Vocal sacs do not act as visual cues in acoustically guided courtship in Cope’s gray treefrog (*Hyla chrysoscelis*)

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

Hongyu Li

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

Advised by Mark A. Bee

June 2020
Acknowledgements

I am grateful to all who have helped me in different forms during my graduate study. The Department of Ecology, Evolution and Behavior has provided financial support for my research and personal living, as well as various resources to complete my study. I particularly appreciate the guidance from my advisor Dr. Mark Bee. He is a brilliant scientist and I have gained a lot because of him. He also put much work into helping me grow and improving this thesis project. I also received help from my committee and other faculty members. I’d like to thank Dr. Marlene Zuk, Dr. Paloma Gonzalez-Bellido, Dr. Emilie Snell-Rood, Dr. Peggy Nelson, Dr. Trevor Wardill, and Dr. Peter Marchetto for providing advice or feedback on my research.

It was a pleasure for me to work with a group of smart and generous peers in the lab and the EEB community. I appreciate the scientific and emotional support from Saumya Gupta, Lata Kalra, Katie LaBarbera, Norman Lee, James Tumulty, and Jessie Tanner. In addition, this work would not be possible without the hard work done by undergraduate research assistants and volunteers, especially Chang Liu, Yi Fang Tan, Kelley Hoiseth, Becky Marshall, Annika Ruppert, Jordyn Massop, and Miranda Dahl. David Tinetti kindly fabricated the robotic system for this research project. Katrina Schrode did some pioneer work that laid foundation for this project.

Finally, I am deeply grateful for my families and friends, who always have my back.
Abstract

Components in multiple sensory modalities are present in many animal signals, which provides opportunities for receivers to use them as complementary cues in communication, especially in noisy environments that impose difficulty on signal perception. In frogs, it has been suspected that females use the visual byproduct of call production - the inflation of vocal sacs - as a cue in finding individual calling males in loud choruses. This mate recognition and selection behavior was traditionally considered as acoustically guided but recently there has been rising discussion on whether it was a multimodal process. We investigated whether female Cope’s gray treefrog (*Hyla chrysoscelis*) use visual cues in the context of sexual communication to find and select males. We performed playback experiments in a field setting under natural light using robotic frog models as visual stimuli and examined females’ responses. Acoustic stimuli were played back in quiet, in noise, and with ambiguous acoustic features. Despite the various acoustic conditions tested in a realistic lighting environment, we did not find any evidence that females use visual cues in the context of sexual communication. We review previous reports on the use of vocal sacs as visual cues in nocturnal anurans and discuss potential reasons for the stark contrast between those reports and this research.

**Keywords**: multimodal communication, *Hyla chrysoscelis*, vocal sac, gray treefrog, robotic model, multisensory, outdoor experiment, phonotaxis
# Table of Contents

Acknowledgements .......................................................................................................................... i
Abstract ........................................................................................................................................ ii
Table of Contents .......................................................................................................................... iii
List of Tables ................................................................................................................................... iv
List of Figures ................................................................................................................................. v
Introduction ....................................................................................................................................... 1
General materials and methods ....................................................................................................... 4
Experiment 1: Responses to unimodal (acoustic) stimuli ................................................................. 11
Experiment 2: Responses to multimodal stimuli in quiet ................................................................. 14
Experiment 3: Responses to multimodal stimuli with degraded acoustic information ............... 16
Experiment 4: Responses to multimodal stimuli in acoustically noisy environments .................. 19
General discussion ......................................................................................................................... 23
Illustrations ..................................................................................................................................... 29
Bibliography ..................................................................................................................................... 43
List of Tables

Table 1 Experiment 1: Statistical summary for analysis of influential factors for female choices in negative control.......................................................... 35

Table 2 Experiment 4: Statistical summary for analysis of influential factors for female responses not including the irradiance data.......................................................... 40

Table 3 Experiment 4: Statistical summary for analysis of influential factors for female responses including the irradiance as a potential factor ...................................................... 41
List of Figures

Figure 1 The frog models used as visual stimuli in this research .............................................. 29
Figure 2 The schematic and photos of experimental setup for phonotaxis tests ......................... 31
Figure 3 Experiment 1: Female preferences and response latency for unimodal stimuli ............ 33
Figure 4 Experiment 2: Female preferences and response latency for multimodal stimuli ........... 34
Figure 5 Experiment 3: The oscillograms of sinusoidally amplitude-modulated (SAM) calls..... 37
Figure 6 Experiment 3: The modulation depth thresholds that elicited phonotaxis and the response latency for stimuli with the three visual treatments ................................................................. 38
Figure 7 Experiment 4: Female responses to multimodal stimuli in noisy environment ............ 39
Introduction

It has been found in an increasing number of animal communication studies that animal signals contain multiple components that were previously unrecognized (Elias et al. 2005; Ota et al. 2015). When the components stimulate different sensory modalities in receivers, they are likely to bring additional advantages to receivers (Hebets and Papaj 2004). It has been proposed that multimodal signals can be better perceived by receivers due to the signal processing mechanisms that bias multisensory inputs (Rowe 1999), and indeed some psychophysiological evidence to support this idea has been found in humans and other mammals (Lovelace et al. 2003; Ghazanfar 2005; Stanford et al. 2005; Zahar et al. 2009; Bizley and King 2009; Stein 2012). Additionally, cues in different sensory modalities are faced with their particular environmental constraints on signal transmission or their particular sensory constraints on signal perception (Bro-Jørgensen 2010; Brumm 2013; Halfwerk et al. 2014b). Thus, multimodal cues may be complementary to each other in terms of reducing receiver uncertainty about signals (Munoz and Blumstein 2012; Gomes et al. 2016).

These advantages of multimodal cues would be pronounced when the communication environment is noisy for one or more of the modalities. In humans, this has been demonstrated in a daily phenomenon – lip-reading. In several studies, subjects’ perception of speech was improved when the speech was presented with videos of lip movement, and such improvement was especially significant when the environment is acoustically noisy (Ross et al. 2007; Zion Golumbic et al. 2013; Van Engen et al. 2019). Similarly, in non-human animals, it has been reported in a few studies that receivers rely more on multimodal cues when there is noise in the environment than when it is noise-free (Rhebergen et al. 2015; Gomes et al. 2016; Munoz and Blumstein 2020). As communication rarely happens in noise-free environments, whether and how animals take advantage of multimodal cues constitutes an important aspect for us to understand animals’ communication behaviors.

An interesting case of use of multimodal cues lies in frogs. Many frogs are thought to specialize in acoustic communication. However, when a male frog calls, there is another visually appealing piece. Males’ vocal sacs - a structure made of a skin membrane under mouths - inflate to serve as an impedance matching device when calling. The dynamically inflating (hereby referred as dynamic) vocal sacs almost certainly evolved for efficient sound transmission, but meanwhile they create the byproduct of a visual component obligately tied with calling. In a diurnal dart-poison frog species, using a speaker and a frog model, Narins et al. (2003, 2005) found that a male would
exhibit agonistic behaviors towards the model only when male calls and the vocal sac movement were both presented. This was the first indication that the vocal sac served as visual cues in anuran communication.

We focus on another category of behaviors that are traditionally thought of as being solely guided by acoustics for many anuran species – their courtship behavior. In nocturnally breeding frogs, males produce advertisement calls at night during breeding seasons, which attract females to approach the source of the sound - a behavior referred to as “phonotaxis”. Females recognize and select conspecific males by certain acoustic features in their calls. The frog sexual communication is one of the scenarios where the benefits of multimodal components are likely to contribute to receivers’ perception of signals. Females listen in a loud chorus, which potentially encourages recruiting help from visual cues. At the same time, there is growing acknowledgement of the exceptional nocturnal vision of frogs (King et al. 1993; Cummings et al. 2008; Yovanovich et al. 2017), suggesting that vision is an important sensory facility for nocturnal frogs. Altogether, our understanding of communication behaviors of frogs leads us to the hypothesis that visual cues are involved in acoustically guided courtship in addition to calls, and that a good candidate for the visual cue is the dynamic vocal sac of a calling male.

It has been reported in three anuran species so far that vocal sacs are used by females as visual cues in acoustically-guided courtship (Taylor et al. 2008; Gomez et al. 2010; Laird et al. 2016). However, previous work in these species has either used video playbacks or artificial light sources to illuminate a frog model and, therefore, suffer from the potential drawback of unrealistic visual displays. The small sample sizes in some research also hindered the robustness of the results. Besides, research in more species is needed to confirm whether it is a general phenomenon for females to use males’ vocal sacs as visual cues in frogs. Considering that vocal sacs are present in almost all anuran species and that there is growing interest on their roles in communication besides amplifying sound (Sztatecsny et al. 2010; Starnberger et al. 2014), this question is worth investigation to understand frog communication.

