic energy. Most of the acoustic energy is concentrated in regions around 900 Hz, 2700 Hz, and 3000 Hz, in natural conspecific calls and around 450 Hz, 1800 Hz, and 2250 Hz in natural heterospecific calls. Based on the spectra of natural calls (Gerhardt, 1974a; Oldham and Gerhardt, 1975), synthetic advertisement calls were generated to have spectral components characteristic of *H. cinerea* calls (blue horizontal bars) and *H. gratiosa* calls (red horizontal bars). Three different sets of heterospecific spectral components were generated. One set ($f_0 = 450$ Hz) was harmonically related to the conspecific spectral components ($\Delta F = 0\%$). The heterospecific components in the other two sets were mistuned, i.e., had their frequencies shifted either upward (495, 1980, and 2475 Hz) or downward (405, 1620, and 2025 Hz) by 10% (i.e., $\Delta F = \pm10\%$) so that they were rendered inharmonic with the conspecific spectral components.

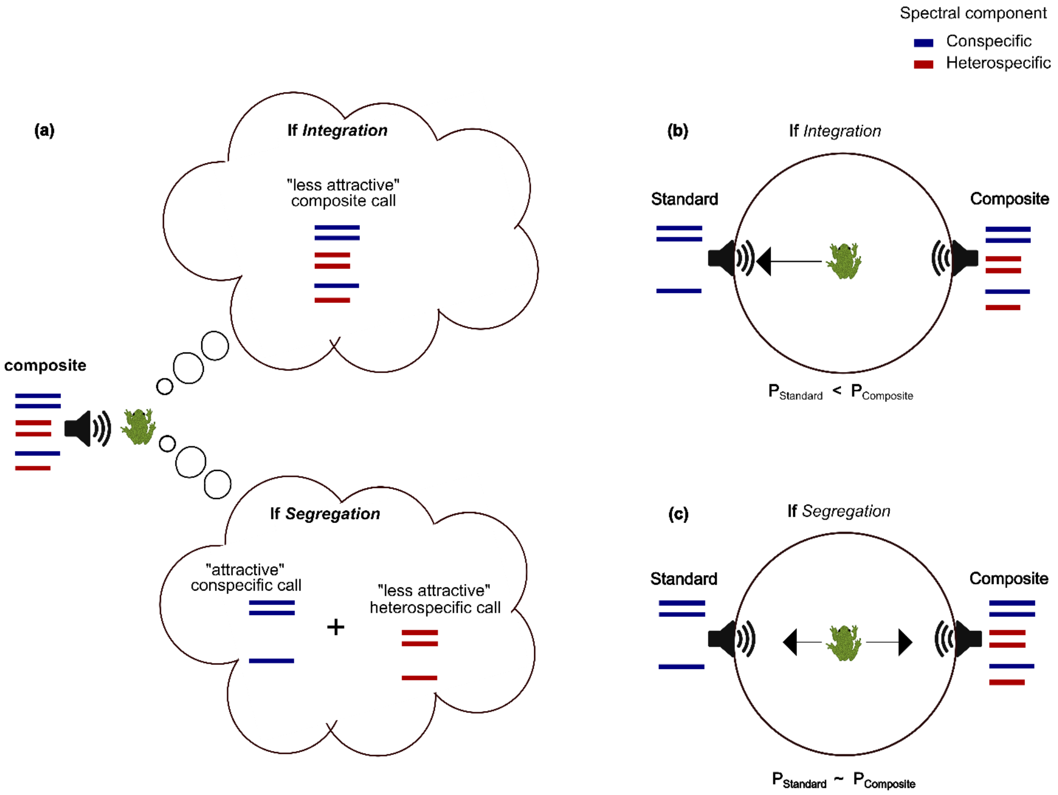

Figure 1-2. Integration versus segregation determines the behavioral attractiveness of the two alternative calls. a) If subjects integrated the conspecific (blue bars) and heterospecific components (red bars), they would perceive a single call which is less attractive. If subjects segregated these two kinds of spectral components, they would perceive two distinct calls, an attractive conspecific call (identical to the standard call), and a less attractive heterospecific call. Consequently, in a two-alternative choice test, b) integration of conspecific and heterospecific components would result in a greater preference for the standard call against the less attractive composite call, and c) segregation of the conspecific and heterospecific components would result in an equal preference for the standard and composite call based on their identical conspecific spectral components.

Methods

Subjects

Ninety gravid females of *H. cinerea* collected from Bowens Mill hatchery (Ben Hill County, GA, USA) were used as subjects in this study. Females were collected in amplexus at night (between 2100 and 2200 h) during June and July of 2019 and 2022. After collection, females were separated from their mates, transferred to separate containers, and subsequently tested in multiple phonotaxis tests (between 2200 h and 0500 h). There is little evidence of carryover effects when testing female frogs in multiple phonotaxis tests (Akre and Ryan, 2010; Gerhardt, 1981). Females and their mates were released at the collection site within 24 hours of collection.

Stimulus and experimental design

Synthetic calls (44.1 kHz, 16 bit) were generated in MATLAB R2018B (Mathworks, Natick, MA, USA) and had spectral and temporal properties based on analyses of the natural advertisement calls of *H. cinerea* and *H. gratiosa* (Gerhardt, 1974a; Oldham and Gerhardt, 1975). We conducted a series of two-alternative phonotaxis tests in which females were allowed to choose between stimuli that acoustically simulated two males calling in perfect alternation, each at a rate of 40 calls/min. One stimulus consisted of repeated standard calls; the alternative stimulus consisted of repeated composite calls (Figure 1-3). The standard call had a frequency spectrum that was typical of conspecific calls and is attractive to females (Gerhardt, 1974b). It comprised three spectral components (900, 2700, and 3000 Hz) that were 250 ms in duration (Figure 1-3). The composite call combined the same three conspecific spectral components of the standard call with three heterospecific spectral components that were 150 ms in duration. The frequency of each heterospecific spectral component was determined by the level of ΔF (Figure 1-3; Table 1-1). The shorter duration of the heterospecific components allowed us to manipulate, across treatments, the temporal onset asynchrony between the conspecific and heterospecific components while keeping the overall duration of the composite call consistent at 250 ms (Figure 1-3; Table 1-1). The duration of the composite call was kept consistent across treatments to minimize any behavioral constraints imposed by subjects' preferences for call duration *per se* (Gerhardt, 1987). The chosen durations of 250 ms and 150 ms for the conspecific and heterospecific components, respectively, fall within the natural range of call duration for both species (Gerhardt, 1974a; Oldham and Gerhardt, 1975). Note that, in designing the composite call, we encountered a trade-off between maintaining consistent call duration and consistent temporal 'offset' synchrony between the conspecific and the heterospecific components. As such, one limitation of this design was that, any shifts in the relative temporal onsets of the conspecific and heterospecific components introduced shifts in their relative temporal offsets (see below; Figure 1-3). We chose to maintain a consistent duration of composite call over a consistent degree of temporal offset synchrony because temporal offset synchrony is believed to have a weak effect on perceptual organization (we elaborate on this argument further in the ‘Discussion’ section). Hereafter, any value of ΔT, strictly corresponds to the temporal onset asynchrony between the heterospecific components and the conspecific components.

In all stimulus calls, each spectral component had a random starting phase and an amplitude envelope shaped by inverse exponential rise and fall times of 25 ms and 50 ms, respectively. The starting phase of each spectral component in each stimulus call was re-randomized at the start of every night of testing, such that subjects tested on different nights heard different exemplars of the stimuli. At the beginning of each night of testing, each separate spectral component in each stimulus call was calibrated to a sound pressure level (SPL re 20 µPa; LCF) of 79 dB SPL at a distance of 1 m; therefore, at 1 m, the overall level of the standard call (3 components) was 83.8 dB SPL and that of the composite call (6 components) was 86.8 dB SPL. These values fall within the natural range of variation in sound pressure levels recorded for both *H. cinerea* and *H. gratiosa* (Gerhardt 1975).

In total, there were 15 different choice tests based on the factorial combination of five levels of temporal asynchrony (ΔT = 0 ms, 25 ms, 50 ms, 75 ms, 100 ms) and three levels of inharmonicity (ΔF = 0%, +10%, -10%) (Table 1-1). We manipulated temporal asynchrony in the composite call by delaying the onset of its heterospecific components by ΔT relative to the onset of its conspecific components (Table 1-1; Figure 1-3). Hence, at ΔT = 0 ms, all spectral six components had synchronous onsets and at ΔT = 100 ms they had asynchronous onsets and, consequently, synchronous offsets (Figure 1-3). We manipulated harmonicity by either increasing or decreasing the frequency of each heterospecific spectral component by 10% (Figure 1-1; Figure 1-3). Because there were no significant differences in responses that depended on increasing versus decreasing frequency (see Appendix 1), we hereafter combine these two treatments as ΔF = ±10%. Thus, at ΔF = 0%, all six spectral components were harmonically related and shared a common f_0 of 150 Hz. But at ΔF = ±10%, harmonicity within the three conspecific components and within the three heterospecific components was maintained, but the conspecific and heterospecific components were no longer harmonically related to each other (Table 1-1; Figure 1-3). We tested

ΔT within subjects and ΔF between subjects (n = 30 each for ΔF = 0%, +10%, and -10%; total n = 90), and we randomized the order in which different levels of ΔT were tested for each subject.

Table 1-1. The 15 factorial combinations of ΔT and ΔF tested and the corresponding frequencies of conspecific and heterospecific spectral components

Treatment		**Conspecific components**	**Heterospecific components**
ΔF	**ΔT**		
0%	0 ms	900, 2700, and 3000 Hz	450, 1800, and 2250 Hz
	25 ms		
	50 ms		
	75 ms		
	100 ms		
-10%	0 ms	900, 2700, and 3000 Hz	405, 1620, and 2025 Hz
	25 ms		
	50 ms		
	75 ms		
	100 ms		
+10%	0 ms	900, 2700, and 3000 Hz	495, 1980, and 2475 Hz
	25 ms		
	50 ms		
	75 ms		
	100 ms		

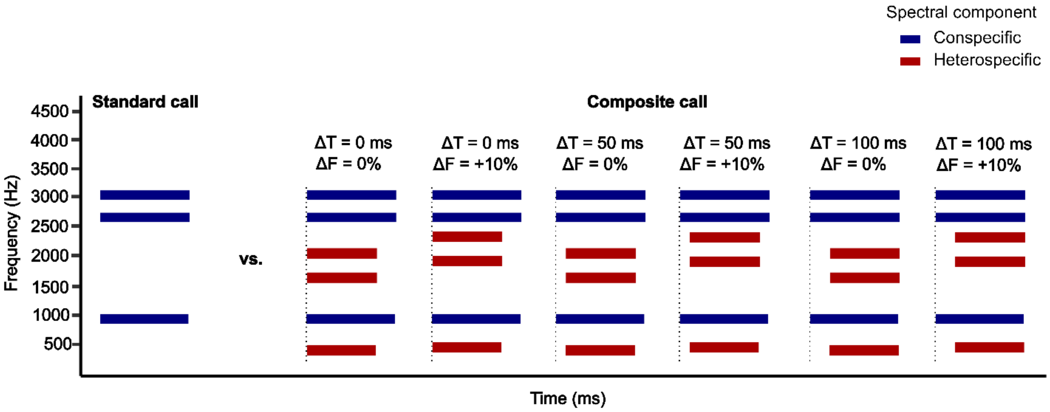

Figure 1-3. Experimental treatments. Various manipulations of composite call (enlarged here) shown for six treatments (3 levels of ΔT and 2 levels of ΔF) in which the temporal onset and frequencies of the heterospecific components were altered. Note that, the onset of the heterospecific components is delayed, relative to the conspecific components, in each treatment where $\Delta T \neq 0$ ms. Consequently, there are corresponding shifts in the relative timing of the offsets of the conspecific and heterospecific components. Here, only a subset of treatments and only single calls for each treatment are shown, see text for additional detail on all the treatments used in this study as well as for the organization of multiple identical calls within each treatment.

Testing protocol

Phonotaxis tests were performed in a circular arena (2-m diameter and 60-cm height) that was located at the field site, approximately 300 m away from the nearest pond from which the subjects were collected. The arena’s floor was made of foam mats and its walls were constructed out of hardware cloth, covered in black fabric, to create a visually opaque but acoustically transparent barrier between the subjects and the speakers. Two speakers (Gallo Acoustics, Scotland, UK) were placed 180° apart from each other behind the arena wall, facing inward towards a subject’s release point at the center of the arena. Stimuli were broadcast using Adobe Audition 3.0 (Adobe Systems Inc. San Jose, CA, USA) installed on an HP ProBook 450 G6 (HP Inc., Palo Alto, CA, USA). Sound was output through a MOTU M4 sound card (MOTU, Inc., Cambridge, MA, USA), amplified by a Digital 2 Channel MOSFET amplifier (Dual Electronics, Corp., FL, USA), and broadcast through the two speakers. Sound pressure levels were calibrated using a Larson Davis Model 831 sound level meter (Larson Davis Inc., Depew, NY) connected to a microphone placed 1 m away from the speaker at the approximate position of a subject’s head at the release point. Across subjects, we randomized both the speakers assigned to broadcast the standard versus the composite calls as well as the type of call that was broadcast first.

At the beginning of a test, a single subject was placed inside a release cage fixed at the center of the arena floor and given an acclimation period of 60 s after which the stimulus broadcast began. During stimulus broadcast, the standard and composite calls alternated in time such that equal periods of silence preceded and followed each call. After two repetitions of both the standard and composite calls, the lid of the release cage was lifted remotely using a pulley system, and a timer was simultaneously started. After this point, the alternating calls continued while the subject was free to move inside the arena. A choice was recorded if the subject entered a 10-cm semicircular response zone in front of one of the speakers within 5 min. A no-response was recorded if the subject did not leave the release cage within 3 min or did not make a choice within 5 min. On average, females responded within 45.5 s of being released.

Statistical analysis

Statistical analyses were performed in R studio version 2023.12.1 ($\alpha = 0.05$ in all analyses). We used two-tailed binomial tests at each combination of ΔT (0 ms, 25 ms, 50 ms, 75 ms and 100 ms) and ΔF (0% and ±10%) to determine whether the proportion of subjects choosing the standard call exceeded the null expectation of 0.50. We assessed differential attractiveness of the standard and composite calls as functions of ΔT and ΔF by fitting a GEE model with logit-link function and an exchangeable correlation structure using the *geepack* package (Hardin and Hilbe, 2012). Wald-statistic was used to assess the differential attractiveness. Each subject's choice was scored as a binary response variable (1 = standard call, 0 = composite call). The independent variables were ΔT, ΔF, and their interaction. Since both ΔT and ΔF were coded as categorical variables, the output of the GEE model reported the effects of any combination of ΔT and ΔF in relation to a reference condition, which we specified as the test where $\Delta T = 0$ ms and $\Delta F = 0\%$.

Results

Subjects had a significant 4:1 preference ($p < 0.01$) for the standard call (proportion = 0.80) over the composite call (proportion = 0.20) when the conspecific and heterospecific spectral components of the latter had synchronous onsets ($\Delta T = 0$ ms) and were harmonically related ($\Delta F = 0$ %) (Figure 1-4, Table 1-2). In stark contrast, the proportions of subjects choosing the standard call were not significantly different from the null expectation of 0.50 when the conspecific and heterospecific spectral components of the composite call had asynchronous temporal onsets ($\Delta T \neq 0$ ms), were inharmonic ($\Delta F \neq 0\%$), or both (Figure 1-4, Table 1-2).

Compared to the reference test (ΔT = 0 ms, ΔF = 0%), introducing temporal onset asynchrony in the composite call while maintaining its harmonicity (i.e., ΔT ≠ 0 ms, ΔF = 0%) elicited a significant decrease in the proportion of subjects choosing the standard call for values of ΔT equal to 25 ms, 50 ms, and 75 ms, but not for ΔT equal to 100 ms. Additionally, introducing inharmonicity in the composite call while maintaining synchronous onsets (i.e., ΔF ≠ 0%, ΔT = 0 ms) also elicited a significant decrease in the proportion of subjects choosing the standard call compared to the reference test (Table 1-3). Finally, introducing both temporal onset asynchrony and inharmonicity in the composite call additively decreased (no significant interaction) the proportion of subjects choosing the standard call in all treatments except when ΔT was 25 ms and ΔF was ±10%. In that test, there was a significant interaction between ΔT equal to 25 ms and a ΔF of ±10% since the proportion of subjects choosing the standard call was higher than that predicted from an additive decrease imposed by ΔT of 25 ms and ΔF of ±10% (Table 1-3).

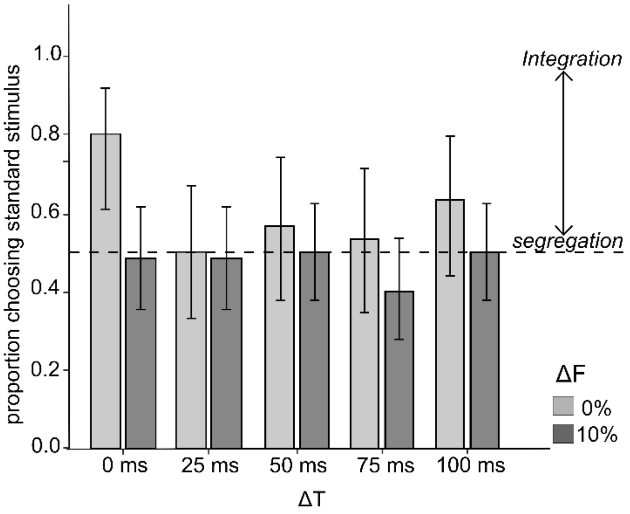

Figure 1-4. Proportion of subjects choosing the standard call across all combinations of temporal onset asynchrony (ΔT) and inharmonicity (ΔF). The dashed horizontal line indicates the null expectation of 0.50 in a two-alternative choice tests. Gray and black bars indicate ΔF = 0% and ΔF = ±10%, respectively. Error bars represent 95% exact binomial confidence intervals.

Table 1-2. Results of phonotaxis tests showing the proportions of subjects (±95% exact binomial confidence intervals) that chose the standard call over the composite call as a function of temporal onset asynchrony (ΔT) and inharmonicity (ΔF). Also shown are P values from two-tailed binomial tests of the hypothesis that the proportion **choosing the standard call did not differ from the null expectation of 0.50.**

ΔT	**ΔF**	**N**	**Proportion choosing the standard call**	**CI**	***P* value**
0 ms	0%	30	0.80	0.61-0.92	< 0.01*
	±10%	60	0.48	0.39-0.65	0.90
25 ms	0%	30	0.50	0.33-0.67	0.99
	±10%	60	0.48	0.35-0.61	0.90
50 ms	0%	30	0.57	0.38-0.74	0.60
	±10%	60	0.50	0.38-0.62	0.99
75 ms	0%	30	0.53	0.35-0.71	0.90
	±10%	60	0.40	0.28-0.54	0.20
100 ms	0%	30	0.63	0.44-0.80	0.20
	±10%	60	0.50	0.38-0.62	0.99

Table 1-3. Results of the Generalized Estimating Equation (GEE) model. The effects of temporal onset asynchrony (ΔT), inharmonicity (ΔF), and their interaction are reported in relation to the reference test where ΔT = 0 ms and ΔF = 0%. 'Condition' represents any manipulation in ΔT, ΔF or the interactive effects of different combinations of ΔT and ΔF. Interactions are shown by the '×' symbol. 'β' indicates the estimate value for each condition. 'Wald X^2' and 'P value' correspond to the results of the Wald-statistics.

