Received: 18 November 2020 | Accepted: 17 February 2021

DOI: 10.1111/1365-2656.13465

RESEARCH ARTICLE

Journal of Animal Ecology BRITISH ECOLOGICAL SOCIETY

Multisensory modalities increase working memory for mating signals in a treefrog

Bicheng Zhu[1] | Ya Zhou[1,2] | Yue Yang[1] | Ke Deng[1] | Tongliang Wang[3] | Jichao Wang[3] | Yezhong Tang[1] | Michael J. Ryan[4] | Jianguo Cui[1]

[1]Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Ministry of Education Key Laboratory for Tropical Plant and Animal Ecology, College of Life Sciences, Hainan Normal University, Haikou, China

[4]Department of Integrative Biology, University of Texas, Austin, TX, USA

Correspondence
Jianguo Cui
Email: cuijg@cib.ac.cn

Funding information
Youth Innovation Promotion Association of the Chinese Academy of Sciences, Grant/Award Number: 2012274; CAS "Light of West China" Program, Grant/Award Number: 2016XBZG_XBQNXZ_A1_003 and 2017XBZG_XBQNXZ_B_013; National Natural Science Foundation of China, Grant/Award Number: 31572275, 31772464 and 32001095

Handling Editor: Elizabeth Derryberry

Abstract

1. Animal choruses, such as those found in insects and frogs, are often intermittent. Thus, females sampling males in the chorus might have to remember the location of the potential mates' calls during periods of silence. Although a number of studies have shown that frogs use and prefer multimodal mating signals, usually acoustic plus visual, it is not clear why they do so. Here we tested the hypothesis that preference for multimodal signals over unimodal signals might be due to multimodal signals instantiating longer memories than unimodal signals, particularly during the inter-chorus intervals.
2. We tested this hypothesis in serrate-legged small treefrogs *Kurixalus odontotarsus* whose males produce advertisement calls accompanied by conspicuous vocal sac inflation. Females were tested with acoustic and acoustic + visual (video of inflating-deflating vocal sac) mating calls.
3. We found that females prefer multimodal calls over unimodal, audio-only calls. Furthermore, multimodal calls are still preferred after a silent period of up to 30 s, a time that spans the average silent period of the chorus. This was not true of unimodal calls.
4. Our results demonstrate that a multimodal signal can engage longer working memory than a unimodal signal, and thus female memory might favour the evolution of multimodal signals in males through sexual selection. Selection might also favour female preference for multimodal signals if longer memory facilitates mate searching and assessment. Our study does not allow us to elucidate the sequence of evolution of this trait and preference.

KEYWORDS

active time, *Kurixalus odontotarsus*, mate assessment, multimodal signal, sexual selection, signal evolution, vocal sac, working memory

1 | INTRODUCTION

In many species of animals, especially insects and frogs, males gather together to acoustically advertise to females (Gerhardt & Huber, 2002). Although choruses can last much of the night, calling is not continuous and often interspersed with bouts of silence. Thus, females who are searching for males and comparing males to one another would benefit from the ability to remember the location of the males' calls over bouts of silence. In at least one species of frog, more complex acoustic stimuli instantiate longer-term memory than less complex calls (Akre & Ryan, 2010).

The complexity of a male's sexual display can be increased by adding additional sensory modalities. Multimodal displays are composed of signals and cues in more than one sensory modality, which are widespread in humans and other animals. Multimodal signals can profoundly impact cognitive processes in humans and other species (Halfwerk et al., 2019; McGurk & MacDonald, 1976). For example, multisensory modalities activated different subregions of the hippocampus to activate a number of memory-based processes in humans (Kesner, 2015). Multisensory stimuli are even applied to improving the short-term memory of dyslexic children (Saputra & Nugroho, 2015). Animals also have multiple senses through which they can detect the multimodal information. For instance, hunting bats *Trachops cirrhosus* prefer to attack model frogs emitting multisensory cues, in this case acoustic call and call-induced water ripples (Gomes et al., 2016; Halfwerk et al., 2014).

Acoustic signals produced by most frogs are accompanied by the inflation and deflation of a conspicuous vocal sac. The vocal sac is a sound radiator that likely evolved in response to selection for increased calling efficiency (e.g. increasing the call rate; Pauly et al., 2006). Previous studies revealed that frogs have excellent behavioural and retinal visual sensitivities corresponding to nocturnal light intensities, even under moonless conditions (approx. 0.004 lux; Cummings et al., 2008; Leslie et al., 2020; Underhill & Höbel, 2017). Diurnal species have greater sensitivity in the photopic domain while nocturnal species have 10-fold to 100-fold greater sensitivity in the scotopic domain. Physiological analysis suggests rod-specific adaptations drive scotopic thresholds in nocturnal species (Rosencrans et al., 2018). Given that nocturnal species should be able to perceive the dynamic vocal sac inflation, this cue can function as a component of the male's multimodal sexual display under natural nocturnal conditions (Taylor et al., 2008). In addition, the anuran vocal sac also has been co-opted for seismic and even chemical signalling (Halfwerk, Jones, et al., 2014; Starnberger et al., 2013, 2014).

Understanding the specific advantages of multimodal signal evolution is challenging. Numerous female animals prefer multimodal mating signals over unimodal signals, but it is not clear why they do so. Although the advantage of multimodal signals to the signaller has been demonstrated in some species (Halfwerk et al., 2014; Narins et al., 2003; Taylor & Ryan, 2013), the advantage to the receiver of preferring multimodal mating signals over unimodal signals has only been hypothesized (Candolin, 2003; Hebets & Papaj, 2005; Moller & Pomiankowski, 1993; Partan, 2013). One possible advantage of multimodal mating signals to receivers is that male multimodal sexual signals may instantiate longer memories in females than unimodal signals. We investigated the effect of male multimodal sexual signals on female memory duration in the serrate-legged small treefrogs *Kurixalus odontotarsus*.

Working memory refers to the ability to encode, maintain and flexibly manipulate information no longer present in the environment (Bizon et al., 2012). A classic method of assessing working memory duration in animals is the spatial delayed response task, which requires the animal to remember the location of a food reward concealed in one of two wells across a brief delay (Croxson et al., 2012; Rodriguez & Paule, 2009). Akre and Ryan (2010) used a similar method to assess the duration of female working memory of acoustic signals with numbers of call components, or complexity, in the túngara frog. They demonstrated that additional components, specifically chucks, increased the active time (the period over which a signal influences a receiver's response to that signal) of male mating calls. Other factors may also affect the active time associated with simple and complex signals, such as enhanced motivation to approach the speaker broadcasting complex calls and the potential 'behavioural mnemonic' effect (Akre & Ryan, 2010); that is, retaining orientation towards a sound after it ceases could maintain an attractive signal's location information (Akre & Ryan, 2010). Considering that multimodal cues can facilitate signal detection and recognition (Rowe, 1999), we tested the hypothesis that multimodal sexual signals increased the duration of working memory in females in the context of mate choice.