In the present study, we investigated whether female Cope’s gray treefrogs (Hyla chrysoscelis) use the visual presence of males, especially the movement of their vocal sacs, as cues in sexual communication at night in an outdoor setting using frog models. Although the acoustic communication of H. chrysoscelis has been extensively studied, there is no research on the role of vocal sacs as visual cues in this species so far. The outdoor setting of this study is noteworthy because it departs from most previous studies of frog vocal sacs as a visual cue, which have been conducted in the laboratory using video playbacks or robots presented under artificial lighting.
conditions (Rosenthal et al. 2004; Taylor et al. 2008, 2011a, 2011b, 2017; Gomez et al. 2009, 2010, 2011; Richardson et al. 2010; Taylor and Ryan 2013; Troïanowski et al. 2014; Laird et al. 2016; Stange et al. 2017). Because the spectrum and intensity of nocturnal light inevitably vary with changes in moon phase and cloud cover, we conducted each of our experiments over the span of at least half of a lunar cycle during the mating season. By doing so, we attempted to sample the full range of illumination female frogs would be expected to experience naturally when making their mate-choice decisions. Our outdoor setting allowed the frog model to appear visually realistic due to the natural illumination. Also, the natural environment provided in our study allowed the results to better reflect frogs’ behaviors in the field.

We examined female responses to sexual signals under various acoustic conditions with four separate experiments. In Experiment 1, we attempted to replicate results of female responses to unimodal acoustic signals established in the laboratory so that we could validate our experimental setup and protocol. In Experiment 2, we tested whether multimodal stimuli affect female responses to calling individuals in a quiet acoustic environment. In Experiments 3 and 4, we introduced unreliability to the acoustic signals by degrading the acoustic features of calls (Experiment 3) and creating an acoustically noisy environment (Experiment 4). We hypothesized that females would be more likely to rely on visual cues when signals were degraded and in a noisy acoustic environment if they attended to visual stimuli at all.
General materials and methods

Subjects

A total of 203 females of the western mtDNA lineage of *H. chrysoscelis* (Ptacek et al. 1994) was used as subjects in this study. Females were found in amplexus and collected at night (2100 to 0100 h) between mid-May and early-July in both 2018 and 2019 in ponds located in the Carver Park Reserve (Carver County, MN, USA) and the Tamarack Nature Center (Ramsey County, MN, USA). Females were kept with their mates in separate plastic containers at approximately 2 °C to delay egg laying until the time they were tested (within 72 hours after capture). Females and their mates were returned to their collection locations after testing within 5 days of capture. On the night of testing, we housed subjects in small plastic containers placed inside a Styrofoam box that blocked external light and sound. Subjects were placed in this box for at least 30 min prior to testing to allow their body temperature to reach ambient temperature and to ensure their eyes were dark-adapted. The nights we conducted behavioral tests coincided with nights when gray treefrogs were actively calling and breeding at our collection sites. Our collecting, handling, and testing processes followed the protocol approved by the University of Minnesota’s Institutional Animal Care and Use Committee (1701-34456A, approved March 3, 2017).

Acoustic stimuli

Acoustic stimuli used in all experiments were synthesized in MATLAB R2017a (MathWorks, Natick, MA, USA). With the exception of Experiment 3 (see below), each experiment involved broadcasting a synthetic “standard” advertisement call designed to simulate the acoustic features of an average *H. chrysoscelis* advertisement call, as described in Schrode et al. (2012). Each pulse in the standard call was composed of two phase-locked sinusoids with frequencies (and relative amplitudes) of 1.25 kHz (-11 dB) and 2.5 kHz (0 dB). Each call consisted of multiple pulses separated by silent inter-pulse intervals of equal duration, with a constant pulse duty cycle of 0.5 for all calls. Each individual pulse was shaped by modifying its amplitude envelope with a rise and fall times that were constant proportions (0.31 and 0.51, respectively) of the pulse duration. In natural calls, the pulse repetition rate (pulse per second, pps) varies linearly and directly with temperature (Ward et al. 2013). Therefore, we created separate versions of the standard call for use in tests conducted at different temperatures. Pulse rates were adjusted to different temperatures as described by Platz and Forester (1988). Separate versions of the standard call having temperature-corrected pulse rates, pulse durations, and inter-pulse intervals were made.
for 16°C (40.2 pps, 12.5 ms), 18°C (44.5 pps, 11 ms), 20°C (48.8 pps, 10 ms), 22°C (53.1 pps, 9.5 ms), and 24°C (57.4 pps, 8.5 ms). The number of pulses in each standard call was determined so that the overall call duration was always as close to 600 ms as possible. Each standard call had a linear rise time over the first five pulses of the call. Prior to conducting a playback experiment, we measured the air temperature and used the appropriate temperature-adjusted standard call as the stimulus. Standard calls were always played back at a repetition rate of 11.4 calls per min. Additional details about other acoustic stimuli are included in the relevant sections below.

We used Audacity® (Audacity Team 2019), Audition v3.0 (Adobe Inc., San Jose, CA, USA), or Goldwave (Goldwave Inc., St John’s, Newfoundland, Canada) to broadcast acoustic stimuli from an Acer E5-471 laptop (Acer Inc., New Taipei City, Taiwan) or Lenovo Yoga 720 laptop (Lenovo Group Limited, Beijing, China). Digit signals were sent to a NBA-200U USB External 7.1 Channel Audio Adapter (Vantec Thermal Technologies, Fremont, CA, USA) and transformed to mono analog sound inputs to amplified speakers (SRS-XB10 or SRS-XB12, Sony, Sony City, Minato, Tokyo, Japan). The sound pressure levels (SPL, re 20 µPa, slow, C-weighted) of acoustic stimuli were calibrated at least once each night prior to conducting behavioral tests using a sound level meter (Casella CEL-430/2, Enviro-Equipment Inc., Pineville, NC, USA) with the microphone tip placed 50 cm away from speakers. We attenuated the level of acoustic stimuli in software until the measured SPLs were within ±0.5 dB of the desired SPLs, with the exception of SPLs below 76 dB. Below 76 dB, the measurement of SPLs was often heavily influenced by wind and ambient noise, making calibration difficult. From our observations, adjustments to sound levels made in software were linear and yielded the specified dB change (±0.5 dB) in the sound output from speakers over ranges we could reliably measure with the sound level meter. Therefore, for SPLs below 76 dB, we adjusted SPLs to the desired level in software. The sound level meter was calibrated (Casella CEL-110/1) at least once each week.

Visual stimuli

Robotic frog models were constructed to serve as visual components of the stimuli (Fig. 1). The frog models were 3D printed with PCA plastic using an Original Prusa i3 MK3S 3D printer (Prusa Research, Prague, Czech Republic). The 3D model was created by modifying a pre-existing treefrog template (https://www.thingiverse.com/thing:182144) with Meshmixer (Autodesk Inc., San Rafael, California, USA) and Tinkercad (Autodesk Inc.) to mimic a male *H. chrysoscelis* in calling position (Fig. 1a). As part of our modification, the chin of the model was removed to create
space for the installation of an elastic vocal sac. A tunnel was created inside the model for passage of an air tube for inflating the vocal sac. The snout-to-vent length (SVL) of the frog models was set to 38.74 mm, based on the average measurement of 44 males (Ward et al. 2013). The final model file we used for our 3D printed gray treefrog is publicly available (https://www.thingiverse.com/thing:3831458).

The 3D frog models were colored to match real frogs. In 2010, we followed procedures described by Cummings et al. (2008) to measure the reflectance spectrum of the dorsal and ventral surfaces of 12 males and the inflated vocal sac of 38 males. We collected calling males from our collection sites at night and returned them to the lab, where we took measurements that same night. For measurements of real frogs, the animals were deeply anesthetized by submersion in buffered MS-222. Once anesthetized, we glued the mouth and nares shut with super-glue, then used a syringe to push air into the vocal sac until it was fully inflated. Reflectance measurements were taken from the center of the vocal sac and the center of the dorsal surface (i.e., the back) and ventral surface (i.e., the abdomen). All measurements of a single frog were completed in under 10 min. We kept frogs moist during measurements to facilitate cutaneous respiration, and we euthanized the frog with an overdose of MS-222 immediately after completing measurements.

We used an Ocean Insight S2000 Miniature Fiber Optic Spectrometer, a PX-2 Pulsed Xenon Lamp as the light source, a R200-7-UV/VIS reflection probe as the integrated optical fiber for light input and output, and the SpectraSuite software for data recording (Ocean Insight, Largo, Florida, USA). To measure reflectance, we pointed the probe against the surface being measured. A rubber ring (probe holder) surrounded the probe end, which sealed a space between the probe end and the surface. The probe holder blocked external light from entering the probe and ensured the probe was positioned perpendicular to the surface from a fixed distance for every measurement. We set the white reference for reflectance measurement by pointing the integrated probe against an Ocean Insight WS-1 reflectance standard with the PX-2 light source turned on and set the dark reference by pointing the probe against a dark surface with the PX-2 light source turned off. Based on our measurements of real males, we mixed acrylic paint (MyArtscape, Seattle, Washington, USA) and Dr. Ph. Martin’s ink (Salis International Inc., Oceanside, CA, USA) to obtain the approximate colors for different body parts. We painted frog models with brushes and colored vocal sacs by dipping them in the color mixture. The reflectance of painted 3D frog models with inflated vocal sacs were measured for comparison. As illustrated in Figure 1, the reflectance values of our frog models generally fell within ±1 SD of the mean reflectance measured from real frogs.
A robotic system was built to drive the inflation and deflation of the vocal sac (Fig. 2a). The vocal sac, which was crafted from a latex finger cot, was tied to one end of a 4 m long silicone tube that went through the tunnel in the 3D model and was connected to a syringe on the other end. A command signal was sent from the computer via the audio adapter to a Nano v3.0 board (Keywishbot, Shenzhen, China), providing analog input to control a programmable servo motor (HS-7940TH or HSG-8315BH, Hitec RCD USA, Poway, CA, USA), whose rotation was transmitted to linear movement of the syringe’s plunger by a rack-and-pinion or slider-crank mechanism. The motor rotated upon receipt of the command signal to push the plunger in to inject a previously calibrated amount of air into the inflating the vocal sac. The motor retracted the syringe to the original position when there was no input from the command signal. Code for robotic system operation was generated in Arduino (Arduino, Somerville, MA, USA). During tests, we aligned the timing of the command signal with the acoustic stimuli played back in audio software to synchronize the inflation of the vocal sac with calls. We could turn the robotic system on and off by playing or muting, respectively, the command signal.