Condition	**β**	**Wald X^2**	***P* value**
ΔT = 25 ms	-1.39	5.35	0.02*
ΔT = 50 ms	-1.11	4.37	0.04*
ΔT = 75 ms	-1.25	5.08	0.02*
ΔT = 100 ms	-0.83	1.95	0.16
ΔF = ±10%	-1.44	7.64	0.01*
ΔT = 25 ms × ΔF = ±10%	1.39	3.90	0.05*
ΔT = 50 ms × ΔF = ±10%	1.18	3.46	0.06
ΔT = 75 ms × ΔF = ±10%	0.91	1.97	0.16
ΔT = 100 ms × ΔF = ±10%	0.90	1.74	0.19

Discussion

The primary goal of this study was to test the hypothesis that females of *Hyla cinerea* exploit temporal synchrony and harmonicity as perceptual cues in forming auditory objects of vocal communication signals. The data are consistent with this hypothesis. The main result from this study is that heterospecific spectral components rendered the composite call less attractive than the standard call *only* when they were temporally synchronous with (ΔT = 0 ms) and harmonically related to (ΔF = 0%) the conspecific spectral components. When the conspecific and heterospecific spectral components of the composite call had asynchronous onsets, inharmonic relationships, or both, the composite call was equally as attractive as the standard call, which included only the conspecific spectral components. Our interpretation of these results is that temporal synchrony and harmonicity promoted the perceptual integration of the conspecific and heterospecific spectral components of the composite call into a single auditory object that was less attractive than a call consisting of only conspecific components. Onsets that were asynchronous by as little as 25 ms or shifts in frequency on the order of 10% promoted the perceptual segregation of the conspecific and heterospecific spectral components into different auditory objects. Because heterospecific calls are neither attractive nor inherently aversive in the presence of conspecific calls, onset asynchrony and inharmonicity reduced each female's choice to one between two equivalent standard calls, for which there was no preference. Based on these results, we conclude that female green treefrogs exploit synchronous temporal onsets and harmonicity in forming auditory objects of vocal communication signals.

Role of temporal onset synchrony in auditory object formation

Multiple studies in humans have demonstrated temporal synchrony as a robust cue for auditory object formation underlying the perception of speech perception as well as non-communication sounds. Bregman and Pinker (1978) and Dannenbring and Bregman (1978) presented human subjects with a sequence of a two-tone complex. When the temporal onsets of the two tones in the complex were synchronous, the tone complex was perceived as a single rich tone with a unified timbre. In contrast, when the temporal onset of the lower tone was shifted to lead the upper tone of the complex, the two tones were segregated and the distinct timbre was no longer perceived. Further investigations on vowel perception (Darwin, 1981; Darwin, 1984; Darwin and Ciocca, 1992; Darwin and Sutherland, 1984; Hukin and Darwin, 1995; Roberts and Moore, 1991) demonstrated that temporal onset synchrony effects the tendency of a single formant to be perceptually integrated with the rest of the vowel and, in turn, influence the perceived phonetic category of the vowel. For vowel sounds of smaller duration (~ 50 ms) a temporal onset asynchrony as small as 30 ms was sufficient for perceptual segregation of a single formant from the rest of the vowel, while for longer vowel sounds, larger asynchrony was required. More recently, Elhilali et al. (2009) and Micheyl et al. (2013a) demonstrated that two sequences were integrated and heard as a 'single' sequence when their constituent tones shared temporal synchrony. Shifting the temporal synchrony between those tones promoted the segregation of sequences thereby causing the percept of two distinct sequences.

Thus far, only two studies have investigated if non-human animals can exploit temporal synchrony in forming auditory objects of communication signals. Neilans and Dent (2015) presented budgerigars with synthetic conspecific contact calls and demonstrated that increased temporal asynchrony between overlapping conspecific signals promoted their perceptual segregation, thereby causing them to be heard as separate signals. Further, Gupta and Bee (2020) demonstrated, in Cope's gray treefrogs, that the low and high conspecific spectral components were integrated when they were temporally synchronous. Our study corroborates both of these findings and thus contributes to the emerging evidence for the temporal synchrony as a robust auditory object formation cue in non-human animals. More studies in future would be required to assess the role of temporal synchrony in perceptual organization in other non-human animal species.

It is also worth looking at the present findings in the light of male-male signaling dynamics in nature. Males in dense acoustic aggregations frequently adjust the relative timing of their signals in two broad ways. First, commonly seen in birds, frogs and insects, the relative signal timings are asynchronous so as to approximate alternation or minimal signal overlap between the signals of different signalers (reviewed Gerhardt and Huber, 2002; Hall, 2009; Klump and Gerhardt, 1992; Thorpe, 1975). Second, a relatively uncommon signaling pattern is one in which the relative signal timings are synchronized so that there is maximal signal overlap between the signals of different individuals (Grafe, 1999; Legett et al., 2020; Legett et al., 2021; Tuttle and Ryan, 1982; Wells and Schwartz, 1984). Signal alternation might serve to facilitate a signaler's assessment of a neighbor's call, minimize degradation of species-typical temporal information required for mate choice or exploit a female's perceptual bias to prefer a leading male's calls (Greenfield et al., 1997; Schwartz, 1987). On the other hand, signal synchronization may serve to attract females from a distance, avoid predators by potentially impairing their ability to localize individual signalers, or to jam the signals of neighboring males (Grafe, 1999; Greenfield, 1994; Greenfield et al., 1997; Legett et al., 2021; Tuttle and Ryan, 1982). There is emerging evidence that adaptive significance of different signaling dynamics may relate to the perceptual processes within the receivers (Bosch and Márquez, 2002; Grafe, 1999; Höbel and Gerhardt, 2007; Minckley and Greenfield, 1995; Reichert et al., 2024; Snedden and Greenfield, 1998). Neilans and Dent (2015) and the present study demonstrated that temporal asynchrony between signal produced by different individuals promotes the perceptual segregation of these signals by the receiver, making the signals easier to recognize. Given the diversity of inter-male calling dynamics across animal taxa, more work would be needed to investigate how the temporal patterns of such signaling interactions relate to perceptual organization processes in conspecific and heterospecific receivers.

One limitation of this study was that it did not control for temporal offset asynchrony. Here, since the conspecific and heterospecific components had unequal durations, any change in their relative temporal onsets was coupled with a change in their relative temporal offsets. For instance, at a temporal 'onset' asynchrony of 0 ms, the conspecific and heterospecific spectral components had a temporal 'offset' asynchrony of 100 ms. We speculate that any behavioral effects imposed by the uncontrolled temporal offsets in this study were marginal because of two reasons. First, at least in humans, asynchronous temporal onsets have a significantly greater effect on promoting perceptual segregation than asynchronous temporal offsets. Darwin (1984) and Madsen and Moore (2014) demonstrated that sounds were more likely to be segregated when they exhibited asynchronous onsets compared to when they exhibited asynchronous offsets. Second, the present results also indirectly suggest that perceptual segregation is primarily driven by asynchrony in temporal onsets. A significantly greater than chance proportion of subjects segregated conspecific and heterospecific spectral components when their temporal onsets were asynchronous, regardless of the relative timings of their temporal offsets. Recall that, even when the temporal offsets of the conspecific and heterospecific components were synchronized ($\Delta T = 100$ ms; Figure 1-3, Table 1-2), subjects equally preferred the standard and composite calls, indicating that they segregated the spectral components even when their temporal offsets were synchronous. However, we did observe that the attractiveness to the standard call did not significantly differ between when the temporal onsets of conspecific and heterospecific components were synchronous ($\Delta T = 0$ ms) versus when their temporal offsets were synchronous ($\Delta T = 100$ ms; Table 1-3). The present design was limited by the tradeoff between maintaining a consistent duration of the composite call versus a consistent temporal offset synchrony of the conspecific and heterospecific components within the composite call. Follow-up studies could adopt an alternate design in which the offsets of the heterospecific and conspecific spectral components can be kept synchronous throughout and the onset asynchrony can be manipulated by shifting the duration of the heterospecific spectral components.

Role of Harmonicity in auditory object formation

Research using synthetic tone-complexes and speech sounds reveals that harmonicity mediates perceptual organization of concurrent spectral components. Darwin (1981), Darwin and Gardner (1986) and Gardner et al. (1989) demonstrated that mistuning a single formant within a vowel causes the formant to be perceptually segregated from the rest of the vowel, leading to an overall change in the perceived timbre of the vowel and, in turn, the perception of the phonetic category of the vowel. Additionally, Moore et al. (1985, 1986) and Hartmann et al. (1990) demonstrated that mistuning single components within a harmonic complex causes them to be perceptually segregated and heard separately from the tone-complex. A mistuning as small as 3% could be detected by subjects in the form of perceived changes in the amplitude envelop of the tone complex; a mistuning above 8% leads to complete perceptual segregation.

Even though signals in non-human animals are frequently characterized by harmonic complexes that resemble vowel sounds (Hoeschele, 2017; Kershenbaum et al., 2016), few studies have investigated how harmonicity impacts perceptual organization of communication signals. Lohr and Dooling (1998) used stimuli modelled on natural calls and showed that both zebra finches and budgerigars are highly sensitivity to the mistuning of single harmonics within calls. Simmons and Bean (2000) found that the evoked calling response of male bullfrogs was significantly lower when particular harmonics within the call were mistuned, compared to when all the frequencies within the signal were harmonic. Klinge and Klump (2010, 2009) demonstrated that gerbils were remarkably sensitive to mistuning of individual components within a tonal complex. In those studies, shifts in behavioral responses were observed even when inharmonicity was introduced by very small shifts (< 1 Hz) in the frequency of individual components. Our findings, demonstrating that perceptual segregation is promoted upon mistuning of the specific spectral components corroborates these previous findings on birds, frogs and rodents and adds to the evidence for the prevalence for harmonicity mediated perceptual organization in non-human animals. In conjunction with our findings on temporal synchrony, these findings on harmonicity lend support to the emerging view that parallel perceptual processes may govern auditory object formation in humans and non-human animals.

To our knowledge, this study is the first systematic investigation of harmonicity as an auditory object formation cue in frogs. There exists no single consensus on the behavioral significance of the prominent harmonic structures of the vocalizations used by many frogs. Using a physiological procedure called reflex modification, Simmons, (1988) demonstrated that males of *H. cinerea* exhibit sensitivity to shifts in the harmonic structure of two-tone complexes consisting of conspecific frequencies. However, behavioral investigations in *H. cinerea* using female phonotaxis by Gerhardt et al., (1990) and male evoked calling behavior by Simmons et al., (1993) demonstrated no significant differences in either phonotaxis or male evoked calling upon manipulating the harmonic relatedness of the conspecific spectral components. One key difference between these two behavioral studies and the present study could be related to the test design. While Gerhardt et al., (1990) and Simmons et al., (1993) measured differential responsiveness to conspecific calls based on inharmonicity, the present study investigated perceptual segregation of conspecific and heterospecific signal components based on inharmonicity. Recall that in this study, conspecific spectral components were always harmonic with each other and instead their inharmonicity with the heterospecific spectral components was manipulated. Taking all of the above findings into consideration, we speculate that harmonicity may play a role in perceptual organization underlying species recognition but may not be behaviorally meaningful in the context of mate discrimination or male-male competition with conspecifics.

Multiple cue interactions during auditory object formation

Multiple cues may be differentially weighed, act in an additive fashion, or interact during auditory object formation. Most of our understanding of the effects of multiple cues on object formation comes from human psychoacoustic studies. Singh and Bregman (1997) showed that both common amplitude rise times and spectral similarity additively mediate integration of sounds. Elhilali et al. (2009) found that regardless of having considerable frequency separation (which promotes segregation), tones were integrated when they shared temporal synchrony. This finding indicated that temporal synchrony overrides spectral separation in mediating perceptual integration. Further, similar to the present study, Micheyl et al. (2013b) manipulated temporal coherence and harmonicity and found an additive effect of the two cues on perceptual organization. Recent findings from non-human animals suggest similar patterns for multiple cues mediating auditory object formation. Itatani and Klump (2020) showed that European starlings rely on the additive effects of common spatial origin and spectral similarity for perceptual organization, though spectral similarity had a larger effect compared to common spatial origin. Dent et al. (2016) showed, in zebra finches and budgerigars, stronger effects of spatial location and sound intensity than spectral cues in detecting a missing syllable from a stimulus bird song. Schwartz and Serratto Del Monte (2019) found no additional improvement in segregation when sounds that were spatially separated were also separated in their frequencies. The present study adds to this small but growing literature on auditory object formation in the presence of multiple cues in non-human animals. It is worth expanding the present findings by incorporating more auditory object formation cues. In future work, besides manipulating temporal onset asynchrony and inharmonicity, spatial location of the conspecific and heterospecific spectral components could also be manipulated using the present study design.

Perceptual organization and neuroethological theories of auditory processing

An important finding to emphasize from this study was that sounds were integrated into the same auditory object *only if* they shared commonalities across multiple acoustic attributes, that is, both synchronous temporal onsets and harmonic relationships. It is worth discussing these findings in the light of how natural signals are structured and neuroethological theories of auditory processing. Natural signals often possess multiple spectro-temporal commonalities like harmonicity and slow temporal fluctuations often correlated across frequency components (Kershenbaum et al., 2016; Nelken et al., 1999; Singh and Theunissen, 2003; Suga, 1992). These spectro-temporal commonalities are borne out of natural physics of sound production. Neuroethological theories of auditory processing argue that neural systems in animals have evolved to process behaviorally relevant communication signals (Theunissen and Elie, 2014). These theories are supported by the discovery of neurons that respond best to features present in conspecific signals (Doupe and Konishi, 1991; Feng et al., 1990; Fitzpatrick et al., 1993; Grace et al., 2003; Römer, 2016). Specifically in frogs, some neurons are maximally tuned to the sound frequencies that are present in the conspecific signals (reviewed in Feng et al., 1990; Simmons, 2013), selective to the combination of low and high frequency conspecific components (Fuzessery and Feng, 1983; Lee et al., 2017; Megela, 1983), or respond only to species-specific temporal patterns (Edwards et al., 2002; Gupta et al., 2021).

Based on the neuroethological theories and the supporting evidence, auditory object formation may be based on the design principles of natural signals. That is, both temporal synchrony and harmonicity are required to promote perceptual integration because signal components in natural signals possess both of these regularities. This possibility could be tested by integrating various bioacoustic, behavioral and neurophysiological approaches in the same study system. For instance, the behavioral findings of the present study on *H. cinerea* could be compared against biologically realistic values of temporal synchrony and harmonicity present within and between signals of different males. Further, neurophysiological responses to the present experimental stimuli could be acquired to estimate the neural mechanisms of auditory object formation. Auditory processing at both the peripheral and central levels could be possible candidates for future neurophysiological investigations. Three distinct population of auditory neurons have been identified in the VIII nerve of *H. cinerea*, a low frequency (< 500 Hz) tuned, a mid-frequency (500-1200 Hz) tuned, and a high frequency (3100-3800 Hz) tuned population (Capranica et al., 1976; Moffat and Capranica, 1974). Ehret et al. (1983) found that responses of certain low frequency tuned VIII nerve fibers in *H. cinerea* exhibited a ‘two-tone’ suppression wherein certain inhibitory tones (comprising frequencies typical of heterospecific signals) suppressed the responses of those neurons to their excitatory frequencies. Further, neural recordings from the Inferior Colliculus and Thalamus by Fuzessery and Feng, (1983) in *Rana pipiens*, Megela, (1983) in *Rana catesbeiana*, and Mudry and Capranica (1987) and Lee et al., (2017) and in *Hyla cinerea* showed the presence of certain combination-sensitive neurons that exhibited increased sensitivity at two distinct frequency regions and respond either maximally or exclusively to the combination of low and high frequencies present in the conspecific calls. Mudry and Capranica (1987) demonstrated that combination sensitive neurons also exhibit suppression upon added stimulation by certain inhibitory frequencies (1200-1800 Hz; frequently reflected in the natural signals of heterospecific and also used as one of the heterospecific components in this study). This neural suppression of peripheral and central neurons in response to added stimulation by inhibitory frequencies is reflected behaviorally too, as demonstrated in studies by Gerhardt (1974b) and Gerhardt and Höbel (2005). In those studies, adding a heterospecific spectral component to a conspecific call starkly reduced its attractiveness. We hypothesize that auditory objects formation may be partially encoded by changes in the response patterns of the VIII nerve fibers and the combination sensitive neurons. Future studies could measure how shifts in temporal synchrony and harmonicity (comparable to the values used in this study) between the conspecific and heterospecific spectral components may introduce changes in the suppression patterns of the VIII nerve fibers and the combination sensitive neurons.

Chapter 2

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

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 gray 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 gray 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 suggest the hypothesis that any perceptually salient acoustic difference can be exploited as a cue for stream segregation is not supported in all species. We discuss these findings in the context of cues used for species recognition and auditory streaming. [1]

---

[1] This chapter is the peer reviewed version of the following article: Kalra, L., Altman, S. & Bee, M.A. Perceptually salient differences in a species recognition cue do not promote auditory streaming in eastern grey treefrogs (*Hyla versicolor*). *J Comp Physiol A* (2024). https://doi.org/10.1007/s00359-024-01702-9

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 (Remez, 2021; Repp, 1988) or recognizing a melody in an orchestral piece (Dowling, 2012; McDermott and Oxenham, 2008) 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, 1994). 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, 1994), 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, 1994; Bregman and Campbell, 1971; Moore and Gockel, 2002; Moore and Gockel, 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 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 1994; 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 (Hyland Bruno and Tchernichovski, 2019; Winn et al., 1981), 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 (Bee and Micheyl, 2008; Gerhardt and Huber, 2002; Greenfield, 2005). 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 gray treefrog, *Hyla versicolor*, we used the ABAB stimulus paradigm to test the hypothesis that perceptually salient acoustic differences promote auditory streaming. The eastern gray 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 (Figure 2-1a) 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 (Figure 2-1b) (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 (Figure 2-1a). 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 (Figure 2-1c) (Gerhardt and Doherty 1988; Ptacek et al. 1994; Gupta et al. 2021). Pulse amplitude rise and fall patterns – together described as “pulse shape” – facilitate species recognition 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 (Figure 2-1c) over pulses that have faster rise times and an overall shape that more closely resembles the heterospecific pulses of *H. chrysoscelis* (Figure 2-1d). 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 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 (Figure 2-1a) over the faster pulse rate of *H. chrysoscelis* calls (Figure 2-1b), 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 broadcasted 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 (Figure 2-1e) had the pulse duration and rise time typical of conspecific pulses (Figure 2-1c) and were expected to be preferred by the females. The “B” pulses (Figure 2-1f) 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 (Figure 2-1d). 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 (Figure 2-2). 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 (Figure 2-2).

