Male within-individual variability in a sexual signal component and its impact on female choice

Thierry Lengagne, a Yann Voituron, a and Doris Gomez b
aUMR5023 Ecologie des Hydrosystèmes Naturels et Anthropisés, Université de Lyon; ENTPE, CNRS, Université Lyon 1, 6 rue Raphaël Dubois, 69622 Villeurbanne, France and bCentre of Evolutionary and Functional Ecology, UMR CNRS 5175, 1919 route de Mende, 34090 Montpellier, France

Received 21 December 2015; revised 24 June 2016; accepted 28 June 2016.

A growing body of literature deals with the variability and repeatability of signal components over long time periods. Surprisingly, variations of signal quality over minutes or hours are less studied although most of the behavioral decisions of the receiver are taken at this time scale. Despite potentially strong implications on theoretical developments linked to sexual selection and communication, within-individual variability in a signal component and its possible consequence on accuracy of female choice has never been thoroughly investigated. Focusing on call dominant frequency (DF) in the European tree frog, Hyla arborea, we showed that frequency variability is due to a warm-up effect on the beginning of the call sequence but not to an exhaustion effect at the end of the sequence. Nevertheless, the great majority of male within-variability at the night scale is due to sudden discontinuities with independent temporal patterns from one individual to another. Secondly, we simulated female mate choice decisions with simple rules based on DF. Within-individual variability in DF the proportion of beneficial choice decreases up to 30% in the worst case. In addition, to overcome these temporary variations in male signal, we emphasize a weak advantage supplied by increasing sampling duration. The costs of being selective are assumed to increase with time sampling. We suggest that females may benefit from assessing several signal components simultaneously in short samples instead of disproportionately increasing sampling duration.

Key words: anurans, call dominant frequency, mate choice, multicomponent signal, sampling, sexual selection, simulation, static signal component, within-individual variability.

INTRODUCTION

Phenotypic variability is the basis of all selection processes (natural, artificial, and sexual; Darwin 1859). However, while between-individual variability raises the potential for selection, within-individual variability may impair it (Boake 1989). Such a relation between and within-individual level is an emerging question especially regarding the understanding of the behavioral processes and their evolution (Digemanse and Dochtermann 2013). In a sexual selection context, females often select their mates on the basis of signal components (Andersson 1994). These signal components may have evolved to transmit honest information about male condition, to exploit a sensorial bias or for signal efficiency during both propagation and discrimination task (Zahavi 1977; Guilford and Dawkins 1991; Endler and Basolo 1998).

A growing body of literature deals with the variability and repeatability of signal components over long time periods (e.g., Hoefler et al. 2009; Svensson et al. 2009; Deb et al. 2012). The context of communication is more likely to change over long time periods, and signal variability over long time periods may respond to these context changes. Indeed, several signal components present in males are plastic in regards to environmental changes and females are able to adjust their preferences accordingly (e.g., social context: Martinez-Rivera and Gerhardt 2008; physical environment: Gerhardt 1978; time during the breeding season: Griffith and Sheldon 2001, Sisneros et al. 2004). By contrast, variations of male signal quality over shorter time periods (minutes or hours) are less studied. Over such time periods, variations in emitted signals may correspond to varying signaling strategies: males can temporarily increase the attractiveness of some signal components or increase the overall amount of signals produced (Castellano 2009; Brepson et al. 2013). Such variations over relatively short time periods may help them to maintain their attractiveness for females. For instance,
in a context of sexual conflict between males and females, males can invest in costly signal components when in presence of females; reciprocally, females can overcome these temporary variations in male attractiveness by increasing the time invested in assessment of male quality. Despite potentially strong implications on theoretical developments linked to sexual selection and communication, such kind of within-individual variability in a signal component and its possible consequence on accuracy of female choice has been poorly investigated.