The serrate-legged small treefrog *K. odontotarsus* is a tropical species that has been well studied in acoustic communication and sexual selection (Zhu et al., 2016; Zhu, Wang, Brauth, et al., 2017; Zhu, Wang, Zhao, et al., 2017). Males produce multisensory courtship calls by vocalizing and presenting their inflating and deflating (i.e. dynamic) vocal sac as a visual cue (Figure 1). Females base their mate choices on male advertisement calls, which consist of several wideband frequency notes (Figure 1; Zhu, Wang, Zhao, et al., 2017). Most male

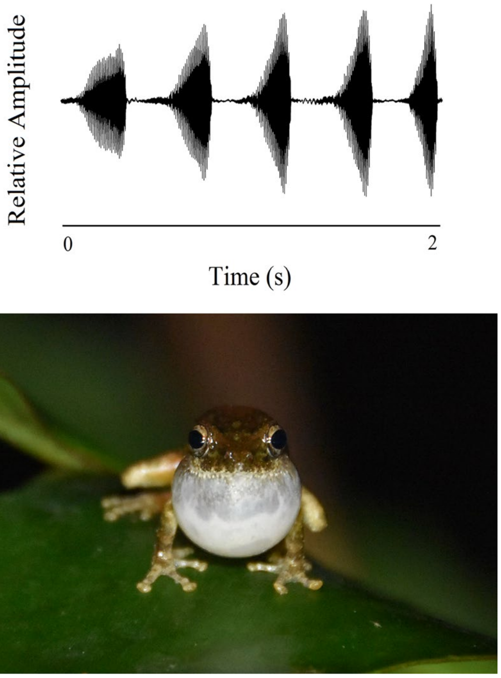

FIGURE 1 Waveform of a call and photo of the vocal sac of a serrate-legged small treefrog. The waveform (top) of an advertisement call with five notes (abbreviated as '5A') and vocal sac (bottom) of a male serrate-legged small treefrog. Male frogs usually produce five-note advertisement calls with an average duration of 2 s, although they can produce advertisement calls consisting of one to nine notes. The vocal sac inflates and deflates for each note at a mean frequency of 4.1 times per second

serrate-legged small treefrogs call at night, and a few call during the day as well. Since there are many shelters in the tropical rainforest, females sometimes can hear but not see advertising males.

Video-playbacks have proven to be an effective method to explore the perception of multimodal signals in numerous animal taxa, including some nocturnal anurans (Reichert & Höbel, 2015; Rosenthal et al., 2004; Smith & Evans, 2008; Uetz & Roberts, 2002). Using the video-playback technique, we presented manipulated acoustic stimuli and/or simultaneous video sequences to female serrate-legged small treefrogs. We investigated the hypothesis that multimodal signals are more effective than unimodal signals in eliciting female responses through their efficacy in engaging working memory, which may allow females to find potential mates more reliably during silent periods between choruses. We tested two predictions of this hypothesis: (1) the multimodal signal is more attractive than a unimodal signal to females and (2) multimodal signals instantiate longer working memory compared to unimodal signals.

2 | MATERIALS AND METHODS

2.1 | Experimental procedures

All experiments were completed in Hainan, China at the Mt. Diaoluo National Nature Reserve between May and August 2016, 2018 (during the natural breeding season of *K. odontotarsus*). Gravid females (average body length, 47 mm) were captured by hand from the wild and held separately in tanks with a meshed cover (30 cm × 30 cm × 30 cm) which contained water and branches with foliage under natural light and temperature condition until they were tested. Subjects were held in the tanks for an average of 1 hr between collecting and testing. Each female completed a series of binary phonotaxis tests in a 150 cm × 150 cm × 120 cm sound-attenuating chamber without a roof between 20:00 hr and 2:00 hr (temperature: 22.7 ± 1.1°C, relative Humidity: 86.3 ± 4.9%). The chamber was placed in a darkened sound-attenuating house (4.0 m × 2.7 m × 3.0 m). The behaviour of the females was observed using a wide angle lens video system (Woshida, 84H10P; Shenzhen Woshida Technology Ltd.) equipped with an infrared light source. After all female frogs completed the experiment, we used the sucker-clipping method to prevent recapture of the same subjects in the wild on subsequent nights, following the Guidelines for the Use of Live Amphibians and Reptiles in Field Research (Beaupre et al., 2004). We only clipped part of the soft tissue (i.e. sucker) from the frog's smallest fingertip to minimize harm. After the sucker-clipping marking, the wound was wiped with 75% alcohol to avoid infection. Subsequent observations proved that such treatment did not affect the activities (e.g. climbing, ovulation) of female frogs in the field. After about 2–3 months the regrown sucker is smaller than the original and easy to distinguish. Subjects were released immediately in the same location of capture after testing, after being held in the tanks for an average of 2 hr between testing and release.

2.2 | Stimulus presentation

We filmed calling male serrate-legged small treefrogs in the field at night using a digital camera (D7200; Nikon, Japan) connected to a shotgun microphone (MKE 400; Sennheiser, Germany). We kept the camera and microphone facing the front of the male frog (as shown in Figure 1), and the distance from the camera to the calling male was about 1 m. We used an adjustable flashlight (Fenix UC30; Shenzhen Langheng Electronics Co., Ltd.) to provide the faint light when filming (camera lens: 80–100 mm; frame rate: 29 frames/s). The light intensity measured at the male's calling site measured with a light metre (TES-1339, TES Electrical Electronic Corp.) was about 4.40 lux (range: 2.72–6.91 lux) with the flashlight. The audio sampling rate was 48 kHz. The temperature was 21.8 ± 0.76°C (mean ± *SD*) when recording in the field, similar to the temperature at which females were tested.

To generate a multimodal signal, we placed a 20.2 cm × 13 cm LCD monitor (EYOYO, 082IPS; Shenzhen Gempanel Technology Ltd.; 1,920 × 1,200 resolution; 75 Hz refresh rate) that showed the entire frog with its dynamic vocal sac (i.e. inflating and deflating) above each speaker (Edifier, R1200TII; Beijing Edifier Technology Ltd.), which formed a triangle with the frog such that the angle of the screens relative to the release point was 60° (Figure 2). The speakers broadcast the acoustic signal and the monitor synchronously presented the visual stimulus. Screens had a large visual angle (170°) to allow animals to view stimuli from a larger angle of positions.

To ensure the parameters of the two screens were the same, we used the colour calibration function of Windows to adjust the brightness, contrast ratio and colour balance of the two screens. Meanwhile, we calibrated the colour of the screens using a colorimeter (Datacolor Spyder 5 Elite; Datacolor Trading (Shanghai) Co., Ltd.) and the brightness of the screens using a screen luminance metre (SANPOMETER, SM208; Shenzhen Xinbaorui Instrument Co., Ltd.). The brightness of one screen that showed the static frog was 75.5 cd/m^2, and the brightness of the other screen that broadcast the calling frog was 77.1 cd/m^2. The light intensity measured at the release point (1 m away from the screen) was 0.06 lux, and 8.59 lux at 10 cm away from both screens, which coincides with the intensity of nocturnal light in the wild where serrate-legged small treefrogs occur (Deng et al., 2019). To minimize pseudoreplication, six audio-visual stimuli derived from six different calling males were used. After adjustments, six audio-visual stimuli with different recording background had the same light intensity. When each female frog participated in the experiment, one of the six audio-visual stimuli was randomly selected. In total, 164 female frogs participated in the experiment.