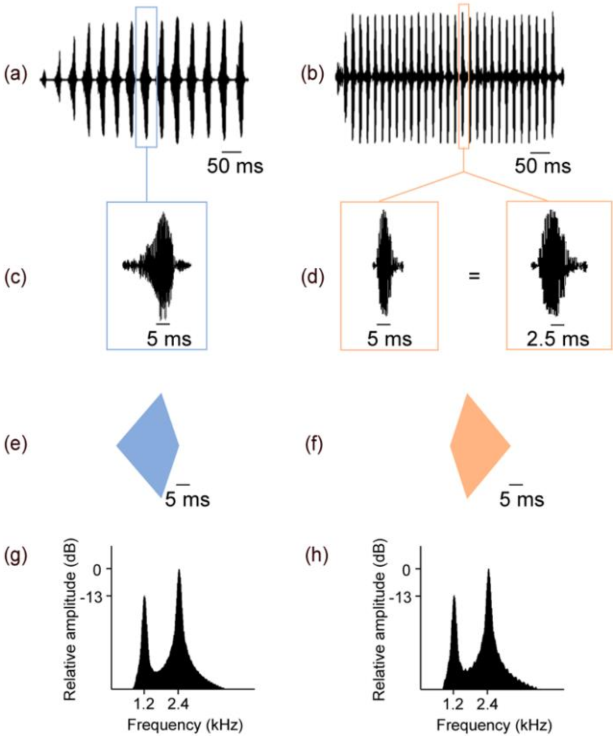

Figure 2-1. Natural and synthetic signals of *H. versicolor* and *H. chrysoscelis*. **a** Oscillogram of a natural advertisement call of *H. versicolor*. **b** Oscillogram of a natural advertisement call of H. chrysoscelis depicting a faster pulse rate compared to *H. versicolor*. **c** A highlighted 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 magnified two-fold to highlight that the pulse shape (relative rise and fall patterns) in *H. chrysoscelis* is 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].

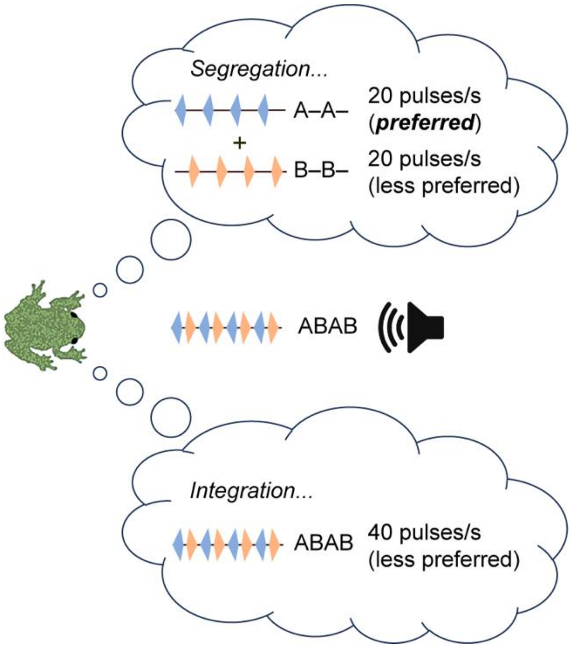

Figure 2-2. Protocol for testing auditory streaming. We broadcasted interleaved pulsatile sequences of A (in blue) and B (in orange) pulses (ABAB) to female *H. versicolor*. If pulse rise time differences were sufficient to promote segregation of sounds, we expected the females to perceive two distinct 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.

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 minutes 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 broadcasted alternating synthetic models of the two species' calls (as 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 hours 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, U.S.A.) connected to Marantz PMD620 or PMD670 recorders (D&M Professional, Itasca, IL, U.S.A.). 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 (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 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 shape, which we characterized by measuring pulse rise time (ms, time from pulse onset to peak amplitude) and pulse fall time (ms, time from pulse peak amplitude to offset). To place measures of pulse shape 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 spectral components (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 (Figure 2-1g, 1h). 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 sound pressure level (either 80 dB or 100 dB). We used these sound pressure 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 (Figure 2-1e) 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 (Figure 2-3a) 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 (Figure 2-3b) 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 (Figure 2-1f) 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 expectations from 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 (Figure 2-4a). As discussed below, this prediction was supported by the data, thus allowing us to use the perceptually salient 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 (Figure 2-3c). 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," Figure 2-3d) or all B pulses ("BBBB," Figure 2-3e). The final stimulus was created by interleaving pairs of A ("AA—," Figure 2-3f) and B ("BB—," Figure 2-3g) pulses so that, like the ABAB stimulus, this "AABB" (Figure 2-3h) 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 was comprised of 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 ms 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, 2005b; Gerhardt and Doherty, 1988; Gerhardt and Schul, 1999). 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 (Figure 2-4b).

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 (Figure 2-4c).

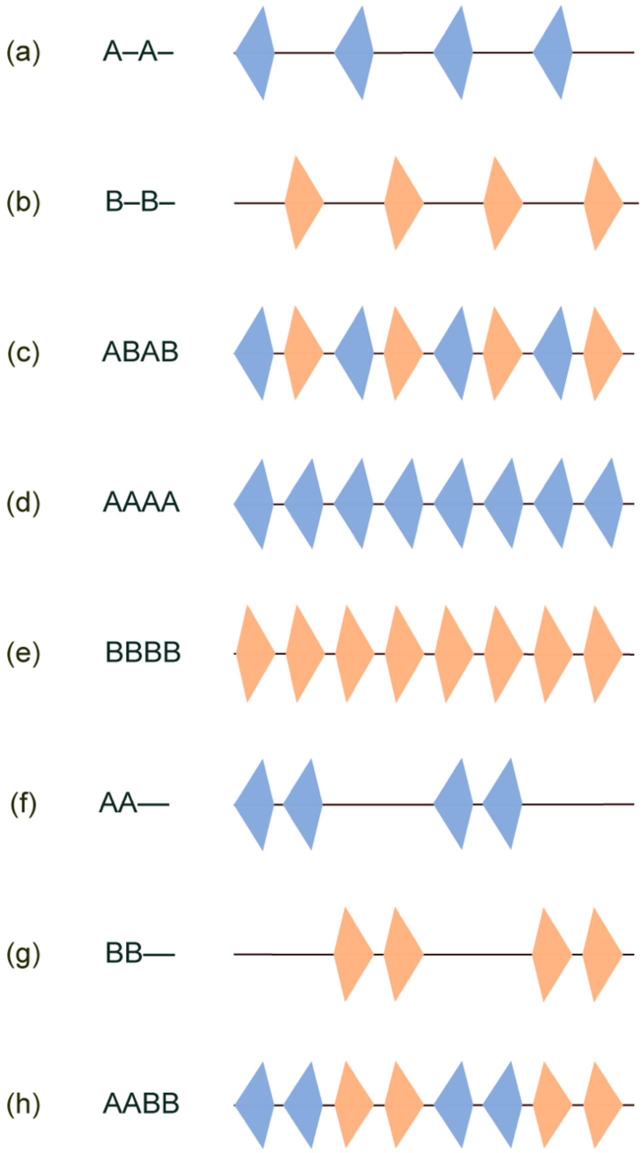

Figure 2-3. Schematic segments of the acoustic stimuli for behavioral experiments. **a** "A–A–," and **b** "B–B–" were constructed as sequences of A and B pulses, respectively, repeating at a rate of 20 pulses/s and having a regular inter-pulse interval of 30 ms. **c** "ABAB" was constructed by temporally 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 ms 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.

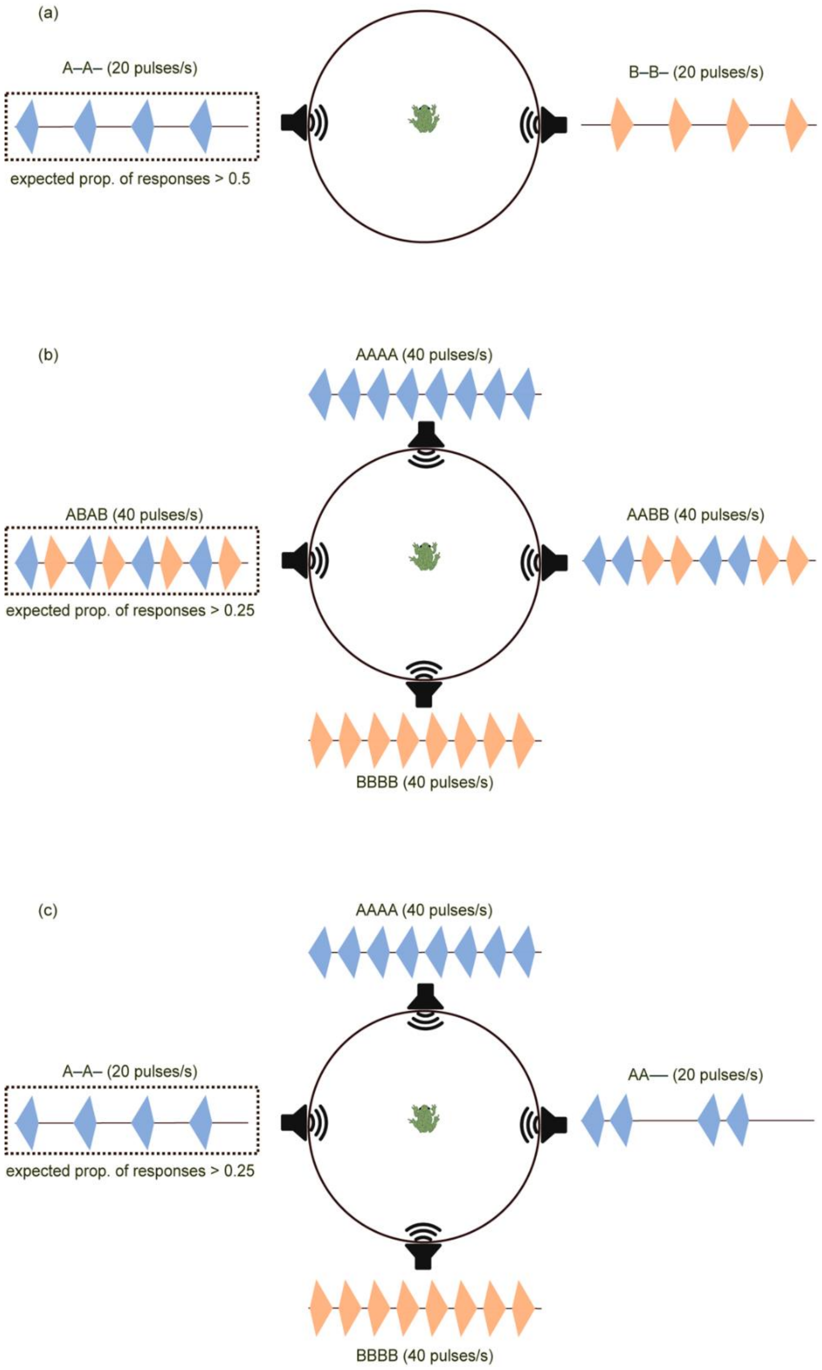

Figure 2-4. Design and predictions for the behavioral experiments. **a** Perceptual salience test. If the rise time differences between pulses A and B are perceptually salient, subjects were expected to prefer A–A– stimulus more than the chance probability of 0.5. **b** Auditory streaming test. If perceptually 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.

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 and separated by 90° 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. For four-choice tests, alternative stimuli were broadcast through four different speakers while for two-choice tests, alternative stimuli were broadcasted through two speakers located 180° apart. Since all the alternatives had the same call rate of 10 calls/min, the consecutive calls within each alternative stimulus sequence were separated by a silence of 5.23 s. Further, the temporal order of broadcasting different alternatives was such that each call of a given alternative stimulus was also separated by an equal silent interval from the calls of the alternative stimulus preceding and succeeding it. As such, there was no temporal overlap between the calls of the same as well as different alternatives (Figure 2-5). We further controlled for any leader/follower relationships by randomizing, across subjects, the order in which the very first calls of different stimuli were broadcast in a playback. Additionally, we also randomized which speaker was assigned for broadcasting any given alternative stimulus to ensure that the alternative stimuli weren't always broadcast from the same spatial location around the arena.

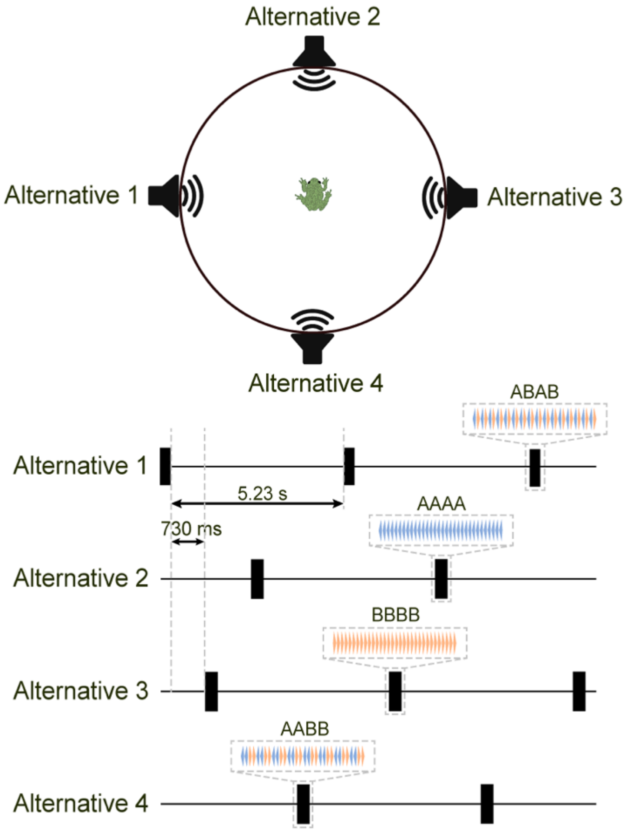

Figure 2-5. Setup for broadcasting alternative stimuli in the auditory streaming test. Here, only one example of the random temporal orders and speaker assignments (spatial location) for broadcasting the alternative stimuli is shown. The temporal order and speaker assignments for broadcasting alternative stimuli was randomized across subjects. Across all the tested temporal order and assigned speaker combinations for the broadcast, any given call was always separated by a silence of 730 ms from the call of a different alternative preceding or succeding it (as shown between the calls of alternative 1 and alternative 3) and by a silence of 5.23 s from the subsequent calls within the same alternative stimulus (as shown between the two consecutive calls of alternative 1).

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 through an overhead IR camera mounted directly over the test arena. 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.

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 2-1. 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 out of 43) and 91% (39 out of 43) of the females tested responded by making a choice at sound pressure levels of 100 dB and 80 dB, respectively. We predicted that females 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 data were consistent with this prediction. The proportion of subjects choosing 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 chose A–A– ($n = 42$, $p < 0.001$) and at 80 dB, approximately 92% of subjects chose A–A– ($n = 39$, $p < 0.001$) (Figure 2-6a). 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 out of 43) and 70% (30 out of 43) of the females tested responded by making a choice at sound pressure levels of 100 dB and 80 dB, respectively. We predicted females 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 the chance probability of 0.25 (exact binomial test, $\alpha = 0.05$) at either of the signal levels. At 100 dB, approximately 15% of subjects chose ABAB ($n = 33$, $p = 0.231$) and at 80 dB, approximately 37% of subjects chose ABAB ($n = 30$, $p = 0.143$) (Figure 2-6b).

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 out of 43) and 72% (31 out of 43) of the females tested responded by making a choice at sound pressure levels of 100 dB and 80 dB, respectively. We predicted that females 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 a chance probability of 0.25 (exact binomial test, $\alpha = 0.05$) at both signal levels. The percentage of subjects choosing A–A– was approximately 81% at 100 dB ($n = 36$, $p < 0.001$) and 61% at 80 dB ($n = 31$, $p < 0.001$) (Figure 2-6c).

Table 2-1 Descriptive statistics of acoustic properties of *H. versicolor* advertisement calls ($n = 30$ males) recorded in Minnesota and standardized to a temperature of 20º C. The range of temperatures at which males were recorded was 10.2 º to 29.0º 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

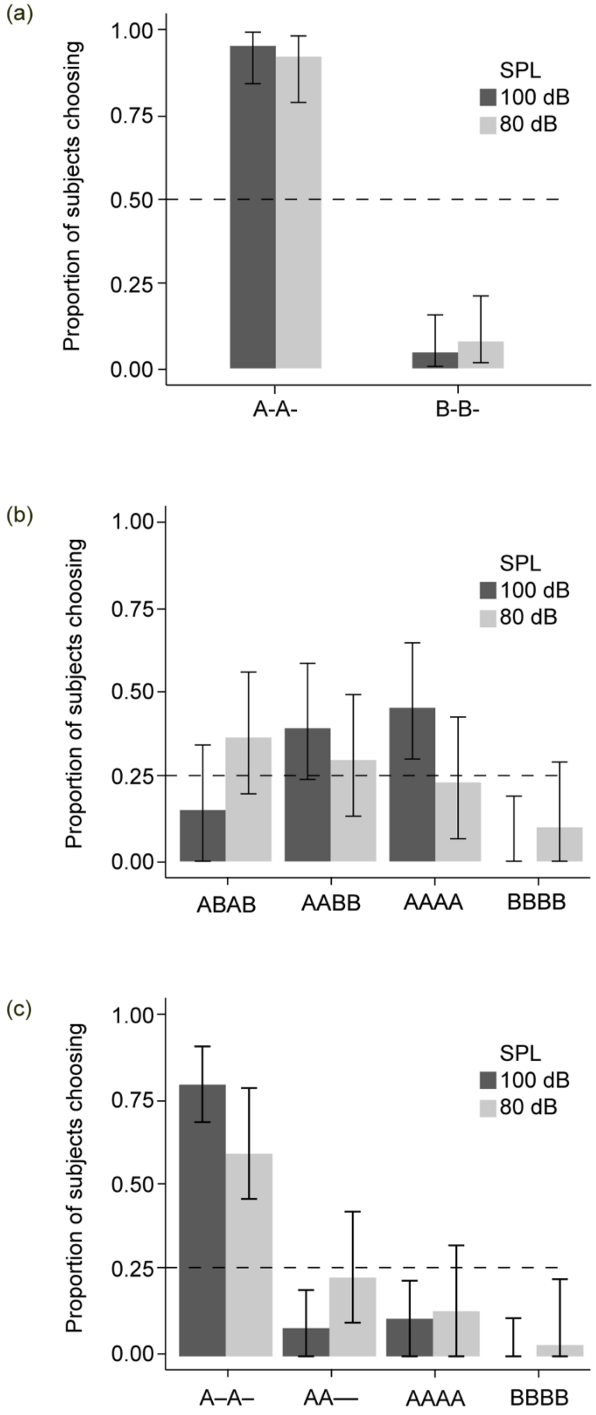

Figure 2-6. Results for behavioral experiments. Black and gray bars indicate the proportions of subjects choosing a given stimulus at 100 dB and 80 dB, respectively. a Results for the perceptual salience test. Error bars depict exact 95% binomial confidence intervals (CIs). b Results for the auditory 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.

