# Automated acoustic monitoring of endangered Common spadefoot toad populations reveals patterns of vocal activity

Running title : Acoustic monitoring of *Pelobates fuscus*

**Guillaume Dutilleux[a] <guillaume.dutilleux@ntnu.no>; Charlotte Curé[b]**

- a Acoustic Research Centre, Department of Electronic Systems, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), O.S. Bragstads plass 2, Trondheim, Norway
- b Cerema, Ifsttar, UMRAE, F-67035 Strasbourg, France

Keywords : acoustic monitoring, automated species detection, underwater sounds, amphibian, *Pelobates fuscus*

## SUMMARY

1. In the context of global amphibian decline, monitoring and restoration programs are important. Acoustic monitoring is a possible approach for underwater vocalizing species like the rapidly declining European Common spadefoot toad (*Pelobates fuscus*). In this study our aim was to design a dedicated software detector to be used in combination with programmable audio recorders to process the large amount of data generated by long term acoustic monitoring, and to use it for investigating the seasonal and circadian patterns of P. fuscus vocal activity.
2. The software detector targets advertisement calls of the species. Based on acoustic analysis of that call, we developed a detector that utilises both frequency and time features of the calls. Data collected during 3 breeding seasons in 4 known or potential *P. fuscus* breeding sites of northeastern France, was used to build a ground truth in order to test the performance of the detector. Then, we used the detector for analyzing 4 acoustic monitoring campaigns conducted in 2 different sites over 2 breeding seasons to gain insight into the seasonal and circadian patterns of vocal activity of this species.
3. Evaluation of the *P. fuscus* call detector against a ground truth returned false positive rates below 1.5 % and true positive rates ranging from 53 to 73 %. These figures are compatible with long term monitoring of the presence of the species. Running the software detector on standard hardware, the computation time for post-processing the 360 hours of a typical 3-month monitoring campaign was less than 1 day.
4. The seasonal pattern of *P. fuscus* underwater vocal activity is more complex than previously recognized. Over the whole ostensible 3-month breeding season, the actual time window for vocalizing and breeding can last from a few days up to several weeks, and may be split into clearly distinct episodes. When vocalizations occurred both at night- and daytime, the circadian vocal activity of *P. fuscus* occasionally

proceeded uninterrupted for 24 hours but usually a several hour lull occurred immediately prior to sunset. When vocalizations occurred both at night- and daytime, the vocal activity pattern followed a bimodal distribution with a nocturnal highest peak of activity and a second peak occurring in the morning.

5. Our results demonstrate that it is feasible to monitor presence of *P. fuscus* in northeastern France using a dedicated software detector combined with programmable audio recorders. Based on the outcomes of the detector applied to long-term audio data sets, we reveal temporal patterns of the vocal activity of the species and subsequently provide recommendations for attended and unattended acoustic monitoring.

# Introduction

Amphibian decline was identified in the 1980s as a global mass extinction (Collins *et al.*, 2003) which prompted implementation of monitoring and restoration programs. In this context, classical monitoring methods (Sutherland, 2006) based on visual cues (e.g. observations of the species, presence of eggs) can be difficult to implement, particularly in small-sized species vulnerable to human disruption and for which habitats are difficult to access (e.g. remote location, at dark, in ponds). Since amphibian species are vocally active, an alternative method is to assess their presence by recording their sound. Recent advances in electronics, signal processing and computer science have enabled recording and analysis of long sequences of environmental audio signals, providing a tool to document the presence of species, to identify individuals, and to monitor populations (Chesmore *et al.*, 2008). This approach has the advantage of being non-intrusive and relatively low-cost. Moreover, the use of programmable recorders has considerably reduced the constraints of working in the field, allowing automatic acoustic recording in a given environment over a long period of time without requiring the presence of a human operator (Weir *et al.*, 2005; Brandes, 2008). The potential of such bioacoustic monitoring methods has been investigated and successfully harnessed in anuran species (Waddle *et al.*, 2009; Steelman & Dorcas, 2010). Indeed, vocalizations play a major role in anuran life. The analysis of their acoustic characteristics can be useful for taxonomic studies (Köhler *et al.*, 2017) and the study of temporal calling patterns can provide insights into the biology of the species such as behavioral rhythms (Cui *et al.*, 2011).

In the present study, we aimed to develop reliable automatic acoustic monitoring of an endangered lowland anuran species, the European Common spadefoot toad, *Pelobates fuscus (Pelobatidae),* in order to assess the presence/absence of the species and to investigate the seasonal and circadian patterns of the vocal activity. This species used to be common from western Europe to western Siberia (Eggert *et al.*, 2006). It is now listed in the Annex IV of the European Habitats directive (European Council, 1992) and the Appendix II of Bern Convention (Council of Europe, 1979). Although classified as *Least Concern* on the IUCN Red List of threatened species (IUCN, 2009), the populations of the species are overall decreasing and particularly in the western part of its distribution range (Eggert *et al.*, 2006). Reintroduction programs for the species have been tried or are ongoing in Italy and Germany (Giovannini *et al.*, 2014; ten Haagen *et al.*, 2016). In France *P. fuscus* was common in the northern half of the country except Britain in the 19th century (Eggert *et al.*, 2006). It is now classified as *Endangered* in the national Red List 2015 (IUCN, 2015). At the time of writing, only three highly isolated and very small populations are known to remain in the country (Lescure & de

Massary, 2012) and a regional action plan dedicated to monitoring and restoration of the species has been ongoing in Northeastern France since 2012.

Several aspects of the biology of *P. fuscus* make the species difficult to document in the field by conventional means, in addition to the particularly small size of the remaining populations in some areas. First, *P. fuscus* is fossorial, *i.e.* it spends most of its life underground and emerges for reproduction (Noellert, 1990), and is mainly nocturnal. Second, the breeding period is difficult to predict and the breeding activity which takes place in shallow water bodies is secretive (Eggert & Guyétant, 2003), *i.e.* adult *P. fuscus* remain invisible in the depths of the water bodies. Therefore, visual observation of the species is difficult to achieve and may only be conducted at night when individuals are on their way to and from the reproduction sites. Third, although terrestrial vocalizations have been documented both for adults and recently for juveniles (ten Haagen *et al.*, 2016), most vocal emissions of the species occur underwater and can hardly be heard from the littoral (Frommolt *et al.*, 2008). The vocal repertoire of the species has been described and contains four to six different call types (Müller, 1984; Andreone & Piazza, 1990) for which the functions are still not well established (Seglie *et al.*, 2013; Frommolt *et al.*, 2008). Among those vocalizations, the species-specific advertisement call is by far the most frequently produced. It is emitted by both males and females during the breeding season (Noellert, 1990; Frommolt *et al.*, 2008) and has a relatively simple acoustic structure, making this vocalization a good target for bioacoustic monitoring of the species (Frommolt *et al.*, 2008).

The authors have implemented acoustic monitoring of *P. fuscus* with programmable recorders in several water bodies of the northeasternmost part of France since 2011 (Dutilleux *et al.*, 2012, Gosset *et al.* 2012; Curé *et al.*, 2013; Curé & Dutilleux, 2014). This work led to the development of the first version of a software semi-automatic detector (Dutilleux & Curé, 2016) for *P. fuscus* advertisement calls, based on spectral features. This detector was fairly efficient at detecting this call type but suffered from too large a false alarm rate. On large data sets, numerous false alarms mean impractical time requirements of a human operator checking the detections.

In this paper we present a redesigned dedicated detector for *P. fuscus* advertisement calls, which utilises both frequency- and time-domain features and delivers presence information and estimations of call counts. For this application the desired characteristics of the detector are 1) a low enough false positives rate so that human post-processing time is practicable and 2) a high enough true positive rate that it is impossible to miss presence of the species over the whole breeding season. Since we aim to focus on presence-absence information, false negatives are less critical in the context of long term monitoring. The objectives of our study were i) to test the efficiency of the detection process on several sites against a ground truth made of manually annotated acoustic recordings and to evaluate both computation- and human operator- time, and ii) to apply the detector to long term field monitoring campaigns conducted over two successive years on two studied sites to reveal circadian and seasonal patterns in vocal activity of *P. fuscus*. In conclusion we discuss recommendations for monitoring techniques in the context of conservation plans.

# Methods

## Studied populations and sites

Among the three remnant *P. fuscus* populations in France, two are located in Alsace, *i.e.* the eastern part of the Grand Est administrative region in France. There, *P. fuscus* breeds typically in minimum 1 m deep water bodies that can reach 80 m$^2$ in surface area. The breeding season generally occurs somewhere between late March and late May and lasts for more than one month (Eggert, 2003). However in the studied area, the *P. fuscus* breeding season is slightly shifted, occurring between April and late June (Vacher & Dutilleux, 2010). Two phreatic study sites were selected in the northeasternmost part of the region: Mothern and Sauer's Delta (the latter is named Sauer for convenience). Both sites belong to the Rhine river's floodplain and are 4 km apart as the crow flies. These sites were chosen because of the past and recent evidence of species presence (Curé *et al*., 2013; Curé & Dutilleux, 2014), and because the water bodies were known to be permanent during the breeding season of the species although they often dry out during the winter. Moreover, these water bodies are of limited surface area allowing monitoring of the species vocalizations underwater without the need for multiple hydrophones. Sauer is a roughly circular water body with an average diameter of 10 m depending on water depth. Located in a floodable meadow nature reserve it is surrounded by a gravel road on one side and by a *Phragmites australis* (Poaceae) reed bed elsewhere. This water body serves as reproduction site for other amphibian species: *Rana dalmatina* (Ranidae), the *Pelophylax sp.* complex (Ranidae) and *Hyla arborea* (Hylidae). The breeding season of *R. dalmatina* takes place before the one of *P. fuscus* and the species disappears from the water bodies when its breeding season is over (Godinat, 2010). *Pelophylax sp.* and *H. arborea* are likely to vocalize during the breeding season of *P. fuscus* (Vacher, 2010; Buchel, 2010). Mothern is a crescent-shaped water body whose largest dimension is 50 m. It is surrounded by beeches on the edge of a forest. It is bounded on one side by a forest lane. Mothern is a reproduction site for the same amphibian species as Sauer but *Pelophylax sp*. There is no significant road or rail infrastructure within a distance of 400 m from each site.

To evaluate the performance of the detector, we used acoustic recordings collected in Mothern and Sauer in 2015 and 2016, and we also included recordings gathered in 2017 in two different water bodies located at Leutenheim (named Leutenheim 1 and 2), located 16 km southwest of Sauer. These water bodies are considered suitable breeding habitats for *P. fuscus* and one individual of the species had been observed there in 2013.