The study of vision is challenging in any system because it is difficult to present stimuli that replicate the salient visual features of objects in nature as perceived by the study organism (Reichert et al., 2014; Rosenthal, 1999). Because the spectra vary widely in tropical rainforests and our field observation found that the body colour of male serrate-legged small treefrogs varies greatly, a few males even call at noon, it is impossible to match the natural spectrum with

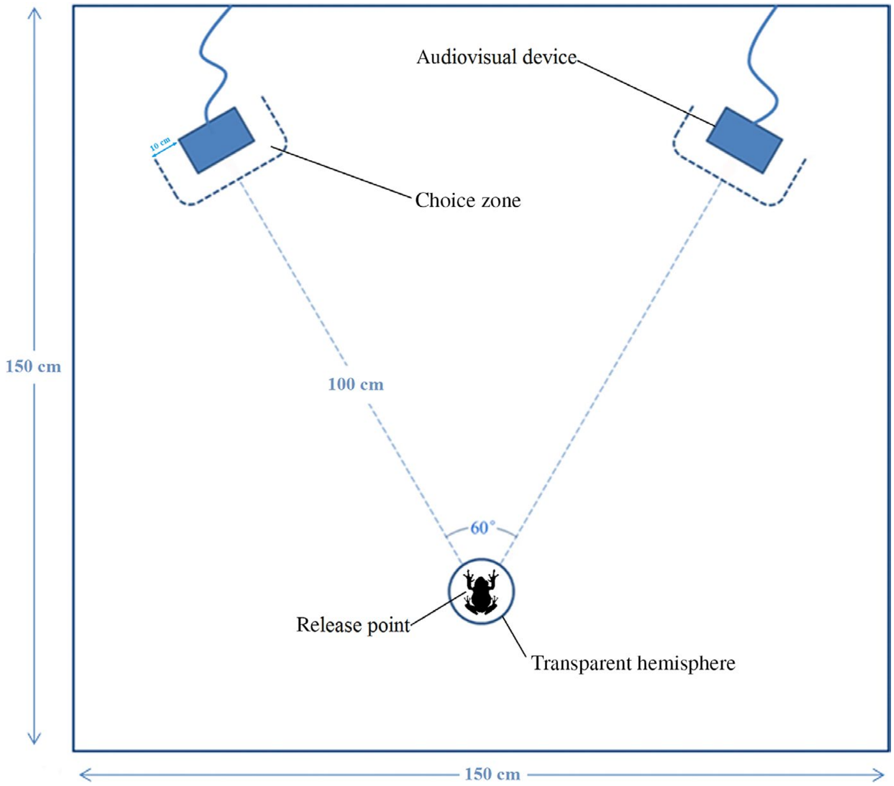

FIGURE 2 The schematic diagrams of the phonotaxis test arena (top view). We restrained a female frog under a transparent hemisphere that was placed between two audio-visual devices, which were separated by 100 cm and formed a triangle with a 60° separation relative to the release point in a 150 cm × 150 cm sound-attenuating chamber. Female frogs could choose between two stimuli, a unimodal signal presenting only a sound or a video playback of dynamic vocal sac and a multimodal signal presenting a synchronized combination of a sound and a dynamic vocal sac

an experimental spectrum. Using a similar approach with túngara frogs, data obtained from video playback experiments (Rosenthal et al., 2004) were statistically indistinguishable from the results of analogous experiments using robotic frogs (Taylor et al., 2008). Adobe Premiere Pro CS6 (Adobe, USA) was utilized to construct the audio and video tracks. Images were scaled so they would appear life-sized (head length: 12.1 mm, the largest diameter of inflating sac: 15.7 mm) on the playback monitor. We played back videos (saved as MPEG files, 25 frames/s, 1,440 × 800 pixel resolution, 8 bit depth, 48 kHz audio sampling rate) using the Windows Media player onto an Acer portable computer. The frame rate (25 Hz) of the video playback is much higher than the frequency (4.1 Hz) at which the vocal sac of serrate-legged small treefrog inflates. In addition, the quality of the video playback will not be compromised, as the refresh rate (75 Hz) of the monitors is higher than the frame rate of video playback. Thus, the serrate-legged small treefrogs studied in our video playback should not perceive the video as 'stuttering'. The speakers were connected to the computer through audio cables and the monitors were connected to a HDMI splitter which was connected to the portable computer via a HDMI video cable. Videos were broadcast from the computer located outside of the testing chamber to two monitors placed inside the chamber using the split screen mode. Also see Figure S1 for more details.

Since five-note and four-note advertisement calls (abbreviated as '5A', '4A') are most common in natural recordings (Zhu, Wang, Brauth, et al., 2017), we designed the base stimuli that included the 1.6-s call '4A' and the 2-s call '5A' (Figure 1). We broadcast the test stimuli antiphonally with 5-s inter-stimulus intervals. The stimulus pair was presented in a loop so that the intervals between call 4A and call 5A were 5 s and the two stimuli were played the same number of times.

To eliminate the possibility that females were merely attracted by light/images from the screens (phototaxis), we displayed a background video of a male frog's calling location (as if they were blocked by leaves in the tropical rainforest) on the monitor as a control when audio-4A or audio-5A was playing (Figure S1). We emphasize again that there are many situations in nature when the female can hear a male frog but not see its vocal sac. For each phonotaxis test, we placed a female frog under a visually and acoustically transparent hemisphere with some small holes at the bottom (12.8 cm in diameter, 1.3 mm thick). We randomized which speaker/monitor was used to present the multimodal stimuli, the last call played before the silent delay, and the leading call after the silent delay. We presented females with unimodal or multimodal stimuli, which were played back at the 'fast' root-mean-square amplitude of 80 dB SPL (re: 20 µPa, Z-weighted). The sound pressure level of each test call was calibrated using a sound level meter (AWA 6291; Hangzhou Aihua Instruments Co.).

2.3 | Phonotaxis trials

In the signal attraction tests (a classic binary phonotaxis test), we designed two stimulus pairs, visual-5A (sac inflation without the vocalization) versus audio-5A (Table 1, Signal attraction test, stimulus pair 1), and audio-5A (Table 1, Signal attraction test, stimulus pair 2, stimulus 1) versus multi-5A (stimulus pair 2, stimulus 2) to investigate female preferences for the isolated visual signal and the multimodal signal relative to the acoustic-only signal (Table 1). Before each

TABLE 1 Experimental design and the stimuli contained in each test

Experiment content	Stimulus pair	Stimulus 1		Stimulus 2		Comparison method
		Audio	Visual	Audio	Visual	
Signal attraction test	1	—	5A	5A	—	Intra-pair
	2	5A	—	5A	5A	Intra-pair
Working memory test	1	4A	—	5A	—	Inter-pair
	2	4A	—	5A	5A	
Preference strength test	1	4A	—	5A	—	Inter-pair
	2	4A	—	5A	5A	

Note: The signal attraction test was designed to investigate female preference for the isolated visual signal and the multimodal signal relative to the acoustic signal, through intra-stimulus pair comparison (e.g. in stimulus pair 1, stimulus 1 vs. stimulus 2). In the working memory test, based on a pure acoustic signal (audio-4A), we compared the duration of working memory between a unimodal acoustic signal (audio-5A) and a multimodal signal (multi-5A), through inter-stimulus pair comparison (i.e. stimulus pair 1 vs. stimulus pair 2). In the preference strength test, we tested whether female frogs showed different levels of motivation for approaching the speaker broadcasting unimodal or multimodal calls through comparing female preference strength between unimodal and multimodal stimulus pairs (i.e. stimulus pair 1 vs. stimulus pair 2).

phonotaxis test, the female frog was restrained under the transparent hemisphere for 60 s to acclimate, during which both devices alternately broadcast identical audio-4A (the habituation phase). Then, the hemisphere was raised to release the female as two devices began to broadcast visual-5A versus audio-5A (Table 1, Signal attraction test, stimulus pair 1) or audio-5A versus multi-5A alternately (Table 1, Signal attraction test, stimulus pair 2; the selection phase).