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-choice tests. However, there was no evidence that this salient acoustic 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 2-1) 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-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 the same 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-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 be 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. Therefore, in planning this study, we viewed the four-alternative choice design to be a more robust empirical test of streaming than a two-alternative choice test that predicted 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 (≥ 70%) chose one of the four stimuli in our auditory streaming test (including the ABAB and AABB stimuli), which suggests B pulses were not inherently aversive. Results from an additional four-alternative choice test (see Appendix 2) 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 this other work, 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 streaming.

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 the 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, which 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 (Anstis and Saida, 1985; Bregman, 1978; Deike et al., 2012; Micheyl et al., 2005). 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 gray 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 gray 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. (1994 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.

Chapter 3
Evidence of Rhythmic Masking Release in Cope's gray treefrog

Abstract

Auditory streaming is the process of perceptually assigning overlapping sound sequences in the environment to their respective sources. Sounds produced by the same source must be integrated and perceived as a single auditory stream while those produced by different sources must be segregated into distinct auditory streams. Understanding of how auditory streaming mediates signal perception in complex acoustic environments remains limited in non-human animals. Here, in Cope's gray treefrog, we studied auditory streaming of multiple sequences using the Rhythmic Masking Release (RMR) paradigm from human psychoacoustics. In this paradigm, the distinct temporal rhythm of a Target sequence gets masked in presence of the identical pulses of a Distractor sequence. A release from masking (RMR) is induced when an added Captor sequence perceptually captures the Distractor into an auditory stream, separate from the Target. Here, we modelled the Target as a repeated pulsatile call exhibiting a species-typical rhythm (pulse rate). Using phonotaxis (movement towards sound), we tested how Target is perceived in presence of pulsatile Distractor and Captor sequences. Our findings provide evidence for Rhythmic Masking Release in frogs. Behavioral responses to the Target went down when it was presented with its pulses temporally interleaved with the pulses of the Distractor, indicating a masking of the attractive temporal rhythm. However, upon the addition of the Captor to the Target and Distractor complex, there was a significant increase in the behavioral responses to the Target compared to when it was interleaved with only the Distractor, indicating a release from masking of the Target.

Introduction

Acoustic communication in humans and other animals involves the production and perception of sound sequences (Kershenbaum et al., 2016). Perceiving sound sequences in complex acoustic scenes is challenging because multiple signalers often produce signals concurrently (Brumm and Slabbekoorn, 2005; Gerhardt and Huber, 2002; Greenfield, 2005). Receivers must perceptually organize temporally overlapping signals into distinct auditory streams, each corresponding to perceptual representations of the signals produced by different individuals (Hulse, 2002; McDermott and Oxenham, 2008; Remez, 2021). Termed 'auditory streaming,' the process of perceptually assigning different sounds to the same or different streams depends on the degree of similarity in their acoustic features (reviewed in Bee et al., 2010; Bregman, 1990; Darwin, 1997; Dent and Bee, 2018; Moore and Gockel, 2012; Shamma et al., 2011). For example, sounds having the same fundamental frequency, spectral and temporal modulation patterns, onset/offset times, and spatial origin tend to be perceptually integrated into the same stream. Sounds having sufficiently large differences in these same features tend to be perceptually segregated into different streams. While auditory streaming has been extensively investigated in the contexts of human speech and music perception, its role in nonhuman acoustic communication has received much less attention (Bee and Micheyl, 2008; Dent and Bee, 2018; Hulse, 2002).

Frogs represent a taxonomic group for which auditory streaming is important in evolutionarily consequential contexts, such as reproduction and aggression (Bee, 2015; Dent and Bee, 2018; Gerhardt and Huber, 2002). Frogs communicate in these contexts using sound sequences consisting of repeated calls, which are themselves frequently composed of sequences of shorter acoustic elements, such as repeated pulses or different note types (Gerhardt, 1991; Koehler et al., 2017; Narins and Capranica, 1978; Pettitt et al., 2012; Ryan, 1985; Toledo et al., 2015). Moreover, frogs often communicate in complex acoustic scenes, such as breeding choruses, that are characterized by the presence of temporally and spectrally overlapping sound sequences produced by multiple males, often from multiple different species (Feng and Schul, 2006; Wells and Schwartz, 2007). Female frogs recognize, assess, localize, and discriminate among potential mates in such acoustic scenes by displaying phonotaxis behavior, which involves selectively approaching a preferred mate (Gerhardt, 1982; Gerhardt, 1995; Gerhardt and Bee, 2007). In many species, the temporal features of sound sequences, such a call's pulse rate or its total number of pulses, form the acoustic basis for species recognition and intrasexual discrimination (reviewed in Gerhardt and Huber, 2002).

Previous studies of auditory streaming in Cope's gray treefrogs (*Hyla chrysoscelis*) suggest that, like humans, frogs can exploit similarities and differences in some spectral and temporal features, as well as spatial origin, in assigning concurrent sounds to the same or different streams (Bee, 2010; Farris et al., 2002, 2005; Gupta and Bee, 2020; Nityananda and Bee, 2011 *but also see* Bee and Riemersma, 2008; Kalra et al., 2024; Schwartz and Gerhardt, 1995; Schwartz and Serratto Del Monte, 2019). In this species, the temporal rhythm of the call's pulse rate (approximately 40 to 60 pulses/s) is an important species recognition cue, because both faster and slower pulse rates are less attractive and elicit little or no phonotaxis from females (Blair, 1958; Gerhardt, 1978; Gupta and Bee, 2020; Ward et al., 2013a). Pulse rate is considered a "static" call property because it exhibits low levels of variation within males (Gerhardt, 1991); it also exhibits low levels of variation between males (Ward et al., 2013a). Experiments investigating auditory streaming in this species have successfully employed two temporally interleaved pulse sequences (ABAB…) to measure perceptual integration or segregation as a function of the acoustic differences between the two sequences (A–A–… and B–B–…). In a previous investigation of spectral proximity as an auditory streaming cue, Nityananda and Bee (2011) temporally interleaved the pulses of a 'Target' call, recognized by its species-typical pulse rate, with a continuous sequence of pulses produced at the same rate in a behaviorally neutral 'Distractor'. When the Target and Distractor were similar in frequency (i.e., high spectral proximity), the two sequences became perceptually integrated into a single stream. Consequently, recognition of Target call based on the rhythm conveyed by its pulse rate was impaired, as evidenced by reduced phonotaxis behavior. Phonotaxis to the Target increased as a function of the frequency difference (spectral separation) between Target and Distractor, indicating frequency separation promoted the assignment of Target and Distractor into distinct streams.

In this study of *H. chrysoscelis*, we build on the findings of Nityananda and Bee (2011) to adapt the Rhythmic Masking Release (RMR) paradigm (Turgeon et al., 2002, 2005) from human psychoacoustic studies to investigate auditory streaming of multiple sound sequences in a frog. The RMR paradigm involves the auditory streaming of three pulsatile sequences: a Target having a distinct temporal rhythm (Figure 3-1a), a Distractor spectrally proximate to the Target (Figure 3-1b), and a Captor that is temporally coherent with the Distractor but spectrally separated from the Target and Distractor (Figure 3-c, e). Temporally interleaving the Target and Distractor can impair the ability of human listeners to identify the Target when its temporal rhythm is masked by the Distractor (Figure 3-1d). However, adding the Captor to the Target and Distractor complex can cause the masked rhythm of the Target to perceptually re-emerge (Figure 3-1f). This release from rhythmic masking is believed to occur because the temporal coherence of the Captor and Distractor promotes their assignment to the same auditory stream and their segregation from a distinct Target stream. The extent to which RMR is experienced by nonhuman animals, and whether it could be a useful paradigm to study auditory streaming in animals, remain largely unexplored. Testing Rhythmic Masking Release in non-human animals may shed light on how auditory streaming of multiple sequences impact recognition of important temporal patterns within biologically relevant signals.

Here, the Target stimulus was a repeated advertisement call having the species-typical pulse rate that elicits positive phonotaxis from females. The Distractor stimulus was a continuous sequence of pulses having the same frequency and pulse rate as the Target. The Captor stimulus was also a continuous pulse sequence in which the pulses were temporally coherent with those in the Distractor but were an octave apart in frequency from the Target and Distractor. We conducted phonotaxis tests in which the Target was presented by itself or with its pulses temporally interleaved with those of the Distractor, the Captor, or both the Distractor and Captor. The percentage of female subjects exhibiting positive phonotaxis was used to measure recognition of the Target. Based on Nityananda and Bee (2011), we expected the spectral proximity of the Distractor to mask recognition of the Target’s pulse rate, and thus reduce phonotaxis. If females experienced RMR, we predicted the addition of the Captor to the Target + Distractor complex would cause a release from masking of the Target, and hence elicit greater phonotaxis. This prediction follows from the observation that, as in humans (Elhilali et al., 2009; Shamma et al., 2013), temporal coherence promotes auditory grouping in gray treefrogs (Gupta and Bee, 2020). Hence, we expected their temporal coherence to bind the Distractor and Captor into the same auditory stream, allowing the Target to be more easily segregated and recognized.

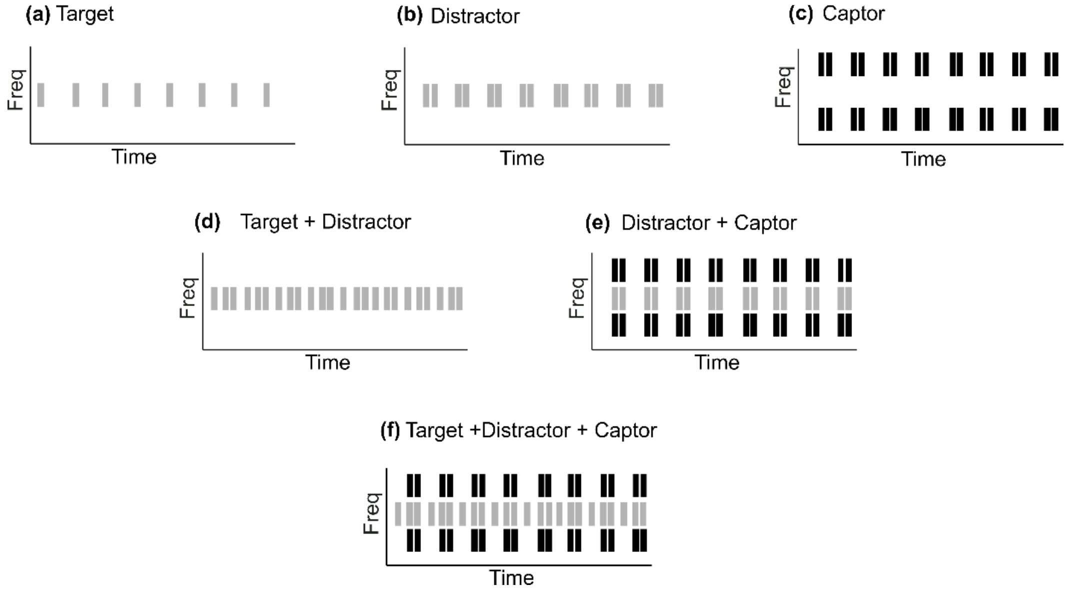

Figure 3-1. Components of the RMR paradigm (Turgeon et al., 2002b) used in humans. **a** Target sequence of pulses with a distinct repetition pattern or temporal rhythm. **b** Distractor sequence with pulses spectrally proximate as that of Target. **c** a Captor sequence of pulses spectrally separated from the Target and Distractor. **d** Perceptual masking of Target's rhythm upon temporal interleaving with Distractor. **e** The Distractor and Captor shown together to highlight the temporal coherence shared between their pulses. **f** Addition of Captor to Target and Distractor complex elicits Rhythmic Masking Release as the perception of the Target's temporal rhythm re-emerges.

Methods

Study system and Subjects

During their breeding season, males of *H. chrysoscelis* advertise to attract females by producing a series of pulsatile calls repeated, on average, at rate of about 11 calls/min. The pulses within each call repeat at an average pulse rate of about 50 pulses/s at 20°C (Ward et al., 2013a). Each pulse has two harmonically related spectral peaks at approximately 1.25 kHz and 2.5 kHz. The lower and higher spectral peaks correspond to the frequencies of peak sensitivity of the two sensory papillae in the gray treefrog inner ear that transduce airborne sound frequencies, the amphibian papilla and basilar papilla, respectively (Gerhardt and Schwartz, 2001; Hillery and Narins, 1984; Schrode et al., 2014; Zakon and Wilczynski, 1988). In two-alternative choice tests, females prefer synthetic calls having both spectral peaks ('bimodal' signals) over alternatives having only the lower or upper spectral peak ('unimodal' signals), but they nevertheless exhibit robust phonotaxis to unimodal signals in no-choice tests (Bee, 2010; Gerhardt, 2008; Gupta and Bee, 2020; Nityananda and Bee, 2011).

Gravid females of the Western clade of *H. chrysoscelis* (Booker et al., 2022) were used as subjects in this study. Animal collections occurred at night (2200 to 0100 h) during the breeding season between mid-May and early July in 2020, 2021, and 2022, from wetlands in the Carver Park Reserve (Carver County, MN, USA), the Hyland Lake Park Reserve (Hennepin County, MN, USA), and the Tamarack Nature Center (Ramsey County, MN, USA). Subjects were collected in amplexus with a male and returned to the laboratory, where they were held at approximately 4°C to delay oviposition until testing occurred (typically within one day) (Gerhardt, 1995). Pairs were placed in an incubator 30 minutes prior to testing to allow their body temperatures to reach 20±1°C. Subjects were separated from their mates for testing, then returned to their mates and placed in the incubator between subsequent tests for a minimum of 5 min to preserve responsiveness and maintain body temperature. All pairs were returned to their collection site within 48 hours of the completion of experiments.

Because *H. chrysoscelis* breeds syntopically at some of our field sites with its morphologically indistinguishable sister species (*H. versicolor*), subjects collected from those sites were tested in an initial ‘positive control’ test to confirm their species identity. In this test, subjects were given a choice between two alternating synthetic calls modeled after the two species’ advertisement calls (as in Gupta et al., 2021; Kalra et al., 2024). The acoustic features of these stimuli were based on bioacoustic analyses of natural calls (Kalra et al., 2024; Ward et al., 2013a). Only subjects that approached the *H. chrysoscelis* stimulus in this test were used in subsequent behavioral tests.

General protocol

Our testing protocols generally followed those described in previous studies (Bee and Schwartz, 2009; Bee and Vélez, 2018; Bee et al., 2012; Vélez et al., 2013). All behavioral tests were conducted in an acoustically transparent and visually opaque circular arena (2-m diameter, 60-cm height) with walls made of hardware cloth covered with black fabric. The arena was located on the carpeted floor of a hemi-anechoic chamber (length × width × height: 2.8 × 2.3 × 2.1 m; Industrial Acoustics Company, IAC, North Aurora, IL, USA). Speakers (Orb Audio, Sherman Oaks, CA, USA) were placed behind the arena wall and facing inward toward the center of the arena. An acoustically transparent cage was also located at the center of the arena that enabled us to remotely release subjects from outside the chamber at the beginning of a behavioral test. Because female preferences for some call features vary with temperature (Gerhardt & Huber, 2002), the anechoic chamber was maintained at a constant temperature of 20±1°C.

Acoustic stimuli were played through Adobe Audition 3.0 (Adobe Systems Inc., San Jose, CA, USA) running on a Dell Optiplex 980 PC (Dell Computer Corporation, Round Rock, TX, USA) system. The output from Adobe Audition was delivered through a MOTU model 16A sound card (MOTU, Inc., Cambridge, MA, USA), amplified using a Crown XLS 1000 High-Density Power Amplifier (Crown International, Los Angeles, CA, USA), and broadcast through the assigned speaker. The sound pressure level (SPL, LCF, re 20 µPa) of each stimulus was calibrated by placing a Bruël and Kjær Type 4950 microphone connected to a Bruël and Kjær Type 2250-L sound level meter (Bruël and Kjær, Nærum, Denmark) 1 m away from each speaker at the approximate position of a subject's head at the release point.

At the start of a test, a single subject was placed inside the release cage for 105 s, which included a silent period of 60 s for acclimation followed by 45 s of stimulus playback. After 105 s, a timer was simultaneously started with lifting the lid of the release cage. From this point onwards, the subject was free to move around inside the arena while the stimuli continued to play. A response was recorded if the subject entered a small response zone (10-cm radius) directly in front of the speaker within the first 5 min of the test. A no-response was recorded if the subject did not exit the release cage within 3 min or exited the release cage but either first touched the arena wall in the semicircle opposite the stimulus speaker or did not enter the response zone within 5 min.

Acoustic stimuli

Synthetic acoustic stimuli were generated using MATLAB R2019b (Mathworks, Natick, MA, USA) at a sampling rate of 44.1 kHz and a bit-depth of 16. The Target, Distractor and Captor each consisted of sequences of pulses produced at the same rate of 50 pulses/s. The temporal features of all pulses were identical and approximated the average pulse from a *H. chrysoscelis* call (20-ms pulse period, 10-ms pulse duration, 3.4-ms inverse exponential rise-time; 5.8-ms exponential fall-time) (Ward et al., 2013a). The pulses within each stimulus had identical frequency spectra consisting of a single spectral peak of either 1.25 kHz or 2.5 kHz, depending on the specific design of each test.

The Target simulated a calling male and was constructed as a sequence of identical pulsatile calls separated by silent intervals and repeated at a rate of 11 calls/min, which approximates the average call rate at 20°C (Ward et al., 2013a). Each repeated call (Figure 3-2a) consisted of 30 identical pulses (590 ms call duration), which approximates the average call duration (Ward et al., 2013a). In contrast to the Target, both the Distractor and Captor consisted of a continuous sequences of pulses instead of temporally separated calls (Figure 3-2b-c). Continuous pulse sequences like the Distractor and Captor are behaviorally neutral and elicit neither positive nor negative phonotaxis from females (Nityananda and Bee, 2011), a result replicated in the present study (see Appendix 3).