In our experiments, we compared responses across three visual treatment conditions that included presenting a dynamic model (DM), a static model (SM), or no model (NM). For the DM conditions, the model’s vocal sac inflated and deflated synchronously with the onset and offset, respectively, of the acoustic stimulus with which it was paired. In the SM condition, the model’s vocal sac did not inflate and deflate with the sound, but instead remained partially inflated to mimic the natural level of inflation of male vocal sacs between calls. The NM conditions consisted of presenting acoustic stimuli alone. Hereafter, we note the visual stimuli as DM, SM, or NM, and we indicate the presence ("+"+) or absence ("-"-) of a paired acoustic stimulus using superscripts (e.g., DM+ or SM-).

Phonotaxis experiments

The same general protocol was used to conduct four different phonotaxis experiments, which are described in more detail in subsequent sections. All experiments were conducted in the field at night under natural nocturnal light, when the sun was more than 10° below the horizon. For testing, we selected a flat surface at one of our collection sites so that testing conditions would be as close as possible to the frogs’ natural habitats and to the environmental conditions under which the frogs normally breed. The selected site was sufficiently far (approximately 760 m) from breeding ponds where frogs called such that no chorus could be heard by human listeners at the
site, but other aspects of the environment (temperature, humidity, illumination) were as similar as possible to nearby ponds where breeding occurred. There was no canopy coverage to cause shadow at the site. Each experiment was conducted over the span of at least half of a lunar cycle during the breeding season.

Phonotaxis tests were conducted using a hexagonal arena with a side length of 0.58 m. The arena was constructed from a 1.27-cm-thick rubber mat used as flooring that was surrounded by 37-cm-high black mesh walls (Fig. 2). Adjacent to the test arena, an infrared (IR) sensitive camera (Sony HDR-XR520, IR light source Sony HVL-HIRI) was set up on a tripod so that videos of each test could be viewed remotely. A workstation was set up 3 m away from the test arena, where a tester remotely controlled the presentation of stimuli from the computer, watched subjects on a video monitor (Padarsey, China), and recorded the outcomes of each test in a field notebook. During testing, we attempted to minimize any movement, sound, or artificial light near the test arena. A 0.23 m³ light-blocking cube was constructed around the computer and video monitor by interlocking gym flooring tiles (0.61 m × 0.61 m, BalanceFrom, Los Angeles, CA, USA) to house all light-emitting electronics at the workstation. Between tests, subjects were given “time outs” lasting at least 5 min inside the Styrofoam box where subjects were housed. We used dim red light in a restricted way between tests for operational needs to guarantee frogs’ eyes remained dark adapted.

Individual subjects were tested in up to 13 tests but were never tested more than once in exactly the same test to avoid pseudoreplication. At the beginning of each test, a single subject was placed at one end of the test arena under an acoustically and visually transparent release cage fashioned from a transparent plastic cup with holes. Depending on the experimental design, subjects were tested in either no-choice tests or two-alternative choice tests. In a no-choice test, acoustic stimuli were played from a single speaker that was 1 m away from the release point. In tests where the visual stimulus was presented (i.e., the DM or SM conditions), a frog model was placed directly in front of the speaker. Acoustic signals were broadcast after the subject had acclimated inside the release cage for 1 min. After two calls were broadcast, the lid of the release cage, which was suspended from a 50 cm string, was lifted by a second tester who otherwise sat motionless next to the test arena during the test. The subject was allowed to move freely in the arena for up to 5 min. A response was scored, and the test was completed, if the subject entered a 10-cm wide response zone in front of the speaker, a response criterion similar to those used commonly in frog phonotaxis experiments conducted in laboratory sound chambers. Otherwise, the test was terminated if the subject did not enter the response zone after 5 min or remained at the release site for 3 min after
being released. For each test, we recorded whether the subject entered the response zone (test outcome) and the time required to enter the response zone (response latency). For a two-alternative choice test, the testing protocol was similar except that alternating acoustic stimuli were broadcast through two separate speakers that were 60° apart. The test outcome (i.e. the choice of the subject) was recorded as the first response zone it entered. Calls broadcast from the two speakers were interleaved in time, such that there were equal periods of silence preceding and following each broadcast call. To control for any potential side bias, we switched the physical position of the two speakers across nights; across tests conducted within a night, we randomly determined the side from which each stimulus was presented and the stimulus that was presented first. As reported below (Experiment 1a), no side bias was observed to occur using our setup and protocols.

Data analysis

All data analysis was conducted using R (R Core Team 2019). The alpha level for determining statistical significance was 0.05. General data analysis strategies are described below; additional details specific to each experiment are described later in corresponding sections. In each experiment, we aimed to test a minimum sample of \( n = 20 \) subjects per treatment condition, as this is a sample size used frequently in phonotaxis studies of multimodal perception in frogs (e.g., Gomez et al. 2010; Halfwerk et al. 2014a; Halfwerk et al. 2014b; Laird et al. 2016; Richardson et al. 2010; Rosenthal et al. 2004a; Taylor et al. 2007; Taylor et al. 2008; Taylor et al. 2011b; Taylor and Ryan 2013). In all but one test, our sample size exceeded this minimum sample size. Final sample sizes were usually determined based on factors such as the availability of gravid females, weather conditions, and breeding season duration.

For two-alternative choice tests (Experiment 1 & 2), we used two-tailed binominal tests to determine whether the proportions of subjects choosing each alternative deviated significantly from 0.5. For the two subsets of subjects that chose the two different alternatives, we compared the distributions of their response latency using a non-parametric method – the two-sample Anderson-Darling test – with R package twosamples. In an Anderson-Darling test, the null hypothesis is that two samples are drawn randomly from the same combined population and thus have the same distribution. Results from this test are significant when the difference between two samples’ cumulative distribution functions (represented by an AD-statistic value) is larger than that between two random subsamples of the combined population.
For no-choice tests (Experiments 3 & 4), we examined what factors significantly influenced test outcome and the response latency by comparing models with different explanatory terms. We fitted our data with linear regression models, or variations thereof, using the maximum likelihood method with R packages *lme4* (Bates et al. 2015) and *survival* (Therneau and Grambsch 2000; Therneau 2015). Different variations of linear regression models were adopted to deal with particular types of data. First, when there were categorical explanatory factors in models, generalized regression models were fitted in which each level of a certain factor was encoded as a dummy variable with two values (1 for being true and 0 for false). Second, for experiments where individual subjects were tested in multiple experimental conditions, mixed-effect models were used to partition the between-individual variation. Third, to analyze the variable test outcome, logistic regressions were fitted to account for the binary distributions (response versus no response). Forth, for the variable response latency, Weibull regressions were fitted to account for asymmetry in the distributions of response latency (Zhang 2016). Weibull regression is a robust method to describe the distribution of time it takes for an event to happen when the chance of the event at any moment is associated with the amount of time that has already passed. In our case, the tendency for females to respond to stimuli at any moment changed (in most cases increasing) as they sampled the stimuli for a longer time. After fitting the models, we ran likelihood ratio tests (Chi-square distribution) between the full model and a reduced model that excluded a certain factor to evaluate whether the factor increased the model’s goodness of fit, that is, whether the factor contributed to explaining the pattern of subjects’ responses. A p-value was obtained based on the deviance difference between the two models.
Experiment 1: Responses to unimodal (acoustic) stimuli

Our objective in Experiment 1 was to validate our general experimental approach for testing subject behavior at night in an outdoor test arena by assessing whether use of our setup and protocol would allow us to replicate the outcomes of phonotaxis tests conducted previously under more controlled conditions in laboratory sound chambers. To this end, we performed three separate phonotaxis experiments (1a-1c) that examined responses to unimodal acoustic stimuli in the absence of visual stimuli using two-alternative choice tests. All acoustic stimuli were presented at the sound pressure level (SPL) of 88 dB at 50 cm in this experiment to reflect the natural sound levels of calling males (Gerhardt 1975).

