Aquat. Living Resour. 29, 204 (2016)
© EDP Sciences 2016
DOI: 10.1051/alr/2016015
www.alr-journal.org

Aquatic
Living
Resources

Automatic acoustic recognition of shad splashing using a smartphone

Daniel Diep[1], Hervé Nonon[2], Isabelle Marc[1,a], Isabelle Lebel[3] and Frédéric Roure[4]

[1] Ecole des Mines d'Alès, Laboratoire de Génie Informatique et Ingénierie de Production, Parc scientifique Georges Besse, 30000 Nîmes, France
[2] DIVULCO Résidence Fontcarrade 2-E5, 728 rue de Fontcarrade, 34070 Montpellier, France
[3] Association Migrateurs Rhône Méditerranée (MRM), Z.I. Nord, rue André Chamson, 13200 Arles, France
[4] GECO Ingénierie, Le Clavelet, Port Fluvial, Route de Bagnols, 30290 Laudun L'Ardoise, France

Received 30 November 2015; Accepted 17 April 2016

Abstract – Monitoring the numbers of shad (*Alosa fallax rhodanensis*, Rhodanian twaite shad) at their reproduction sites in the Rhone basin is an important step for measuring inter-annual changes in their population size. Manual counting involves listening to detect the sounds of splashes produced by shad during spawning. This is a costly operation, requiring high resource levels under difficult working conditions. In order to automatically estimate the number of migrating shad in rivers, an acoustic signal analysis method is proposed. It is based on short-term spectral analysis, combined with a Gaussian mixture model. Implemented on a smartphone, the application provides a number of advantages, such as mobility, audio recording, spawning detection and counting, and means of communication. The results obtained are very promising, and the deployment of such a device is expected to be of great help for counting shads and locating their spawning sites.

Keywords: signal processing / audio recording / shad spawning detection / *Alosa fallax rhodanensis* / smartphone application

1 Introduction

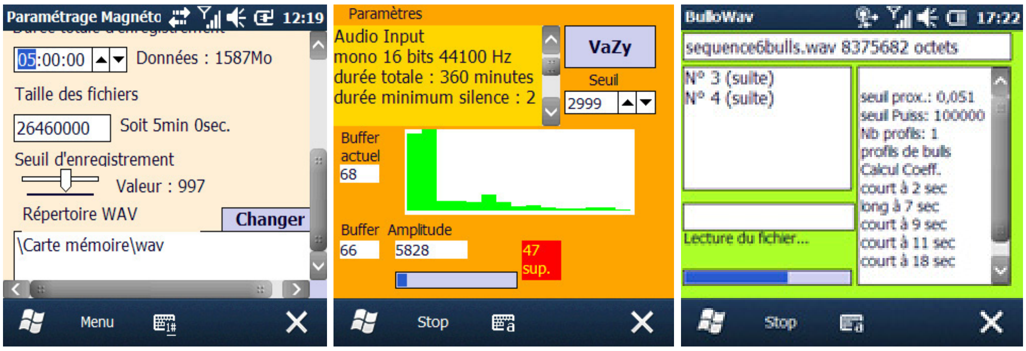
The twaite shad (*Alosa fallax rhodanensis,* Rhodanian twaite shad) is a migratory fish species living primarily in the sea which swims upstream in rivers in spring to spawn. In Europe, this species has declined considerably since the mid-20th century due to overfishing, pollution and obstacles to migration, and for this reason is now given considerable legal protection (Kirchhofer et al. 2012). In France, the Committee for the Management of Migratory Fish (COGEPOMI) in the Rhone-Mediterranean basin has undertaken a series of actions aimed at restoring the presence of migratory species in rivers (Lebel et al. 2007). In particular, monitoring the numbers of shad at their spawning sites in the Rhone basin is an important step for measuring inter-annual changes in their population size. The monitoring provides information on the vulnerability of the species, threatened by dams, pollution and the deterioration of spawning grounds, and also assesses the effectiveness of structures such as sluices and fish passes, created to facilitate their annual upstream migration.

Shad spawn at night close to the water surface, turning quickly and noisily at the time of spawning and emitting a characteristic sound lasting a few seconds known as a "spawning splash". The current method is manual counting from the river bank by an observer who listens and counts the splashes (Lebel et al. 2001; Chanseau et al. 2004).

This manual counting method is highly restrictive (inconvenient times, tedious work) and costly in terms of human resources. Recently, thanks to technological advances in the field of multimedia, particularly audio media, counting devices using microphones and portable recorders have been developed. However, they still require considerable intervention on the part of the operator, including installation, monitoring of the equipment, shut down, listening to the recordings, splash identification and counting.