For each trial, we scored whether a ‘choice’ or ‘no choice’ had been made by the female frog. We considered a choice to have been made when the female approached within 10 cm of a speaker (i.e. the choice zone) within 10 min of stimulus onset without simply following the wall. If she failed to enter the choice zone within 10 min, or did not leave the release point within 5 min, we considered this to indicate that no choice had been made.

In the working memory test (Table 1), we designed specialized phonotaxis experiments to test whether multimodal signals instantiated longer working memory than unimodal acoustic signals as in the experiment with túngara frogs (Akre & Ryan, 2010). In the working memory tests, we studied female frogs' choice behaviour after three silent periods, 0, 15 and 30 s, for both unimodal (audio-5A versus audio-4A; Table 1, Working memory test, stimulus pair 1) and multimodal (multi-5A versus audio-4A; stimulus pair 2) stimulus pairs. Similar to the reference stimulus ‘whine’ call in túngara frogs, audio-4A was chosen as a reference stimulus in the experiments to determine working memory duration. As a reference stimulus, audio-4A is attractive to females, but its attraction is significantly lower than the target stimulus audio-5A/multi-5A (Table 1).

The working memory test contained four phases. First, two playback devices alternately broadcast identical audio-4A for 60 s (the habituation period). For the next 60 s, one presented the audio-4A while the other presented either audio-5A or multi-5A (the stimulation period). After a silent period (the delay period), the female frog was released, and both devices began to play identical audio-4A (the selection period). Detailed phases and operation were shown in Table S1. If most females chose the speaker that had previously broadcast the target stimulus (i.e. audio-5A or multi-5A), we took this to indicate that the silent period was within this call's active time, that is, it was remembered by the female. We further studied the active time of multimodal calls by testing female preference after silent periods of 45, 60 and 120 s.

We also tested female preference strength (a population measure of the proportion of females that respond to a particular stimulus), and measured female latency to make a choice (a measure of an individual's motivation to choosing behaviour) for unimodal and multimodal stimulus pairs, to test whether female frogs showed different levels of motivation for approaching the speaker broadcasting unimodal or multimodal calls. In the preference strength tests, two stimulus pairs were presented, audio-4A versus audio-5A (Table 1, Preference strength test, stimulus pair 1) and audio-4A versus multi-5A (stimulus pair 2), to compare female preference strength between unimodal and multimodal stimulus pairs through comparison between groups (Table 1). Similar to above signal attraction test, the preference strength test is a classic binary phonotaxis test and only contains the habituation period and the selection period as well.

To test the ‘behavioural mnemonic’ effect, we recorded the original orientation of each female frog in each test when the hemisphere was raised. This effect occurs when the female's original orientation to the speaker predicts her choice. Females completed a block of trials in random order and were tested only once with any given stimulus. To avoid experimental fatigue, each female frog participated in no more than five tests and was allowed a 3-min break between consecutive tests. During the interval of each test, we mopped the ground to keep the arena moist and to eliminate possible chemical cues.

2.4 | Natural chorus activity

We recorded three natural choruses of serrate-legged small treefrogs over three nights to measure the durations and silent intervals. Using three Song Meter Digital Voice Recorders SM4 (Wildlife Acoustics, Inc.), we recorded choruses continuously for 12 hr (18:00–06:00 hr) with temperatures ranging from 16 to 24°C. We chose the first five bouts and intervals from each 60-min period between 19:30 and 23:30 hr, which included most of the chorus activity for that night.

2.5 | Statistical analysis

The waveform of the male frog's advertisement call was visualized using free software PRAAT (Boersma, 2002). The results of the phonotaxis experiments were visualized using Sigmaplot 11.0 software (Systat Software Inc.). Considering that the generalized estimation equation (GEE) has advantages in statistical analysis of discrete repeated measurement data (such as the binary variable) and the data with missing values, we used the GEE procedure in SPSS 17 and created a binary logistic model of speaker choice predicted by signal modality (unimodality, multimodality) and silent period durations to determine the influence of signal modality on speaker choice after different silent period durations. Our statistical analysis controlled for the fact that most females completed multiple tests but that few completed all conditions using the repeated subject function. The pairwise comparisons are based on the Estimated Marginal Means pairing of the initial scale of the dependent choice. The exact binomial test was used to analyse the results of the signal attraction tests. The Fisher's exact test and two-tailed paired *t* test were used to analyse female preference strength and latency to choose between unimodal and multimodal stimulating pairs, respectively. The duration and silent intervals of the choruses were analysed using Adobe Audition 3.0 (Adobe, USA). All data are expressed as the mean ± *SD*, and $p < 0.05$ was considered to be statistically significant.

2.6 | Ethics note

All applicable international, national and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were approved by the Animal Care and Use Committee of Chengdu Institute of Biology, CAS.

3 | RESULTS

3.1 | Signal attraction tests

We conducted phonotaxis experiments to investigate the differences in female preferences for audio-5A over visual-5A and over multi-5A (Table 1). First, for the visual versus audio comparison, a total of 34 gravid females were utilized and 8 subjects failed to make a choice. We found that 80.8% of females chose audio-5A over isolated visual signals (21 vs. 5; exact binomial test: $p < 0.001$, $n = 26$). Then, we tested whether audio-visual signals are more attractive than acoustic-only signals. For the multi-5A versus audio-5A comparison, 34 gravid females were utilized and 4 subjects failed to make a choice. We found that 73.3% of females chose multi-5A over audio-5A (22 vs. 8; exact binomial test: $p = 0.005$, $n = 30$).

3.2 | Working memory tests

After a silent period of either 15 or 30 s, females preferred to choose the speaker that originally broadcast multi-5A, but there was no preference for the speaker that originally broadcast audio-5A when the two speakers began to play identical audio-4A (Figure 3). The repeated-measures logistic regression with the GEE model revealed a significant effect of signal modality on speaker choice ($p = 0.003$; see Tables S2 and S3). The silent period × signal modality interaction was not significant (Table S2). Pairwise comparisons showed that results at each delay period differed (near-) significantly between unimodal and multimodal calls (0 s: 19 vs. 11, $p = 0.058$, $n = 30$; 15 s: 21 vs. 9, $p = 0.03$, $n = 30$; 30 s: 21 vs. 9, $p = 0.049$, $n = 30$; see Table S4). Only one subject failed to make a choice after 0 and 15 s delay, and four subjects failed to make a choice after 30 s delay. These results show that multimodal signals result in memories of longer duration after 15 and 30 s intervals of silence.