Methods

Experiment 1a served as a negative control in which we presented the same standard call as the two alternatives in a two-alternative choice test (i.e., NM* versus NM*). This experiment served to determine whether our outdoor setup created any directional response bias that could influence subjects’ behaviors as they exhibited phonotaxis in the absence of a visual component. Subjects ($n = 34$) were expected to choose the two identical alternatives in this experiment in equal proportions, which we evaluated using a two-tailed binomial test. In addition, we used separate generalized logistic regression models to examine the possibility that two aspects of our experimental setup and protocol – the physical speaker that broadcast each stimulus and the stimulus that was broadcast first – might have caused bias in female responses.

Experiments 1b and 1c served as positive controls in which we examined whether we could replicate known preferences of female gray treefrogs established in previous laboratory experiments. In Experiment 1b, we tested for reliable species discrimination. In a two-alternative choice test, we gave subjects ($n = 16$) a choice between the standard (conspecific) call and an alternating synthetic (heterospecific) call of the eastern gray treefrog, *Hyla versicolor*. As morphologically indistinguishable sibling species that form a cryptic diploid-tetraploid species complex (Ptacek et al. 1994), *H. chrysoscelis* (diploid) and *H. versicolor* (tetraploid) occur in sympathy and breed syntopically across much of their shared geographic range in North America including in central Minnesota where this study was conducted. Females of *H. chrysoscelis* rely on acoustic features of males’ calls, primarily pulse rate, to discriminate between the two species, and such acoustic discrimination is robust in laboratory experiments (Bush et al. 2002; Schul and Bush 2002; Gerhardt 2005; Bee and Riemersma 2008; Lee et al. 2017a). The synthetic *H. versicolor* call
was generated using the same procedures used to generate conspecific calls. The call consisted of two frequency components (1.2 kHz [-5dB] and 2.4 kHz [0 dB]), had a pulse rate of 25 pps, a 50% pulse duty cycle, a 20-ms pulse duration, and a 600-ms call duration (Gerhardt and Doherty 1988). The pulse rise and fall times were 0.67 and 0.33 of the pulse duration, respectively, and the call was shaped with a linear rise time of 100 ms.

In Experiment 1c, we tested for reliable intraspecific discrimination based on individual differences in the duration of conspecific calls. Previous laboratory studies have shown that females of *H. chrysoscelis* prefer males that produce longer calls with more pulses, all else being equal (Gerhardt et al. 1996; Bee 2008; Vélez et al. 2013; Ward et al. 2013; Tanner et al. 2017). In a two-alternative choice test, we gave subjects (*n* = 23) a choice between two alternating calls with pulse numbers that were +1 SD above and -1SD below the population mean pulse number of 30 pulses (i.e., 34 pulses and 26 pulses, respectively) (Ward et al. 2013). These two stimuli were created by modifying the 30-pulse standard call so that they had the appropriate number of consecutive pulses.

**Results and discussion**

In Experiment 1a (the negative control), the proportion of subjects choosing each of the identical standard (conspecific) calls (i.e., left versus right speaker) did not differ from 0.5 (Fig. 3; two-tailed binomial: *P* = 0.230). Regression models showed that neither the specific speaker used (*χ*^2^(1) = 2.124, *P* = 0.145) nor the order of stimulus presentation (*χ*^2^(1) = 0.504, *P* = 0.478) had a significant influence on the choice subjects made (Table 1). The response latency also did not differ significantly between choices of the two alternatives (Fig. 3; AD-test statistic = 0.199, *P* = 0.126).

Given a choice between conspecific and heterospecific calls in Experiment 1b, 100% of subjects chose the conspecific call (Fig. 3; two-tailed binomial: *P* < 0.001). A comparison of response latency was not possible for Experiment 1b because all 16 subjects chose the same stimulus (Fig. 3). Experiment 1b is the only experiment reported in this study with a sample size smaller than *n* = 20.

In Experiment 1c, the proportion of subjects choosing the longer conspecific call (34 pulses) over the shorter conspecific call (26 pulses) was significantly greater than 0.5 (Fig. 3; two-tailed binomial: *P* = 0.035). There was no significant difference in response latency between choices of the two stimuli (Fig. 3a; AD-test statistic = 0.145, *P* = 0.642).

Together, results from Experiment 1 established the validity of our experimental setup and protocol for conducting phonotaxis tests in the field under nocturnal illumination and confirmed
that our procedures could replicate known female preferences established in previous laboratory studies in response to unimodal acoustic stimuli.
**Experiment 2: Responses to multimodal stimuli in quiet**

In Experiment 2 our objective was to assess the extent to which the visual component in a multimodal stimulus influences female preference in quiet using two-alternative choice tests. In all choice tests, both alternatives comprised the same acoustic component consisting of the standard call (similar to Experiment 1a). Across choice tests, we varied the visual components that were being compared. To assess whether a visual component influenced female preferences, we gave subjects a choice between a unimodal acoustic stimulus with no model (NM\(^*\)) and a multimodal stimulus in which the visual component consisted of either the static model (SM\(^*\)) or the dynamic model (DM\(^*\)). We also gave subjects a choice between two multimodal stimuli consisting of the static model (SM\(^*\)) versus the dynamic model (DM\(^*\)) to assess whether dynamic movements of the vocal sac concurrent with sound influenced female preferences.

**Methods**

Three two-alternative choice tests (NM\(^*\) versus SM\(^*\), NM\(^*\) versus DM\(^*\), and DM\(^*\) versus SM\(^*\)) were performed with acoustic stimuli calibrated to a SPL of 88 dB at 0.5 m from the speakers. Previous research has suggested that SPL may impact female preferences for multimodal stimuli (Taylor et al. 2011b); therefore, we replicated the choice test of DM\(^*\) versus SM\(^*\) at the additional SPLs of 82 dB and 76 dB (at 0.5 m). The level of background noise that we recorded at our testing site at the times of testing ranged between 40 to 50 dB SPL. Thus, the minimum signal-to-noise ratio expected during a test was approximately +26 dB, which is well above masked behavioral response thresholds measured in previous laboratory studies of *H. chrysoscelis* (Bee and Schwartz 2009; Vélez and Bee 2011; Nityananda and Bee 2012; Lee et al. 2017b). Because background noise levels in a breeding chorus typically average about 70 dB SPL and range between 60 and 80 dB SPL (Tanner and Bee 2019), we consider the conditions for Experiment 2 to be relatively quiet.

**Results and discussion**

In this experiment, the proportions of subjects choosing any particular alternative in a choice test did not differ significantly from chance expectations (0.5). At 88 dB (Fig. 4a), 56.3% subjects chose SM\(^*\) over NM\(^*\) (\(n = 48\), two-tailed binomial: \(P = 0.471\)), 56.3% subjects chose DM\(^*\) over NM\(^*\) (\(n = 48\), two-tailed binomial: \(P = 0.471\)), and 47.3% subjects chose DM\(^*\) over SM\(^*\) (\(n = 74\), two-tailed binomial: \(P = 0.728\)). At lower SPLs (Fig. 4b), the proportions of subjects choosing
DM* over SM* were 39.3% at 82 dB (n = 28, two-tailed binomial: P = 0.345) and 40.6% at 76 dB (n = 32, two-tailed binomial: P = 0.377).

At 88 dB, the response latency did not differ between the alternatives for SM* vs NM* (Fig. 4a; AD-test statistic = 0.041, P = 0.646) or DM* vs NM* (Fig. 4a; AD-test statistic = 0.082, P = 0.291). In contrast, response latency differed significantly between choices of DM* vs SM* at 88 dB (Fig. 4a; AD-test statistic = 0.121, P = 0.019). The mean (48.4 s) and median (34.0 s) response latencies for choosing DM* were shorter than those for choosing SM* (75.1 s and 56.5 s). However, this pattern of shorter latencies for choices of DM* compared with SM* was not observed at lower sound playback levels of 82 dB (Fig. 4b; AD-test statistic = 0.277, P = 0.103) or 76 dB (Fig. 4b; AD-test statistic = 0.281, P = 0.054).

In summary, females did not prefer a multimodal stimulus (DM* or SM*) over a unimodal acoustic stimulus (NM*). In tests with a dynamic versus a static vocal sac, (DM* vs SM*), females also did not show a preference. Although the latency in response to DM* was shorter than that for SM* at 88 dB, the difference was not large (26.7 s on average), and the effect was not replicated at lower SPLs of 82 dB and 76 dB. In addition, if an inflating vocal sac had a robust effect in reducing response latencies, we also should have observed shorter latencies in response to the dynamic model in the test of DM* versus NM*, but no such difference was observed. Because female preferences were largely independent of the presence of visual components in multimodal stimuli under quiet conditions, we designed Experiments 3 and 4 to test the general hypothesis that visual cues have more pronounced effects on subject responses under less favorable acoustic conditions.