To make further progress, the COGEPOMI has set up a study to investigate the possibility of automating shad monitoring by means of audio recordings. The objective of the first 4-year study (2004 to 2008) was to design a field device for recording and automatically counting the audio signals of shad splashes. In a second phase (2010–2014), a prototype based on a smartphone application providing shad counts online was developed and deployed, and its performances evaluated. The first advantage of this approach is improved count accuracy, as automatic recordings can cover the whole spawning period. Secondly, it reduces the tediousness of nocturnal monitoring

[a] Corresponding author: isabelle.marc@mines-ales.fr



05:00:00 Données : 1587Mo

Taille des fichiers

26460000 Soit 5min 0sec.

Seuil d'enregistrement

Valeur : 997

Répertoire WAV

\Carte mémoire\wav

Audio Input
mono 16 bits 44100 Hz
durée totale : 360 minutes
durée minimum silence : 2

Seuil 2999

Buffer actuel 68

Buffer 66

Amplitude 5828

47 sup.

sequence6bulls.wav 8375682 octets

N° 3 (suite)
N° 4 (suite)

Lecture du fichier...

seuil prox.: 0,051
seuil Puiss: 100000
Nb profils: 1
profils de bulls
Calcul Coeff.
court à 2 sec
long à 7 sec
court à 9 sec
court à 11 sec
court à 18 sec

Fig. 1. Smartphone user interface developed to automatically count shad spawning splashes.

tasks and considerably decreases costs in terms of human resources. Thirdly, it enables the monitoring to be systematised and performed on a regular basis. Finally, the equipment can be used to search for new shad spawning sites. The work presented in this article describes the technological choices and the methods involved in the design of a portable monitoring device, from audio capture to shad splash counting.

Mobile technology has been a major component of the design process. Smartphones can now combine a large number of functions in a few square centimetres, including audio recording, high storage capacity, power autonomy, wireless communication and GPS. The idea was to benefit from the smartphone's range of features and high level of integration, rather than to develop a new device from scratch.

Using a smartphone to replace experimental instrumentation also introduces a number of limitations: limited characteristics of the audio sensor, insufficient power autonomy, and poor computing capability. In addition, as a new consumer electronics device, smartphones are subject to rapid obsolescence, which makes software development and maintenance rather difficult.

2 Materials and methods

2.1 Motivation

Few studies have used acoustic records to monitor shad or related species (allis shad, twaite shad, etc.), and even fewer have carried out automatic recognition of acoustic signals. Trouilhet et al. (1993) and Coustaux et al. (1994) trialled spectral analysis and classification by neural networks. However, the results showed a lack of robustness and excessive sensitivity to environmental noise.

The work presented here follows a different stepwise approach, guided by the following pragmatic considerations: (i) Design of a mobile device, suitable for different field conditions, as most spawning zones are difficult to access; (ii) capture and storage of good-quality audio recordings as a basic function; and (iii) detection and counting of shad spawning acts (SSAs) first to be used simply as an aid for manual counting, then to be tested and improved before being implemented on the portable device.

An overview of this work is presented in Diep et al. (2013), with the principle of the method for detecting SSAs being described in Diep et al. (2013). Details of the implementation and further results are presented here.

2.2 Data acquisition

Various sensors were tested for this study, ranging from simple recorder microphones to long-range microphones designed for wildlife sound recording, including hydrophones commonly used to listen to marine mammals. The Sony ECM PP1C outdoor microphone was found to be a good compromise between acceptable performance in terms of range and coverage of a spawning zone, and moderate size, being equipped with a 17 cm diameter parabolic dish. Unfortunately, Sony stopped manufacturing this microphone and a substitute was created combining an Olympus ME52W microphone with a 3D printed tailor-made parabolic dish.

A standard commercial smartphone was employed as a recording device for the application. The Samsung GT-B7350 offers basic smartphone functions such as 3G cellphone service, WiFi and Bluetooth communication, a micro SD memory card, and operates under Windows Mobile 6.5. To maintain high audio quality, a sampling rate of 44 100 Hz was chosen. To adapt the input and impedance level, an external preamp was inserted between the microphone and the smartphone.

2.3 Smartphone application

A dedicated software application was developed to meet the requirements of the device. Figure 1 shows various screen shots of the software application. The smartphone application has the following components and functionalities:

- Autonomy: an external battery was wired to attain at least seven days of power autonomy, extendable with the addition of photovoltaic cells. Storage autonomy is about the same with a 32 Gb SD card, and can easily be extended by signal compression, e.g. MP3 (not tested for detection) or by deletion of blank signals.
- Task scheduler: this component enables applications to be started and stopped at pre-defined times. It allows the smartphone to record for four hours each night for example and remain on standby for the rest of the day.