Anurans have been a well-studied model group in the field of sexual acoustic communication during the last decades (Bee et al. 2015); they have notably allowed some theoretical progress concerning the variability of signal components and their potential for selection (Gerhardt 1991). Call dominant frequency (DF) is one of the acoustic signal components most studied in anurans and insects (Gerhardt and Huber 2002). In most of the species studied to date, DF is considered as among the most static signal components. Yet, a few observations suggest that signal variation is greater when analyzed from calls recorded over a long rather than a short period (Gerhardt 1991; Castellano and Giacoma 1998; Howard and Young 1998). In the European tree frog, Hyla arborea, recent experimental evidence reveals that males show important within-individual variability in DF at the scale of an entire night (Lengagne et al. 2015). Indeed, this study emphasizes an average 116 Hz of call DF variability and one male exhibit a frequency differences as high as 470 Hz during the course of the night. In this species, females have been shown experimentally to prefer calls with low DF over call with a DF close to the population mean (Richardson et al. 2010). Several non-mutually exclusive reasons might explain why females preferentially select calls with low DF in this species. By doing so, females may select the largest males (Marquez and Tejedo 1990; Friedl and Klump 2002) and potentially obtain subsequent indirect benefits from pairing with these males (Voituron et al. 2012). Females may also avoid mating with males of closely related species (Castellano et al. 2002). Such preference could also be due to a preexisting sensory bias for low frequency calls potentially leading to indirect benefits for females (Ryan et al. 1990). Hence, in the present study, we focused on acoustic signals emitted by H. arborea to investigate the existence of signal component variability and its consequences on mate choice accuracy and mate choice strategy.

In this study, 1) we measured the temporal variability of frequency in males at the scale of an entire night to determine the temporal structuration, if any, of male within-variability in DF and assess the potential links between frequency variability and morphological parameters; 2) by using simulation analyses, we assessed the potential impact of this variability on the accuracy of female choice; 3) by exploring various mate quality assessment rules, we assessed to what extent females could counteract male signal variability and overcome their potential inaccuracy.

**MATERIALS AND METHODS**

**Capture and housing conditions**

Individuals were obtained from a population located near Lyon, France, on the Crémieu Plateau (5°21'77"E, 45°44'20"N). During the breeding season, 36 males were captured individually by hand when calling at the pond (3 capture sessions along the breeding season, 12 males captured per session). They were then transported in individual dark boxes to the housing room of the laboratory. Males were then immediately placed in individual terraria (25 × 17 × 15 cm) with a water-filled basin and a tree branch. Males were fed ad libitum with domestic crickets (Acheta domestica) for 3 to 4 days before the experiment for acclimation. Males were exposed to natural photoperiod and chorus noise was broadcast each night from 08:00 PM to 02:00 AM. Temperature was kept at 22 ± 1 °C. After acclimation, males were recorded individually during a whole night. After the recording, males were weighed and measured (femur-tibia length) and, for each individual, a scaled mass index (SMI, size-independent condition indicator) was computed according to the methodology proposed by Peig and Green (2010). All males were released at the pond the night after recording.

**Recording**

We recorded a total of 36 males, in 3 independent sessions during the breeding season. In each session, 12 males were thus simultaneously recorded during one night (Lengagne et al. 2015). However, due to a technical issue, 4 males of the first session and 2 males of the second session were recorded the day after the other males of the same session. Each terrarium was placed in an individual semi-anechoic chamber the day of the recording at 08:00 PM. Terraria had plastic mesh sides and lid instead of glass sides in order to avoid sound reflectance. Calls were recorded with a Sony ECM-T6® microphone placed at the entrance of each chamber 30 cm from the calling male and plugged in a Roland R-44® four-channel recorder or a Tascam US 144® soundcard connected to a notebook computer. To ensure that a change in body posture did not result in a shift in the dominant frequencies, we conducted a pilot study. Three frogs were recorded in the field with a 4-channel Roland R44 recorder and we showed that the DF did not change with the position of the microphone (front, back, left and right side). Therefore the position of the animals did not impact the values of DF (see the Supplementary Information S1 for details).

Each semi-anechoic chamber was open on the front side to avoid stressing males. Males were stimulated by broadcasting during the night a repeated 2-min chorus noise, via an amplified loudspeaker (KH pas-1000®) connected to a CD player, placed in front of the chambers. This stimulating recording was an artificial chorus noise, constituted of calls from several isolated males H. arborea, in which no male was louder than any other to avoid the risk of the tested male responding to the calls of a particular competitor (Richardson et al. 2008). Although recording chambers were open on their front side, they were not placed face to face. As a consequence, sounds perceived by a focal male from the other adjacent males were very weak compared to the volume of the stimulating records that were broadcasted. The 36 males emitted on average 14 546 ± 8644 calls (mean ± standard deviation [SD]; range: 561–31 641 calls) grouped in 672 ± 383 call bouts (mean ± SD; range: 32–1317 bouts) during 2.96 ± 1.07 h from the first to the last call (mean ± SD; range: 0.22–4.23 h). Details on procedure can be found in Lengagne et al. (2015).