## *P. fuscus* advertisement call structure and propagation

The advertisement call of *P. fuscus* was the target of the automated detector. When produced, this highly stereotyped call consists of a series of evenly spaced short bursts (Fig. 1). Each burst is made of the succession of 2 (or sometimes 3) relatively similar "cloc" notes. Each note is made of a first small amplitude transient clearly separated from the 2 main transients at the end of the note.  Most of the acoustic energy is located between 700 Hz and 1200 Hz for male *P. fuscus* (Frommolt *et al*., 2008).

Unlike many anurans, *P. fuscus* do not perform choruses *i.e.* when several individuals simultaneously call together (Eibl-Eibesfeldt, 1956). Thus, vocalizations from one individual are less likely to be blurred by vocalizations from conspecifics, which is desirable for pattern recognition by the automated detector.

Regarding sound propagation (See for instance (Medwin, 2006; Lurton, 2010) for more details), underwater *P. fuscus* vocalizations propagate in shallow water which acts as a high-pass filter (Forrest *et al*., 1995). Previous experiments in relation to *P. fuscus* monitoring have shown that with 50 cm water depth the cut-off frequency is about 1 kHz (Frommolt *et al*., 2008). Therefore, from a distant receiver, depending on water depth the vocalizations are likely to be either strongly attenuated or heavily distorted in addition to normal geometrical divergence. However, the purpose of species acoustic monitoring is not to collect every single vocalization of each individual and a reasonable assumption is that, over time, each *P. fuscus* will wander through the water body and vocalize at different positions. Given the low reverberation characteristics of the shallow waters in which *P. fuscus* breeds, we assumed that the time structure of vocalizations was stable and so that a detected vocalization could not be an echo of a previous vocalization.

Based on this, we were confident that a suitable detector for this species should integrate both time-domain and frequency domain features of vocalizations.

## Acoustic monitoring protocol

Acoustic monitoring of the species was conducted using SM2 SongMeter programmable audio field recorder connected to a single hydrophone of type HTI-96 (both from Wildlife Acoustics Inc., Maynard, USA). Each SM2 was attached on a pole fixed in the ground close to each study site for the duration of the campaign. Each hydrophone was attached to a second graduated pole used as a levelling rod for water depth measurement. Acoustic recordings were sampled and stored without any data reduction in PCM format at 16 kHz and with 16-bit resolution. No calibration of the audio acquisition channels was performed.

Acoustic monitoring was conducted on both studied sites during the 2015 and 2016 breeding seasons of the species. To ensure to cover the whole breeding season, SM2 recorders operated continuously from late March to late June. For the 4 campaigns (Mothern 2015, Sauer 2015, Mothern 2016 and Sauer 2016), SM2 recorders were programmed to record for 5min every half hour. At least every fortnight, the recorded data was collected, water depth was measured and the clock and recording system were checked.

Once collected, the acoustic recording data was post-processed using a software detector specifically developed for the freely available Scilab (ESI Group, Rungis, France) environment to detect *P. fuscus* advertisement calls. The audio fragments detected by the software detector were then listened to by an operator for verifications. The whole post-processing was carried out on a 2015 laptop (2.2 GHz Intel core i7) with SSD hard disk running macOS Sierra 10.12.3 and Scilab 6.0.

## Principles of automated *P. fuscus* detection

The software detector we developed is based on time-domain signal processing and implements routinely used sound descriptors although in different contexts of environmental noise monitoring such as road traffic and community noise assessment (ISO, 2003). The detector processes each 5-min audio file independently. For each detection a 1s  audio fragment is extracted from the input file. For each input file the detector stores (1) a detection count and (2) an audio file which is the concatenation of all the 1s fragments extracted. The latter file features a discontinuous time axis with the intervals between detections removed, in order to accelerate the occasional verification by a human operator.

### Frequency and signal/noise ratio features

The flow chart of the detector is given in Fig. 2. The detector computes the sound level of two different frequency bands. The first one named signal frequency band is the octave band of centre frequency 800 Hz (frequency band: [566-1131 Hz]). The signal frequency band roughly corresponds to the frequency range of *P. fuscus* advertisment calls and is the result of the optimization of a frequency-domain detector previously developed by the authors (Dutilleux & Curé, 2016). The second one named background noise frequency band [200-566Hz] corresponds to a lower frequency range in which the target call is much less likely to have significant energy. As a general rule in acoustics, sound generation at larger wavelengths - here 2.65 to 7.5m assuming 1500 m/s for sound speed in freshwater - is due to larger sources (Beranek, 1986). This background noise frequency band is very unlikely to be used by insects or anurans. Some fishes vocalize in this band, however fish are unlikely to be present in these temporary waters, and moreover it well known that *P. fuscus* does not coexist with predatory fish (IUCN, 2009) and for some common European herbivorous fish the fry feed on amphibian larvae (Griffiths, 1996). Lungfish do not exist in Europe. These elements support the assumption of absence of animal vocalization in this frequency band in water bodies where *P. fuscus* was studied. Therefore, the [200-566] Hz range is used as proxy for background noise. In the context of our research, background noise is mostly caused by rain. Due to its transient nature, the sound of a raindrop has a poor frequency localization so that its spectrum will overlap both the background frequency band and the signal frequency one. In the absence of rain, the background noise is very likely to be lower than the self-noise of the recording equipment. These two frequency bands are separated in the time domain using 4th order Butterworth infinite impulse response digital filters (Oppenheim, 1975).

A candidate detection is produced when the recording contains significantly more energy in the signal frequency band than in the background noise frequency band. Instead of applying an absolute threshold, this comparison is done to help eliminate false detections because of rain.

The maximum sound level of the signal frequency band $L_{max}$ is then calculated and compared to the background noise level as follows. The background noise level is estimated by $L_{90}$ fractile sound level which corresponds to the sound level exceeded for 90 % of the duration of the acquisition, here 5 minutes. $L_{90}$ is derived from the short-term equivalent sound levels $L_{eq,\tau}$ with time constant $\tau$ and no overlap. $L_{max}$ is taken as the maximum of the so-called time-weighted sound pressure level $L_\tau(t)$ (IEC, 2013) with exponential time-weighting. $L_\tau(t)$ is an approximation of the signal envelope. Compared to short-term equivalent sound level it is better suited to the estimation of maximum sound levels. By definition, $L_{max}$ and $L_{90}$ are expressed in decibels with respect to a reference value $x_0$ (notation dB re $x_0$). Since, as already mentioned, the recordings post-processed are not calibrated with respect to a reference pressure source, it is not possible to associate the amplitude x of an arbitrary sample in an arbitrary audio file to a value of sound pressure in water. We can only assume a proportionality relationship between sound pressure and amplitude in the file. Here we used $x_0$=1. The absence of calibration is inconsequential for the detection procedure which relies on time and *relative* squared amplitude features.

### Time-domain features

A *P. fuscus* advertisement call will appear as peaks in $L_\tau(t)$. Since such a call is made of several close peaks, the detector checks first whether each candidate peak is strong enough and remains within a certain duration range. Moreover, in $L_\tau(t)$ representation, peaks associated to *P. fuscus* vocalizations appear to have a characteristic ripple on the top (Fig. 1 - top) that helps discriminate them from raindrop-generated peaks. So this feature is also checked.

The detector proceeds by gathering the peaks in groups assuming that two peaks are not in the same group if the time span from one to the other is above a threshold. In order to assign whether a peak group is likely to be a *P. fuscus* advertisement call, the detector checks that (1) the number of peaks in the group is more than one and less than four and (2) the peaks within the group have about the same amplitude.

The numerical values for the different time parameters given in Table 1 have been derived from the observation using the audio software Audacity® (version 1.3.12, Audacity Team) of two high quality recordings performed by one of the authors in 2007 and 2008 in two water bodies of the studied region, one close to but different from the Sauer site discussed here and another one in Brumath, 35 km southwest of the Sauer site. These two recordings were chosen because they were deemed reasonable lower and upper bounds for the duration of a “cloc” and of the time between two “clocs”. A small fraction (less than 1%) of Mothern 2015 and Sauer 2015 recordings was used to manually adjust the decibel thresholds presented in Table 1 in a few test runs of the detector. This fraction contained recordings with rain, *R. dalmatina* or *H. arborea*, the main causes of false positives. This fraction builds the so-called "tuning set". Since the recording chain is not calibrated, the minimum $L_{max}$ of peaks given in Fig. 1 is attached to particular hydrophone and recorder types and particular recorder settings.

Due to the relatively large number of parameters involved in the detector and the fact that satisfying performance has been achieved with this hand-tuned parameter set as will be shown later in this paper, a numerical optimization of parameters has not been performed.

## Assessing the performance of the detector

### *Building a ground truth*

In order to evaluate the performance of the detector, 200 30-sec audio fragments were selected among the 5-min recordings in the available 2015 and 2016 (Mothern and Sauer sites) audio data, and also from the 2017 audio recordings collected in Leutenheim 1 and 2 sites. For each site and each year 200 fragments were selected using uniform random deviates. These files formed the so-called "test set" and represented 10 hours of audio data. The files were played and visually inspected on a spectrogram by using Audacity. They were listened and annotated by one of the authors using circumaural headphones (closed Sennheiser HD 280 Pro (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) or open Beyerdynamic DT990 (Beyerdynamic GmbH, Heilbronn, Germany)) in a quiet listening environment. The annotation consisted mainly in identifying if a file contained vocalizations of *P. fuscus*. In addition, qualitative information was collected about the other types of sounds that could be identified.

### *Comparison between ground truth and detections*

The 30-sec recordings were given as input to the detector. Since the purpose is to evaluate the performance of the automatic part of the detection procedure the last validation step by a human listener was not performed. Through comparison of human-generated ground truth to automated detection results we counted the following 4 standard aspects of detection quality: "true positives", when *P. fuscus* call is present in the file and detection count is above zero; "true negatives", when *P. fuscus* call is not present in the file and the detection count is zero; "false positives" when *P. fuscus* call is not present and the detection count is above zero; "false negatives" when *P. fuscus* call is present in the file whereas detection count is zero.

### Evaluation of computation and human operator time of the semi-automatic *P. fuscus* detection method

In order to assess presence of *P. fuscus*, to investigate the seasonal and circadian vocal activity of the species, and to evaluate computation and human time required for the processing each of the 4 Sauer and Mothern campaigns, we applied the detector on the 4 full datasets (i.e. Mothern 2015, Sauer 2015, Mothern 2016 and Sauer 2016). As part of the method, all the audio fragments selected  by the detector, i.e. potentially having the *P. fuscus* call, were listened to by a human operator in order to extract correct assignments from false alarms.