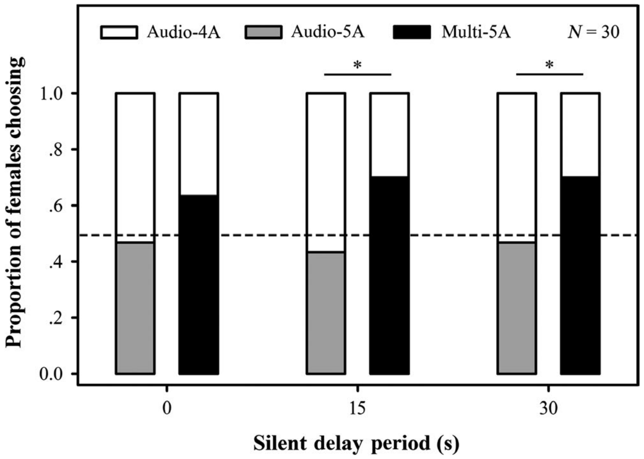

FIGURE 3 Proportion of females who chose the speaker originally broadcasting unimodal or multimodal signal after different silent periods. Bars show the proportion of female frogs that chose the speaker originally broadcasting unimodal 5A (sound only) or multimodal 5A (audio + visual) relative to the audio-4A after a silent period of 0, 15 or 30 s. The dashed line represents chance level (50%). In this experiment, the two speakers initially played identical audio-4A, and then one played the audio-4A (white bars) and the other played either audio-5A (grey bars) or multi-5A (black bars). After a silent period, female frogs preferred to choose the speaker that originally broadcast multi-5A and did not prefer to choose the speaker that originally broadcast audio-5A when the two speakers began to play identical audio-4A again. General estimating equation: $n = 240$ choices, $p = 0.012$; pairwise comparisons with the estimated marginal means contrasts, $^{*}p < 0.05$

We further measured the temporal limit of the multimodal signal's influence on females by adding the 45, 60 and 120 s silent periods. Only one subject failed to make a choice after the 60 and 120 s delay, and three subjects failed to make a choice after 45 s delay. Female preference for the speaker that had initially broadcast multimodal calls was maintained after 30 s and reached chance level (50%) after 45 s (Figure 4).

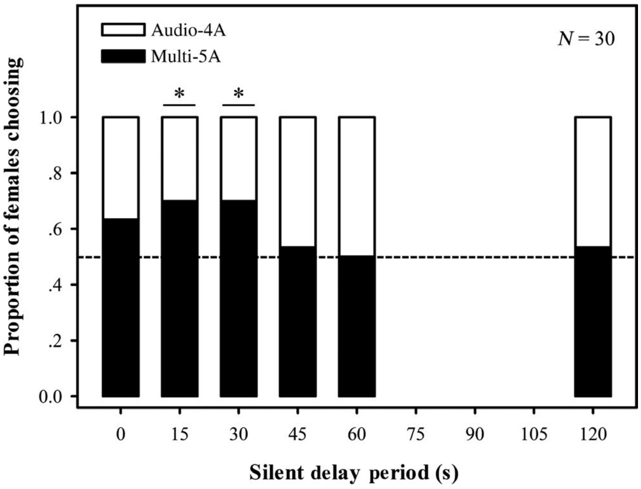

FIGURE 4 Proportion of female frogs who chose the speaker originally broadcasting multimodal signal decreases over time. Bars show the proportion of female frogs that chose the speaker originally broadcasting multimodal calls after variable silent periods. The dashed line shows the null hypothesis of 50% female choice for each speaker. Female preference for the multimodal signal faded after silent periods greater than 30 s and reached chance level after 45 s (*n* = 30 each condition, exact binomial test, *$p < 0.05$)

3.3 | Preference strength tests

An increased active time for multimodal calls could be a result of both different working memory durations and enhanced motivation to approach the speaker broadcasting the multimodal calls. Thus, we recorded two parameters of motivation, female preference strength and latency to make a choice, to test whether female frogs showed different levels of motivation for approaching the speaker broadcasting unimodal or multimodal calls.

We designed phonotaxis experiments to test the difference in female preference strength between the unimodal and multimodal stimulus pairs (Table 1). We found that 64.5% of females chose the audio-5A over the audio-4A (40 vs. 22; exact binomial test: $p = 0.007$, $n = 62$), and 66.7% of females chose the multi-5A over the audio-4A (20 vs. 10; exact binomial test: $p = 0.028$, $n = 30$). There was no difference in preference strength between the unimodal group (i.e. audio-5A vs. audio-4A) and multimodal group, that is, multi-5A versus audio-4A (Fisher's exact test: $p = 0.515$, Figure 5a), although females preferred the multi-5A to audio-5A (22 vs. 8; exact binomial test: $p = 0.005$, $n = 30$). Three female frogs failed to make a choice when the unimodal stimulus pair was played back, and one failed to make a choice when the multimodal stimulus pair was played back. Therefore, there was no difference in the proportion of females that made choices versus the total number of females tested (i.e. female responsiveness) for audio-5A versus audio-4A (95.4%) and multi-5A versus audio-4A (96.8%). Meanwhile, there was no significant difference in the latency to make a choice between the unimodal group (i.e. audio-5A vs. audio-4A) and multimodal group, multi-5A versus audio-4A (unimodal: 160.1 ± 11.5 s; multimodal: 154.5 ± 9.8 s; two-tailed paired *t* test: $p = 0.648$, $n = 120$; Figure 5b). Thus, the

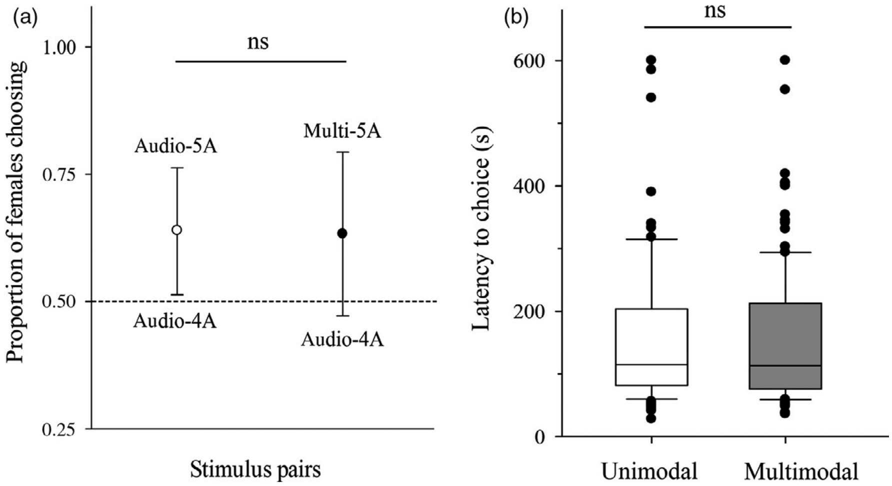

FIGURE 5 Tests for female motivation to unimodal and multimodal stimuli. (a) Female preference strength test. The audio-5A was a five-note advertisement call (sound only); the audio-4A was a four-note advertisement call (sound only); the multi-5A was a five-note advertisement call accompanied with a dynamic vocal sac. The dashed line indicates the proportion of females choosing one of the other alternatives based on chance (50%). Bars represent 95% confidence intervals; ns, not significant. There was no difference in preference strength between the unimodal group (audio-5A vs. audio-4A, open circle) and the multimodal group (multi-5A vs. audio-4A, filled circle), Fisher's exact test, $p = 0.515$. (b) Female latency to choose. There was no significant difference in female latency to make a choice between the unimodal group (audio-5A vs. audio-4A, white box) and the multimodal group (multi-5A vs. audio-4A, grey box), two-tailed paired *t* test, $p = 0.648$

difference in the duration of working memory between the unimodal group and multimodal group was not influenced by different levels of motivation.