Because sound-evoked behaviors in frogs can vary with both a signal's carrier frequency and sound level (Gerhardt, 1987; Gerhardt, 2008; Gupta and Bee, 2023), we attempted to generalize the result of this study in the following two ways. First all tests were replicated (within subjects) using each of two frequencies for the Target, either 1.25 kHz or 2.5 kHz. In any given test, the Distractor's frequency was always the same as that of the Target, while the Captor's frequency was always an octave higher or lower. When the Target frequency was 1.25 kHz, the Distractor and the Captor frequencies were 1.25 kHz and 2.5 kHz, respectively and vice-versa. Second, all tests were replicated (between subjects) using a sound pressure level for calls in the Target of either 67 dB SPL or 73 dB SPL. These values are lower than the SPL of natural calls (Gerhardt 1978). They were chosen in an attempt to reduce simultaneous stimulation of both inner ear papillae with unimodal stimuli (Gerhardt, 2005a) while also exceeding both behavioral and neural response threshold levels (Hillery & Narins, 1984; Bee & Schwartz, 2009; Nityananda & Bee, 2011). In all tests, the Distractor and the Captor were separately calibrated to 67 dB SPL. Because the pulses of the Distractor and Captor were always temporally coherent in tests in which they were presented together, the resulting sound level of the Distractor + Captor complex was 70 dB SPL.

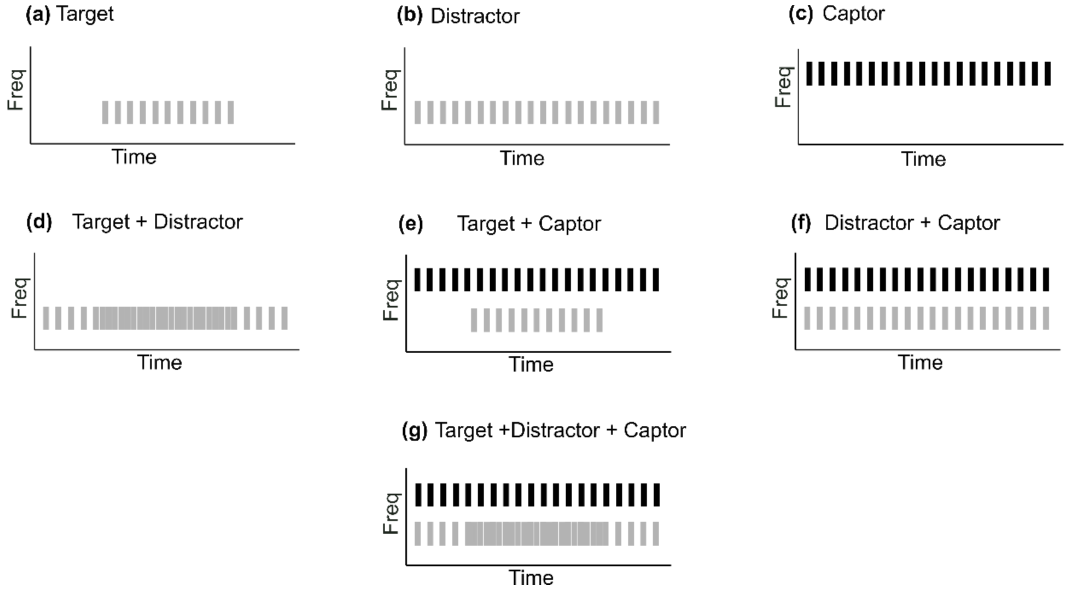

Figure 3-2. Components of the RMR paradigm adapted for *H. chrysoscelis*. **a** Target call with a pulse rate of 50 pulses/s is expected to be behaviorally attractive. Behaviorally neutral continuous sequence of pulses (pulse rate = 50 pulse/s) constituting **b** the Distractor were composed of the same frequency as Target's pulses and **c** the Captor were composed of frequency that was an octave apart from the Target and Distractor. **d** Upon temporal interleaving with the Distractor, perceived pulse rate of the Target becomes 100 pulses/s (unattractive). **e** Temporal interleaving with the spectrally remote Captor induces relatively less masking than that with the Distractor. **f** The Distractor and Captor shown together to highlight the temporal coherence between their respective pulses. g Rhythmic Masking Release was predicted when Target is simultaneously interleaved with both the Distractor and Captor. Panels a-**g** depict a small segment of each stimulus. See text for details on the exact duration of and number of pulses within each stimulus.

Experimental design

Four types of no-choice phonotaxis tests (n = 30-35 per test) were conducted in which either a single stimulus or multiple stimuli were broadcast through a single speaker. In the 'Target only' test, we assessed recognition of the Target in the absence of any concurrent stimuli. Based on earlier work, we expected robust phonotaxis responses in this test (Bee, 2010; Gerhardt, 2008; Gupta et al., 2023; Nityananda and Bee, 2011). In the 'Target + Distractor' test, the Target and Distractor were broadcast simultaneously such that the pulses in each call of the Target were temporally interleaved with the continuous pulses of the Distractor (Figure 3-2d). Based on their spectral proximity (Nityananda and Bee, 2011), we expected the Target and Distractor to be perceptually integrated into a single stream having a composite unattractive pulse rate of 100 pulses/s each time a call in the Target stimulus was broadcast (Figure 3-2d). Consequently, we predicted the percentage of subjects responding in the 'Target + Distractor' test would be significantly lower compared to the ‘Target only’ test. In the ‘Target + Captor’ test, we simultaneously broadcast the Target and Captor such that pulses of each call in the Target were temporally interleaved with the continuous pulses of the Captor. Because the frequency separation between the Target and Captor (one octave) was expected to promote their perceptual segregation (Figure 3-2e ; Nityananda and Bee, 2011), we predicted higher response rates in the ‘Target + Captor’ test compared with the ‘Target + Distractor’ test. However, based on the results of Nityananda and Bee (2011), we still expected a lower response rate in the ‘Target + Captor’ test compared with the ‘Target only’ test. Finally, in the main test to investigate RMR, the ‘Target + Distractor + Captor’ test (Figure 3-2g), calls in the Target were broadcast such that their pulses were temporally interleaved with the continuous and temporally coherent pulses of the Distractor and Captor. Because temporal coherence between the Distractor and Captor pulses (Figure 3-2f) was expected to promote their perceptual integration, and thus their segregation from the Target, we predicted higher response rates in the ‘Target + Distractor + Captor’ test compared with the ‘Target + Distractor’ test. For each subject, we randomized the order in which behavioral tests were performed. Between subjects, we also randomized the speaker assigned to broadcast the stimuli. We also performed a separate set of no-choice tests (n = 34) to confirm the behavioral neutrality of Distractor and Captor (see Appendix 3).

Statistical analysis

Statistical analysis was performed in R version 3.5.1 (R core team). We fitted a Generalized Estimating Equation (GEE) model (Hardin and Hilbe, 2012) with logit link functions and used Wald statistics to test our hypothesis. We used GEE because it is particularly suited for repeated measures taken for the same individual and can be applied to binomial data as well. We specified an exchangeable correlation structure, which assumes that any given pair of observations within a subject are equally correlated. The dependent variable was the outcome of each behavioral test, scored as a binary (1 = response, 0 = no response). Test type (‘Target only’, ‘Target + Distractor’, ‘Target + Distractor + Captor’ or ‘Target + Captor’) was the independent variable. Since test type was a categorical variable, output from the GEE model chose one test type as a reference and reported the effects of the other three test types in relation to the reference. To evaluate all the pairwise comparisons relevant to our predictions, we fitted the same GEE model two ways, with either the ‘Target only’ test or the ‘Target + Distractor’ test specified as the reference.

Although not of primary interest to this study, we also fitted eight additional GEE models in which Target frequency and Target SPL were added as independent variables. These models included all three independent variables and all possible combinations of the three pairwise interactions. We selected the model that best fit the data using a model selection method (see Appendix 3) based on the quasi-likelihood information criterion, QIC, an analog of Akaike Information Criterion (AIC) that is used for GEE models (Cui, 2007).

Results

Overall, differences in response rates across the four test types were consistent with predictions, and the GEE model revealed significant differences between all four test types (Table 3-1). Averaged across both Target frequencies (1.25 kHz and 2.5 kHz) and both Target SPLs (67 dB and 73 dB), approximately 93% of subjects responded in the 'Target only' tests. This high overall rate of phonotaxis in the 'Target only' tests confirms that females recognized unimodal stimulus calls presented alone as the calls of a suitable mate. Compared to the 'Target only' tests, the percentage of subjects responding declined significantly ($p < 0.01$) in the 'Target + Distractor' tests, decreasing from 93% to 14% overall (Figure 3-3a; Table 3-1). Consistent with expectations, this nearly sixfold decrease in response rate suggests the presence of the Distractor disrupted recognition of the call-like pulsatile structure of the Target by masking the rhythm of the species-typical pulse rate. Response rates were also significantly lower compared to the 'Target only' tests in both the 'Target + Captor' test ($p < 0.01$; 93% versus 72% overall) and the 'Target + Distractor + Captor' test ($p < 0.01$; 93% versus 35% overall) (Figure 3-3a; Table 3-1). However, consistent with the expectation that frequency separation would promote perceptual segregation, the percentage of subjects responding in the 'Target + Captor' test was approximately fivefold higher compared to the 'Target + Distractor' test ($p < 0.01$; 72% versus 14% overall) (Figure 3-3a; Table 3-1). The critical comparison to assess RMR was between tests in which the Target was presented with just the Distractor versus those in which it was presented with both the Distractor and Captor. Response rates were significantly higher – by more than twofold – in the 'Target + Distractor + Captor' test compared with the 'Target + Distractor' test ($p < 0.01$; 35% versus 14% overall) (Figure 3-3a; Table 3-1). This result is consistent with the interpretation that, at least for some subjects, the addition of the Captor resulted in some degree of release from the masking caused by the Distractor of the Target's species-typical rhythm. If instead the addition of the Captor to the Distractor had further impaired recognition of the Target, then response rates should have declined. In general, the pattern of results described here was broadly consistent across different levels of Target frequency (Figure 3-3b) and Target SPLs (Figure 3-3c) (also see Appendix 3).

Table 3-1. Output of a GEE model, fitted two ways, to evaluate the effects of test type on phonotaxis responses. 'Reference' represents the test type that was set as a reference while fitting the model. 'Test condition' represents the test type being compared to the reference. 'β' represents the change in the response variable for a given test condition. 'W' and '*p*' represents the results of Wald chi-square tests.

Reference	Test condition	β	W	p
'Target only'	'Target + Distractor'	-4.41	224.34	<0.01
	'Target + Captor'	-1.55	28.15	<0.01
	'Target + Distractor + Captor'	-3.22	162.60	<0.01
'Target + Distractor'	'Target + Captor'	2.85	82.00	<0.01
	'Target + Distractor + Captor'	1.19	18.10	<0.01

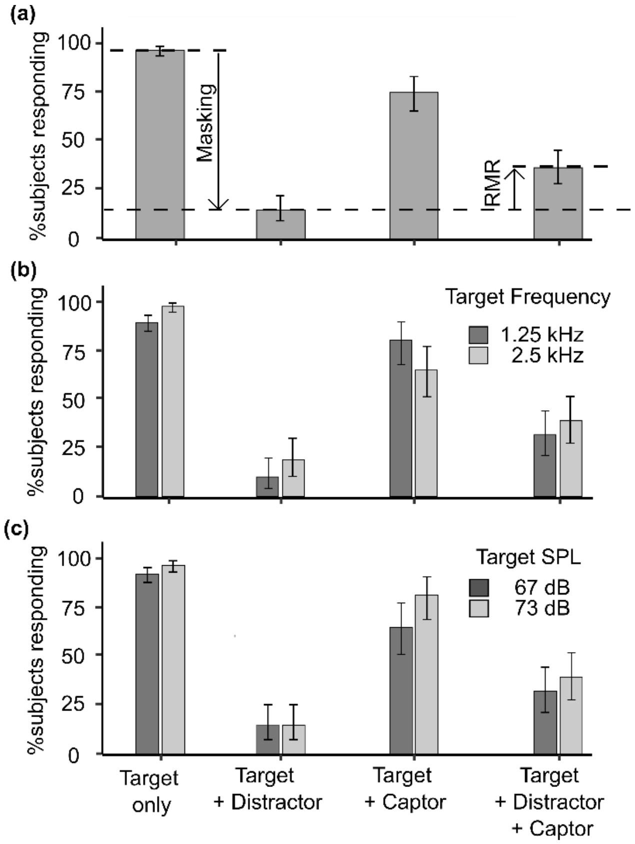

Figure 3-3. Behavioral responses across different test types. **a** Percentage of subjects responding across the four test types averaged over both Target frequencies and both Target SPLs. Masking of the temporal rhythm of the Target is indicated by the downward vertical arrow between the 'Target only' test and the 'Target + Distractor' test. RMR (release from masking) is indicated by the upward vertical arrow between the 'Target + Distractor' test and the 'Target + Distractor + Captor' test. Percentage of subjects responding across the four test types **b** averaged over both Target SPLs shown separately for each Target frequency and **c** averaged over both Target frequencies shown separately for each Target SPL. Errors bars indicate 95% binomial CIs.

Discussion

In this study, our goal was to investigate auditory streaming of multiple sequences using the Rhythmic Masking Release paradigm. We examined how a behaviorally relevant signal (Target), recognized by its species-typical pulse rate, is perceptually organized in presence of additional sound sequences, that is, the Distractor and Captor. We specifically tested if the Target, when masked by the presence of the Distractor, experiences a release from masking upon the addition of the Captor to the Target and Distractor complex. Overall, our findings provide evidence for Rhythmic Masking Release in *H. chrysoscelis.*

In presence of the Distractor, the recognition of the Target degraded due to the masking of its species-typical pulse rate. This apparent masking was elicited because the spectrally proximal Target and Distractor became perceptually integrated and assigned to the same auditory stream that had a perceived pulse rate of 100 pulses/s. Compared to when presented with the Distractor, the recognition of the Target was improved when presented with the spectrally separated Captor. This improved recognition indicated perceptual segregation of the Target and Captor, thereby facilitating the recognition of the Target based on the perceived pulse rate of 50 pulses/s. These results corroborate the findings of auditory streaming studies on insects (Schul and Sheridan, 2006), fish (Fay, 1998; Fay, 2000), birds (Dent et al., 2016; Hulse et al., 1997; Itatani and Klump, 2014) and mammals (Izumi, 2002; Ma et al., 2010; Noda et al., 2013) and contribute to the evidence that non-human animals can exploit spectral proximity as a cue for auditory streaming. More specifically, these findings replicated the results of Nityananda & Bee (2011) on the same species. Replicating this earlier work on auditory streaming in *H. chrysoscelis* was critical for laying the experimental framework to test our hypothesis regarding Rhythmic Masking Release.

The main finding of our study, which extended the results of Nityananda & Bee (2011), was that the addition of the Captor to the Target and Distractor complex elicited Rhythmic Masking Release. Response rates were significantly higher when Target was presented with both the Captor and Distractor compared to when presented with just the Distractor. This increase in response rates between the two tests indicated that, at least for some subjects, the Target was perceptually segregated from the Distractor and Captor and was recognized based on the perceived pulse rate of 50 pulses/s. This Captor induced Rhythmic Masking Release indicates that the temporal coherence between the pulses of the Distractor and Captor promoted their integration into an auditory stream that was segregated from the Target stream. This result corroborates studies in humans demonstrating the temporal coherence promotes integration of sounds during auditory streaming (Elhilali et al., 2009; Micheyl et al., 2013a; Teki et al., 2013). To our knowledge, only one study, so far, has systematically investigated how temporal coherence mediates auditory streaming in a biologically relevant context in a non-human animal. Gupta and Bee (2020) demonstrated that temporal coherence between the spectral components within a signal promotes their binding into the same auditory stream. They presented *H. chrysoscelis* females with two unimodal advertisement calls at 1.25 kHz and 2.5 kHz, respectively. Subjects perceived a single bimodal call (indicating integration) when pulses within the two calls were temporally incoherent and perceived two distinct unimodal calls (indicating segregation) when these pulses were temporally incoherent. The present findings are in line with those by Gupta and Bee (2020) and contribute to the small but growing literature on the role of temporal coherence as an auditory streaming cue in non-human animals.

To our knowledge, the present study is the first to demonstrate Rhythmic Masking Release in a biologically relevant context in a non-human animal. Temporal patterns within signals frequently mediate behavior in non-human animals across diverse contexts like species recognition, mate choice, intrasexual competition, predator avoidance and parental care (see reviews by Gerhardt and Huber, 2002; Kershenbaum et al., 2016; Nowicki and Searcy, 2004; Searcy and Andersson, 1986). More studies in future would be needed to examine how auditory streaming of overlapping sounds in the environment impact the recognition of such temporal patterns in other non-human animals. Experimental paradigms, similar to the one used in the present study could be useful in future investigations.

In humans, the RMR experimental paradigm has been successfully employed to investigate behavioral and neurophysiological correlates of auditory streaming and auditory masking (Farahbod et al., 2023; Sach and Bailey, 2004; Turgeon et al., 2002b; Turgeon et al., 2005). The present study closely adapted the experimental framework of the studies by Turgeon et al. (2002a, 2005) that investigated the impacts of multiple cues on auditory streaming in humans. Broadly, the present findings are in line with those of Turgeon et al. (2002a, 2005) demonstrating a release from masking of the Target due to the integration of temporally coherent Distractor and Captor. However, the magnitude of improvement in the recognition of the Target elicited by the addition of the Captor was much lower in the present study compared to those studies. In Turgeon et al. (2005), subjects were trained to identify two different Target rhythms and were subsequently tested on correctly identifying the Target rhythm in presence of Distractor and Captor. In that study, the percentage of correct recognition of Target when it was presented with temporally coherent Distractor and Captor was close to 100%. Recall that in this study the percentage of subjects recognizing the Target in the ‘Target + Distractor + Captor’ test was close to 35% and, thus, much lower.

These differences in the RMR in our study and the one by Turgeon et al. (2005) could be mediated by differences between frogs and humans with regards to how multiple cues may interact during auditory streaming. Broadly, the RMR paradigm introduces a competition between two perceptual representations. The Distractor could be perceptually integrated either with the Target, because of their *spectral proximity*, or with the Captor, because of their *temporal coherence*. If temporal coherence was a strong cue that could override spectral proximity, most subjects would integrate Distractor and Captor leading to an overall high magnitude of RMR and vice-versa. This is indeed the case for humans wherein, temporal coherence can override other cues like spectral similarity, harmonicity and common spatial location (Elhilali et al., 2009; Micheyl et al., 2013b). In contrast, it is possible that temporal coherence is a relatively weak cue in frogs and therefore a relatively weak RMR was observed in the present study. However, we are limited in making this conclusion in this study because we only manipulated the presence or absence of the temporally coherent Captor and not the degree of temporal incoherence and spectral differences between the sequences. Future studies could simultaneously manipulate multiple acoustic differences between sequences, using the RMR paradigm of the present study, to examine interactions between auditory streaming cues and the consequent impacts on Rhythmic Masking Release.