**Measure of DF variability**

Males produce advertisement calls in series named bouts containing on average 25 calls and lasting 4 s on average (see the Supplementary Information S2 for details on acoustic signal emitted by European tree frog). We extracted DF for each call from the entire night records, using Avisoft Saslab® software. DF was here defined as the frequency with the highest energy, in a frequency
range between 1.8 and 3kHz. Mean power spectra were automatically computed for each call with a Fast Fourier Transform of 256 points and we computed DF values from these spectra. We analyzed these spectra with SONIO, a custom program (http://sites.google.com/site/sonioprogram/ ©Doris Gomez). For each male, we merged all Avisoft output files for one male over one night into one single output file. We then computed a “portrait” file, which compiled for the entire file obtained for a male during one night the mean, the SD and the coefficient of variation (CV) of DF, and, for each call, the number of pulses in the call, along with its start time and end time. We further analyzed the potential temporal structuration in male DF patterns (consequences of a warm-up or an exhaustion of the structures involved in call production concerning DF variability, see the statistics paragraph).

### Female choice simulations

With the 36 recorded males, we artificially built all possible 630 male pairs (40h of recordings from the recording archives of the Madrid Museum (see www.fonozoo.com)). We then simulated female mate choice between 2 potential mates using SONIO. This program allowed to simulate the choice of an “artificial” female: it built all possible pairs between males, and for each pair, it used male portrait files to simulate female choice. Female choice followed 2 steps: the program simulated the mate choice that an artificial female would make based on windows of male calls that she would attend. We made the assumption that the female is averaging over a given number of calls irrespective of either the total sampling time or how the calls are distributed within this sampling window.

Regarding the decision process, we considered that it followed 2 simple rules. 1) DF discrimination was limited by a fixed threshold. Above the threshold, males were perceived as different, while below the threshold, males were perceived as displaying the same DF. 2) If females could detect a difference in mean DF, they systematically preferred the lowest detected mean DF without error. These rules stemmed from relevant biological characteristics.

Firstly, results of mate choice of *Hyla versicolor*’s advertisement call observed by Gerhardt and Watson (1995) suggest that females can resolve small differences in mean values between the calls of different males despite considerable within-male variability (i.e., females discriminate call sequences differing by 5.4% in their pulse number). Secondly, in a 2-choice experimental design, females of *H. arborea* have been shown to be able to discriminate between calls differing in DF by 190 Hz, preferring calls with the lowest DF (Richardson et al. 2010). To take into account that potential differences between females and/or differences between situations could impact the discrimination threshold we considered 3 threshold values: 100, 200, and 400Hz and evaluated their impact on the output of female mate choice. Regarding male call sampling, the artificial female was allowed to sample male calls using different sample sizes (10, 20, 50, 100, 200, 500, 1000, 2000, 5000, and 10 000 consecutive calls corresponding to an approximate time window ranging (10, 20, 50, 100, 200, 500, 1000, 2000, 5000, and 10 000 consecutive calls) to the reference choice based on the sampling of the entire night records for both males. We considered it a beneficial choice if it targeted the same male as the reference choice. If males of the pair did not differ for more than the threshold on the entire night, the choice was considered random by default with 50% of beneficial choice. Each case (pair, discrimination threshold and window size, expressed in number of calls) was simulated 100 times (1 890 000 simulations in total), which provided a proportion of beneficial choices in each situation.

Here, we used a best-of-n (with n = 2) choice with a sequential sampling in absence of background noise and inter-female variability in decision rules. These conditions are a simplification of reality as females can assess more than 2 males simultaneously (Murphy and Gerhardt 2002; Richardson and Lengagne 2010), background noise limits the efficiency of communication (Richardson and Lengagne 2010), and females differ in many aspects related to mate choice like hormonal state (Lynch et al. 2006) or condition (Jennions et al. 1995). Yet, considering a simple evolutionary scenario is a crucial first step to assess the selective forces potentially acting on female choice.