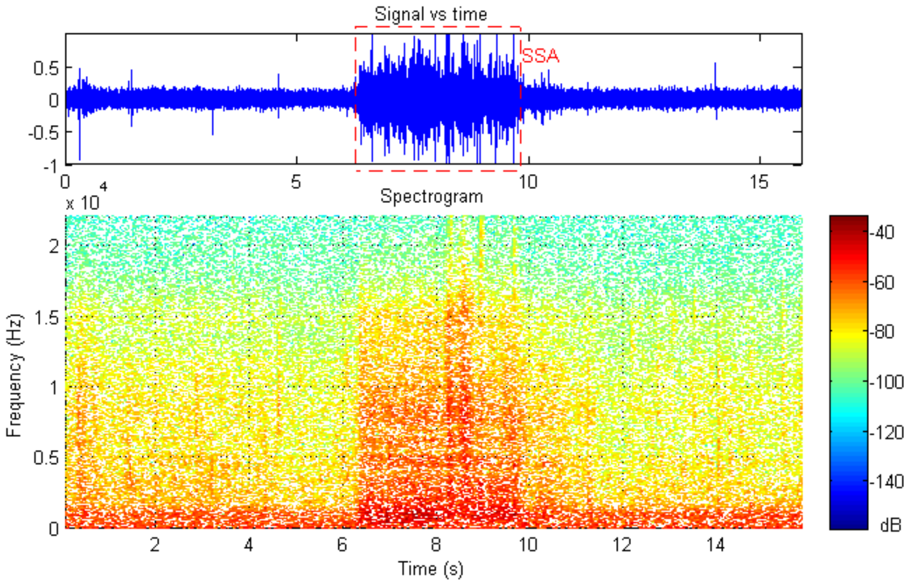


Fig. 2. Signal amplitude and frequency spectrogram of a shad spawning act (SSA). Signal amplitude is expressed in Volts (V) and the spectrogram of PSD (Power Spectral Density) in decibels/Hertz (dB/Hz) at the sampling frequency of 44 100 Hz. PSD = 10 $\log_{10}(P/P_0)$, with power P in V^2/Hz and $P_0 = 1$.

- Remote control via SMS: a functionality designed to monitor the smartphone remotely, for example from the river bank when the smartphone is mounted at the end of a pole or a branch, or from any other location.
- Detection of SSAs: this functionality is implemented on the smartphone and can be activated either online or offline depending on processor load.

2.4 Detection of SSAs

Acoustic signals generated by shad spawning acts are somewhat difficult to analyse and to properly discriminate from other water sounds because of their non-stationary and non-harmonic nature. Physically, they are produced by turbulent flows which are very complicated to model. Water sounds are mainly produced by air bubbles that are trapped in the water and vibrate. Van Den Doel (2005), modelling isolated bubbles with a sound model, showed that numerical simulations were able to reproduce different water sounds. Guyot et al. (2013) used the same physical model to determine the time-frequency localisation of water drops in a spectrogram and to identify different water sounds in an indoor environment.

We used a Short-Term Fourier Transform (Rabiner and Schafer 2010) to analyse spectrograms of SSAs (Fig. 2). Visual inspection of spectrograms of shad splashes indicates that the spectra are located over a rather broad range of frequencies, from 100 to 5000 Hz, and that no regular structures can be identified. Instead, SSA sounds can be viewed as a set of small clusters appearing irregularly in the time-frequency representation (Fig. 3). Based on this observation, the proposed SSA detection method consists of the following steps:

Feature extraction: the audio signal sampled at 44 100 Hz is processed in frames of 93 ms each, which roughly corresponds to the width of a cluster. For each frame, the energy of the signal within the frame is computed using Fast Fourier Transform (FFT) across 10 spectral bands covering a frequency range from about 100 Hz to 5000 Hz on a logarithmic scale. The ten triangular filters are centred on 100, 274, 485, 742, 1055, 1435, 1899, 2426, 3149, and 3984 Hz, corresponding to a linear progression on a mel scale (Fig. 4), which is known as a perceptual scale adapted to human hearing (Rabiner and Schafer 2010). The following standard formula was employed to convert frequency (f) into mel scale (m):

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right).$$

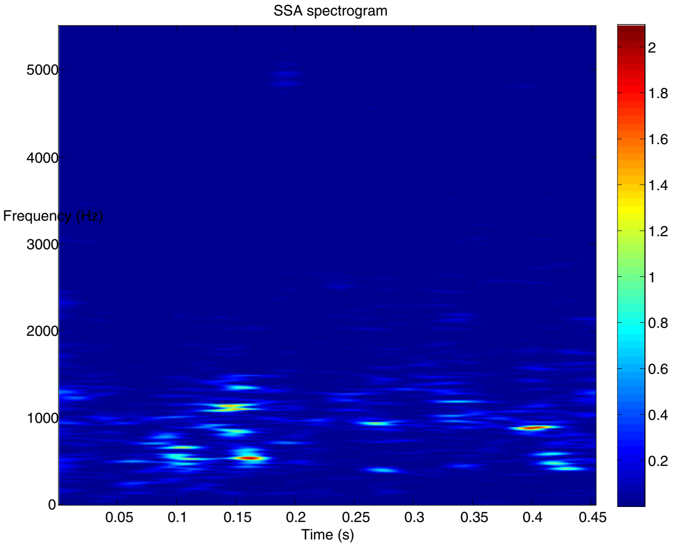


Fig. 3. Spectrogram (partial view, linear scale) of a shad spawning act (SSA). Right hand scale for signal power ($Volt^2$).