15
Experiment 3: Responses to multimodal stimuli with degraded acoustic information

In Experiment 3, the key acoustic feature used by females to recognize conspecific advertisement calls – the pulsatile structure – was degraded to simulate effects on signal structure caused by communicating in acoustically noisy environment (Ryan and Sullivan 1989; Kuczynski et al. 2010). Females of H. chrysoscelis recognize the calls of conspecific males based on pulse rate (Klump and Gerhardt 1987; Gerhardt 1991, 2005; Gerhardt et al. 1996; Bush et al. 2002; Schul and Bush 2002; Bee and Riemersma 2008; Ward et al. 2013; Tanner et al. 2017; Lee et al. 2017a). However, in natural conditions in a chorus, the pulsatile structure, which is essentially the temporal dynamic of sound amplitude within a call, is degraded by background noise. Noise “fills in” the otherwise silent intervals between pulses, causing pulses to be less distinct as amplitude peaks within a call (Ryan and Sullivan 1989; Kuczynski et al. 2010). Thus, females’ perception of calls is potentially impaired in noise because of the degraded pulsatile structure. Following Kuczynski et al. (2010), we simulated this effect of degraded acoustic information using sinusoidally amplitude-modulated (SAM) calls in which we manipulated the depth of modulation between successive peaks in the sinusoidal waveform of the acoustic stimulus. Using these artificially degraded acoustic signals, we measured call recognition thresholds in response to a unimodal acoustic stimulus (NM+) and multimodal stimuli in which the visual component consisted of either the static model (SM+) or dynamic model (DM+). Threshold was determined as the minimum modulation depth that elicited phonotaxis using an adaptive tracking procedure based on that developed by Bee and Schwartz (2009). By comparing recognition thresholds across the NM+, SM+, and DM+ conditions, we tested the hypothesis that receivers rely on the visual components of multimodal stimuli when acoustic information is degraded.

Methods

Females exhibit strong selection on the pulse rate but are lenient to other aspects of pulse structure such as the duration and shape of individual pulses (Gerhardt 2005). Therefore, we could use SAM calls, whose amplitude envelopes simulated the pulsatile structure in natural calls and whose depths of amplitude modulation could be altered to degrade the call’s pulsatile structure, as acoustic stimuli to elicit females’ phonotaxis behaviors (Fig. 5). To create a SAM call, we first created an unmodulated complex tone having the duration of a standard call and consisting of
frequencies (and relative amplitudes) of 1.25 kHz (-11 dB) and 2.5 kHz (0 dB). Then we imposed sinusoidal modulation on its amplitude envelope according to the following equation

\[ s_t = c_t \left( 1 - \frac{d}{2} \left( \cos(2\pi f_m t) + 1 \right) \right) \]

where \( s_t \) is the resulting SAM tone, \( c_t \) is the unmodulated complex tone, \( f_m \) is the rate of sinusoidal amplitude modulation (50 Hz), and \( d \) is the modulation depth. We chose \( f_m = 50 \) Hz to match the pulse rate of the standard call (50 pps). We defined \( d \) as the difference in amplitude between the peak and trough of the modulated envelope expressed as a proportion of the peak amplitude of the waveform. Hence, a value of \( d = 0.0 \) produced an unmodulated envelope and a value of \( d = 1.0 \) produced a fully modulated amplitude envelope that simulated the pulsatile structure of natural calls. SAM calls had the same rise time as that of standard calls and had the fall time of 20 ms (equivalent to one pulse period). SAM calls with modulation depths from 0.1 to 1 were created in 0.1 steps. Kuczynski et al. (2010) showed that SAM calls are effective in eliciting female phonotaxis.

We measured a recognition threshold – the minimum modulation depth that elicited phonotaxis – for each of the three visual treatment conditions (NM*, SM*, and DM*). According to our hypothesis, we predicted that when visual components were presented subjects would show phonotaxis responses to SAM calls with shallower amplitude modulation and thus have lower thresholds for signal recognition. In addition, we predicted that the response latency to stimuli with degraded acoustic structure would be shorter when visual components were present.

Each subject went through three rounds of adaptive tracking procedures to find the individual thresholds for the three visual treatment conditions (Bee and Schwartz 2009). The testing order of the three treatments was randomized for each individual. Each round of the adaptive tracking procedure started with a test where the modulation depth of the signals presented was 0.5. Depending on whether the subject being tested responded in the test or not, the modulation depth for the next test increased (if it did not respond) or decreased (if it responded) by a step (0.2 for the second test, 0.1 for all subsequent tests). The procedure continued until the subject behaved differently to two consecutive modulation depths (0.1 step), of which the subject would respond to the deeper one but not the shallower one. The threshold was then defined as the lowest modulation depth the subject responded to. The thresholds were usually found within three tests. The response latency for the lowest modulation depths the subject responded to for each visual treatment was also recorded. We then decided whether visual treatment affected the results with regression.
models. Since each subject was tested with its specific modulation depths, the comparisons of response latency across visual treatment conditions for the same modulation depths were not possible.

**Results and discussion**

The average recognition thresholds that elicited phonotaxis responses from females for NM+, SM+, and DM+ were 0.427, 0.419, and 0.427 respectively (Fig. 6; \( n = 29 \) to 30 per visual treatment). Visual treatment did not significantly affect the recognition thresholds (\( \chi^2(2) = 0.057, P = 0.972 \)) or response latency (Fig. 6; \( \chi^2(2.1) = 2.944, P = 0.253 \)). The thresholds measured in this experiment correspond with previous study. In Kuczynski et al. 2010, females’ response rate was similar to the chance level at the depth of \( d = 0.33 \) but increased significantly at \( d = 0.57 \). [Note that our measure of modulation depth \( d \) corresponds to theirs \( m \) in their equation 2) according to the equation \( m = dl/(2-d) \).

This experiment showed that, as in Experiment 2, neither the presence of a static multimodal stimulus nor one with a dynamic vocal sac measurably altered behavioral responses relative to those elicited by a unimodal stimulus. Hence, these data do not support the hypothesis that receivers rely on the visual components of multimodal stimuli when acoustic information is degraded.
Experiment 4: Responses to multimodal stimuli in acoustically noisy environments

In Experiment 4, we investigated the extent to which females use visual components in acoustically noisy environments using no-choice tests. We broadcast synthesized chorus noise along with calls and created various signal-to-noise ratios (SNRs) by varying the SPL at which we broadcast calls. Subjects’ responses were examined under different visual treatment conditions. In a study of lip-reading in humans, the visual presentation of lip movements improved speech recognition in noise, especially when the SNR was low (Ross et al. 2007). Therefore, we hypothesized that visual components can improve females’ ability to recognize the signal in acoustically noisy environments, and such improvement is more profound as the SNR decreases. According to our hypothesis, we expected that higher proportions of subjects would show phonotaxis to multimodal stimuli than unimodal stimuli and the response latency would be shorter, especially when the SNRs were low.

Methods

Synthetic chorus noise was played through the speaker in addition to standard calls in no-choice tests to create an acoustically noisy environment. The noise used was the same as the ones in Vélez and Bee (2011). Briefly, chorus samples of at least 1.5 min were recorded from 25 different choruses with a microphone during the breeding seasons between 2007 and 2010. The choruses were recorded from 5 cm above water or ground (where females usually assessed males), at least 4 m away from the closest calling males, when the choruses were at peak. The recordings’ frequency spectra were analyzed, and an experimental noise was generated by filtering white noise so that it had the long-term frequency spectrum of natural chorus noise. Five noise exemplars were generated and used in experiments to avoid potential artificial effects caused by any particular noise exemplar.

Calls were played at five different SPLs ranging from 64 dB to 88 dB in 6-dB steps. The SPL of the noise was held constant at 76 dB across all tests, which is within the amplitude range of the natural chorus (Swanson et al. 2007; Tanner and Bee 2019). The decibel difference between the SPLs of calls and noise defined five signal-to-noise ratios (SNRs), which ranged from -12 dB to +12 dB in 6-dB steps. Since we considered the noise as the acoustic background in which subjects were to be tested, we commenced broadcast of the noise after the first 30 sec of silence during the 1 min acclimation time at the beginning of each test. For each subject, we first performed a “reference test,” in which we broadcast the standard call at 88 dB in the absence of noise to verify
the female was responsive. Subjects that responded in this reference test were randomly assigned to one of the three visual treatments (NM, SM, or DM). Each subject went through five tests, each at one of the aforementioned SNRs, plus one additional test in which no call was presented (NM', SM', or DM'). The order of these six tests was randomized for each individual. In the NM' condition, the animal was released as usual, but no acoustic or visual stimulus was presented. For the treatments in which no calls were presented (SM' and DM'), visual stimuli were presented as in other tests, but the channel broadcasting the acoustic stimulus was muted in software. In the DM' treatment, the vocal sac of the model inflated and deflated at the same times it would have had a call been present as in the DM' treatment. A response was recorded if the subject entered the response zone even though no call was presented. For subjects that did not respond in the last test ($n = 57$), we performed an additional reference test to verify the subject maintained responsiveness through the end of all tests. Subjects that failed this reference test (5 of 57) were excluded from the dataset.

Additionally, during this experiment we measured the irradiance of the testing environment to examine whether the effect of visual components was influenced by the level of ambient light. The measurement of irradiance was possible due to the availability of light measuring instrument in the year this experiment was conducted. The irradiance was measured every five minutes by pointing a photomultiplier SPM068 (International Light Technologies, Peabody, Massachusetts, USA) equipped with a wide-angle diffuser (W series, International Light Technologies) vertically to the sky and recording the measurements with the light meter ILT5000 (International Light Technologies). Irradiance was analyzed on a log scale. The log irradiance for each test of a subject was recorded as the reading during the test or the average log irradiance of the two flanking readings (within 5 mins of the test) if no measurement was available during a specific test.