### Statistical analyses

We first explored male DF patterns and first hypothesized that variability in DF would be greater within call bouts at the beginning and/or at the end of the night, due respectively to a warm-up or an exhaustion of the structures involved in call production (i.e., trunk muscle, laryngeal apparatus and/or vocal sac). To test the warm-up hypothesis, we first computed the CV of DF for each call bout. We then compared intra-bout variability of the first call bouts (one by one) to the mean intra-bout variability computed on the subsequent call bouts (except those thought to be involved in the exhaustion effect). Likewise, to test the exhaustion effect we compared intra-bout variability of the last call bouts (one by one) to the mean intra-bout variability computed on the previous call bouts (except those thought to be involved in the warm-up effect). These tests were performed using paired t-tests for each male separately. Tests were corrected using Bonferroni’s corrections for multiple comparisons.

Correlations between the CV computed on the entire night and morphological characteristics were investigated using linear models.

We then tested for the existence of temporal structuration in DF patterns during the course of the night. For this purpose, we used statistical methods designed for multivariate time series, using the MARSS package (Holmes et al. 2013). Such analyses are particularly suited 1) to include non-independent data points (DF values are correlated from time t to time t+1), 2) to detect temporal structuration in time series (males may increase or decrease in DF over time), and 3) and to reveal correlations between time series (the variations shown by a male may depend on the variations shown by other males measured simultaneously). MARSS models consider that observations—DF values computed from recordings—express, with variations, an underlying hidden process (also called hidden trajectory). In the model, we need to specify the error associated to observations and given all males were measured with the same protocol, we assumed observation error was equal for all males. Male hidden processes or trajectories show patterns of temporal autocorrelation, trends, and process errors. The pattern of temporal autocorrelation describes how DF values at time t are related to DF values at time t+1; males can share the same pattern of temporal autocorrelation or
have their specific pattern of temporal autocorrelation. The trend describes the general increase or decrease with time of DF values. The process error describes the unexplained error in male trajectories—males can share the same error or have their specific error—and describes whether errors are correlated between males or if they are independent from one male to the other. In practice, we first homogenized the structure of DF time series to get comparable time points for all males; we computed the mean DF value for each minute spent calling (from minute 1, the beginning of the record, to minute 264) and a “NX” for each minute without any call for a given male. Second, we considered all records performed during the same night as a multivariate time series. We then computed MARSS analyses and estimated parameters using expectation-maximization algorithm and Kalman filter/smoother, which are particularly robust tools for such analyses. We used Akaike Information Criterion (AIC) to select the model with best fit. We computed 3 types of models: 1) models where males from the same session shared the same temporal autocorrelation, trend and process error; 2) models where males from the same session shared the same trend and process error, but have different temporal autocorrelation, and 3) models where males from the same session had different temporal autocorrelation, trend and process error. We tested whether process errors were correlated or not between males. If males did not share any parameter (temporal autocorrelation, trend, uncorrelated error), we considered that DF values varied erratically in males.

Regarding the simulation of female mate choice, we analyzed the probability of choosing the male with lowest DF using generalized linear mixed models with a binomial error and males as random factors. As fixed effects, we tested the discrimination threshold, the number of calls sampled, the difference in DF between males, the variability in DF and all second order interactions between these factors. To characterize the variability in DF shown by the pair, we computed the sum of the CVs of the 2 males. We used a maximum likelihood approach with a Laplace approximation and selected the best model by minimizing the Akaike’s Information Criterion (Bolker et al. 2009). These statistical analyses were performed using R® software v.2.15.1 (R Development Core Team 2012) and the packages “lme4” (Bates et al. 2011) for generalized linear mixed models and “nlme” (Pinheiro et al. 2009) for linear mixed models and males as random effects. To characterize variability in DF shown by the pair, we computed the sum of the CVs of the 2 males. We used a maximum likelihood approach with a Laplace approximation and selected the best model by minimizing the Akaike’s Information Criterion (Bolker et al. 2009). These statistical analyses were performed using R® software v.2.15.1 (R Development Core Team 2012) and the packages “lme4” (Bates et al. 2011) for generalized linear mixed models and “nlme” (Pinheiro et al. 2009) for linear mixed models.