# Results

## Content of the ground truth

### *Description of the underwater soundscape*

The test set contained samples of a variety of biotic, abiotic and anthropogenic sounds that were expected to be heard. Despite sound transmission from air to water suffering substantial power loss, the recordings featured some noticeable airborne sounds.

Abiotic sounds were mostly isolated water drops or rainfall of various intensities over the whole test set. In several files heavy rain caused saturation. Throughout the files, many shock noises occurred and were likely to be attributable to wind and animals moving underwater. Among anthropogenic noises, faint church bells, train klaxon and pass-bys could be heard very early in the morning probably due to an atmospheric temperature inversion.

Various animal sounds were identified in the recordings in Mothern and Sauer. Most of them were amphibian species vocalizing underwater. Beside the dominant *P. fuscus* vocalizations, *Hyla arborea* and *Pelophylax sp.* vocalizations were also present in the test set. A large variety of birds could be heard underwater even though they were obviously singing in air. At least 6 species of passerine birds could be identified (*Turdus merula (Turdidae), Turdus philomelos (Turdidae), Fringilla coelebs (Fringillidae), Phylloscopus collybita (Phylloscopidae), Acrocephalus scirpaceus (Acrocephalidae), Erithacus rubecula (Muscicapidae)*) plus the Common cuckoo (*Cuculus canorus (Cuculidae)*). Some unidentified invertebrates and insects were also present above 2 kHz.
The biotic soundscape in Leutenheim, was mostly made of invertebrate sounds.

In all of the sites, the ground truth does not reveal any animal vocalizing in the frequency range used for estimating background noise except *P. fuscus* itself as discussed below. Several files, however, contained scratching noises with low frequency components that can be attributed to invertebrates on hard substrate like vegetation, the floor of the water body or the hydrophone itself and its cable. These structure-borne noises are typically periodic with significant intervals of silence so that they do not compromise the detection of *P. fuscus* calls.

## *Pelobates fuscus* calls

Among the 1200 files building the test set, 188 featured the *P. fuscus* advertisement call. One third of the latter is provided by Sauer and 2/3 by Mothern. The ground truth returned 2492 calls in total. No vocalization of *P. fuscus* was detected in any of the two Leutenheim sites.

In Mothern 2016 campaign 47 out of the 200 files of the test set contained *P. fuscus* advertisement calls in the 800 Hz octave band. 5 others featured calls with most of the energy in the background noise frequency band. This was not observed in the 3 other campaigns.

As expected, although other vocalizations types of *P. fuscus* occurred in the test set, the advertisement calls were by far the most frequent. The average number of advertisement calls per 30-sec file featuring the target call is 13.6 ±10.9 (mean±SD).

## Acoustic analysis of the advertisement call

Calls may vary between conspecifics, gender and across breeding sites and populations. Therefore, in order to check the validity of the acoustic parameters implemented in the *P. fuscus* call detector, we conducted an acoustic analysis in the temporal and frequency domains of a sample of 20 calls recorded at different date and time over 2 years of 4 distant breeding sites (see details of the measures on Fig. 3).

The analysis (Table 3) showed that the elapsed time between the two notes of the call is 159±18 ms. The two notes last on average 93±18 msec and 78±20 ms, respectively. Power spectrum analyses revealed that each of the two notes contain a clear spectral energy peak at 758±160Hz and 805±169Hz, respectively. The occurrence of those two successive spectral peaks occurring over the call and the fact that they both match the frequency octave band value centered at 800Hz that we set up for the detector, support the chosen features of our automated detector.

## Validity of the *P. fuscus* detector on the test set

For the test set, true positives, true negatives, false positives and false negatives were calculated separately for each site and year as shown in Table 4. Since *P. fuscus* is not present in the ground truth for Leutenheim 1 and 2, the true positives rate could only be calculated for Mothern and Sauer and ranged from 53 to 73 %. Whatever the site, the odds of false positives were lower than 1.5 %. These results suggest that the detector is reasonably capable of detecting advertisement calls from *P. fuscus* in the corresponding campaigns and that it generated very few false alarms.

## Full-scale operation of the detector

Since the detector reached satisfying performance on the ground truth, it was applied to 4 complete data sets (i.e. Mothern and Sauer, in 2015 and 2016). From one 5-min audio file to another, the processing time varied from 4 s when no peak was above threshold to 1 min when several thousand candidate peak groups were formed and had to pass through the other next steps of the detector (Fig. 2). For Sauer 2015 which is the longest monitoring campaign analysed (in total 456 h of audio recordings; Table 2), offline post-processing by the detector required 13h46, and the 1h20 of audio retained by the detector took less than 3 hours of listening by a human operator for the final verification step.

Whatever the year or site, the water bodies never dried out during the monitored periods although water depth was occasionally below 20 cm at the beginning of Spring. The 2015 data was collected almost without interruption from mid-March to the end of June (Table 2). In Sauer, the 2015 campaign did not suffer any interruption. Only 5 days of data was missing in Mothern between March 21st and 26th 2015, due to a battery power outage. The 2016 campaign in Sauer was shortened because of an oncoming flood that required removal of equipment earlier than planned but provides a long uninterrupted time series of underwater audio data until the end of May. In Mothern the 2016 campaign was disrupted two times from April 13th to 27th and June 9th to 22nd because of battery issues. The total audio data collected represented 347 days.

## Detection confidence

Since the detector returns a call count for each file analyzed, we checked if a higher call count was correlated to a lower error rate at the file level on the 4 complete data sets from Mothern and Sauer. We considered the files with call counts above zero in the full-scale operation case where the files are 5min long. The results are shown in Fig. 4. The error rates can be quite high for low call counts but for call counts above 60 in the worst case, the error rate falls to zero both for Mothern and Sauer and for the two years of monitoring available. The uncertainty of the value for the error rate is however higher for high call count bins because of the lower number of files taken into account, except for Mothern 2015. The higher erroneous call counts are caused by *H. arborea*.

## Vocal activity of *P. fuscus*

### *Seasonal pattern*

In the Mothern 2015 campaign, the breeding season was split into markedly separate periods as shown in Fig. 5. In this campaign there was clearly a first 7-day period with intense vocal activity in April followed by a 2-week lull before a new 4-week long period of intense vocal activity in May. In the same year in Sauer intense vocal activity was shorter since it lasted only for 4 days between April 7th and 10th (Fig. 6). For the Sauer 2016 campaign, the first detection was documented in April, 2 weeks before the peak of activity observed in May with very few repetitions in the meantime (Fig. 7). On the Mothern site, the 2016 campaign documented sporadic vocalizations in late June (Fig. 8), more than one month after the end of the main vocal activity period.

### *Circadian pattern*

From detections patterns observed in the different campaigns analyzed *P. fuscus* was mostly vocal at night (Fig. 5 and 6). In Sauer 2015, the detection of vocalizations stopped right after sunrise and resumed after sunset. For Mothern 2015, the vocalization rate increased abruptly around sunset and remained fairly stable for seven hours as illustrated in Fig. 9. A second peak of vocal activity was observed during daytime around 9:00 followed by a lull in the hours preceding sunset. Continuous vocalizations over 24 hours occurred once, on May 11th .

# Discussion

The aim of the present study was to improve long term acoustic monitoring method of the *Pelobates fuscus*, an endangered amphibian species. Building on a previous study we redesigned an automatic detector of the species advertisement call that reached both a reasonable true positive rate and a low false positive rate that limits the human operator time required for the validation of post-processing to 1 day assuming a typical 3-month monitoring period for 1 site (representing 360 hours of collected audio data). Moreover, the use of this detector in long term studies of two sites and two consecutive years provided new insights into the seasonal and circadian pattern of vocal production of the species, bringing practical recommendations for *P. fuscus* populations monitoring.

## Accuracy of the detector

Preparing a ground truth is a time-consuming task and implies some trade-offs. Here the emphasis was put on a broader sampling of the different campaigns at the expense of the similarity with the monitoring protocol, since 30-sec samples were taken in the ground truth versus 5-min samples in the monitoring. For a fixed total duration of the ground truth, this means that the different performance rates are likely to be more reliable due to the larger number of samples than if we had used 5-min samples. However, with 10 times shorter files in the ground truth configuration, the detector is given less information than in the real monitoring configuration to decide between presence and absence at a particular time. Therefore, the true positive rates and the false positive rates can be somewhat underestimated. Fortunately, the full-scale tests on whole campaigns indicate that the false positive rate remains fairly low.

Although it would have been desirable to include data from other sites, the number of available audio datasets with *P. fuscus* is rather limited. Regarding the risk of pseudoreplication (Williams *et al.*, 2002), it must be kept in mind that a recorder placed on a single site is likely to pick up vocalizations from different individuals at a given time. It will also pick up vocalizations from different individuals visiting the water body at different moments of the breeding season. Moreover, for the same individual the recorder is likely to receive vocalizations from various ranges and depths. The long term monitoring will reflect different propagation conditions as water depth varies throughout the season, it will also integrate the variation of water temperature which is known to influence vocalizations of the target species (Seglie *et al.*, 2013). In addition the monitoring will integrate the variability of background sound.

### *False positives*

In the context of long term monitoring, false positives are one of the main issues of semi-automatic detection methods because their rate is correlated to the time required for the post-processing phase carried out by a human operator who verifies the correct assignment performed by the detector. In our study, the false positives generated by the *P. fuscus* detector on the ground truth data set were either rainfall noise, shock-like sounds or *H. arborea* calls. None of the other biotic sounds contained in the ground truthed recordings induced false positives. However, the false positive rate depends on the soundscape that can differ according to the studied sites and ecological factors so that the generalization of the rates obtained to other sites is not straightforward.

Regarding the abiotic sounds, whereas rainfall can be favorable for stimulating anuran vocal activity (Brauer *et al.*, 2016), it was an occasional cause of false positives. Although rainfall generated overall broadband peaks, in some situations tonal noise and the time domain parameters of rain could accidentally shape the *P. fuscus* call pattern.