We first examined whether the visual components by themselves elicited responses from subjects in the absence of acoustic signals. For this purpose, we isolated the data from the tests in which noise was presented as the only acoustic component (i.e., the NM', SM', and DM' conditions) and examined the effect of visual treatment on response proportions. Next, with the data for tests in which calls were present (i.e., NM*, SM*, and DM*), we analyzed how SNR, log irradiance, visual treatment, and the interaction between visual treatment and each of the other two factors affected response proportions and response latency.

**Results and discussion**
When no call was broadcast in the presence of noise as the acoustic background, the overall proportion of subjects that were scored as exhibiting a response was low (approximately 0.21; Fig. 7). This response rate was similar to the 0.20 “false alarm rate” estimated in Vélez and Bee 2010, meaning a small proportion of subjects entered the response zone not as a response to stimuli but by chance. In addition, the proportions of subjects that responded were similar across the NM, SM, and DM conditions (Fig. 7b; n = 27 to 29 per visual treatment, \( \chi^2(2) = 1.365, P = 0.505 \)). Assuming the response proportion for the NM treatment (0.26) represents the baseline response probability when subjects were not stimulated by an acoustic signal or corresponding visual component, this result confirms that, in an acoustically noisy environment, the visual components were not able to elicit female responses in the absence of the acoustic component.

When acoustic signals were present (n = 24 to 29 per visual treatment at each SNR, 397 tests in total), the comparison between the null model and the model including SNR as a factor showed the response proportions increased significantly as SNR increased regardless of visual treatment (Fig. 7a; \( \chi^2(1) = 16.659, P < 0.001 \)). Adding visual treatment to the model did not yield a better fit (\( \chi^2(2) = 1.364, P = 0.506 \)) nor did adding visual treatment along with their interactions with SNR (\( \chi^2(4) = 2.286, P = 0.683 \)). Together, these analyses indicate that response proportions were only influenced by SNR and not by visual treatment (Fig. 7b; see Table 2 for detailed model comparisons). For the tests in which subjects responded (135 tests in total), response latency was not significantly affected by either SNR (\( \chi^2(0.9) = -0.529, P = N/A \)) or visual treatment (\( \chi^2(-1.8) = -0.052, P = 0.964 \); Fig. 7b). Moreover, the log likelihood of the model slightly decreased (indicated by the negative \( \chi^2 \) values) when either factor was included.

The irradiance data was available for a total of 199 tests in which acoustic signals were presented. Irradiance values ranged between 8.81×10^{-12} W/cm^2 and 8.80×10^{-10} W/cm^2. Building from the model in which SNR was included as an explanatory factor, log irradiance did not significantly explain response proportions (\( \chi^2(1) = 2.948, P = 0.086 \)), nor did the visual treatment (\( \chi^2(2) = 0.851, P = 0.654 \)) or the interaction between them (\( \chi^2(5) = 9.444, P = 0.093 \); Table 3). For the tests where responses were scored (75 tests), response latency was not significantly influenced by log irradiance (\( \chi^2(-0.91) = -0.289, P = 0.550 \)), SNR (\( \chi^2(1.6) = 1.249, P = 0.435 \)), or visual treatment (\( \chi^2(-2.1) = -0.283, P = 0.888 \)). While the degree of variation in our sample of measured light levels was consistent with previous measurements reported by Cummings et al. (2008) and Taylor et al. (2008), our measured levels were smaller by one to two orders of magnitude. We note that irradiance measured in power units (e.g., W/cm^2) represents the cumulative energy of
ambient light across all wavelengths within the sensitivity range of the device. Thus, comparisons of absolute irradiance measurements across different studies may not be appropriate because the values greatly depend on the response curve of the light meter and the spectrum of the light. A solution to this problem would be to use light meters with response curves that are tuned to the spectrum sensitivity of the human eye and to obtain illuminance values measured in photometric units (e.g., lux). This method would yield comparable values across studies, though it would not take into account any species differences in how light levels are perceived. Notably, the units used (W/cm² versus lux) should not change the outcome of our analyses, as we were primarily interested in assessing the potential impacts of variation in irradiance across tests in our study.

In summary, higher proportions of females responded at higher SNRs, but neither the proportion of females responding nor response latency depended on whether a visual stimulus was present, whether the visual stimulus was static or dynamic, or the variation in light level.
General discussion

We examined the role of visual presence of males, especially their dynamic vocal sacs, as visual cues in acoustically guided courtship in *H. chrysoscelis*. The frog models were not found to be used by females as visual cues to recognize or select conspecific males in our experiments, regardless of whether there was movement of the vocal sacs or not.

Research comparisons and methodological analysis

In contrast to what was found in our study, it has been reported in other anuran species that vocal sacs influenced female responses to sexual signals. When comparing these studies, it should be noted that methodological differences exist. Early studies on visual communication in mating behaviors used video playbacks to present visual stimuli (Rosenthal et al. 2004; Gomez et al. 2009). However, video playbacks have inherent drawbacks that cannot be eliminated, such as an inability to accurately reproduce color, a lack of depth cues, and the creation of abnormal contrast between the monitor screen and ambient environment (Fleishman and Endler 2000; Zeil 2000). In one study of the eastern gray treefrogs (*Hyla versicolor*), Reichert et al. (2014) showed how the glowing screen itself was an effective stimulus for females that generated an artificial effect that could have been mistakenly attributed to the dynamic vocal sacs. It has been reported in European tree frogs (*Hyla arborea*) that females exhibit preferences on vocal sac coloration, but such preferences became less prominent if videos were replaced by frog models as visual stimuli (Gomez et al. 2009, 2010).

An improvement to video playback is to use frog models for visual stimuli and conduct lab experiments with artificial light source for illumination. Research on the role of dynamic vocal sacs as visual cues has been conducted using this approach in the túngara frog (*Physalaemus pustulosus*) and American green treefrog (*Hyla cinerea*). In túngara frogs, for example, mixed results have been obtained in two-alternative choice tests conducted by the same research group: females have been shown to have a significant preference for multimodal audiovisual stimuli over unimodal acoustic stimuli (Taylor et al. 2008; Stange et al. 2017), a potential preference for multimodal stimuli (Taylor and Ryan 2013; detailed data not provided), and no preference for multimodal stimuli (Taylor et al. 2011b; Cronin et al. 2019). When the dynamic model was paired with a less preferred call, the dynamic model was unable to reverse females’ preference based on acoustic properties of the calls (Stange et al. 2017). Efforts have also been made to test whether females reject multimodal stimuli with asynchronous auditory and visual components. They did so when the inflation of the
vocal sac lagged behind the call for 50% or 100% of the call duration (Taylor et al. 2011b) but not when the lag was shorter or when the inflation preceded the call (Taylor et al. 2011b, 2017). In summary, the effects of a dynamic vocal sac as a visual cue have been found to be inconsistent across studies and test conditions (including the SPL of call playback, the lighting condition, and the experimental design) in túngara frogs. In American green treefrogs, females significantly preferred a multimodal stimulus over a unimodal acoustic alternative when identical calls were presented for both alternatives (Laird et al. 2016). However, this preference was relatively weak: when the calls for one of the alternatives were lower in frequency – a feature preferred by females – the lower-frequency calls were preferred regardless of the presence of the dynamic model (Laird et al. 2016). The conflicting results across these studies might reflect certain rules for cross-modal interactions. However, caution should be used when interpreting the results for two reasons. First, explanations proposed for some results exclusively apply to the very specific conditions in which the animals were tested. Second, interpretations based only on the statistical significance of results may vary across studies due to small sample sizes. For example, most two-alternative choice tests investigating multimodal communication in túngara frogs and green treefrogs with dynamic models have used sample sizes of 20 animals per test. In a two-tailed binomial test with a sample size of 20 and an expected null proportion of 0.5 choosing each alternative, a 14:6 (0.70) outcome is not statistically significant (P = 0.1153) but an outcome of 15:5 (0.75) is significant (P = 0.0414). Yet two independent proportions of 0.70 and 0.75 (n = 20 each) are not significantly different from each other (z = 0.354; P = 0.7233). Thus, some explanations may be interpretations of spurious results instead of implying general rules of multimodal communication. Overall, vocal sacs were found to affect female preferences, but the effect was not consistent or robust.

Different from the work mentioned above, we conducted our experiments outdoors under natural illumination, which potentially contributed to differences between our results and those reported in previous laboratory studies. Realistic presentations of visual models depend critically on the spectrum of ambient light. By testing animals outdoors at night, our frog models were illuminated by natural ambient light spectra that represent similar conditions of nocturnal illumination under which the frogs make mate choice decisions in nature. To the best of our knowledge, no previous laboratory experiment has successfully replicated the ambient spectrum of natural light found in the research subjects’ natural habitat. Most previous studies of multimodal signaling in anurans have used LED or incandescent lights with no explicit attempt to match the spectrum of natural light when lighting was needed. In one of the only studies using artificial light to report its spectrum (Cummings et al. 2008), the match between the spectrum of natural light and
the artificial light source was not particularly good at short wavelengths. In addition, many laboratory experiments are conducted in small enclosed spaces, such as sound chambers, which can cause unnatural reflections from light-colored floors and walls. Thus, in laboratory experiments, even if the frog models are color painted to match the reflectance of real frogs, the color females perceive may be quite different from how males appear in nature under nocturnal illumination. We suggest all future laboratory studies of audiovisual signals should report the spectrum of artificial light sources used and justify their use relative to the variability of ambient light spectra recorded in the animal’s natural habitat.