RESULTS

Patterns of DF variability

Mean DF was on average 2324 ± 141 Hz (mean ± SD; range: 2250–2803 Hz) and was negatively correlated with body mass ($r = -0.45$, $F_{1,34} = 8.70$, N = 36, $P = 0.0057$). Within-individual CVs computed on DF measures from the entire night were 4.57 ± 2% (Lengagne et al. 2015). Individual CV of DF decreased exponentially when the number of calls emitted during the night increased (Figure 1, $r = -0.54$, $F_{1,34} = 13.82$, $N = 36$, $P = 0.0007$). Nevertheless, there was no relationship between this CV and caller morphology, namely mass ($r = 0.07$, $F_{1,31} = 1.70$, $N = 36$, $P = 0.2$), length ($r = 0.04$, $F_{1,31} = 1.03$, $N = 36$, $P = 0.305$) and SMI ($r = 0.00$, $F_{1,34} = 0.006$, $N = 36$, $P = 0.941$).

We observed that the first 3 call bouts were significantly more variable in DF than the subsequent ones (paired t-test on the 36 males, $P < 0.01$ for the first 3 bouts, $P > 0.05$ for the following bouts) whereas the last call bouts were not significantly more variable than the previous ones (Figure 2). Hence, we observed a warm-up effect but no exhaustion effect on DF variability.

Most of the nightly within-individual variability in DF seemed to occur at large temporal scale with some sudden discontinuities (Figure 3 shows 3 contrasted cases among the 36 males, see the Supplementary Information S3 for all the males). These discontinuities could lead to great changes in DF during the course of the night. Indeed, if we compared the minute with the lowest pitch and the minute with the highest pitch for each male, the difference was on average of 444.2 ± 144.1 Hz (mean ± SD) and could reach 817 Hz in case of pronounced discontinuities.

When analyzing records made during the same night as multivariate time series, patterns of variation in DF were better described with different temporal autocorrelation, trend and process error for each male (Table 1). This means that DF showed no general trend to increase or decrease during the course of the night and that the time at which discontinuities occurred, their direction and their magnitude were independent from one individual to another for males recorded simultaneously. In other words, males recorded

**Figure 1**

Decrease of the CV of dominant frequency computed over the entire night with the number of calls emitted by the male during the night. Each point represents a male ($r = -0.54$, $F_{1,34} = 13.82$, $N = 36$, $P = 0.0007$).

**Figure 2**

(a) Warm-up effect shown through the mean difference (and confidence interval at 95%) between intra-bout variability in DF measured on the focal call bout and mean intra-bout variability computed on the subsequent call bouts. The first 3 call bouts are significantly more variable in DF than the others. (b) Exhaustion effects, the last call bouts are not significantly more variable in DF than the previous ones.
during the same night behaved independently one from another. We thus ran a model gathering all 36 males (and not only the males recorded during the same night) as this increased our power to detect potential patterns common to males in temporal autocorrelation, trend or process error. Again, patterns of variation in DF were better described with different temporal autocorrelation, trend and process error for each of the 36 males (Table 1). Hence, all males behaved independently one from another. It is important to underline that among all computed trend values, none was significantly different from 0. Hence, no male significantly increases or decreases DF during the course of the night.

**Female choice simulations**

The model that best explained the proportion of beneficial choices is presented in Table 2. The discrimination threshold appeared to be the major determinant of the overall proportion of beneficial choices; a lower threshold led to a higher proportion of overall beneficial choices, which is an intuitive result (Table 2, Figure 4).

For a given threshold, the proportion of beneficial choices was highly affected by the difference between the mean DF values of the males; a greater difference led to a higher proportion of beneficial choices (Table 2, Figure 4). When the variability around the mean DF values increased, the proportion of beneficial choices decreased. The interaction between the difference and the variability in DF was negative; hence the negative impact of the variability on the proportion of beneficial choices was enhanced when the difference increased. A large variability of about 20% in DF (sum of the 2 CVs) could decrease the proportion of beneficial choices by up to 30% in the worst case (Figure 4c). On the other hand, when the size of the sample assessed by the female increased, the proportion of beneficial choices increased. The interaction between the difference and the sample size was positive; hence the positive impact of the sample size on the proportion of beneficial choices was enhanced when the difference increased. By increasing the number of calls sampled by the female from 10 to 10 000 calls, the proportion of beneficial choices was enhanced by up to 10% in the best case (Figure 4d). Although the threshold interacted with the other 3 factors, theses predicted effects of sample size and DF variability were almost the same for the 3 tested discrimination thresholds (Figure 4). Distributions of the variability around the mean DF and the difference between the mean DF values observed

![Figure 3](image)