### *True positives and false negatives*

This objective of the present acoustic detector was to provide a tool in order to detect the presence of the species on specific studied sites. A key issue relating to the ratio of true positives versus false negatives is the detection range around the hydrophone that depends on different factors involving the recording equipment and the software detector. A specific study will be needed to further test for these factors, the main ones being the shallow water environment with strong and frequency-dependent attenuation, the internal noise of the hydrophone itself and the SM2 analog preamplifier stage, and the thresholds and principles implemented in the software detector. The detection range is likely to vary over a several-month long campaign with the fluctuation in water depth. This information is essential to fully determine how to sample the sound field in a water body of arbitrary size. The SM2 and its standard hydrophone are relatively cheap equipment with suboptimal self-noise characteristics. Without changing anything in the principles of the software detector, low noise hardware can bring a clear improvement toward reducing the false negative rate. However, it may also imply more false positives.

## Potential improvements of the detector

Results from Fig. 4 suggest that the pattern matched by the detector is fairly species-specific in the group of vocalizing species considered here. They suggest the prospect of a fully automated version of the detector. It could be defined by adding a threshold on call counts at the end of the detection process in order to decide on presence/absence at the file level without the intervention of a human operator.

Adding more *P. fuscus* intrinsic signal features would certainly help reducing false positives, provided that they are reliable and in particular not sensitive to propagation in shallow water. Further work should include a temperature-dependent parameter set in the light of the correlation between water temperature and the inter-onset time of successive notes that was found by Seglie *et al.*, (2013). From a broader perspective it would be desirable to develop a multi-species classifier and to perform an auditory scene analysis (Virtanen *et al*., 2017). Such a classifier would bring positive reasons to reject vocalizations because they would match the characteristics of another species and not only give a bad score with respect to that of *P. fuscus* vocalizations. At the beginning of this research, it would not have been possible to implement machine learning techniques like neural networks (Haykin, 2016) due the lack of publically available recordings of *P. fuscus* vocalizations to form large enough a training set. With the data gathered it becomes possible to implement machine learning.

## Applicability of the detector to other populations

While the time pattern shown in Fig. 1 is taken from a recording performed in Alsace (France), it corresponds rather closely to the one documented in (Frommolt *et al*., 2008) for northeast German *P. fuscus* populations of the nominal species and in (Seglie *et al*., 2013) for the *P. fuscus insubricus* subspecies in Italy. Therefore, it is worth looking at the applicability of the detector to other populations. The acoustic analysis of *P. fuscus insubricus* calls conducted by Seglie *et al.* (2013), and the oscillograms of the German *P. fuscus* population calls provided by Frommolt *et al*. (2008) show that the durations of the call (range: 250-410 msec) and of each notes of the call (range: 70-120 msec) are concordant with our findings (see Table 3) and are therefore compatible with the time parameters of our detector (Table 1). Regarding frequency parameters, (Frommolt *et al.*, 2008) report that maximum of acoustic energy ranges from 700 to 1200 Hz and (Seglie *et al.*, (2013) which again matches our findings (Table 3). As such we suggest that the detector could be used to monitor these remote Italian and German *P. fuscus* populations, although the higher frequency vocalizations may be ignored by the detector. It would be of course possible to choose a higher frequency limit for the *P. fuscus* range in our detector, but the impact on false alarms will depend on the other vocalizing species and remains to be evaluated.

## Use for conservation purposes

The detector developed here is usable for presence/absence monitoring of *P. fuscus*. For endangered species, the relevance of abundance versus presence/absence monitoring was discussed in Joseph *et al.*(2006) for birds. These authors highlight that the monitoring strategy depends on the budget available and recommend as a rule of thumb to use presence/absence methods when fewer than 16 sites are monitored per year. This is very likely to be the case when using autonomous recorders.

The detector can be used to monitor both historical and potential *P. fuscus* breeding sites. Even on historical sites, documenting the presence of the species proves to be difficult by conventional means, because, for economical reasons, the time allotted to each site is too short. The use of automated recorders offers an unmanned permanent attendance on site at a lower cost than conventional field surveys with lower odds of false negatives. Moreover it is even harder to find the species at sites where it has never been documented before. For the same reasons autonomous recording seems a cheaper and more efficient approach than conventional techniques.

### *Full-scale operation*

False positive and true positive rates certainly matter when it comes to evaluating a detector. However, these values relate to the percentage of files with detection with respect to the total number of files analyzed, not to the percentage of audio to be listened to with respect to the total duration of recordings collected in the field.

What is relevant for the human-operator-based validation phase is the number of person-hours implied. Listening time is proportional to the amount of audio recording extracted by the detector which is made of true and false positives. Even though the false positive rate is low, the real breeding period of *P. fuscus* is likely to be short - from a few days to a few weeks, compared to the potential monitoring period that operates typically for 3 months. Therefore, a low proportion of candidate false positives may represent most of the listening time of the human post processing phase of detection. Beyond listening time, the total processing time for such several-month monitoring campaigns may also be an issue for users not having access to high performance computing resources, especially if several populations are to be monitored in parallel.

By running the software detector presented here on standard hardware, the computation time for post-processing the 360 hours of a typical 3-month monitoring campaign was less than 1 day. Therefore, the new detector brings approximately 90% of reduction in overall human and computation time compared to its previous version (Dutilleux & Curé, 2016).

### *Potential for estimating the number of individuals*

The detector discussed here was developed specifically to identify whether *P. fuscus* is present or not under the assumption that it regularly vocalizes underwater. But what about counting individuals vocalizing at the same moment? The data collected contains many instances where *P. fuscus* individuals vocalize in parallel with or without overlapping each other. A trained listener will be able to discriminate 3 or 4 individuals on the basis of the following cues and assumptions: (1) the amplitude received from each individual is fairly constant, (2) the amplitudes from different individuals are mutually different and (3) the spectral composition received is characteristic of the individual and its position. But implementing this in software is not straightforward. A possible approach could be to rely on a correlation to be established between the number of individuals and the number of counts per minute. Further developments could consider introducing at least one additional hydrophone which would give an information about the direction of arrival using phase differences between the two hydrophones placed at a sufficient distance from each other. A third hydrophone would facilitate mapping the individuals on the horizontal plane. These are classical signal processing techniques that have been already implemented with success in other species (Wahlberg *et al*., 2003), and by Frommolt *et al*. (2008) on *P. fuscus* although without automated detection. However, determination of the geometry of such a hydrophone array requires investigation of the active space of *P. fuscus* vocalizations in a shallow water environment where attenuation may increase rapidly with the distance from the source and show a strong frequency dependence.

## Seasonal and circadian rythms of *Pelobates fuscus* vocal activity and recommendations for acoustic monitoring

Long term automated acoustic monitoring allows subsequent study of seasonal pattern and circadian patterns of the vocal activity of the species. Even though we do not have a continuous collection of data, our sampling effort was high enough (5min every 30min) to rely on and extrapolate outcomes on the temporal vocal pattern over 24h and over the breeding season. The absence of detection at one moment does not imply that vocalizations have not been produced at that time. The vocal individuals could have simply been out range of the hydrophone. However, if one can assume that the number of individuals present in a water body remains fairly constant over a 24-hour period and that individuals do not remain immobile in the water body, the absence of detection for several hours may arguably be associated to the absence of vocalizations.

As expected, we found a nocturnal peak of vocal activity that is typical in other anuran species, although the precise timing of such intense nocturnal vocal periods can vary across species (Cui *et al*., 2011). In addition, we observed occasionally a minor diurnal peak of vocal activity. This peak was already reported for *P. fuscus* (van Gelder *et al*., 1971). However we found that this peak was not systematically present.

In terms of monitoring recommendations, although vocalizations at daytime appear to be less likely than the nighttime ones, our results suggest that surveying the species by regular direct observations with hydrophones in the morning is a reasonable approach if one wishes to avoid nighttime fieldwork or automated audio recording. If nighttime work is admissible though, visiting sites right after sunset appears to be the time with highest chance of vocal detection of the species. Moreover, if programmable recorders are to be used for routine monitoring purposes, due to the circadian vocal activity pattern of the species, it is not mandatory to record between sunrise and sunset. This simple change would reduce processing time dramatically. With processing time in mind, refinements of batch processing are worth considering if only a daily proof of presence is sought.

Whereas it has been believed that the actual *P. fuscus* breeding phase lasts for a few days at some stage during the 3 month breeding period, our observations of seasonal activity suggest that the actual breeding season can be split into distinct episodes. Therefore, a documented peak of activity at one moment followed by the apparent disappearance of vocalizations does not mean that the breeding season is over and another peak might actually follow even several weeks after the first one occurred. Also noteworthy is that there were harbingers and latecomers in *P. fuscus* too as some individuals do not emerge synchronously with the main part of the population. Bridges & Dorcas (1999) found that a population of Southern Leopard Frogs (*Rana sphenocephala*) called intensively in July whereas the species was thought to breed only in early Spring and Fall. Altogether, such information provides important insights, revealing that temporal variation in anuran calling activity warrants further investigation and should be considered when developing anuran monitoring programs.

An explanation of such variation in the vocal activity observed in the circadian and seasonal pattern of vocal activity, which is not necessarily reproducible from one site to another nor to one year to another one, might be that vocal activity can be influenced by various environmental factors including water and air temperature, humidity, depth of the water bodies etc. (Oseen & Wassersug., 2002; Saenz *et al*., 2006; Cui *et al*., 2011). Therefore, further investigations on how ecological constraints might influence vocal activity is also important for establishing or refining monitoring protocols according to the context.

# Acknowledgments

This research received the financial support from the Regional Directorate of Environment Planning and Housing in Alsace (DREAL Alsace) until 2014 and from 2015 to 2016 from the French ministry in charge of Ecology. The authors wish to thank Victoria Michel and Jean-Pierre Vacher, BUFO, and Murielle Diss at Sauer's delta natural reserve for their help all along the monitoring. The authors wish to thank the Conservatoire des Sites Alsaciens for having been allowed to instrument the Sauer site and the Mayor of Mothern to have given access to the Mothern site. Many thanks to Laurent Brendel at Cerema-Est who carried out most of the field work with programmable recorders and helped with the fine-tuning of the detector software by producing an extensive ground truth and also to his colleague Thierry Falwisanner who took part in the field work.

# References

Andreone, F. & Piazza, R. (1990). A bioacoustic study on *Pelobates fuscus insubricus* (Amphibia, Pelobatidae), *Bolletino di zoologia*, **57**(4), 341-349.

Beranek, L. (1986). Acoustics, American Institute of Physics

Council of Europe (1979). *Convention on the Conservation of European Wildlife and Natural Habitats*, COE. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680304355 (Accessed 2017.03.25).