Neural mechanisms underlying Rhythmic Masking Release in frogs are presently unknown. In humans, neural processing at the level of primary auditory cortex contributes to auditory streaming and, more specifically, Rhythmic Masking Release. Fishman et al. (2012) used monkeys to investigate the neural correlates of RMR in humans. That study demonstrated that certain neurons in the monkey’s auditory cortex fired in response to both Target and Distractors. However, the responses of those neurons to the Distractor were suppressed when Captors, that shared temporal coherence with the Distractor, were added to the Target and Distractor complex. Fishman et al. (2012) proposed that this selective suppression of neurons upon the presentation of temporally coherent Distractor and Captor facilitates RMR. Insights from auditory perception in frogs suggest the possibility that a mechanism similar to that proposed by Fishman et al. (2012) might facilitate RMR in frogs. Species recognition, frequently mediated by pulse rates in gray treefrogs, is driven by processing within the midbrain region, the Inferior Colliculus (IC). Small lesions of IC can impair the pulse-rate driven phonotaxis behavior in gray treefrogs (Endepols et al., 2003). In some frogs, a considerable proportion of neurons in the IC exhibit combination sensitive responses to the spectral components present in natural calls (Fuzessery and Feng, 1982; Lee et al., 2017; Mudry and Capranica, 1987a). Lee et al. (2017) demonstrated that a major proportion of combination sensitive neurons exhibit maximal firing in response to a single frequency but exhibit suppression when stimulated with a combination of frequencies. We hypothesize that stimulation by the bimodal (1.25 kHz and 2.5 kHz) Distractor and Captor complex induces a suppression in the neuronal responses of the combination sensitive neurons. This selective suppression facilitates the segregation of the unimodal (1.25 kHz or 2.5 kHz) Target from Distractor and Captor stream. Future neurophysiological studies could test this hypothesis by using Target, Distractor and Captor stimuli similar to the ones used in this study.

Conclusions

This dissertation aimed to identify perceptual mechanisms that enable non-human animals to recognize signals of interest in complex acoustic environments. The findings from this work reveal that treefrogs can exploit various spectral and temporal commonalities and differences between sounds to form coherent perceptual representations of communication signals.

Findings from Chapter 1 highlight that receivers may exploit temporal asynchrony and inharmonicity to segregate the concurrent spectral components of overlapping signals. Composite calls, comprising both conspecific and heterospecific spectral components, were perceived as two distinct calls when the conspecific and heterospecific components had asynchronous temporal onsets, inharmonic relationship, or both. These findings suggest the presence of parallel perceptual mechanisms in human and non-human animals to perceptually organize concurrent spectral components of communication signals.

Findings from Chapter 2 provide evidence against the hypothesis that any perceptually salient acoustic difference may promote the segregation of overlapping sound sequences. Interleaved pulsatile sequences, modelled on overlapping advertisement calls, were not perceptually segregated even though the differences in the rise times of their constituent pulses were behaviorally salient. Based on this result, I conclude that this hypothesis, based on the human psychoacoustic findings, does not explain perceptual organization in all non-human animals.

Findings from Chapter 3 demonstrated that spectrally similar, 'Distractor' sequences in the environment have the potential to perceptually mask the species-typical temporal rhythms and impair signal recognition. However, a release from masking of the signal is observed upon the additional presence of a sequence that shares temporal synchrony with the Distractor and thus perceptually captures the Distractor into an auditory stream, separate from the Target signal. These findings suggest the concurrent sound sequences are organized based on both spectral similarity as well as temporal synchrony. Additionally, these findings suggest that more concurrent sequences in an acoustic scene may not always elicit enhanced masking of a signal and instead may even improve signal recognition in certain circumstances.

Taken together with the research on perceptual organization in other animals like humans, primates, rodents, birds, fish and insects, this work on frogs highlight some remarkable similarities and differences in how animals across taxa organize complex acoustic mixtures. Since this work specifically incorporated insights from human perception to test hypotheses of auditory perception in non-human animals, it also provides a bridge between two disconnected fields of research, human psychoacoustics and sensory ecology. Making these connections are exciting when examined in the context of evolution of acoustic communication and sound reception organs in vertebrates. First, vocal communication has evolved multiple times independently in major tetrapod lineages; the earliest origins date as far back as 350 million years ago (Chen and Wiens, 2020). Second, the tympanic middle ear has also evolved multiple times after the diversification of major tetrapod lineages (Christensen-Dalsgaard and Carr, 2008). The findings presented in this dissertation suggest that in spite of the independent evolution of vocal communication and sound reception organs, animals have evolved similar perceptual machineries to solve the analogous problem of communicating in complex acoustic environments.

Further, by unraveling certain principles governing the perceptual machinery within receivers, this work is valuable in estimating how signal structures and signaling behaviors may evolve under the selective pressures of a receiver’s perception. Finally, there is a rapid emergence of the use of computational models like ‘agent-based modelling’ to ask fascinating new questions in animal behavior (DeAngelis and Diaz, 2019). There is also an increasing effort to design more realistic models to understand behavioral decision making. The insights from this work on the auditory perceptual machinery can be incorporated in future designs of computational models, especially the ones investigating mate choice.

Bibliography

Akre, K. L. and Ryan, M. J. (2010). Proximity-dependent Response to Variably Complex Mating Signals in Túngara Frogs (Physalaemus pustulosus). *Ethology* **116**, 1138–1145.

Anstis, S. and Saida, S. (1985). Adaptation to Auditory Streaming of Frequency-Modulated Tones. *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 257–271.

Bee, M. A. (2004). Equipment Review: Sound Ruler Acoustical Analysis: a free, open code, multi-platform sound analysis and graphing package. *Bioacoustics* **14**, 171–178.

Bee, M. A. (2010). Spectral preferences and the role of spatial coherence in simultaneous integration in gray treefrogs (Hyla chrysoscelis). *J. Comp. Psychol.* **124**, 412–424.

Bee, M. A. (2015). Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. *Int. J. Psychophysiol.* **95**, 216–237.

Bee, M. A. and 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.

Bee, M. A. and Riemersma, K. K. (2008). Does common spatial origin promote the auditory grouping of temporally separated signal elements in grey treefrogs? *Anim. Behav.* **76**, 831–843.

Bee, M. A. and Schwartz, J. J. (2009). Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise. *J. Acoust. Soc. Am.* **126**, 2788–2801.

Bee, M. A. and Vélez, A. (2018). Masking release in temporally fluctuating noise depends on comodulation and overall level in Cope's gray treefrog. *J. Acoust. Soc. Am.* **144**, 2354–2362.

Bee, M. A., Micheyl, C., Oxenham, A. J. and Klump, G. M. (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 Neuroethol. Sensory, Neural, Behav. Physiol.* **196**, 543–557.

Bee, M. A., Vélez, A. and Forester, J. D. (2012). Sound level discrimination by gray treefrogs in the presence and absence of chorus-shaped noise. *J. Acoust. Soc. Am.* **131**, 4188–4195.

Bizley, J. K. and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* **14**, 693–707.

Blair, W. F. (1958). Call Difference as an Isolation Mechanism in Florida Species of Hylid Frogs. *Florida Acad Sci* **21**, 32–48.

Booker, W. W., Gerhardt, H. C., Lemmon, A. R., Ptacek, M. B., Hassinger, A. T. B., Schul, J. and Lemmon, E. M. (2022). The Complex History of Genome Duplication and Hybridization in North American Gray Treefrogs. *Mol. Biol. Evol.* **39**, msab316.

Bosch, J. and Márquez, R. (2002). Female preference function related to precedence effect in an amphibian anuran (Alytes cisternasii): tests with non-overlapping calls. *Behav. Ecol.* **13**, 149–153.

Bregman, A. S. (1978). Auditory streaming is cumulative. *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 380–387.

Bregman, A. S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bregman, A. S. and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Exp. Psychol.* **89**, 244–249.

Bregman, A. S. and Pinker, S. (1978). Auditory streaming and the building of timbre. *Can. J. Psychol. Can. Psychol.* **32**, 19.

Bregman, A. S., Ahad, P., Kim, J. and 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.

Bregman, A. S., Ahad, P. A. and 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.

Broadbent, D. E. and Ladefoged, P. (1957). On the Fusion of Sounds Reaching Different Sense Organs. *J. Acoust. Soc. Am.* **29**, 708–710.

Brokx, J. P. L. and Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *J. Phon.* **10**, 23–36.

Brumm, H. and Slabbekoorn, H. (2005). Acoustic Communication in Noise. In *Advances in the Study of Behavior*, pp. 151–209. Academic Press.

Bush, S. L., Gerhardt, H. C. and 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.

Cai, H., Screven, L. A. and Dent, M. L. (2018). Behavioral measurements of auditory streaming and build-up by budgerigars (Melopsittacus undulatus) . *J. Acoust. Soc. Am.* **144**, 1508–1516.

Capranica, R. R., Llinas, R. and Precht, W. (1976). Auditory system: morphology and physiology. *Handb. Frog Neurobiol.*

Carl Gerhardt, H. (1978). Temperature coupling in the vocal communication system of the gray tree frog, Hyla versicolor. *Science (80-.).* **199**, 992–994.

Chen, Z. and Wiens, J. J. (2020). The origins of acoustic communication in vertebrates. *Nat. Commun.* **11**, 369.

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.* **25**, 975–979.

Christensen-Dalsgaard, J. and Carr, C. E. (2008). Evolution of a sensory novelty: tympanic ears and the associated neural processing. *Brain Res. Bull.* **75**, 365–370.

Christison-Lagay, K. L. and Cohen, Y. E. (2014). Behavioral correlates of auditory streaming in rhesus macaques. *Hear. Res.* **309**, 17–25.

Cui, J. (2007). QIC program and model selection in GEE analyses. *Stata J.* **7**, 209–220.

Dannenbring, G. L. and Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Percept. Psychophys.* **24**, 369–376.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Q. J. Exp. Psychol. Sect. A* **33**, 185–207.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *J. Acoust. Soc. Am.* **76**, 1636–1647.

Darwin, C. J. (1997). Auditory grouping. *Trends Cogn. Sci.* **1**, 327–333.

Darwin, C. J. (2008). Spatial Hearing and Perceiving Sources BT - Auditory Perception of Sound Sources. In (ed. Yost, W. A.), Popper, A. N.), and Fay, R. R.), pp. 215–232. Boston, MA: Springer US.

Darwin, C. J. and Ciocca, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *J. Acoust. Soc. Am.* **91**, 3381–3390.

Darwin, C. J. and Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *J. Acoust. Soc. Am.* **79**, 838–845.

Darwin, C. J. and Gardner, R. B. (1987). Perceptual separation of speech from concurrent sounds. In *The psychophysics of speech perception*, pp. 112–124. Springer.

Darwin, C. J. and Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *J. Acoust. Soc. Am.* **103**, 1080–1084.

Darwin, C. J. and Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Q. J. Exp. Psychol.* **36**, 193–208.

DeAngelis, D. L. and Diaz, S. G. (2019). Decision-making in agent-based modeling: A current review and future prospectus. *Front. Ecol. Evol.* **6**, 237.

Deike, S., Heil, P., Böckmann-Barthel, M. and Brechmann, A. (2012). The build-up of auditory stream segregation: A different perspective. *Front. Psychol.* **3**,.

Dent, M. L. and Bee, M. A. (2018). Principles of Auditory Object Formation by Nonhuman Animals.pp. 47–82. Springer, New York, NY.

Dent, M. L., Martin, A. K., Flaherty, M. M. and Neilans, E. G. (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.

Diekamp, B. and Gerhardt, H. C. (1995). Selective phonotaxis to advertisement calls in the gray treefrog Hyla versicolor: behavioral experiments and neurophysiological correlates. *J. Comp. Physiol. A* **177**, 173–190.

Doupe, A. J. and Konishi, M. (1991). Song-selective auditory circuits in the vocal control system of the zebra finch. *Proc. Natl. Acad. Sci.* **88**, 11339–11343.

Dowling, W. J. (2012). Music perception. *Oxford Handb. Audit. Sci. Hear.* **12**, 4473–4482.

Edwards, C. J., Alder, T. B. and Rose, G. J. (2002). Auditory midbrain neurons that count. *Nat. Neurosci.* **5**, 934–936.

Ehret, G. and Gerhardt, H. C. (1980). Auditory masking and effects of noise on responses of the green treefrog (Hyla cinerea) to synthetic mating calls. *J. Comp. Physiol. □ A* **141**, 13–18.

Ehret, G., Moffat, A. and Capranica, R. R. (1983). Two-tone suppression in auditory nerve fibers of the green treefrog (hyla cinerea). *J. Acoust. Soc. Am.* **73**, 2093–2095.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J. and Shamma, S. A. (2009). Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. *Neuron* **61**, 317–329.

Endepols, H., Feng, A. S., Gerhardt, H. C., Schul, J. and Walkowiak, W. (2003). Roles of the auditory midbrain and thalamus in selective phonotaxis in female gray treefrogs (Hyla versicolor). *Behav. Brain Res.* **145**, 63–77.

Farahbod, H., Rogalsky, C., Keator, L. M., Cai, J., Pillay, S. B., Turner, K., LaCroix, A., Fridriksson, J., Binder, J. R., Middlebrooks, J. C., et al. (2023). Informational Masking in Aging and Brain-lesioned Individuals. *J. Assoc. Res. Otolaryngol.* **24**, 67–79.

Farris, H. E., Rand, A. S. and Ryan, M. J. (2002). The effects of spatially separated call components on phonotaxis in túngara frogs: Evidence for auditory grouping. *Brain. Behav. Evol.* **60**, 181–188.

Farris, H. E., Rand, A. S. and Ryan, M. J. (2005a). The effects of time, space and spectrum on auditory grouping in túngara frogs. *J. Comp. Physiol. A. Neuroethol. Sens. Neural. Behav. Physiol.* **191**, 1173–1183.

Farris, H. E., Rand, A. S. and Ryan, M. J. (2005b). The effects of time, space and spectrum on auditory grouping in túngara frogs. *J. Comp. Physiol. A. Neuroethol. Sens. Neural. Behav. Physiol.* **191**, 1173–1183.

Fay, R. R. (1998). Auditory stream segregation in goldfish (Carassius auratus). *Hear. Res.* **120**, 69–76.

Fay, R. R. (2000). Spectral contrasts underlying auditory stream segregation in goldfish (Carassius auratus). *JARO - J. Assoc. Res. Otolaryngol.* **1**, 120–128.

Feng, A. S. and Schul, J. (2006). Sound Processing in Real-World Environments BT - Hearing and Sound Communication in Amphibians. In (ed. Narins, P. M.), Feng, A. S.), Fay, R. R.), and Popper, A. N.), pp. 323–350. New York, NY: Springer New York.

Feng, A. S., Hall, J. C. and Gooler, D. M. (1990). Neural basis of sound pattern recognition in anurans. *Prog. Neurobiol.* **34**, 313–329.

Fenton, B., Jensen, F. H., Kalko, E. K. V. and Tyack, P. L. (2014). Sonar Signals of Bats and Toothed Whales. In (ed. Surlykke, A.), Nachtigall, P. E.), Fay, R. R.), and Popper, A. N.), pp. 11–59. New York, NY: Springer New York.

Fishman, Y. I., Micheyl, C. and Steinschneider, M. (2012). Neural mechanisms of rhythmic masking release in monkey primary auditory cortex: Implications for models of auditory scene analysis. *J. Neurophysiol.* **107**, 2366–2382.

Fitzpatrick, D. C., Kanwal, J. S., Butman, J. A. and Suga, N. (1993). Combination-sensitive neurons in the primary auditory cortex of the mustached bat. *J. Neurosci.* **13**, 931–940.

Fuzessery, Z. M. and Feng, A. S. (1982). Frequency selectivity in the anuran auditory midbrain: single unit responses to single and multiple tone stimulation. *J. Comp. Physiol.* **146**, 471–484.

Fuzessery, Z. M. and Feng, A. S. (1983). Mating call selectivity in the thalamus and midbrain of the leopard frog (Rana p. pipiens): Single and multiunit analyses. *J. Comp. Physiol.* **150**, 333–344.

Gardner, R. B., Gaskill, S. A. and Darwin, C. J. (1989). Perceptual grouping of formants with static and dynamic differences in fundamental frequency. *J. Acoust. Soc. Am.* **85**, 1329–1337.

Gerhardt, H. C. (1974a). Behavioral Isolation of the Tree Frogs, Hyla cinerea and Hyla andersonii. *Am. Midl. Nat.* **91**, 424.

Gerhardt, H. C. (1974b). The significance of some spectral features in mating call recognition in the green treefrog (Hyla cinerea). *J. Exp. Biol.* **61**, 229–241.

Gerhardt, H. C. (1975). Sound pressure levels and radiation patterns of the vocalizations of some North American frogs and toads. *J. Comp. Physiol. ■ A* **102**, 1–12.

Gerhardt, H. C. (1976). Significance of two frequency bands in long distance vocal communication in the green treefrog. *Nature* **261**, 692–694.

Gerhardt, H. C. (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.

Gerhardt, H. C. (1982). Sound pattern recognition in some North American treefrogs (Anura: Hylidae): implications for mate choice. *Am. Zool.* **22**, 581–595.

Gerhardt, H. C. (1987). Evolutionary and neurobiological implications of selective phonotaxis in the green treefrog, Hyla cinerea. *Anim. Behav.* **35**, 1479–1489.

Gerhardt, H. C. (1991). Female mate choice in treefrogs: static and dynamic acoustic criteria. *Anim. Behav.* **42**, 615–635.

Gerhardt, H. C. (1995). Phonotaxis in Female Frogs and Toads: Execution and Design of Experiments. In *Methods in Comparative Psychoacoustics*, pp. 209–220.

Gerhardt, H. C. (2001). Acoustic communication in two groups of closely related treefrogs. In *Advances in the Study of Behavior*, pp. 99–167. Academic Press.

Gerhardt, H. C. (2005a). Acoustic spectral preferences in two cryptic species of grey treefrogs: Implications for mate choice and sensory mechanisms. *Anim. Behav.* **70**, 39–48.

Gerhardt, H. C. (2005b). Advertisement-call preferences in diploid-tetraploid treefrogs (Hyla chrysoscelis and Hyla versicolor): Implications for mate choice and the evolution of communication systems. *Evolution (N. Y).* **59**, 395–408.

Gerhardt, H. C. (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.