**Figure 3**

Variations in DF during the course of the night (each circle represents the mean DF for a call bout) for 3 contrasted cases: the male on the top had a relatively constant DF whereas the male on the back had a highly variable DF at night scale. The male in the center was an intermediate case. Solid and dashed lines represent respectively the fitted values (state estimates) of the best multivariate time series model (computed over the 36 males’ records) and their 95% confidence intervals.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Corrected Akaike Information Criteria (AICc) corresponding to each multi-time series auto-regressive model</strong></td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Observation error variance</td>
</tr>
<tr>
<td>Trend/ process variance</td>
</tr>
<tr>
<td>Hidden state trajectory</td>
</tr>
<tr>
<td>Session 1 (N = 8)</td>
</tr>
<tr>
<td>Session 2 (N = 10)</td>
</tr>
<tr>
<td>Session 3 (N = 12)</td>
</tr>
<tr>
<td>All sessions (N = 36)</td>
</tr>
</tbody>
</table>

Models with different hidden state trajectories, trends and process variances for each individual are systematically better supported by the data.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimated parameters (estimate ± SE) of the fixed effects predicting the proportion of beneficial choice</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Wald z test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-5.99E+00±3.41E-01$ *</td>
</tr>
<tr>
<td>Sample</td>
<td>$-5.21E-04±6.02E-05$ *</td>
</tr>
<tr>
<td>Difference in DF</td>
<td>$4.87E-02±2.31E-03$ *</td>
</tr>
<tr>
<td>Variability</td>
<td>$2.45E-01±3.17E-02$ *</td>
</tr>
<tr>
<td>Threshold 200</td>
<td>$-5.87E+00±3.87E-01$ *</td>
</tr>
<tr>
<td>Threshold 400</td>
<td>$-2.44E+01±2.02E+00$ *</td>
</tr>
<tr>
<td>Difference in DF × Variability</td>
<td>$-2.23E-03±1.99E-04$ *</td>
</tr>
<tr>
<td>Difference in DF × Sample</td>
<td>$5.35E-06±4.98E-07$ *</td>
</tr>
<tr>
<td>Sample × Threshold 200</td>
<td>$-5.08E+04±5.65E-05$ *</td>
</tr>
<tr>
<td>Sample × Threshold 400</td>
<td>$-1.61E-03±1.60E-04$ *</td>
</tr>
<tr>
<td>Difference × Threshold 200</td>
<td>$5.14E-03±3.32E-03$ *</td>
</tr>
<tr>
<td>Difference in DF × Threshold 400</td>
<td>$2.25E-02±4.47E-03$ *</td>
</tr>
<tr>
<td>Variability × Threshold 200</td>
<td>$1.76E-01±2.03E-02$ *</td>
</tr>
<tr>
<td>Variability × Threshold 400</td>
<td>$6.73E-01±7.28E-02$ *</td>
</tr>
</tbody>
</table>

Difference in DF is expressed in Hz, Variability is expressed in % and Sample is expressed in number of calls. Discrimination threshold is considered as a factor with 3 levels, the reference level is 100 Hz. SE, standard error.

*P < 0.001.
for the 630 tested male pairs are given in Figure 5. From all possible pairs, only 7.6% differ by more than 350 Hz, 25% differ by 200 to 350 Hz, and 67.4% differ by less than 200 Hz in their mean DF value (Figure 5). The variability around the mean DF value is between 5% and 15% for 94.5% of all pairs (Figure 5).

**DISCUSSION**

**Patterns of variability in DF**

As previously observed (Marquez and Tejedo 1990; Castellano et al. 2002; Friedl and Klump 2002; Richardson et al. 2010), our study provides evidence that mean DF is negatively correlated with male body mass in *H. arborea*, even if this relationship is relatively weak. Our results also confirm that within-individual variability in DF is greater when measured over a long time period than when assessed over a short time period, and are thus in accordance with previous results on *H. arborea* (Castellano et al. 2002; Friedl and Klump 2002; Lengagne et al. 2015) and other anuran species (e.g., Gerhardt 1991; Castellano and Giacoma 1998; Howard and Young 1998). Given the patterns of variability observed in this study, we can explain the low variability generally observed on short time periods. We observed a warm-up effect but no exhaustion effect on DF variability. However, the warm-up effect concerns only the first 3 call bouts (about 0.6% of the bouts emitted) and