Brandes, T. S. (2008). Automated sound recording and analysis techniques for bird surveys and conservation, *Bird conservation international*, **18**, 163-173.

Brauer, C. L., Donovan, T. M., Mickey, R. M., Katz J. & Mitchell B. R. (2016). A comparison of acoustic monitoring methods for common anuran of the northeasten United States, *Wildlife Society Bulletin*, **40**(1), 140-149.

Bridges, A. S. & Dorcas, M. E. (2000). Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. Copeia 2000:587–592.

Buchel, E. (2010). La rainette verte, *in* (Vacher J.P. Thiriet J. ed.) *Atlas des amphibiens et des reptiles d'Alsace*, BUFO.

Chesmore, D. (2004). Automated bioacoustics identification of species, *Anais da Academia Brasileira de Ciências*, **76**(2), 435-440.

Chesmore, D., Frommolt, K.-H., Wolff, D., Bardeli, R. & Huebner, S. (2008). *Computational bioacoustics: New tools for assessing biological diversity*, side Event at the ninth meeting of the Conference of the Parties (COP 9). Bonn, Germany.

Collins, J. P. & Storfer, A. (2003). Global amphibian declines: sorting the hypotheses, *Diversity and Distributions*, **9**, 89-98.

Cui, J., Song, X., Guangzhan, F., Xu, F., Brauth, S. E. & Tang, Y. (2011). Circadian rhythm of calling behavior in the emei music frog (*Babina daunchina*) is associated with habitat temperature and relative humidity. Asian Herpetological Research 2011, 2(3): 149-154;

Curé, C., Gosset, M. & Dutilleux, G. (2013). Suivi acoustique du Pélobate brun - rapport de campagne 2013, Cerema.

Curé, C. & Dutilleux, G. (2014). Suivi acoustique du Pélobate brun - rapport de campagne 2014, Cerema.

Dutilleux, G., Michel, V. & Vacher J.P.  (2012). Suivi acoustique du pélobate brun en Alsace, Campagne 2011. CETE de l'Est – Bufo.

Dutilleux, G. & Curé, C. (2016). Un système de détection automatique pour le suivi d'un amphibien menacé, le Pélobate brun (*Pelobates fuscus*), Proc. *Congrès Français d'Acoustique*, Le Mans, France, 2342-2347.

Eibl-Eibesfeldt, I. (1956). Vergleichende Verhaltensstudien an Anuren; 2. Zur Paarungsbiologie der Gattungen Bufo, Hyla, Rana und Pelobates. *Verh. dt. zool. Ges.* **19**, 315-23.

Eggert, C. & Guyétant, R. (2003). Reproductive behaviour of spadefoot toads (*Pelobates fuscus*): daily sex ratios and males' tactics, ages, and physical condition. *Canadian Journal of Zoology*, **81**, 46-51.

Eggert, C., Cogalniceanu, D., Veith, M., Dzukic, G. & Taberlet P. (2006). The declining Spadefoot toad, *Pelobates fuscus (Pelobatidae)*: paleo and recent environmental changes as a major influence on current population structure and status, *Conservation genetics*, **7**, 185-195.

Godinat, G. (2010). La grenouille agile, *in* (Vacher J.P. Thiriet J. ed.) *Atlas des amphibiens et des reptiles d'Alsace*, BUFO.

Gosset, M., Dutilleux, G., Michel, V. & Vacher J. P. (2012). Suivi acoustique du pélobate brun en Alsace, Cam- pagne 2012, 31 p, octobre 2012.

European Council, (1992). Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora, EU, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1992L0043:20070101:EN:pdf (Accessed 2017.03.25).

Forrest, T. G., Miller, G. L., & Zagar, J. R. (1993). Sound propagation in shallow water: implications for acoustic communication by aquatic animals. *Bioacoustics*, **4**,259-270.

Frommolt, K. H., Kaufman, M., Mante, S. & Zadow, M. (2008). Die Lautäußerungen der Knoblauchkröte (*Pelobates fuscus*) und Möglichkeiten einer akustischen Bestandserfassung der Art, *Rana*, **Sonderheft 5**, 101-112.

Giovannini, A., Seglie, D. & Giacoma, C. (2014). Identifying priority areas for conservation of spadefoot toad, *Pelobates fuscus insubricus* using a maximum entropy approach, _Biodiversity conservation, **23**, 1-13.

Griffiths, R. A. (1996). Newts and salamanders of Europe, Academic press.

Haykin, S. (2016). Neural networks and learning machines, 3rd edition. Pearson.

IEC International Electrotechnical Commission. (2013). *IEC 61672-1 Electroacoustics - Sound level meters - Part 1: Specifications*, IEC.

IUCN. (2009) *The IUCN Red List of Threatened Species - Pelobates fuscus*, IUCN, http://www.iucnredlist.org/details/16498/0 (Accessed 2017.03.27).

IUCN. (2015). _La Liste rouge des espèces menacées de France - Amphibiens de France Métropolitaine, IUCN, http://uicn.fr/wp-content/uploads/2015/09/Tableau_Liste_rouge_Amphibiens_de_France_metropolitaine.pdf (Accessed 2017.03.27).

ISO. (2016). *ISO 1996-1 Acoustics — Description, measurement and assessment of environmental noise — Part 1: Basic quantities and assessment procedures*, ISO.

Joseph, L., Field, S.A., Wilcox, C. & Possingham H.P. (2006). Presence-absence versus abundance data for monitoring threatened species. Conservation Biology, **20**, 1679-1687.

Köhler, J., Jansen, M., Rodriguez, A., Kok, P. J. K., Toledo, L. F., Emmerich M., … Vences M. (2017). The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice, *Zootaxa* **4251**.

Lescure, J. (1984). La répartition passée et actuelle des Pélobates (Amphibiens Anoures) en France, Bulletin de la Société Herpétologie de France, **29**, 45–59.

Lescure, J., & de Massary, J. C. (Eds.). (2012). Atlas de répartition des amphibiens et des reptiles de France, *Biotope*.

Lurton, X. (2010). *An introduction to underwater acoustics - Principles and applications*, 2nd edition Springer Verlag.

Medwin, H. (2006). *Sounds in the Sea: From Ocean Acoustics to Acoustical Oceanography*, Cambridge.

Mellinger, D. K., Stafford, K. M., Moore, S., Dziak, R. P., & H. Matsumoto. 2007. Fixed passive acoustic observation methods for cetaceans. Oceanography 20:36–45.

Müller, B. (1984). Bio-akustische und endokrinologische Untersuchungen an der Knoblauchkröte *Pelobates fuscus fuscus* (LAURENTI, 1768) (Salientia: Pelobatidae). *Sala-mandra*, **20**, 121-142.

Müller, M. (2007). *Information Retrieval for Music and Motion*, Springer Verlag.

Nöllert, A. (1990). *Die Knoblauchkröte*, Ziemsen Verlag, Wittenberg.

Oppenheim, A. V., Schafer, R. W. (1975). *Digital signal processing*, Prentice Hall.

Oseen, K. L. & Wassersug, R. J. (2002). Environmental factors influencing calling in sympatric anurans. Oecologia, 133: 616-625.

Saenz, D., Fitzgerald L. A., Baum K. A. & Conner R. N. (2006). Abiotic correlates of anuran calling phenology: the importance of rain, temperature and season. Herpetological Monographs, 20: 64-82.

Seglie, D., Gauna, A., & Giacoma, C. (2013). Description of the male advertisement call of *Pelobates fuscus insubricus (Anura, Pelobatidae)*, with general notes on its acoustic repertoire, *Bulletin de la Société Herpétologique de France*, **145-146**, 61-72.

Steelman, C. K., & Dorcas M. E. (2010). Anuran calling survey optimization: developing and testing predictive models of anuran calling activity. *Journal of Herpetology*, **44**, 64-68.

Sutherland, W.J. (Ed.). (2006). *Ecological census techniques*, Cambridge.

ten Hagen, L., Rodríguez, A., Menke, N., Göcking, C., Bisping, M., Frommolt, K. H,… Vences, M. (2016). Vocalizations in juvenile anurans: common spadefoot toads (*Pelobates fuscus*) regularly emit calls before sexual maturity, *The Science of Nature*, **103**:75, 8p.

Vacher, J.P. & Dutilleux G. (2010), Pélobate brun, *in* (Vacher, J. P. & Thiriet, J. ed.) *Atlas des amphibiens et des reptiles d'Alsace*, BUFO.

Vacher, J. P. (2010). La grenouille de Lessona, la grenouille verte et la grenouille rieuse, *in* (Vacher, J. P. & Thiriet, J. ed.) *Atlas des amphibiens et des reptiles d'Alsace*, BUFO.

Van Gelder, J. J., & Hoedemaekers H. C. M. (1971). Sound Activity and Migration During the Breeding Period of Rana temporaria L., R. arvalis Nilsson, Pelobates fuscus Laur. and Rana esculenta L., *Journal of Animal Ecology*, **40**(3), 559-568.

Virtanen, T., Plumbley, M.D., & Ellis D.  (2017) Computational Analysis of Sound Scenes and Events, Springer.

Waddle, J. H., Thigpen, T. F., & Glorioso, B. M. (2009). Efficacy of Automatic Vocalization Recognition Software for Anuran Monitoring. *Herpetological Conservation and Biology* **4**(3), 384-388.

Wahlberg, M., Toogard, J., & Møhl; B. (2003). Localising bitterns *Botaurus stellaris* with an array of non-linked microphones, *Bioacoustics*, **13**, 233-245.

Weir, L. A., Royle, J. A. , Nanjappa, P. & Jung, R. E. (2005). Modeling anuran detection and site occupancy on North American Amphibian Monitoring Program (NAAMP) routes in Maryland. *Journal of Herpetology*, **39**,627–639.

Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations*. Academic Press.