Gerhardt, H. C. and Bee, M. A. (2007). Recognition and localization of acoustic signals. In *Hearing and sound communication in amphibians*, pp. 113–146. Springer.

Gerhardt, H. C. and Doherty, J. A. (1988). Acoustic communication in the gray treefrog, Hyla versicolor: evolutionary and neurobiological implications. *J. Comp. Physiol. A* **162**, 261–278.

Gerhardt, H. C. and Höbel, G. (2005). Mid-frequency suppression in the green treefrog (Hyla cinerea): Mechanisms and implications for the evolution of acoustic communication. *J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol.* **191**, 707–714.

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

Gerhardt, H. C. and Schul, J. (1999). A quantitative analysis of behavioral selectivity for pulse rise-time in the gray treefrog, Hyla versicolor. *J. Comp. Physiol. - A Sensory, Neural, Behav. Physiol.* **185**, 33–40.

Gerhardt, H. C. and Schwartz, J. J. (2001). Auditory tuning and frequency preferences in anurans. *Anuran Commun.* 73–85.

Gerhardt, H. C., Allan, S. and Schwartz, J. J. (1990). Female green treefrogs (Hyla cinerea) do not selectively respond to signals with a harmonic structure in noise. *J. Comp. Physiol. A* **166**, 791–794.

Gerhardt, H. C., Dyson, M. L., Tanner, S. D. and Murphy, C. G. (1994). Female treefrogs do not avoid heterospecific calls as they approach conspecific calls: Implications for mechanisms of mate choice. *Anim. Behav.* **47**, 1323–1332.

Gerhardt, H. C., Martínez-Rivera, C. C., Schwartz, J. J., Marshall, V. T. and Murphy, C. G. (2007). Preferences based on spectral differences in acoustic signals in four species of treefrogs (Anura: Hylidae). *J. Exp. Biol.* **210**, 2990–2998.

Grace, J. A., Amin, N., Singh, N. C. and Theunissen, F. E. (2003). Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophysiol.* **89**, 472–487.

Grafe, T. U. (1999). A function of synchronous chorusing and a novel female preference shift in an anuran. *Proc. R. Soc. London. Ser. B Biol. Sci.* **266**, 2331–2336.

Greenfield, M. D. (1994). Cooperation and conflict in the evolution of signal interactions. *Annu. Rev. Ecol. Syst.* **25**, 97–126.

Greenfield, M. D. (2005). Mechanisms and Evolution of Communal Sexual Displays in Arthropods and Anurans. *Adv. Study Behav.* **35**, 1–62.

Greenfield, M. D., Tourtellot, M. K. and Snedden, W. A. (1997). Precedence effects and the evolution of chorusing. *Proc. R. Soc. London. Ser. B Biol. Sci.* **264**, 1355–1361.

Gridi-Papp, M. (2007). Sound ruler. *Acoust. Anal. Version 0.9* **6**,.

Griffiths, T. D. and Warren, J. D. (2004). What is an auditory object? *Nat. Rev. Neurosci.* **5**, 887–892.

Gupta, S. and Bee, M. A. (2020). Treefrogs exploit temporal coherence to form perceptual objects of communication signals. *Biol. Lett.* **16**, 20200573.

Gupta, S. and Bee, M. A. (2023). Female preferences for the spectral content of advertisement calls in Cope's gray treefrog (Hyla chrysoscelis). *J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol.* **209**, 31–45.

Gupta, S., Alluri, R. K., Rose, G. J. and Bee, M. A. (2021). Neural basis of acoustic species recognition in a cryptic species complex. *J. Exp. Biol.* **224**, jeb243405.

Gupta, S., Kalra, L., Rose, G. J. and Bee, M. A. (2023). Informational Masking Constrains Vocal Communication in Nonhuman Animals. *bioRxiv*.

Hall, M. L. (2009). A review of vocal duetting in birds. *Adv. Study Behav.* **40**, 67–121.

Hardin, J. W. and Hilbe, J. M. (2012). Generalized estimating equations. 2nd edn Boca Raton. *FL Chapman Hall/CRC*.

Hartmann, W. M. and Johnson, D. (1991). Stream Segregation and Peripheral Channeling. *Music Percept.* **9**, 155–183.

Hartmann, W. M., McAdams, S. and Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *J. Acoust. Soc. Am.* **88**, 1712–1724.

Hillery, C. M. and Narins, P. M. (1984). Neurophysiological evidence for a traveling wave in the amphibian inner ear. *Science (80-.).* **225**, 1037–1039.

Höbel, G. (2015). Sexual differences in responses to cross-species call interference in the green treefrog (Hyla cinerea). *Behav. Ecol. Sociobiol.* **69**, 695–705.

Höbel, G. and Gerhardt, H. C. (2003). Reproductive character displacement in the acoustic communication system of green tree frogs (Hyla cinerea). *Evolution (N. Y).* **57**, 894–904.

Höbel, G. and Gerhardt, H. C. (2007). Sources of selection on signal timing in a tree frog. *Ethology* **113**, 973–982.

Hoeschele, M. (2017). Animal pitch perception: Melodies and harmonies. *Comp. Cogn. Behav. Rev.* **12**, 5.

Hua, X., Fu, C., Li, J., de Oca, A. N. M. and Wiens, J. J. (2009). A revised phylogeny of Holarctic treefrogs (genus Hyla) based on nuclear and mitochondrial DNA sequences. *Herpetologica* **65**, 246–259.

Hukin, R. W. and Darwin, C. J. (1995). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Percept. Psychophys.* **57**, 191–196.

Hulse, S. H. (2002). Auditory scene analysis in animal communication. *Adv. Study Behav.* **31**, 163–200.

Hulse, S. H., MacDougall-Shackleton, S. A. and Wisniewski, A. B. (1997). Auditory Scene Analysis by Songbirds: Stream Segregation of Birdsong by European Starlings (Sturnus vulgaris). *J. Comp. Psychol.* **111**, 3–13.

Hyland Bruno, J. and Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behav. Processes* **163**, 53–59.

Itatani, N. and Klump, G. M. (2009). Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J. Neurophysiol.* **101**, 3212–3225.

Itatani, N. and Klump, G. M. (2014). Neural correlates of auditory streaming in an objective behavioral task. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10738–10743.

Itatani, N. and Klump, G. M. (2020). Interaction of spatial and non-spatial cues in auditory stream segregation in the European starling. *Eur. J. Neurosci.* **51**, 1191–1200.

Izumi, A. (2002). Auditory stream segregation in Japanese monkeys. *Cognition* **82**, B113–B122.

Jones, D. L., Jones, R. L. and Ratnam, R. (2014). Calling dynamics and call synchronization in a local group of unison bout callers. *J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol.* **200**, 93–107.

Kalra, L., Altman, S. and Bee, M. A. (2024). Perceptually salient differences in a species recognition cue do not promote auditory streaming in eastern grey treefrogs (Hyla versicolor). *J. Comp. Physiol. A.*

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., et al. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol. Rev. Camb. Philos. Soc.* **91**, 13–52.

Klinge, A. and Klump, G. M. (2009). Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans. *J. Acoust. Soc. Am.* **125**, 304–314.

Klinge, A. and Klump, G. (2010). Mistuning detection and onset asynchrony in harmonic complexes in Mongolian gerbils. *J. Acoust. Soc. Am.* **128**, 280–290.

Klump, G. M. and Gerhardt, H. C. (1992). Mechanisms and Function of Call-Timing in Male-Male Interactions in Frogs. In *Playback and Studies of Animal Communication* (ed. McGregor, P. K.), pp. 153–174. Boston, MA: Springer US.

Koehler, J., Jansen, M., Rodriguez, A., Kok, P. J. R., Toledo, L. F., Emmrich, M., Glaw, F., Haddad, C. F. B., Roedel, M.-O. and Vences, M. (2017). The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice. *Zootaxa* **4251**, 1–124.

Ladefoged, P. and Disner, S. F. (2012). *Vowels and consonants*. John Wiley & Sons.

Lee, N., Schrode, K. M. and Bee, M. A. (2017). Nonlinear processing of a multicomponent communication signal by combination-sensitive neurons in the anuran inferior colliculus. *J. Comp. Physiol. A* **203**, 749–772.

Lee, N., Christensen-Dalsgaard, J., White, L. A., Schrode, K. M. and Bee, M. A. (2021). Lung mediated auditory contrast enhancement improves the signal-to-noise ratio for communication in frogs. *Curr. Biol.* **31**, 1488–1498.

Legett, H. D., Aihara, I. and Bernal, X. E. (2020). Signal synchrony and alternation among neighbor males in a Japanese stream breeding treefrog, Buergeria japonica. *Curr. Herpetol.* **39**, 80–85.

Legett, H. D., Aihara, I. and Bernal, X. E. (2021). The dual benefits of synchronized mating signals in a Japanese treefrog: attracting mates and manipulating predators. *Philos. Trans. R. Soc. B* **376**, 20200340.

Lohr, B. and Dooling, R. J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (Taeniopygia guttata) and budgerigars (Melopsittacus undulatus). *J. Comp. Psychol.* **112**, 36.

Ma, L., Micheyl, C., Yin, P., Oxenham, A. J. and Shamma, S. A. (2010). Behavioral measures of auditory streaming in ferrets (Mustela putorius). *J. Comp. Psychol.* **124**, 317–330.

MacDougall-Shackleton, S. A., Hulse, S. H. and Ball, G. F. (1998). Neural bases of song preferences in female zebra finches (Taeniopygia guttata). *Neuroreport* **9**, 3047–3052.

Madsen, S. M. K. and Moore, B. C. J. (2014). Effects of compression and onset/offset asynchronies on the detection of one tone in the presence of anothera). *J. Acoust. Soc. Am.* **135**, 2902–2912.

McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* **19**, R1024–R1027.

McDermott, J. H. and Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Curr. Opin. Neurobiol.* **18**, 452–463.

McGregor, P. K. (2005). *Animal communication networks*. (ed. McGregor, P. K.) Cambridge University Press.

Mecham, J. S. (1965). Genetic relationships and reproductive isolation in southeastern frogs of the genera Pseudacris and Hyla. *Am. Midl. Nat.* 269–308.

Megela, A. L. (1983). Auditory response properties of the anuran thalamus: nonlinear facilitation. In *Advances in vertebrate neuroethology*, pp. 895–899. Springer.

Micheyl, C. and Oxenham, A. J. (2010). Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hear. Res.* **266**, 36–51.

Micheyl, C., Tian, B., Carlyon, R. P. and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* **48**, 139–148.

Micheyl, C., Hanson, C., Demany, L., Shamma, S. and Oxenham, A. J. (2013a). Auditory stream segregation for alternating and synchronous tones. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1568.

Micheyl, C., Kreft, H., Shamma, S. and Oxenham, A. J. (2013b). Temporal coherence versus harmonicity in auditory stream formation. *J. Acoust. Soc. Am.* **133**, EL188–EL194.

Middlebrooks, J. C. and Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. *J. Neurosci.* **33**, 10986–11001.

Miller, C. T. and Bee, M. A. (2012). Receiver psychology turns 20: Is it time for a broader approach? *Anim. Behav.* **83**, 331–343.

Minckley, R. L. and Greenfield, M. D. (1995). Psychoacoustics of female phonotaxis and the evolution of male signal interactions in Orthoptera. *Ethol. Ecol. Evol.* **7**, 235–243.

Moffat, A. J. M. and Capranica, R. R. (1974). Sensory Processing in the Peripheral Auditory System of Treefrogs (Hyla). *J. Acoust. Soc. Am.* **55**, 480.

Moore, B. C. J. and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. united with Acust.* **88**, 320–333.

Moore, B. C. J. and Gockel, H. E. (2012). Properties of auditory stream formation. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 919–931.

Moore, B. C. J., Peters, R. W. and Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *J. Acoust. Soc. Am.* **77**, 1861–1867.

Moore, B. C. J., Glasberg, B. R. and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* **80**, 479–483.

Mudry, K. M. and Capranica, R. R. (1987a). Correlation between auditory evoked responses in the thalamus and species-specific call characteristics: I. Rana catesbeiana (Anura: Ranidae). *J. Comp. Physiol. A* **160**, 477–489.

Mudry, K. M. and Capranica, R. R. (1987b). Correlation between auditory thalamic area evoked responses and species-specific call characteristics. II. H. Hyla cinerea (Anura: Hylidae). *J. Comp. Physiol. A.* **161**, 407–416.

Narins, P. M. (2003). *Acoustic Communication in Insects and Anurans. Common Problems and Diverse Solutions*. Chicago: University of Chicago Press.

Narins, P. M. and Capranica, R. R. (1978). Communicative significance of the two-note call of the treefrogEleutherodactylus coqui. *J. Comp. Physiol.* **127**, 1–9.

Neilans, E. G. and Dent, M. L. (2015). Temporal coherence for complex signals in budgerigars (Melopsittacus undulatus) and humans (Homo sapiens). *J. Comp. Psychol.* **129**, 174.

Nelken, I., Rotman, Y. and Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* **397**, 154–157.

Nityananda, V. and Bee, M. A. (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**,.

Noble, G. K. and Hassler, W. G. (1936). Three Salientia of Geographic Interest from Southern Maryland. *Copeia* **1936**, 63.

Noda, T., Kanzaki, R. and Takahashi, H. (2013). Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. *PLoS One* **8**, e83544.

Noorden, L. P. A. S. van (1975). Temporal Coherence in the Perception of Tone Sequences, Institute for Perception. *Inst. Percept. Res.* **Ph. D.**,.

Nowicki, S. and Searcy, W. A. (2004). Song function and the evolution of female preferences: why birds sing, why brains matter. *Ann. N. Y. Acad. Sci.* **1016**, 704–723.

Oldham, R. S. and Gerhardt, H. C. (1975). Behavioral Isolating Mechanisms of the Treefrogs Hyla cinerea and H. gratiosa. *Copeia* **1975**, 223.

Pettitt, B. A., Bourne, G. R. and Bee, M. A. (2012). Quantitative acoustic analysis of the vocal repertoire of the golden rocket frog (Anomaloglossus beebei). *J. Acoust. Soc. Am.* **131**, 4811–4820.

Platz, J. E. and Forester, D. C. (1988). Geographic Variation in Mating Call among the Four Subspecies of the Chorus Frog: Pseudacris triseriata (Wied). *Copeia* **1988**, 1062.

Pressnitzer, D., Sayles, M., Micheyl, C. and Winter, I. M. (2008). Perceptual Organization of Sound Begins in the Auditory Periphery. *Curr. Biol.* **18**, 1124–1128.

Prestwich, K. N. (1994). The energetics of acoustic signaling in anurans and insects. *Integr. Comp. Biol.* **34**, 625–643.

Ptacek, M. B., Gerhardt, H. C. and Sage, R. D. (1994). Speciation by Polyploidy in Treefrogs: Multiple Origins of the Tetraploid, Hyla versicolor. *Evolution (N. Y).* **48**, 898.

Reichert, M. S., Luttbeg, B. and Hobson, E. A. (2024). Collective signalling is shaped by feedbacks between signaller variation, receiver perception and acoustic environment in a simulated communication network. *Philos. Trans. B* **379**, 20230186.

Remez, R. E. (2021). Perceptual organization of speech. *Handb. Speech Percept.* 3–27.

Repp, B. H. (1988). Integration and segregation in speech perception. *Lang. Speech* **31**, 239–271.

Roberts, B. and Moore, B. C. J. (1991). The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony. *J. Acoust. Soc. Am.* **89**, 2922–2932.

Robillard, T., Höbel, G. and Carl Gerhardt, H. (2006). Evolution of advertisement signals in North American hylid frogs: vocalizations as end-products of calling behavior. *Cladistics* **22**, 533–545.

Römer, H. (2016). Matched filters in insect audition: tuning curves and beyond. *Ecol. Anim. Senses Matched Filters Econ. Sens.* 83–109.

Ryan, M. J. (1985). *The túngara frog: a study in sexual selection and communication*. University of Chicago Press.

Sach, A. J. and Bailey, P. J. (2004). Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Percept. Psychophys.* **66**, 1379–1387.

Schrode, K. M., Ward, J. L., Vélez, A. and Bee, M. A. (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.

Schrode, K. M., Buerkle, N. P., Brittan-Powell, E. F. and Bee, M. A. (2014). Auditory brainstem responses in Cope's gray treefrog (Hyla chrysoscelis): effects of frequency, level, sex and size. *J. Comp. Physiol. A* **200**, 221–238.

Schul, J. and Bush, S. L. (2002). Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. *Proc. R. Soc. B Biol. Sci.* **269**, 1847–1852.

Schul, J. and Sheridan, R. A. (2006). Auditory stream segregation in an insect. *Neuroscience* **138**, 1–4.

Schwartz, J. J. (1987). The function of call alternation in anuran amphibians: a test of three hypotheses. *Evolution (N. Y).* **41**, 461–471.

Schwartz, J. J. and Gerhardt, H. C. (1989). Spatially mediated release from auditory masking in an anuran amphibian. *J. Comp. Physiol. A* **166**, 37–41.

Schwartz, J. J. and Gerhardt, H. C. (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, J. J. and Marshall, V. T. (2006). Forms of call overlap and their impact on advertisement call attractiveness to females of the gray treefrog, hyla versicolor. *Bioacoustics* **16**, 39–56.

Schwartz, J. J. and Serratto Del Monte, M. E. (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.

Schwartz, J. J., Bee, M. A. and Tanner, S. D. (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, J. J., Buchanan, B. W. and Gerhardt, H. C. (2002). Acoustic interactions among male gray treefrogs, Hyla versicolor, in a chorus setting. *Behav. Ecol. Sociobiol.* **53**, 9–19.

Searcy, W. A. and Andersson, M. (1986). Sexual selection and the evolution of song. *Annu. Rev. Ecol. Syst.* 507–533.

Shamma, S. A., Elhilali, M. and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* **34**, 114–123.

Shamma, S., Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., Pressnitzer, D., Yin, P. and Xu, Y. (2013). Temporal coherence and the streaming of complex sounds. In *Advances in Experimental Medicine and Biology*, pp. 535–543. Springer.

Simmons, A. M. (1988). Selectivity for harmonic structure in complex sounds by the green treefrog (Hyla cinerea). *J. Comp. Physiol. A* **162**, 397–403.

Simmons, A. M. (2013). “To Ear is Human, to Frogive is Divine”: Bob Capranica’s legacy to auditory neuroethology. *J. Comp. Physiol. A* **199**, 169–182.

Simmons, A. M. and Bean, M. E. (2000). Perception of mistuned harmonics in complex sounds by the bullfrog (Rana catesbeiana). *J. Comp. Psychol.* **114**, 167.

Simmons, A. M., Buxbaum, R. C. and Mirin, M. P. (1993). Perception of complex sounds by the green treefrog, Hyla cinerea: envelope and fine-structure cues. *J. Comp. Physiol. A* **173**, 321–327.