The 'behavioural mnemonic' may also affect the active time associated with unimodal and multimodal stimuli. However, our experiments excluded the possibility of a behavioural mnemonic effect, as did experiments by Akre and Ryan (2010). The hemisphere was transparent, so we could directly observe female frog's movement and postural orientation under infrared illumination. The two speakers were placed at one side of the female frogs and formed a triangle with a 60° (not 180°) separation relative to the release point. The vast majority of female frogs were facing straight ahead when the hemisphere was lifted up. Therefore, female serrate-legged small treefrogs appear not to use a behavioural mnemonic to retain attraction to multimodal calls.

Female preference strength, latency to making a choice and orientation upon release did not differ significantly between the unimodal group (audio-5A vs. audio-4A) and the multimodal group (multi-5A vs. audio-4A). Therefore, we deduced that the increased active time of multimodal calls was a result of an increased duration of working memory.

3.4 | Natural inter-chorus intervals

We identified the period over which females might benefit from retaining attraction to males' calls by measuring the inter-chorus intervals of male treefrogs. From recordings of natural choruses, we found that male choruses (n = 3 chorus groups, 180 bouts) are separated by variable silent intervals with 27.0 ± 1.1 s (mean ± *SD*, n = 180; see Figure S2). As the mean chorus interval in nature is 27.0 s, the active time of the multimodal signal spans the average inter-chorus silent intervals.

4 | DISCUSSION

Our study provides three main findings: (a) female frogs showed a preference for the multimodal signal compared to the unimodal signal, (b) multimodal sexual signals instantiate longer working memory compared to unimodal auditory signals and (c) the active time of the multimodal signal spans the average inter-chorus silent intervals. These results support our hypothesis that multimodal signals are more effective than unimodal signals in eliciting female responses through their efficacy in enhancing the duration of working memory, which may allow females to find potential mates more reliably during silent periods between choruses.

We found that the vocal sac can serve as a visual cue and modulate mate choice when combined with an acoustic signal, but such an isolated visual signal is not as effective for mate attraction as an acoustic signal. Our findings that female serrate-legged small treefrogs prefer multimodal calls over unimodal calls are consistent with the results in the túngara frog and the green treefrog (Laird et al., 2016; Taylor & Ryan, 2013). Meanwhile, weaker preference to the isolated visual sac demonstrates that the acoustic component of a frog's call is its dominant feature for mate attraction. The visual component of the vocal sac mainly facilitates detection and localization of mating partners (Reichert & Höbel, 2015; Taylor et al., 2011).

It is undeniable that video playbacks are a widely used but difficult to implement methodology for the study of anuran vision (Taylor et al., 2008). One key reason is that the visual system of most anurans is poorly understood. Even so, there is still extensive experimental evidence indicating that the video playback method is effective (Reichert & Höbel, 2015; Rosenthal et al., 2004; Smith & Evans, 2008). Reichert et al., (2014) reported that a nocturnal frog (the grey treefrog, *Hyla versicolor*) was responsive to the visual stimulus of a glowing monitor. During the video playback experiment, we displayed a background video on the monitor as a control to eliminate the bias of possible phototactic response. There was no difference in female responsiveness, female preference strength and the latency to make a choice between the unimodal group (audio-5A vs. audio-4A) and the multimodal group (multi-5A vs. audio-4A), which also suggested that the video playback method had a limited impact on our experimental results.

When females chose between speakers broadcasting identical unimodal calls, they preferred the speaker originally broadcasting multimodal calls. Previous studies showed that recent experience can influence later preference (Schlupp et al., 1994; White & Galef, 1998). Our study suggests that later female choice between unimodal calls of the same males can be influenced by recent exposure to multimodal calls of certain males. It emphasizes the need to consider not only how a multimodal call influences signal attraction but also how this influence is maintained once the multimodal signal ceases. It is interesting that females do not demonstrate a preference for the audio-5A stimuli for silent intervals between 0 and 30 s, even though females prefer audio-5A to audio-4A. One possible explanation is that female serrate-legged small treefrogs may have similar durations of working memory for audio-5A and audio-4A. Our result is consistent with the study of túngara frogs showing that females do not demonstrate a preference for the whine-chuck call for silent intervals between 0 and 30 s, even though females prefer the whine-chuck to the whine (Akre & Ryan, 2010). Indeed, we also find it interesting that females do not demonstrate a preference for the multi-5A over audio-4A when the female is tested directly after the playback (0 s delay interval), even though the p value is close to 0.05 ($p = 0.058$). This is similar to the results of the túngara frog (Akre & Ryan, 2010), suggesting that this is probably not a species-specific phenomenon. We can only speculate as to why this is so. One possible explanation is that the female preference may increase after a delay because that situation is more representative of nature. Another is that the presence of a silent interval influences some aspects of working memory. But this is just speculation; nevertheless, it is intriguing that the same results were reported for a similar study with túngara frogs.

Although the function of multimodal signals in sexual selection has been widely studied (Narins et al., 2003; Taylor & Ryan, 2013), why females usually prefer multimodal signals over unimodal signals and how multimodal signals affect mating preference from the memory perspective is still unknown. Previous studies focused on the function of multimodal signals. Both backup-signal and multi-message hypotheses overly focus on message coding by the sender rather than on the perceptual processes of the receiver (Hebets et al., 2016). Our study is the first to measure the active time of a multimodal signal relative to the receiver's working memory and to reveal its potential role in multimodal signal evolution. Females retained attraction to a multimodal signal after a silent period of up to 30 s, which indicated that the active time of multimodal calls is longer than that for unimodal calls. Since the visual signal by itself is not meaningful to the female, then it could be that the interaction of the visual signal and the auditory signal that contributes to the longer memory. It also could be that the visual cue when presented with the audio becomes salient and it is then the visual cue that influences memory. In other words, our results demonstrate that preference for multimodal signals to unimodal signals might be due to multimodal signals instantiating longer memories than unimodal signals.

Meanwhile, the inflating vocal sac engages working memory processes to a greater degree when combined with the acoustic signal versus the acoustic cue alone, which may explain why many male serrate-legged small treefrogs choose conspicuous positions for calling from where they can display multimodal signals regardless of high predation risk.

Human perception of stimuli in multisensory modalities can positively influence signal detection and memory (Bahrick et al., 2004). Although multimodal communications are widespread both in humans and nonhuman animals, whether multimodal signals affect receiver's memory in nonhuman animals was not yet clear. We demonstrated that the duration of working memory is enhanced by a multimodal signal in frogs. Thus, similar to human beings using multisensory learning to improve memory (Saputra & Nugroho, 2015), male frogs emitted multimodal calls that could enhance the duration of the female frogs' working memory. This ability to influence the duration of working memory may be a general characteristic of multimodal signals, which may help us understand their occurrence and evolution.

The duration of the female's working memory for multimodal signals corresponds to the inter-bout silence of natural choruses. Animal choruses, such as those found in frogs, are often intermittent. Females usually integrate calls to assess and search for potential mates (Schwartz et al., 2004). Females sampling males in the chorus might have to remember the potential mates' calls during periods of silence. Prolonged searching time may increase predation risk of female frogs or even alter female mate choice (Rand et al., 1997; Ryan et al., 1982). We found that females could reduce searching time by maintaining their memory of the multimodal calls through the silent inter-chorus intervals. This finding sheds light on the potential benefits for reproductive females selecting males producing multimodal signals, by potentially decreasing mate search and assessment costs. Therefore, selection should also favour female preference for multimodal signals as longer memory of a more complex male's call should facilitate mate searching and assessment.