# Tables

Table 1 : Parameters of the detector of *P. fuscus* advertisement call.

| **Parameter** | **Value** | **Unit** |
|---|---|---|
| a- Time constant τ for sound levels | 0.15 | s |
| b- Minimum peak duration (Fig. 1) | 0.06 | s |
| c- Maximum peak duration (Fig. 1) | 0.25 | s |
| d- Maximum delay between one peak and another in a group (Fig. 1) | 0.5 | s |
| e- Minimum difference Ltau-L90 | 5 | dB re 1 |
| f- Minimum difference LtauPf-LtauLF | 4 | dB re 1 |
| g- Minimum Lmax of peaks | 22 | dB re 1 |
| h- Maximum standard deviation of peak levels in a group | 4 | dB re 1 |

Table 2 : Overview of the automated acoustic recording data collected on the 4 studied sites. Leutenheim1 and 2 recording files were 60-min long whereas Mothern and Sauer files lasted 5min. Part of data collected in these 4 sites were used for testing the performance of the detector. The full audio dataset gathered in Mothern and Sauer campaigns was then processed with the detector to monitor *P. fuscus* presence and to gain information of the vocal activity patterns of the species.

| **Site** | **Year** | **Start date of recording (dd/mm)** | **Stop date of recording (dd/mm)** | **# Days of recording** | **# Files** | **Total file size (GB)** | **Total recording time (hours)** |
|---|---|---|---|---|---|---|---|
| Mothern | 2015 | 11/03 | 23/06 | 101 | 4736 | 45 | 395 |
| Mothern | 2016 | 30/03 | 12/07 | 76 | 3658 | 35 | 305 |
| Sauer | 2015 | 11/03 | 03/07 | 115 | 5470 | 52 | 456 |
| Sauer | 2016 | 30/03 | 25/05 | 55 | 2641 | 25 | 220 |
| Leutenheim 1 | 2017 | 10/04 | 18/05 | 39 | 266 | 51 | 266 |
| Leutenheim 2 | 2017 | 10/04 | 18/05 | 39 | 266 | 51 | 266 |

Table 3: Acoustic analysis conducted on N=20 calls of *P. fuscus* collected over 2 years on 4 distant breeding sites.

|  | Mean | SD | Min | Max |
|---|---|---|---|---|
| Dur1 (s) | 0.093 | 0.018 | 0.067 | 0.131 |
| Dur2 (s) | 0.078 | 0.020 | 0.057 | 0.140 |
| Int (s) | 0.159 | 0.018 | 0.125 | 0.185 |
| FAmax1 (Hz) | 758 | 160 | 431 | 1078 |
| FAmax2 (Hz) | 805 | 169 | 570 | 1054 |

Table 4 : Detection results of the detector against the ground truth (N=1200 30sec audio files coming from the 4 studied sites).

| **Site** | **Year** | **True Positives** | **True Negatives** | **False Positives** | **False Negatives** | **Total # of files** |
|---|---|---|---|---|---|---|
| Mothern | 2015 | 52 | 122 | 3 | 23 | 200 |
| Mothern | 2016 | 25 | 150 | 3 | 22 | 200 |
| Sauer | 2015 | 25 | 166 | 0 | 9 | 200 |
| Sauer | 2016 | 20 | 165 | 3 | 12 | 200 |
| Leutenheim 1 | 2017 | 0 | 198 | 2 | 0 | 200 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Leutenheim 2 | 2017 | 0 | 199 | 1 | 0 | 200 |
| All | All | 122 | 1000 | 12 | 66 | 1200 |

# Figure Legend

Figure 1. Typical advertisement call of *P. fuscus* and relevant time parameters for the automated detector. On top the typical $L_{tau}$ pattern for a single “cloc”.

Figure 2. Flow chart of the automatic part of the *P. fuscus* detector. Parallelograms stand for data or (intermediate) results, rectangles for processes. BP = band pass LF = Low Frequency, Pf = P. fuscus, # = number of. Letters on the left refer to rows in Table 1.

Figure 3. Acoustic variables measured on an advertissment call of *Pelobates fuscus* using Avisoft SAS LabPro software (version 4.39, Avisoft Bioacoustics, Glienicke, Germany) software. Spectrogram (top panel) and amplitude envelope (middle panel) showing the two notes (“cloc-cloc”) of a typical call. Temporal parameters Dur1 (duration of the first note), Dur2 (duration of the second note) and Int (interval duration between the two notes) were measured on the amplitude envelope. In the frequency domain, we measured for each note the frequency of maximum amplitude (FAmax1 and FAmax2, respectively), i.e. carrying most energy. These two frequency measured parameters were obtained from the power spectrum (Hamming window, FFT length: 1024). Power spectrum of the first note of the call (bottom panel) shows the detected spectral peak (black cross) corresponding to the frequency of maximum amplitude of the note 1, FAmax1. Similarly, FAmax2 value was taken from the power spectrum of the second note of the call.

Figure 4: Confidence of the detector on the available long-term experimental data. The error rate is defined at the file level as the number of false positives divided by the total number of files with non-zero call count. The horizontal axis is adjusted to the maximum call count per 5-min file in the data set.

Figure 5. Validated detections of *P. fuscus* during Sauer 2015 campaign. Validated means after validation by a human operator. The color code indicates the number of detections.

Figure 6. Validated detections of *P. fuscus* during Mothern 2015 campaign. Validated means after validation by a human operator.

Figure 7. Validated detections of *P. fuscus* during Sauer 2016 campaign. Validated means after validation by a human operator.

Figure 8. Validated detections of *P. fuscus* during Mothern 2016 campaign. Validated means after validation by a human operator.

Figure 9. Circadian distribution of *P. fuscus* advertisement call counts based on Mothern 2015 data (n=78986).

# Figures

Figure 1

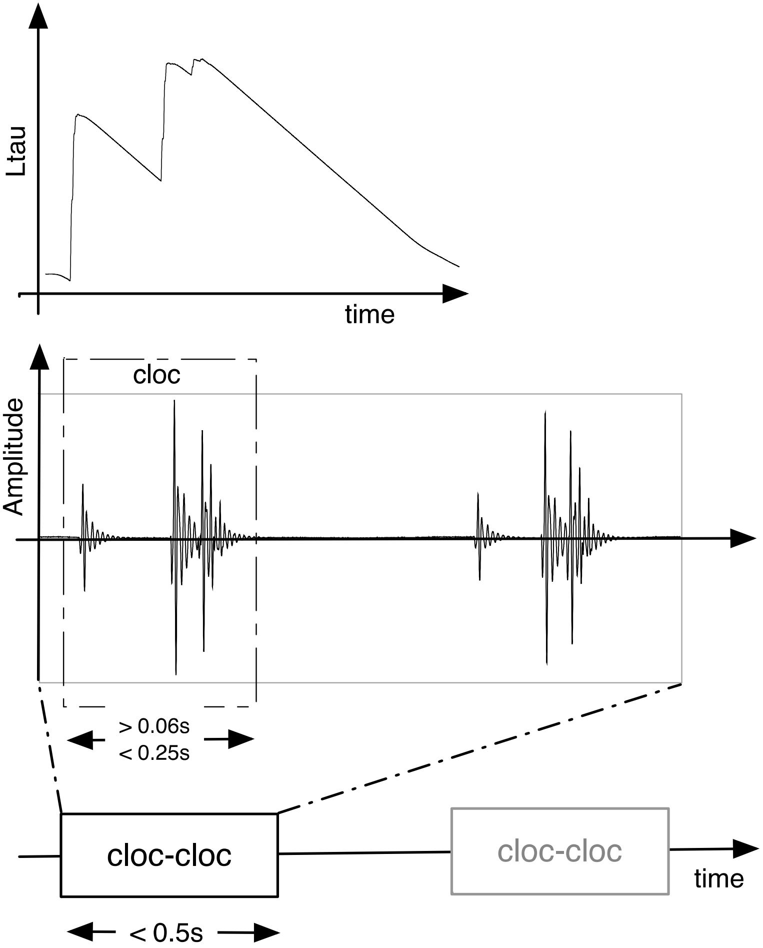

Ltau
time
Amplitude
cloc
> 0.06s
< 0.25s
cloc-cloc
cloc-cloc
time
< 0.5s

Figure 2

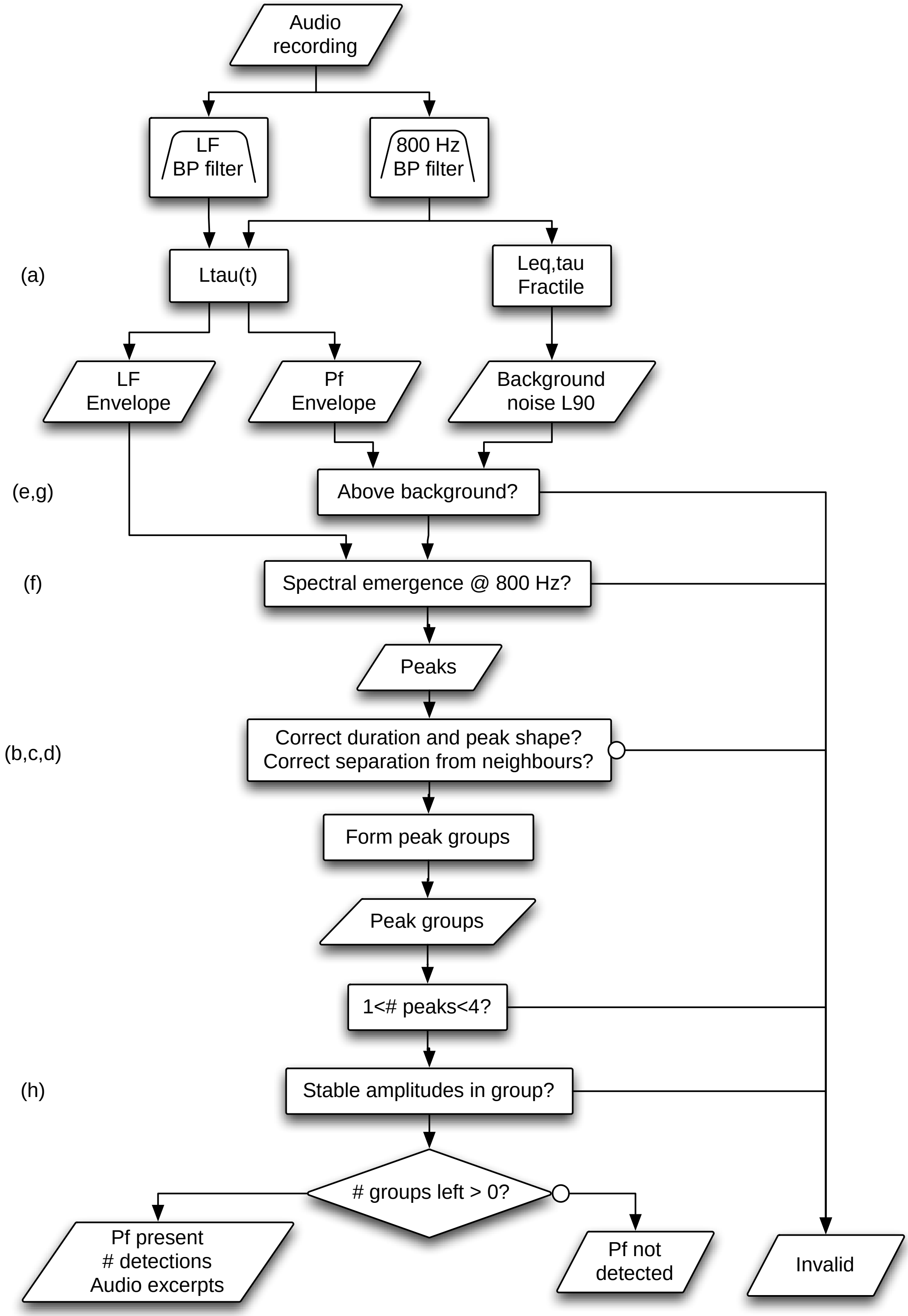

Audio recording
LF BP filter
800 Hz BP filter
(a)
Ltau(t)
Leq,tau Fractile
LF Envelope
Pf Envelope
Background noise L90
(e,g)
Above background?
(f)
Spectral emergence @ 800 Hz?
Peaks
(b,c,d)
Correct duration and peak shape?
Correct separation from neighbours?
Form peak groups
Peak groups
1<# peaks<4?
(h)
Stable amplitudes in group?
# groups left > 0?
Pf present
# detections
Audio excerpts
Pf not detected
Invalid