Singh, P. G. and Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *J. Acoust. Soc. Am.* **102**, 1943–1952.

Singh, N. C. and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* **114**, 3394–3411.

Snedden, W. A. and D GREENFIELD, M. (1998). Females prefer leading males: relative call timing and sexual selection in katydid choruses. *Anim. Behav.* **56**, 1091–1098.

Snyder, J. S., Alain, C. and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* **18**, 1–13.

Stratman, K. D., Oldehoeft, E. A. and Höbel, G. (2021). Woe is the loner: Female treefrogs prefer clusters of displaying males over single "hotshot" males. *Evolution (N. Y).* **75**, 3026–3036.

Suga, N. (1992). Philosophy and Stimulus Design for Neuroethology of Complex-Sound Processing. *Philos. Trans. R. Soc. London. Ser. B. Biol. Sci.* **336**, 423–428.

Teki, S., Chait, M., Kumar, S., Shamma, S. and Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *Elife* **2013**,.

Theunissen, F. E. and Elie, J. E. (2014). Neural processing of natural sounds. *Nat. Rev. Neurosci.* **15**, 355–366.

Thompson, S. K., Carlyon, R. P. and 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.

Thorpe, W. H. (1975). The biological significance of duetting and antiphonal song. *Acta Neurobiol. Exp. (Wars).* **35**, 517–528.

Toledo, L. F., Martins, I. A., Bruschi, D. P., Passos, M. A., Alexandre, C. and Haddad, C. F. B. (2015). The anuran calling repertoire in the light of social context. *Acta Ethol.* **18**, 87–99.

Turgeon, M., Bregman, A. S. and Ahad, P. A. (2002a). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *J. Acoust. Soc. Am.* **111**, 1819–1831.

Turgeon, M., Bregman, A. S. and Ahad, P. A. (2002b). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *J. Acoust. Soc. Am.* **111**, 1819–1831.

Turgeon, M., Bregman, A. S. and Roberts, B. (2005). Rhythmic masking release: Effects of asynchrony, temporal overlap, harmonic relations, and source separation on cross-spectral grouping. *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 939–953.

Tuttle, M. D. and Ryan, M. J. (1982). The role of synchronized calling, ambient light, and ambient noise, in anti-bat-predator behavior of a treefrog. *Behav. Ecol. Sociobiol.* **11**, 125–131.

Vélez, A., Linehan-Skillings, B. J., Gu, Y., Sun, Y. and Bee, M. A. (2013). Pulse-number discrimination by Cope's gray treefrog (Hyla chrysoscelis) in modulated and unmodulated noise. *J. Acoust. Soc. Am.* **134**, 3079–3089.

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.

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.

Ward, J. L., Love, E. K., Vélez, A., Buerkle, N. P., O'Bryan, L. R. and Bee, M. A. (2013a). Multitasking males and multiplicative females: Dynamic signalling and receiver preferences in Cope's grey treefrog. *Anim. Behav.* **86**, 231–243.

Ward, J. L., Buerkle, N. P. and Bee, M. A. (2013b). Spatial release from masking improves sound pattern discrimination along a biologically relevant pulse-rate continuum in gray treefrogs. *Hear. Res.* **306**, 63–75.

Weber, T. and 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.

Wells, K. D. and Schwartz, J. J. (1984). Vocal communication in a neotropical treefrog, Hyla ebraccata: advertisement calls. *Anim. Behav.* **32**, 405–420.

Wells, K. D. and Schwartz, J. J. (2007). The behavioral ecology of anuran communication. In *Hearing and sound communication in amphibians*, pp. 44–86. Springer.

Winn, H. E., Thompson, T. J., Cummings, W. C., Hain, J., Hudnall, J., Hays, H. and Steiner, W. W. (1981). Song of the humpback whale - Population comparisons. *Behav. Ecol. Sociobiol.* **8**, 41–46.

Yao, J. D., Bremen, P. and Middlebrooks, J. C. (2015). Emergence of spatial stream segregation in the ascending auditory pathway. *J. Neurosci.* **35**, 16199–16212.

Zakon, H. H. and Wilczynski, W. (1988). The physiology of the anuran eighth nerve. *Evol. Amphib. Audit. Syst.* 125–155.

Appendix 1

Supplementary Material to Treefrogs exploit temporal onset synchrony and harmonicity in forming auditory objects of vocal communication signals

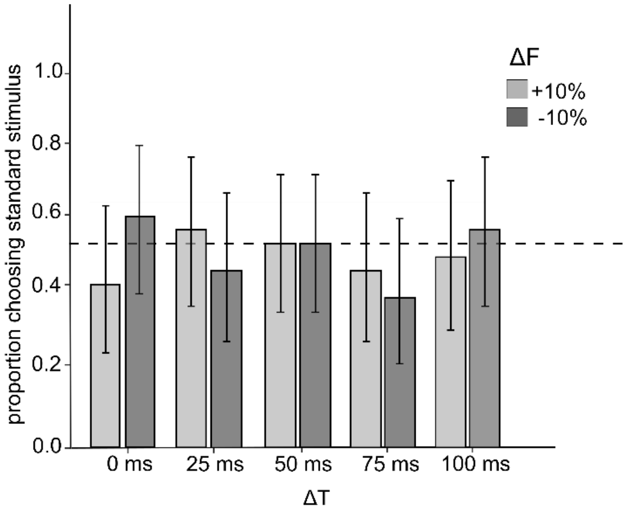

Figure A1-1. Proportion of subjects choosing the standard call across all combinations of temporal onset asynchrony (ΔT) compared between ΔF = +10% and ΔF = -10%. The dashed horizontal line indicates the null expectation of 0.50 in a two-alternative choice tests. Gray and black bars indicate ΔF = +10% and ΔF = -10%, respectively. Error bars represent 95% exact binomial confidence intervals.

Table A1-1 Results of the Generalized Estimating Equation (GEE) model. Here, only the data for which ΔF = ±10% was analyzed to compare ΔF = +10% and ΔF = -10%. The effects of temporal onset asynchrony (ΔT), inharmonicity (ΔF), and their interaction are reported in relation to the reference test where ΔT = 0 ms and ΔF = -10%. 'Condition' represents any manipulation in ΔT, ΔF or the interactive effects of different combinations of ΔT and ΔF. Interactions are shown by the '×' symbol. 'β' indicates the estimate value for each condition. 'Wald X^2' and 'P value' correspond to the results of the Wald-statistics.

Condition	**β**	**Wald X^2**	***P* value**
ΔT = 25 ms	-0.54	1.01	0.32
ΔT = 50 ms	-0.27	0.25	0.62
ΔT = 75 ms	-0.82	3.15	0.08
ΔT = 100 ms	-0.14	0.09	0.76
ΔF = +10%	-0.67	1.65	0.20
ΔT = 25 ms × ΔF = +10%	1.10	2.16	0.14
ΔT = 50 ms × ΔF = +10%	0.67	0.94	0.33
ΔT = 75 ms × ΔF = +10%	0.95	1.99	0.16
ΔT = 100 ms × ΔF = +10%	0.41	0.36	0.55

Appendix 2
Supplementary Material to Perceptually salient differences in a species recognition cue do not promote auditory streaming in eastern gray treefrogs (*Hyla versicolor*)

Methods

Protocols used for collecting subjects, creating acoustic stimuli, and testing choice behvaior are as described in the main text.

Experimental design

We conducted a four-alternative choice test to investigate the preferences of females when encountering a trade-off between two key signal recognition features, pulse shape and pulse rate. The first alternative was a 40 pulses/s sequence of A pulses (AAAA), which had the preferred pulse shape (typical of conspecific pulses) but a less preferred pulse rate (typical of heterospecific calls). The second alternative was a 20 pulses/s sequence of B pulses (B–B–), which had an attractive pulse rate (typical of conspecific calls) but a less preferred pulse shape (typical of heterospecific pulses). The third alternative was a 20 pulses/s sequence of B pulses (BB—), with both a less preferred shape (typical of heterospecific pulses) and pulse timing. The fourth alternative was a 40 pulses/s sequence of B pulses (BBBB), with both a less preferred shape and pulse rate (typical of heterospecific calls). The two key stimuli to investigate the trade-off between pulse shape and pulse rate were AAAA and B–B–. We expected subjects to significantly prefer AAAA if pulse shape is more important than pulse rate in choosing a mate. In contrast, we expected subjects to significantly prefer B–B– if pulse rate is more important than pulse shape in choosing a mate.

Results

At 100 dB, females significantly preferred the AAAA stimulus (n = 38, p = 0.001) and at 80 dB females significantly preferred B–B– stimulus (n = 33, $p < 0.001$) (Figure A2-1). Together, these results confirm that stimuli having either pulse shapes or rates typical of heterospecific calls are not only not aversive but elicit positive phonotaxis in the absence of alternatives possessing both conspecific pulse shapes and rates.

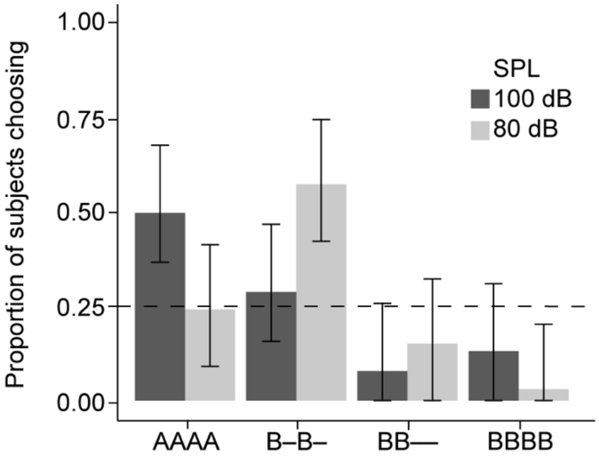

Figure A2-1. Results for the four-alternative choice test. Black and gray bars indicate the proportions of subjects choosing a given stimulus at 100 dB and 80 dB, respectively. Error bars depict 95% multinomial CIs. Horizontal dashed lines depict the chance probability for the experiment.

Appendix 3
Supplementary Material to Evidence of Rhythmic Masking Release in Cope's gray treefrog

Controls for testing behavioral neutrality of the Distractor and Captor

Experimental design

Three no-choice tests were performed on subjects (n = 34) in a random order to confirm that the Distractor, the Captor and a combination of both the stimuli were behaviorally neutral to the subjects. These three 'behavioral neutrality' tests constituted broadcasting continuous pulsatile sequences of 1.25 kHz, 2.5 kHz or both 1.25 kHz and 2.5 kHz, respectively. Additionally, on the same subjects, we conducted three identical positive control tests (described in section IIA) wherein subjects were expected to prefer a stimulus modelled on the conspecific advertisement signal.

For each of the six tests (three behavioral neutrality tests and three positive control tests), a subject's phonotaxis response as well as, wherever applicable, the first point of contact with the arena wall was recorded. Subsequently, the angle of a subject's first point of contact with the arena wall in relation to the active speaker was computed. All the behavioral neutrality test constituted broadcasting the stimuli from a single speaker which was assigned as the active speaker for calculating the points of contact. All the positive control tests were two-alternative choice tests and the speaker broadcasting the conspecific advertisement signal was assigned as the active speaker.

If a stimulus was behaviorally neutral, we expected no significant positive or negative phonotaxis from the subjects. Therefore, we expected the first points of contact with the arena wall to be randomly distributed around the circular arena. In contrast, if a stimulus was behaviorally attractive or repulsive, we expected positive or negative phonotaxis, respectively from the subjects. Consequently, we expected the points of contact to be clustered in proximity or far away from the active speaker, respectively.

Data analysis

The angular values for the three no-choice tests and three positive control tests were analyzed using circular statistics. A Rayleigh's test was done to test if the distribution of angles around the active speaker was uniform or clustered. Here, we first specified the position on the arena wall behind which the active speaker was placed as 0º. Then, based on its angular separation from the active speaker, we assigned a subject's 'first point of contact' to the arena wall into a 15 º bin between 0º-359º.

Results and discussion

The percentage of subjects responding in the behavioral neutrality tests was very low (0% for 1.25 kHz continuous pulsatile sequence, 8.8% for 2.5 kHz continuous pulsatile sequence, and 0% for both 1.25 kHz and 2.5 kHz sequences together). For many tests that resulted in a no-response, the subjects did not exit the release cage. The first point of contact with the arena wall could only be recorded in tests in which the subject exited the release cage and touched the arena wall. Hence, the sample size for circular statistics represented a subset of the actual frogs tested (the proportion of total tests that contributed to the angular data is described as 'k' values in Figure A3-1a-g).

The first points of contact with the arena wall were clustered around the active speaker in all the three positive control tests ($p < 0.01$ for all the three tests) (Figure A3-1a-c). The clustering signified that the subjects were attracted to the active speaker. In contrast, the clustering of the first points of contact with the arena wall was marginally significant (Figure A3-1d) in tests constituting the broadcast of 1.25 kHz continuous pulsatile sequence ($p = 0.06$). However, the clustering in this case was not in the direction of or opposite to the active speaker. Additionally, the first points of contact with the arena wall were not significantly clustered (Figure A3-1e-f) in tests constituting the broadcast of either the 2.5 kHz continuous pulsatile sequence ($p = 0.26$) and the 1.25 kHz + 2.5 kHz continuous pulsatile sequences ($p = 0.32$). The lack of a significant clustering of responses in a direction that was either toward or away from the speaker suggests that the Distractor and the Captor, when presented by themselves as well as in combination with each other, were behaviorally neutral to the subjects. These findings replicated the results of Nityananda and Bee, (2011) demonstrating that continuous pulsatile sequences lacking the temporal organization into short calls are behaviorally neutral.

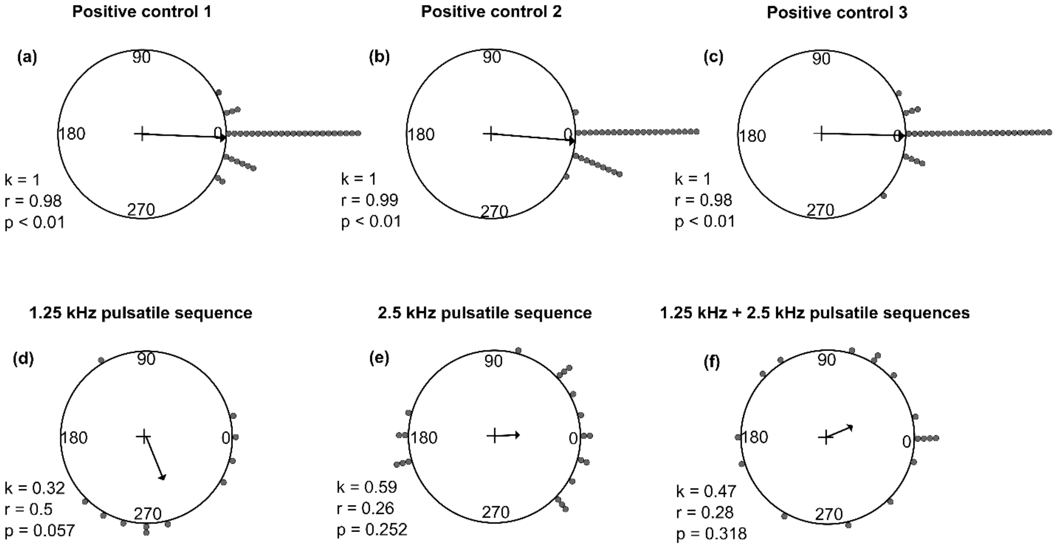

Figure A3-1. Distribution of subjects' first point of contact with arena wall for the a first positive control test, b second positive control test, c third positive control test, d 1.25 kHz pulsatile sequence, e 2.5 kHz pulsatile sequence and f 1.25 kHz + 2.5 kHz pulsatile sequences. Circles represent the arena and the active speaker is depicted at the "0" mark on the circumference. Individual first points of contact in relation to the active speaker are shown by black dots. The length of the arrowed lines depicts the vector ('r' values) and the direction of the arrowed lines depict the circular mean of the distribution. 'k' values in each panel depict the proportion of tests for which subjects left the release cage and made contact with the arena wall. The *p* values depict the p value from the Rayleigh test.

Model selection to evaluate the effects of Target frequency and SPL

Table A3-1. The quasi-likelihood values and the quasi-likelihood information criterion for all eight fitted models. Across models the predictor variables considered included the test type (i.e., the 'Target only', 'Target + Distractor', 'Target + Captor' and 'Target + Distractor + Captor' test), Target frequency (1.25 or 2.5 kHz) and Target SPL (67 dB or 73 dB). Interaction terms are indicated by the '×' sign between different independent variables.

Model	Quasi-likelihood	QIC
Test type + Target frequency + Target SPL + Test type × Target SPL +Test type:Target frequency + Target frequency × Target SPL	-336.72	701.97
Test type + Target frequency + Target SPL + Test type × Target SPL + Test type × Target frequency	-336.87	699.98
Test type + Target frequency + Target SPL + Test type × Target frequency + Target frequency × Target SPL	-337.96	698.13
Test type + Target frequency + Target SPL + Test type × Target SPL + Target frequency × Target SPL	-344.39	710.88
Test type + Target frequency + Target SPL + Test type × Target SPL	-344.45	708.68
Test type + Target frequency + Target SPL + Test type × Target frequency	-338.01	695.82
Test type + Target frequency + Target SPL + Target frequency × Target SPL	-345.53	706.83
Test type + Target frequency + Target SPL	-345.55	704.49

Table A3-2. Summary of the model with the lowest QIC (Test type + Target frequency +Target SPL + Test type × Target frequency). Because the chosen model encompassed an interaction between test type and Target frequency, the effect (averaged over all Target SPLs) of any combination of test type and Target frequency was reported in comparison to a reference condition (in this case, 'Target only' test at a Target frequency of 1.25 kHz). The Target frequency for a test condition is specified inside a parenthesis. Interaction terms are indicated by a '×' sign between specific test type and Target frequency. Also, because the chosen model did not incorporate any interaction between the Target SPL and the other two factors, it reported a main effect of the Target SPL (averaged across all test types and Target frequencies and reported in comparison to the reference Target SPL of 67 dB).

Reference	Condition	β	W	p
Target (1.25 kHz) only'	'Target (1.25 kHz) + Distractor'	4.35	81.78	<0.01
	'Target (1.25 kHz) + Captor'	0.67	2.82	0.09
	'Target (1.25 kHz) + Distractor + Captor'	2.90	69.94	<0.01
	'Target (2.5 kHz) only'	1.38	12.30	<0.01
	'Target (1.25 kHz) + Distractor' × 2.5 kHz	0.65	0.87	0.35
	'Target (1.25 kHz) + Captor' × 2.5 kHz	2.16	14.25	<0.01
	'Target (1.25 kHz) + Distractor + Captor' × 2.5 kHz	1.06	3.78	0.05
67 dB SPL	73 dB SPL	0.49	5.61	0.02