In addition, females may behave differently in laboratory and in the field. Although direct comparisons of females’ strategy of mate selection in the field and in the laboratory is lacking, there is evidence that females’ behavioral strategies to select males can change when the environmental conditions vary. For example, mate selection in túngara frogs is affected by factors such as predation risk and ambient light level (Baugh and Ryan 2010; Bonachea and Ryan 2011a, b, c; Rand et al. 1997). Particularly, whether females prefer multimodal stimuli can be altered by varying illumination levels (Cronin et al. 2019), making it critically important to provide appropriate testing conditions in studies involving vision. The influence of environmental conditions on mate selection is species specific in anurans, as the light level does not affect female choices in the eastern gray treefrogs (Underhill and Höbel 2017). Our experiments eliminated the potential issues caused by artificial light sources and laboratory experiments. Though our experiments were subject to naturally fluctuating environmental conditions across nights, the variation of natural light within and between nights did not influence females’ responses in Experiment 4. While laboratory and field experiments have their own advantages, we consider results from field experiments to be more ecologically relevant in this case, especially when the understanding on how well laboratory experiments reflect what happens in nature is still lacking. Besides our study, attempts have been made to conduct research outdoor with natural illumination in squirrel treefrogs (Hyla squirella) (Taylor et al. 2007, 2011a). Female squirrel treefrogs preferred calls presented with dynamic models over unimodal calls but the preference was most likely due to the presence of lateral body stripes on the model instead of the vocal sac. Females were shown to discriminate between stripes of different sizes but do not discriminate between dynamic and static models. These results supported involvement of visual cues in courtship in anurans but cast doubt on dynamic vocal sacs as being the effective visual cues.

In summary of current knowledge on the role of vocal sacs as visual cues in acoustically guided courtship behaviors in nocturnal anurans, mixed results were found. Note, in none of the
studied species, results were replicated by more than one research team or with different methodologies, making it hard to conclude whether different results are due to species differences or methodology differences. It is unknown how common vocal sacs are used as visual cues in acoustically guided courtship in anurans. Our research is the first report of vocal sacs not serving as visual cues. However, we suspect that the null effect of vocal sacs is under-represented in the published literature.

**Biological explanations**

Considering the biological explanations for the null effect of visual cues in our study, the most concerning questions are whether the nocturnal vision of frogs is good enough to see vocal sacs at night and whether female frogs are visually attentive at night. It is well known in multiple anuran taxa that frogs have extraordinary nocturnal vision. Anurans have a dual-rod system that allows for color vision through the antagonistic interaction of two types of rods whose photopigments have maximal absorbance around 433 nm and 502 nm (Kojima et al. 2017; Mohun and Davies 2019). High visual sensitivity and color sensitivity is found in both visually guided predation and courtship behaviors (Cummings et al. 2008; Yovanovich et al. 2017). Though the visual sensitivity of *H. chrysoscelis* is unknown, it was shown in a congener, *H. cinerea*, through microspectrometry and optomotor experiments, that their visual sensitivity was high enough to detect objects of the frogs’ brightness below the lowest nocturnal light, and it is also likely that females are able to discriminate colors under dim light (King et al. 1993; Veilleux and Cummings 2012). Though it is not known whether females attend to the visual presentation of males, research in *H. versicolor* suggested that vision is somewhat involved in mate selection. Presenting calls with LED flashes influenced female preferences for calls (Reichert and Höbel 2015; Reichert et al. 2016). Together, these studies suggest that the possibility exists that nocturnal frogs use vision in the context of sexual communication.

Our results reinforce the theory that additional signal components are not inherently valuable for receivers, even in cases like ours where signal components are readily available in multiple sensory modalities and receivers have the sensory facilities to take advantage of them. Economic models state that the value of a signal component depends on how much it influences receivers’ estimation of the situation (Ernst and Banks 2002; Wilson et al. 2013; Rubi and Stephens 2016), or their estimation of payoffs if benefits and costs of various responses are factored in (Munoz and Blumstein 2020). In our research, the most likely conditions for females to use visual
cues were provided by reducing the reliability (defined as how well a signal indicates a certain situation by Rubi and Stephens (2016) – in our case the presence of conspecific males) of acoustic signals with degraded pulsatile structures and background noise. One economic explanation for results in this study might be that the reliability of vocal sacs as visual cues is lowered by the fact that they can also suggest the presence of the morphologically similar sympatric species *H. versicolor*. Perhaps the lower reliability renders vocal sacs a less valuable visual cues for females to refine their estimations and females evolutionarily develop the strategy to only use acoustic signals to recognize males.

**The role of vocal sacs in frog communication**

Despite scattered reports of vocal sacs serving as visual cues, we are still far from understanding the implications of these examples. For example, when preferences for calls presented with dynamic models over unimodal acoustic signals were found, it was often assumed that dynamic models added to the attractiveness of acoustic signals (Richardson et al. 2010; Taylor et al. 2011a; Laird et al. 2016; Stange et al. 2017). However, signal components do not necessarily combine additively, and unpredictable preference landscapes for multimodal stimuli can arise (Smith and Evans 2013; Ronald et al. 2017). Likewise, to explain females’ responses to some multimodal stimuli with asynchronous visual and auditory components, a number of untested assumptions were required concerning females’ criteria for deciding whether the two components constitute one multimodal stimulus to be accepted or an unrealistic multimodal stimulus to be rejected (Taylor and Ryan 2013; Taylor et al. 2017). Because many unpredicted results were found, complicated assumptions are needed to explain the outcomes, and the resulting interpretations were mostly study-specific. To form structured knowledge on the role of vocal sacs, research should be designed to aim for descriptive data that can fit into existing theories of multimodal interaction categorization (Richardson et al. 2010; Smith and Evans 2013), information use (Rubi and Stephens 2016), evolutionary function (Hebets and Papaj 2004; Bro-Jørgensen 2010), and sensory integration (Lee et al. 2019).

Taking a step back from searching for ways that vocal sacs are involved in communication besides acoustics, it might be worthwhile considering whether vocal sacs have evolved to be less conspicuous in communication. The predation risk for males increases when they call (Tuttle and Ryan 1981; de Silva et al. 2015), and the dynamic vocal sacs can be potential cues not only for females but also for unintended predators (Rhebergen et al. 2015; Gomes et al. 2016). In some frog
species, including *H. chrysoscelis*, males’ vocal sacs are pigmented in dark colors while females’ throats lack pigmentation and have the same coloration (often whitish) as the rest of the abdomen. Thus, it is possible that vocal sacs have evolved to be camouflaged when calling. The investigation on what vocal sacs do not signal instead of what vocal sacs can signal may be fruitful.

This paper discussed the role of vocal sacs as visual cues in acoustically guided courtship in treefrogs. We encourage research with larger sample sizes, with careful visual presentation and testing conditions, in different species, and by different research teams to establish the common ground of whether vocal sacs serve as visual cues. We also encourage research that is designed to link observations to established multimodal frameworks. More broadly, the role of vocal sacs in other behavioral contexts, such as agonistic behaviors (Narins et al. 2005), and sensory modalities, such as chemoreception (Starnberger et al. 2013) are also being discovered and calling for researchers’ attention (Starnberger et al. 2014).
Illustrations
Because of the access to a spectrum meter in 2019, we repainted the frog models for part of Experiment 2 (SPL 76 dB and 82 dB) and Experiment 4, which were conducted in 2019. (a) shows the 3D design of the model, the real frogs, and frog model 2019. (b) from top to bottom shows the reflectance spectrum comparisons of frog models and real male frogs for the vocal sac, the back, and the belly respectively. The darker gray areas and lighter gray areas represent the standard deviations and ranges of measurements from 38 real males for the vocal sac and 12 for the back and the belly.
Figure 2 The schematic and photos of experimental setup for phonotaxis tests

The system included three modules – the workstation, the robotic system, and the testing area. From the workstation, digital audio signals were played with the laptop and sent to the audio adapter. The audio adapter translated the digital signals to multichannel analog signals. Acoustic signals for sound playback were sent to speakers in the testing area while the command signal used to control the vocal sac movement was sent to the input pin of the control board in the robotic system. Within the robotic system, a 7.4 v lithium-polymer battery was connected to the control board and the motor as a power source. Upon receipt of the command signal, the control board sent input to the motor to control its movement. The rotational movement of the motor was converted to the linear movement of a syringe through one of the two mechanical structures and pumped air in and out of the DM’s vocal sac. The mechanical structures and the integrated robotic system are shown in photos. In the testing area, the arena arrangement for two-choice tests (NM+ vs DM+) is shown. In no-choice tests, the speaker and release point were positioned on the opposite edge of the arena, with a distance of 1 m between them.