Figure 3

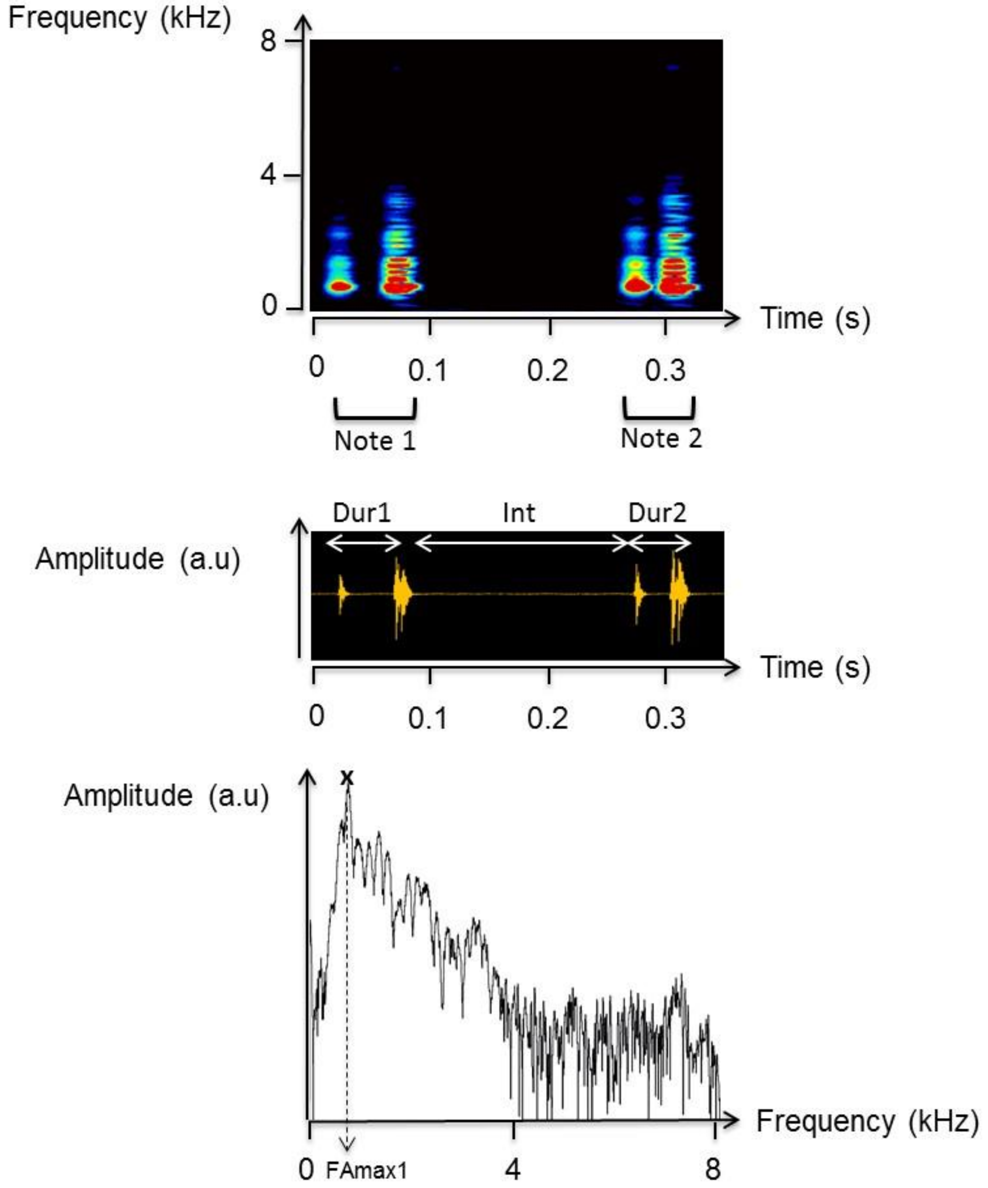

Frequency (kHz)
8
4
0
Time (s)
0
0.1
0.2
0.3
Note 1
Note 2
Dur1
Int
Dur2
Amplitude (a.u)
Time (s)
0
0.1
0.2
0.3
Amplitude (a.u)
x
Frequency (kHz)
0
FAmax1
4
8

Figure 4

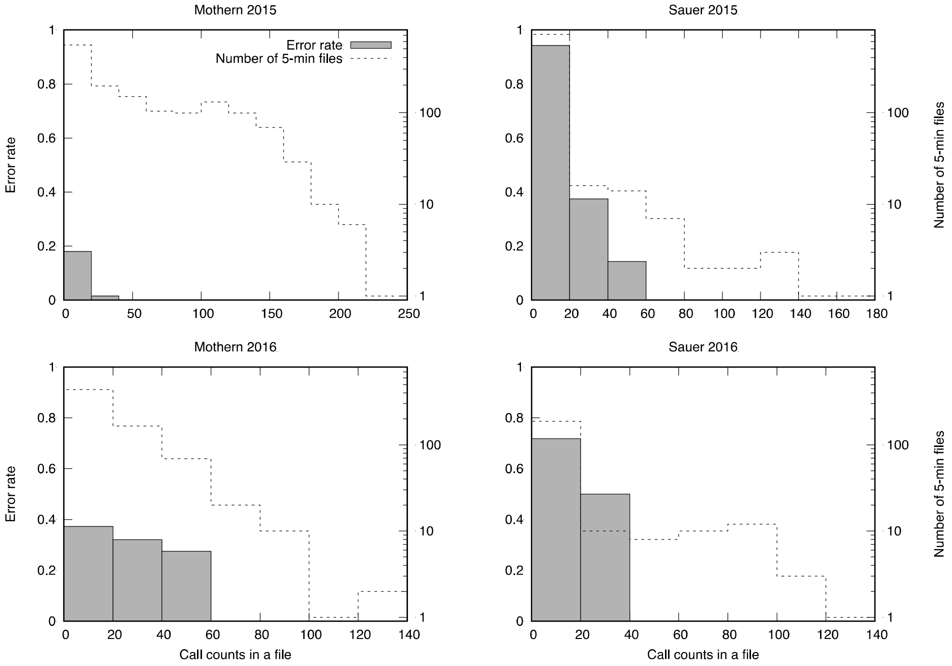

Mothern 2015
Sauer 2015
Mothern 2016
Sauer 2016
Error rate
Number of 5-min files
Error rate
Number of 5-min files
Call counts in a file
0
0.2
0.4
0.6
0.8
1
1
10
100
0
50
100
150
200
250
0
20
40
60
80
100
120
140
160
180

Figure 5

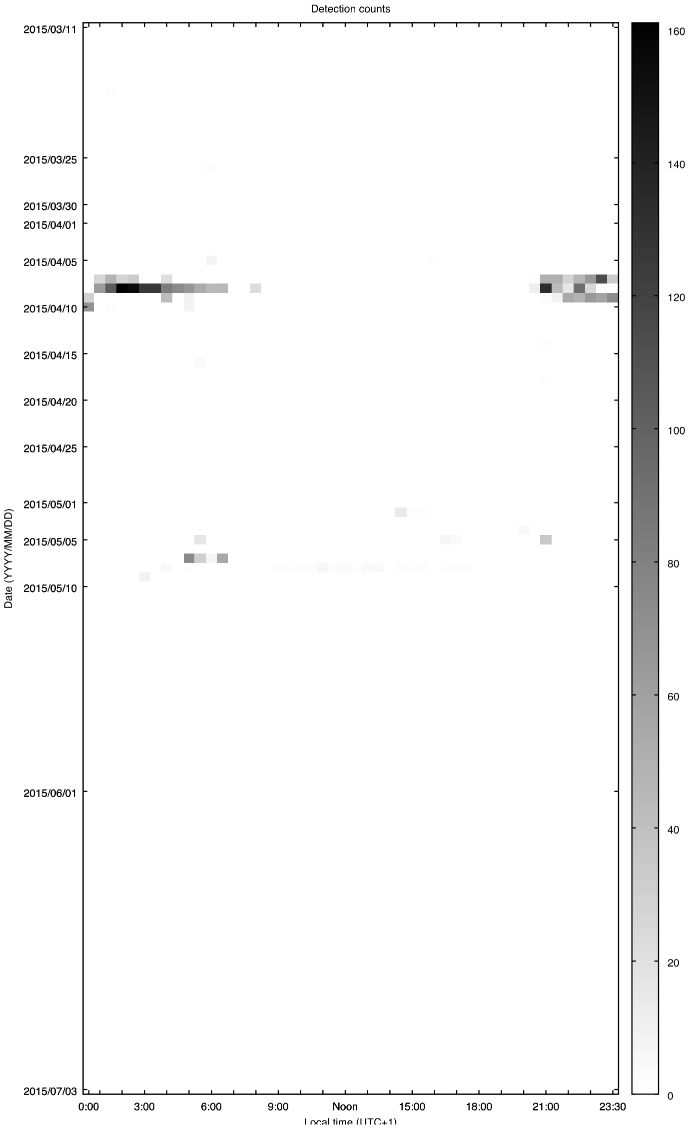

Detection counts
Date (YYYY/MM/DD)
2015/03/11
2015/03/25
2015/03/30
2015/04/01
2015/04/05
2015/04/10
2015/04/15
2015/04/20
2015/04/25
2015/05/01
2015/05/05
2015/05/10
2015/06/01
2015/07/03
0:00
3:00
6:00
9:00
Noon
15:00
18:00
21:00
23:30
Local time (UTC+1)
160
140
120
100
80
60
40
20
0