---

**Figure 4**

Predicted values of the proportion of beneficial choices (a, c and e) depending on the difference between mean DF of the 2 males (“Difference”) and the variability around these mean DF (“Variability”); (b, d and f) depending on the difference between mean DF of the 2 males (“Difference”) and the number of calls sampled by the female (“Sample”). Values are shown for the 3 tested discrimination thresholds: 100 Hz (a, b), 200 Hz (c, d) and 400 Hz (e, f). A 0.5 proportion of beneficial choices means a random choice.
after removing these bouts, the CV of DF decreases only by 1.7% of its value on average. The major part of the frequency variability that we emphasized here is thus due to other processes, namely the sudden discontinuities observed over the night. These discontinuities were observed at a large temporal scale, likely going undetected by short recordings usually used to estimate DF variability. In many anuran species, sound production is due to air flow from the posterior laryngeal chamber, between the vocal folds, into the anterior laryngeal chamber, and out of the larynx through a vertical aperture between the arytenoid cartilages (Pauly et al. 2006). We reveal for the first time in an anuran species that males largely vary in DF over a night. We hypothesize that these variations could be due to the fact that males are unable to strictly maintain the same air flow pressure between the vocal sac and the lung (Pauly et al. 2006) and/or to maintain the same muscle efficiency to stretch the vocal folds through opening the arytenoids (Duellman and Trueb 1994).

The discontinuities we observed over the night seem to occur at different times, in different directions and with different amplitudes for each male. Indeed, the pattern of variability observed in one male seems to be totally independent from those observed in the other males. Moreover, an analysis on the 36 recorded males does not reveal any general trend to increase or decrease DF over the course of the night. These results dismiss the possibility that observed patterns of variation are due to an environmental condition uncontrolled during the experiment. Contrary to what has been observed in some anuran species (e.g., *Leptodactylus alliabris*, Lopez et al. 1988; *Acris crepitans*, Wagner 1989; *Rana clamitans*, Bee and Perrill 1996; *Rana virentis*, Given 1999; *Rana catesbiana*, Bee and Bowling 2002), DF discontinuities are probably not due to changes in social context. Indeed, the social context was standardized by broadcasting a repeated 2-min artificial chorus noise; and even if the males could hear each other (which is improbable, see Material and Methods section) we would expect some synchronization in the discontinuities among males recorded simultaneously. Hence, the patterns of variation over the course of the night observed in the present study seem particularly erratic, an aspect that is likely to hamper their use by females in mate choice.

**Impact of DF variability on female choice**

Since female choice seems generally based on a few minutes listening in anurans (Murphy and Gerhardt 2002; Schwartz et al. 2004, Richardson et al. 2010), females are unlikely to hear the sudden discontinuities in male calls. Hence, with only access to a fragment of male variability, females may take an inappropriate decision when sampling males. Indeed, if we compared the minute with the lowest pitch and the minute with the highest pitch for each male, the difference was on average 444.2 ± 144.1 Hz (mean ± SD) and could reach 817 Hz in case of pronounced discontinuities which is far above the 190 Hz difference at which females are known to be able to discriminate DF (Richardson et al. 2010).

The mate choice simulation conducted in this study shows that the within-individual variability in DF can strongly affect the proportion of beneficial choices made by females according to this call property, from a nearly-full beneficial choice (90%) with a stable DF for both males, to a nearly-random choice (60%) with a highly variable one. In our results, increasing the number of calls sampled by the female from 10 to 10,000 calls (corresponding in average to a sampling time ranging from 1 s to 1.5 h), enhances the proportion of beneficial choices by up to 10% in the best case. Given that the costs of being selective are assumed to increase as the time allowed to the assessment of male condition increases (reviewed in Reynolds and Gross 1990), it appears that increasing the time allowed to sample potential mates is not an appropriate strategy for females to overcome the difficulty of choosing due to the variability in DF. Schwartz et al. (2002, 2004) found similar results concerning mate choice on call duration, calling rate, and pulse effort in *H. versicolor*.

By focusing on the distributions of the difference between mean DF of the 2 males (“Difference”) and the variability around these mean DF (“Variability”) observed for the 630 male pairs, we can observe that the overwhelming majority of pairs (94.5%) have a variability between 5% and 15%, a range for which the estimated

---

**Figure 5**

Upper plot: Distribution of the difference in DF (“Difference”) over the 630 tested male pairs. We computed the mean dominant frequency of a male over the entire night of its recording. We then computed the difference between these 2 values for a given pair of males, and that for all the possible male pairs. This shows that 32.5% of pairs show a difference higher than 200 Hz, and 5% of all pairs show a difference higher than 400 Hz.