Fig. 2a
Fig 2b.
**Figure 3 Experiment 1: Female preferences and response latency for unimodal stimuli**

In the top panel, the dots show the proportions of females choosing the alternatives indicated on the top of the figure with 95% exact binomial confidence intervals indicated by vertical lines. The horizontal dashed lines represent the chance-level proportion (0.5) for females to choose the alternatives. In the lower panels, the boxes show the 25% to 75% range of the response latency, with the lines inside indicating the average values. The vertical lines and dots represent the standard deviations and outliers, respectively.
Figure 4 Experiment 2: Female preferences and response latency for multimodal stimuli

In (a), calls were played back at 88 dB (at 50cm), and the results for pairwise choice tests between the three visual treatment conditions are shown. In (b) are shown the results for choice tests between DM* versus SM* when the calls were played at different sound pressure levels (SPLs). In the top panels, the dots show the proportions of females choosing the alternatives indicated on the top of the figures with 95% exact binomial confidence intervals indicated by vertical lines. The horizontal dashed lines represent the chance-level proportion (0.5) for females choosing randomly. In the lower panels, the boxes show the 25% to 75% range of the response latency, with the lines inside indicating the average values. The vertical lines and dots represent the standard deviations and outliers, respectively.

Fig. 4a
Table 1 Experiment 1: Statistical summary for analysis of influential factors for female choices in negative control

A series of logistic regression models were fitted and compared. Denote the chance for a female to choose a specific alternative (either left or right) as $x$, then the response variable of the model is $\ln \left( \frac{x}{1-x} \right)$. In the “fixed effects” panel, all potential factors evaluated are listed in the left column. For any of the categorical factors, a category was assigned as the reference (indicated in parentheses) to estimate the intercept and to compare the effect of other categories. For each factor, the value in the first row shows the coefficient and the value below in parentheses shows the standard error for the estimation of the coefficient. $n = 34$.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Null model</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−0.48</td>
<td>−0.81</td>
<td>−1.20</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.60)</td>
<td>(0.66)</td>
</tr>
<tr>
<td>Order: lagging</td>
<td>0.52 †</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ref. leading)‡</td>
<td></td>
<td>(0.75)</td>
<td></td>
</tr>
<tr>
<td>Spk: Spk1</td>
<td></td>
<td>1.11 †</td>
<td></td>
</tr>
<tr>
<td>(Ref. Spk2)</td>
<td></td>
<td></td>
<td>(0.79)</td>
</tr>
</tbody>
</table>

Models level evidence

| Deviance       | 45.23  | 44.73  | 43.11  |
| AICc           | 47.36  | 49.12  | 47.50  |

Model comparisons

<table>
<thead>
<tr>
<th>Ref. model</th>
<th>Null model</th>
<th>Null model</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔDeviance (df)</td>
<td>0.50 (1)</td>
<td>2.12 (1)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.478</td>
<td>0.145</td>
</tr>
<tr>
<td>Decision</td>
<td>rejected</td>
<td>rejected</td>
</tr>
</tbody>
</table>

† Factors whose significance is evaluated from corresponding model comparisons.
‡ In a choice test, the alternative presented first between the two is referred to as leading.
Figure 5 Experiment 3: The oscillograms of sinusoidally amplitude-modulated (SAM) calls

Shown are SAM calls with amplitude modulation depths (a/b) of, from top to bottom, 0.2, 0.5 and 1.0.
Figure 6 Experiment 3: The modulation depth thresholds that elicited phonotaxis and the response latency for stimuli with the three visual treatments

The surrounded areas show the distribution of data points and the width of the shape corresponds to the relative probability density at various modulation depths. The dots and bars represent the means and standard deviations, respectively, \((n = 29\) to \(30\) per treatment).
Figure 7 Experiment 4: Female responses to multimodal stimuli in noisy environment

In (a), global response proportions at each SNR regardless of visual treatment are plotted. The vertical lines show the 95% exact binomial confidence intervals. In (b), the response proportions (the top panel; \( n = 24 \) to 29 per visual treatment at each SNR) and response latency (\( n = 4 \) to 17 per visual treatment at each SNR) are broken down by visual treatment. In the lower panel of (b), the boxes show the 25% to 75% range of the response latency, with the lines inside indicating the average values. The vertical lines and dots represent the standard deviations and outliers, respectively.

Fig. 7a

![Figure 7a graph]

Fig. 7b

![Figure 7b graph]
A series of mixed-effect logistic regression models were fitted and compared. Individual frogs were included as random factors in all models to account for the dependence of data from the same individual. The data includes 397 tests in total.

### Table 2 Experiment 4: Statistical summary for analysis of influential factors for female responses not including the irradiance data

A series of mixed-effect logistic regression models were fitted and compared. Individual frogs were included as random factors in all models to account for the dependence of data from the same individual. The data includes 397 tests in total.

<table>
<thead>
<tr>
<th></th>
<th>Null model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.83</td>
<td>−5.377</td>
<td>−5.29</td>
<td>−5.57</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(1.18)</td>
<td>(1.20)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>SPL</td>
<td>0.06†</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Stimulus: DM⁺</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ref. NM⁺)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus: SM⁺</td>
<td>−0.35†</td>
<td></td>
<td>−1.35†</td>
<td></td>
</tr>
<tr>
<td>(Ref. NM⁺)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus×SPL: DM⁺×SPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ref. NM⁺×SPL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus×SPL: SM⁺×SPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ref. NM⁺×SPL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model-level evidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>−244.11</td>
<td>−235.78</td>
<td>−235.09</td>
<td>−234.63</td>
</tr>
<tr>
<td>Deviance</td>
<td>488.21</td>
<td>471.55</td>
<td>470.19</td>
<td>469.26</td>
</tr>
<tr>
<td>AICc</td>
<td>492.24</td>
<td>477.61</td>
<td>480.34</td>
<td>483.55</td>
</tr>
<tr>
<td><strong>Model comparisons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ref. model</td>
<td>Null model</td>
<td>Model 1</td>
<td>Model 1</td>
<td>Model 1</td>
</tr>
<tr>
<td>ΔDeviance (df)</td>
<td>16.66 (1)</td>
<td>1.36 (2)</td>
<td>2.29 (4)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.000045</td>
<td>0.506</td>
<td>0.683</td>
<td></td>
</tr>
<tr>
<td>Decision</td>
<td>accepted</td>
<td>rejected</td>
<td>rejected</td>
<td></td>
</tr>
</tbody>
</table>

† Factors whose significance is evaluated from corresponding model comparisons.
Table 3 Experiment 4: Statistical summary for analysis of influential factors for female responses including the irradiance as a potential factor

The data was included in this analysis only if irradiance measurement was available for when the test done (190 tests). On two nights of the experiments, the irradiance was much higher than the other nights and the tests done on those nights were identified as high leverage data points. We removed the data for those two nights to ensure those datapoints would not over impact the model estimation. A series of mixed-effect logistic regression models were fitted and compared. Individual frogs were included as random factors in all models to account for the dependence of data from the same individual.

<table>
<thead>
<tr>
<th>Null model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−0.64</td>
<td>−5.28</td>
<td>−15.90</td>
<td>−4.92</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(1.70)</td>
<td>(6.73)</td>
<td>(1.73)</td>
</tr>
<tr>
<td>SPL</td>
<td>0.06†</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>log_light</td>
<td></td>
<td>−0.44†</td>
<td></td>
<td>−0.97†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.26)</td>
<td></td>
<td>(0.66)</td>
</tr>
<tr>
<td>Stimulus: DM+</td>
<td></td>
<td>−0.53†</td>
<td>1.32†</td>
<td></td>
</tr>
<tr>
<td>(Ref. NM+)</td>
<td></td>
<td>(0.65)</td>
<td>(19.22)</td>
<td></td>
</tr>
<tr>
<td>Stimulus: SM+</td>
<td></td>
<td>−0.51†</td>
<td>31.56†</td>
<td></td>
</tr>
<tr>
<td>(Ref. NM+)</td>
<td></td>
<td>(0.66)</td>
<td>(19.09)</td>
<td></td>
</tr>
<tr>
<td>Stimulus: DM+×log_light</td>
<td></td>
<td>0.06†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ref. NM+×log_light)</td>
<td></td>
<td>(0.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus: SM+×log_light</td>
<td></td>
<td>1.32†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ref. NM+×log_light)</td>
<td></td>
<td>(0.78)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model-level evidence

41
<table>
<thead>
<tr>
<th></th>
<th>Ref. Model</th>
<th>Null model</th>
<th>Model 1</th>
<th>Model 1</th>
<th>Model 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔDeviance (df)</td>
<td>8.36 (1)</td>
<td>2.95 (1)</td>
<td>0.85 (2)</td>
<td>9.44 (5)</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.004</td>
<td>0.086</td>
<td>0.654</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>Decision</td>
<td>accepted</td>
<td>rejected</td>
<td>rejected</td>
<td>rejected</td>
<td></td>
</tr>
</tbody>
</table>

† Factors whose significance is evaluated from corresponding model comparisons.
Bibliography


---

1 Audacity® software is copyright © 1999-2020 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.


47