Figure 6

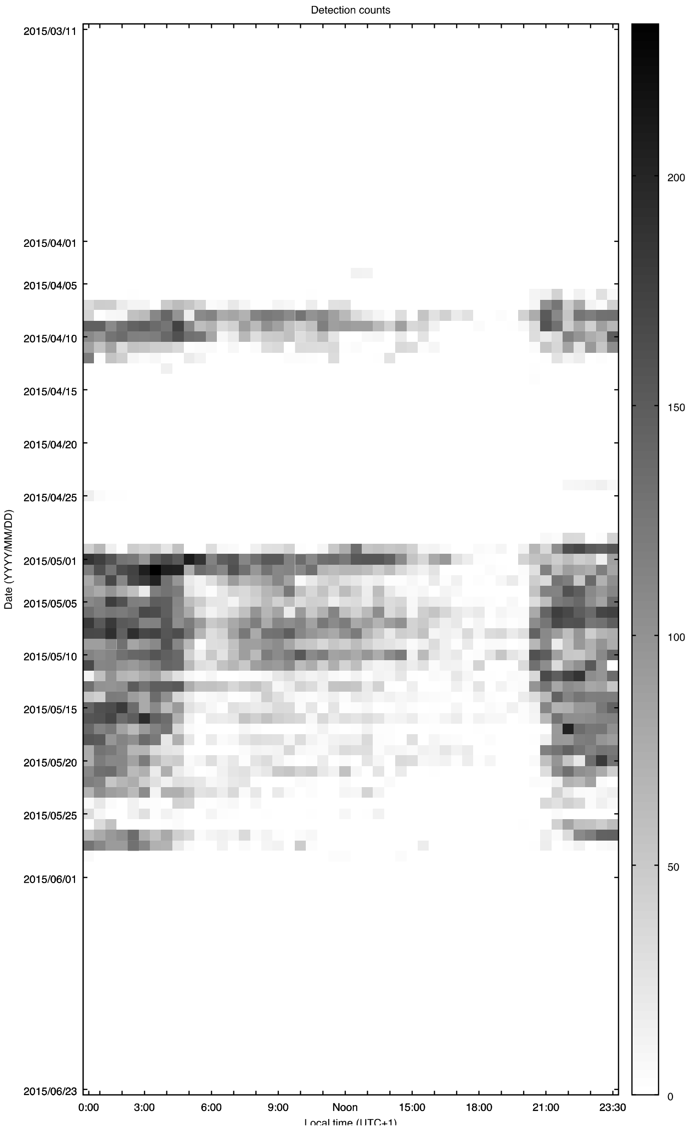

Detection counts
Date (YYYY/MM/DD)
2015/03/11
2015/04/01
2015/04/05
2015/04/10
2015/04/15
2015/04/20
2015/04/25
2015/05/01
2015/05/05
2015/05/10
2015/05/15
2015/05/20
2015/05/25
2015/06/01
2015/06/23
0:00
3:00
6:00
9:00
Noon
15:00
18:00
21:00
23:30
Local time (UTC+1)
200
150
100
50
0

Figure 7

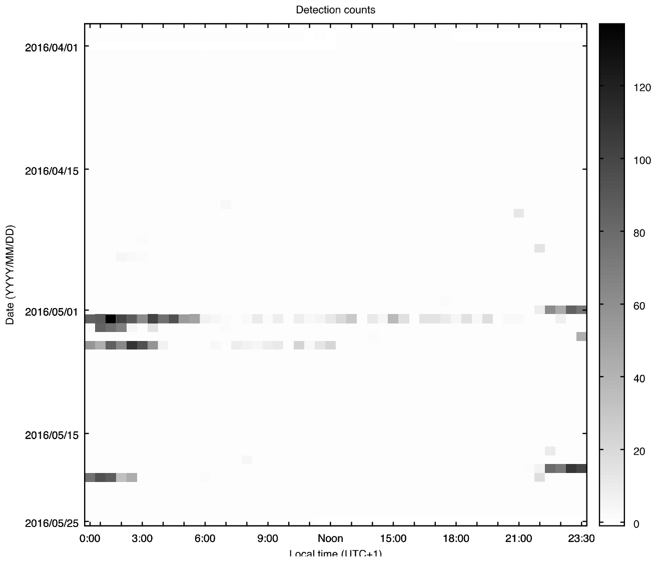

Detection counts
Date (YYYY/MM/DD)
2016/04/01
2016/04/15
2016/05/01
2016/05/15
2016/05/25
0:00
3:00
6:00
9:00
Noon
15:00
18:00
21:00
23:30
Local time (UTC+1)
0
20
40
60
80
100
120

Figure 8

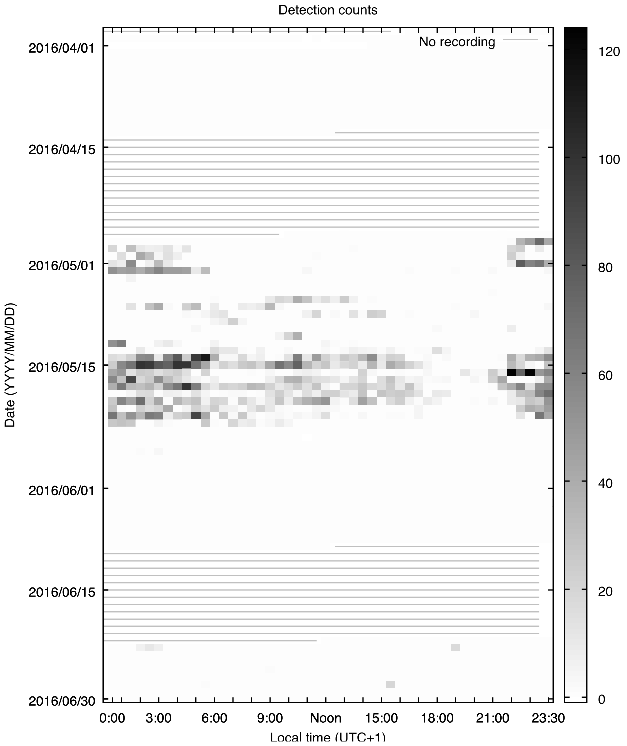
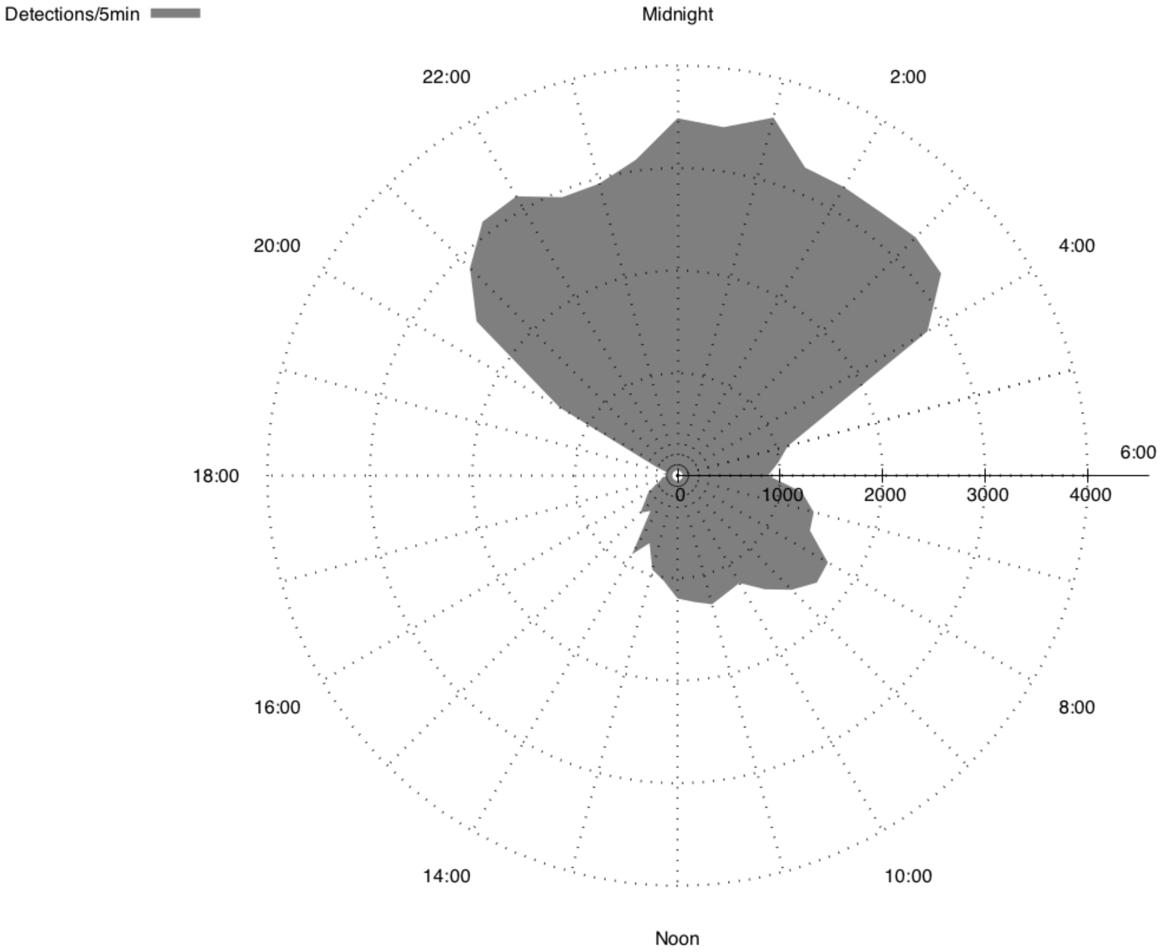

Detection counts
No recording
2016/04/01
2016/04/15
2016/05/01
2016/05/15
2016/06/01
2016/06/15
2016/06/30
Date (YYYY/MM/DD)
0:00
3:00
6:00
9:00
Noon
15:00
18:00
21:00
23:30
Local time (UTC+1)
120
100
80
60
40
20
0

Figure 9


Detections/5min
Midnight
22:00
2:00
20:00
4:00
6:00
18:00
0
1000
2000
3000
4000
16:00
8:00
14:00
10:00
Noon