In conclusion, our study provides a new perspective on the functional relevance of multimodal sexual signals. We not only provide evidence for an increased active time in female response to multimodal signals but also show the biological relevance of this increase, and the benefit it may provide to reproductive females. The finding that the duration of the working memory of females operating in response to multimodal signals spans the average inter-chorus intervals suggests that multimodal signals could be beneficial for females in that it allows females to recall males even during frequent silent periods between choruses. Our study provides a hypothesis (memory-duration increase hypothesis) as to why females prefer multimodal sexual signals.

ACKNOWLEDGEMENTS

We thank the Mt. Diaoluo National Nature Reserve Administration for the permission and support to perform this experiment. This work was supported by National Natural Science Foundation of China (31772464, 31572275, 32001095), Youth Innovation Promotion Association CAS (2012274), CAS 'Light of West China' Program (2016XBZG_XBQNXZ_A1_003, 2017XBZG_XBQNXZ_B_013). The authors declare no competing financial interests.

AUTHORS' CONTRIBUTIONS

B.Z., J.C. and Y.T. conceived the study, designed the experiments, analysed the data and wrote the paper; B.Z. and Y.Z. collected the field data and carried out the female choice experiments; Y.Y., K.D., T.W. and J.W. advised on experiment and discussed the results; J.C. and M.J.R. revised the paper. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

Data available from the Dryad Digital Repository https://doi.org/10.5061/dryad.ksn02v73p (Zhu et al., 2021).

ORCID

Bicheng Zhu https://orcid.org/0000-0003-2803-8206
Yue Yang https://orcid.org/0000-0002-6007-7937
Ke Deng https://orcid.org/0000-0003-2500-6794
Tongliang Wang https://orcid.org/0000-0002-6035-9969
Jichao Wang https://orcid.org/0000-0001-7177-2911
Yezhong Tang https://orcid.org/0000-0002-6302-9308
Michael J. Ryan https://orcid.org/0000-0002-6381-9545
Jianguo Cui https://orcid.org/0000-0001-8746-2803

REFERENCES

Akre, K. L., & Ryan, M. J. (2010). Complexity increases working memory for mating signals. *Current Biology*, *20*, 502–505. https://doi.org/10.1016/j.cub.2010.01.021

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*, 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Beaupre, S. J., Jacobson, E. R., Lillywhite, H. B., & Zamudio, K. (2004). *Guidelines for the use of live amphibians and reptiles in field and laboratory research*. Retrieved from http://www.asih.org/sites/default/files/documents/resources/guidelines

Bizon, J. L., Foster, T. C., Alexander, G. E., & Glisky, E. L. (2012). Characterizing cognitive aging of working memory and executive function in animal models. *Frontiers in Aging Neuroscience*, *4*, 1–14. https://doi.org/10.3389/fnagi.2012.00019

Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *10*, 341–345.

Candolin, U. (2003). The use of multiple cues in mate choice. *Biological Reviews*, *78*, 575–595. https://doi.org/10.1017/S1464793103006158

Croxson, P. L., Kyriazis, D. A., & Baxter, M. G. (2012). Cholinergic modulation of a specific memory function of prefrontal cortex. *Nature Neuroscience*, *14*(12), 1510–1512. https://doi.org/10.1038/nn.2971

Cummings, M. E., Bernal, X. E., Reynaga, R., Rand, A. S., & Ryan, M. J. (2008). Visual sensitivity to a conspicuous male cue varies by reproductive state in *Physalaemus pustulosus* females. *Journal of Experimental Biology*, *211*, 1203–1210. https://doi.org/10.1242/jeb.012963

Deng, K., Zhu, B., Zhou, Y., Chen, Q., Wang, T., Wang, J., & Cui, J. (2019). Mate choice decisions of female serrate-legged small treefrogs are affected by ambient light under natural, but not enhanced artificial nocturnal light conditions. *Behavioural Processes*, *169*, 103997. https://doi.org/10.1016/j.beproc.2019.103997

Gerhardt, H. C., & Huber, F. (2002). *Acoustic communication in insects and anurans*. University of Chicago Press.

Gomes, D. G., Page, R. A., Geipel, I., Taylor, R. C., Ryan, M. J., & Halfwerk, W. (2016). Bats perceptually weight prey cues across sensory systems when hunting in noise. *Science*, *353*, 1277. https://doi.org/10.1126/science.aaf7934

Halfwerk, W., Jones, P. L., Taylor, R. C., Ryan, M. J., & Page, R. A. (2014). Risky ripples allow bats and frogs to eavesdrop on a multisensory sexual display. *Science*, *343*, 413–416. https://doi.org/10.1126/science.1244812

Halfwerk, W., Page, R., Taylor, R., Wilson, P., & Ryan, M. (2014). Cross-modal comparisons of signal components allow for relative-distance assessment. *Current Biology*, *24*, 1751–1755. https://doi.org/10.1016/j.cub.2014.05.068

Halfwerk, W., Varkevisser, J., Simon, R., Mendoza, E., Scharff, C., & Riebel, K. (2019). Toward testing for multimodal perception of mating signals. *Frontiers in Ecology and Evolution*, *7*, 124. https://doi.org/10.3389/fevo.2019.00124

Hebets, E. A., Barron, A. B., Balakrishnan, C. N., Hauber, M. E., Mason, P. H., & Hoke, K. L. (2016). A systems approach to animal communication. *Proceedings of Royal Society B-Biology Sciences*, *283*, 20152889. https://doi.org/10.1098/rspb.2015.2889

Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: Developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology*, *57*, 197–214. https://doi.org/10.1007/s00265-004-0865-7

Kesner, R. (2015). All sensory modalities activate different subregions of the hippocampus to activate a number of memory-based processes. *Acta Physiologica*, *213*, 21–22.

Laird, K. L., Clements, P., Hunter, K. L., & Taylor, R. C. (2016). Multimodal signaling improves mating success in the green treefrog (*Hyla cinerea*), but may not help small males. *Behavioral Ecology and Sociobiology*, *70*, 1517–1525. https://doi.org/10.1007/s00265-016-2160-9

Leslie, C. E., Rosencrans, R. F., Walkowski, W., Gordon, W. C., Bazan, N. G., Ryan, M. J., & Farris, H. (2020). Reproductive state modulates retinal sensitivity to light in female túngara frogs. *Frontiers in Behavioral Neuroscience*, *13*, 293. https://doi.org/10.3389/fnbeh.2019.00293

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1038/264746a0

Moller, A., & Pomiankowski, A. (1993). Why have birds got multiple sexual ornaments? *Behavioral Ecology and Sociobiology*, *32*, 167–176. https://doi.org/10.1007/BF00173774

Narins, P. M., Hödl, W., & Grabul, D. S. (2003). Bimodal signal requisite for agonistic behavior in a dart-poison frog, *Epipedobates femoralis*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 577–580. https://doi.org/10.1073/pnas.0237165100