Lower plot: Distribution of the variability around these mean DF values (“Variability”) observed for the 630 tested male pairs. For each pair, the “Variability” corresponded to the addition of the percentage of variability observed for both males.
effect of variability is relatively strong whatever the considered discrimination threshold. On the other hand, 67.4\% of male pairs differ by less than 200 Hz, supposed to be the mean threshold at which female are able to discriminate between 2 DFs (190 Hz according to Richardson et al. 2010). From all pairs, 25\% differ from about 200 Hz to 350 Hz, the range between which the impact of DF variability is the most important when considering a threshold value of 200 Hz (cf. Figure 4c). Only a small fraction of pairs (7.6\%) differ from more than 350 Hz, suggesting that it may happen rarely that females appropriately discriminate differences in DF in wild population. Taken together, these results provide a good explanation to the lack of female preference for DF when investigated in the wild (e.g., Friedl and Klump 2005; Friedl 2006), while this preferences is obvious in the lab (e.g., Richardson et al. 2010, Richardson and Lengagne 2010). More broadly, unprecedented large scale variability in signal components may hide female preferences when investigated in the field (Wagner and Sullivan 1995). Whatever the potential benefits that females can obtain through the choice of mates calling with a low DF in *H. arborea* (see introduction), considering that the opportunities for accurate selection on DF are rare and that females cannot improve significantly their assessment by increasing the time spent sampling potential mates, it seems puzzling to observe the maintenance of such a preference.

**Why female preference is maintained?**

A key element to understand the maintenance of a female preference despite a risk in reliability of the signal component may rely on a trade-off between DF and other call properties at individual level. Indeed, in *H. arborea*, DF is positively correlated with Within-bout Call Rate, (WCR, Gomez et al. 2011a) meaning that a decrease in DF which potentially improves male attractiveness, is accompanied by a lower WCR, potentially decreasing male attractiveness. Richardson and Lengagne (2010) have shown that females are efficient at discriminating attractive calls in complex acoustic conditions by assessing multiple acoustic signal components (at least DF, call duration and WCR). We can therefore hypothesize that females may guarantee some benefits by maintaining simultaneously a strong preference on several signal components linked together by trade-offs even if these components are non-perfectly reliable, as suggested by Candolin (2003).

By extending the idea developed by Castellano (2009) to the relationships between several signal components, and not only to the relationship between a signal component and the total amount of signal produced, we can provide a more robust explanation to the maintenance of strong female preferences for non-perfectly reliable signal components. This may be particularly true in the European tree frog where females also use the coloration of the vocal sac to discriminate male quality (Gomez et al. 2009, 2010, 2011b). An experimental study has shown an increase of male discrimination accuracy when both acoustic and visual cues are used by females (Gomez et al. 2011b). Hence mate choice is a complex process based on several acoustic and visual cues. Taking into account simultaneously multiple signals and multiples components may thus constitute a strategy used by females to counteract the ability of males to invest differentially in several aspects of signal attractiveness and temporally vary in their investment in call.

The existence of erratic patterns of variation in signal components should be investigated in other sexual communication systems. Nevertheless, the demonstration of its existence in the present study highlights the fact that communication in the context of sexual selection should be framed into a sexual conflict context. This context generates dynamic situations where signal components can be involved in mate choice even if they are not fully reliable (van Doorn and Weissing 2006; Castellano 2009). More largely, considering together signal dynamics and imperfect communication situations shed new light on the potential evolutionary impact of signal variability on decision rules and receiver communication strategies.

**SUPPLEMENTARY MATERIAL**

Supplementary material can be found at http://www.beheco.oxfordjournals.org/

**FUNDING**

We warmly thank Loïc Breptson for its precious help in different stage of this study, Mr Test from Optevoz for giving us access to the field site and Gladys Malide and Angéline Claire for their help in the maintenance of male housing conditions (EcoAquatron animal care structure). This study was conducted in accordance with the current laws in France (decision 2012-086-0019 of the Préfecture de l’Île and permit No. 69266347 of the Direction des Services Vétérinaires).

**Handling editor:** Francisco Garcia-Gonzalez

**REFERENCES**


Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS. 2013. All’s well that begins Wells: celebrating 60 years of Animal Behaviour and 36 years of research on anuran social behaviour. Anim Behav. 85:5–18.