Partan, S. R. (2013). Ten unanswered questions in multimodal communication. *Behavioral Ecology and Sociobiology*, *67*, 1523–1539. https://doi.org/10.1007/s00265-013-1565-y

Pauly, G. B., Bernal, X. E., Rand, A. S., & Ryan, M. J. (2006). The vocal sac increases call rate in the Túngara frog *Physalaemus pustulosus*. *Physiological and Biochemical Zoology*, *79*, 708–719. https://doi.org/10.1086/504613

Rand, A. S., Bridarolli, M. E., Dries, L., & Ryan, M. J. (1997). Light levels influence female choice in túngara frogs: Predation risk assessment? *Copeia*, *1997*, 447–450. https://doi.org/10.2307/1447770

Reichert, M. S., Galante, H., & Höbel, G. (2014). Female gray treefrogs, *Hyla versicolor*, are responsive to visual stimuli but unselective of stimulus characteristics. *Journal of Experimental Biology*, *217*(18), 3254–3262. https://doi.org/10.1242/jeb.106666

Reichert, M. S., & Höbel, G. (2015). Modality interactions alter the shape of acoustic mate preference functions in gray treefrogs. *Evolution*, *69*(9), 2384–2398. https://doi.org/10.1111/evo.12750

Rodriguez, J. S., & Paule, M. G. (2009). Working memory delayed response tasks in monkeys. In J. J. Buccafusco (Ed.), *Methods of behavior analysis in neuroscience* (pp. 247–264). CRC Press. https://doi.org/10.1201/NOE1420052343.ch12

Rosencrans, R. F., Leslie, C. E., Perkins, K. A., Walkowski, W., Gordon, W. C., Richards-Zawacki, C. L., Bazan, N. G., & Farris, H. E. (2018). Quantifying the relationship between optical anatomy and retinal physiological sensitivity: A comparative approach. *Journal of Comparative Neurology*, *526*(18), 3045–3057. https://doi.org/10.1002/cne.24531

Rosenthal, G. G. (1999). Using video playback to study sexual communication. *Environmental Biology Fishes*, *56*(3), 307–316. https://doi.org/10.1023/A:1007597604793

Rosenthal, G. G., Rand, A. S., & Ryan, M. J. (2004). The vocal sac as a visual cue in anuran communication: An experimental analysis using video playback. *Animal Behaviour*, *68*, 55–58. https://doi.org/10.1016/j.anbehav.2003.07.013

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, *58*, 921–931. https://doi.org/10.1006/anbe.1999.1242

Ryan, M. J., Tuttle, M. D., & Rand, A. S. (1982). Bat predation and sexual advertisement in a neotropical anuran. *The American Naturalist*, *119*, 136–139. https://doi.org/10.1086/283899

Saputra, M. R. U., & Nugroho, K. A. (2015). *'Learn-to-read' application for remediation of dyslexic children based on multisensory approach*. 4th International Conference on Instrumentation, Communication, Information Technology and Biomedical Engineering (ICICI-BME). IEEE. https://doi.org/10.1109/ICICI-BME.2015.7401366

Schlupp, I., Marler, C., & Ryan, M. J. (1994). Benefit to male sailfin mollies of mating with heterospecific females. *Science*, *263*, 373–374. https://doi.org/10.1126/science.8278809

Schwartz, J. J., Huth, K., & Hutchin, T. (2004). How long do females really listen? Assessment time for female mate choice in the grey treefrog, *Hyla versicolor*. *Animal Behaviour*, *68*, 533–540. https://doi.org/10.1016/j.anbehav.2003.09.016

Smith, C. L., & Evans, C. S. (2008). Multimodal signaling in fowl, *Gallus gallus*. *Journal of Experimental Biology*, *211*, 2052–2057. https://doi.org/10.1242/jeb.017194

Starnberger, I., Poth, D., Peram, P. S., Schulz, S., Vences, M., Knudsen, J., Barej, M. F., Rödel, M., Walzl, M., & Hödl, W. (2013). Take time to smell the frogs: Vocal sac glands of reed frogs (Anura: Hyperoliidae) contain species-specific chemical cocktails. *Biological Journal of Linnean Society*, *110*, 828–838. https://doi.org/10.1111/bij.12167

Starnberger, I., Preininger, D., & Hödl, W. (2014). From uni- to multimodality: Towards an integrative view on anuran communication. *Journal of Comparative Physiology A*, *200*(9), 777–787. https://doi.org/10.1007/s00359-014-0923-1

Taylor, R. C., Klein, B. A., & Ryan, M. J. (2011). Inter-signal interaction and uncertain information in anuran multimodal signals. *Current Zoology*, *57*, 153–161. https://doi.org/10.1093/czoolo/57.2.153

Taylor, R. C., Klein, B. A., Stein, J., & Ryan, M. J. (2008). Faux frogs: Multimodal signalling and the value of robotics in animal behaviour. *Animal Behaviour*, *76*, 1089–1097. https://doi.org/10.1016/j.anbehav.2008.01.031

Taylor, R. C., & Ryan, M. J. (2013). Interactions of multisensory components perceptually rescue túngara frog mating signals. *Science*, *341*, 273–274. https://doi.org/10.1126/science.1237113

Uetz, G. W., & Roberts, J. A. (2002). Multisensory cues and multimodal communication in spiders: Insights from video/audio playback studies. *Brain Behavior and Evolution*, *59*, 222–230. https://doi.org/10.1159/000064909

Underhill, V. A., & Höbel, G. (2017). Variation in nocturnal light levels does not alter mate choice behavior in female eastern gray treefrogs (*Hyla versicolor*). *Behavioral Ecology and Sociobiology*, *71*, 151. https://doi.org/10.1007/s00265-017-2386-1

White, D. J., & Galef, B. G. (1998). Mate-choice copying in Japanese quail, *Coturnix coturnix japonica*. *Animal Behaviour*, *55*, 545–552. https://doi.org/10.1006/anbe.1997.0616

Zhu, B., Wang, J., Brauth, S. E., Tang, Y., & Cui, J. (2017). The spectral structure of vocalizations match hearing sensitivity but imprecisely in *Kurixalus odontotarsus*. *Bioacoustics*, *26*, 121–134. https://doi.org/10.1080/09524622.2016.1221778

Zhu, B., Wang, J., Zhao, L., Chen, Q., Sun, Z., Yang, Y., Brauth, S. E., Tang, Y., & Cui, J. (2017). Male-male competition and female choice are differentially affected by male call acoustics in the serrate-legged small treefrog, *Kurixalus odontotarsus*. *PEER J*, *5*, e3980. https://doi.org/10.7717/peerj.3980

Zhu, B., Wang, J., Zhao, L., Sun, Z., Brauth, S. E., Tang, Y., & Cui, J. (2016). Bigger is not always better: Females prefer males of mean body size in *Kurixalus odontotarsus*. *PLoS ONE*, *11*, e0149879. https://doi.org/10.1371/journal.pone.0149879

Zhu, B., Zhou, Y., Yang, Y., Deng, K., Wang, T., Wang, J., Tang, Y., Ryan, M. J., & Cui, J. (2021). Data from: Multisensory modalities increase working memory for mating signals in a treefrog. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.ksn02v73p

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zhu B, Zhou Y, Yang Y, et al. Multisensory modalities increase working memory for mating signals in a treefrog. *J Anim Ecol*. 2021;90:1455–1465. https://doi.org/10.1111/1365-2656.13465