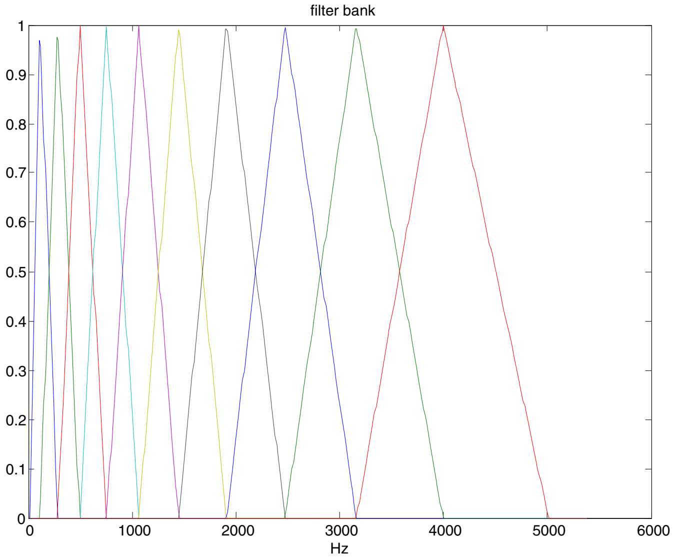


Fig. 4. Bank of ten triangular spectral band filters corresponding to a linear progression on a mel scale used for feature extraction of the shad spawning act signal for each timeframe using Fast Fourier Transform.

Classification: classification of signals in the acoustic domain was carried out using statistical methods such as hidden Markov models (HMM), artificial neural networks (NN) or support vector machines (SVM), which are all supervised learning methods (Rabiner and Schafer 2010). In Diep et al. (2013) an unsupervised classification method based on a Gaussian Mixture Model (GMM) was preferred over supervised classification methods, to take into account the irregularity and compound nature of the SSA sound signal. In a GMM, data are generated by a mixture of Gaussian probability distributions, where each distribution can be associated with a different cluster.

Filtering: an SSA is detected each time a frame of 93 ms has been classified as such. A low pass filter is then necessary to smooth the detection signal. The presence of an SSA is confirmed if the detection exceeds a given threshold θ_T for at least two seconds, which is the generally accepted duration for an actual spawning act.

Simplified SSA detection: in order to implement the detection method on a smartphone with limited computational capabilities, a simplified classification algorithm was chosen. We considered the ten spectral components $T_i\,(k)$, $i = 1\ldots10$, obtained for frame number k and forming vector $T(k)$. We also consider a set of spectral components S_i, $i = 1\ldots10$, extracted from a data set only composed of shad splash signals. The S_i components can be calculated using a clustering method like GMM, or more directly by a simple average. Then we form the Euclidian distance $D(k)$ between vectors $T(k)$ and S

$$D(k) = \sqrt{\sum_i (T_i\,(k) - S_i)^2}$$

and define the SSA detector as the result of the test: $D(k) < \theta_\mathrm{S}$, where θ_S is the threshold corresponding to the radius of a circle around the point defined by the vector S in the domain of spectral components. Vector S is viewed as a spectral signature of SSA, which can be pre-calculated by a learning method for a data set with known SSA presence.

This simplified detector was implemented real-time in the smartphone, the more critical part of the program code being the computation of the FFT.

Thresholding

Briefly, parameterising the counting method mainly consists in adjusting three types of thresholds. These are:

- θ_p, threshold for the minimum signal power, which limits the analysis to signals having sufficient power to be distinguished from ambient noise;
- θ_s, threshold for the maximum spectral distance, which selects signals in a frequency neighbourhood of an SSA;
- θ_t, threshold for the minimum duration of a filtered signal, with signals exceeding this threshold time duration being identified as SSAs.

The first threshold is linked to the instantaneous signal power, or signal energy over a timeframe. Spectral components $T(k)$ must first be normalised (i.e. divided) by the instantaneous signal power, so as to be independent from the amplitude of the received signal. This normalisation enables adaptation and correct identification of SSA signals over a wide range, as the signal power roughly varies in inverse proportion to the distance between the sound source and the receiver. However, signals of low amplitude may be confused with background noise, and thus have to be eliminated using this threshold θ_p, expressed in decibels (dB).

The second threshold θ_s was mentioned above. It defines a basin of attraction around the spectral signature S characterising an SSA and acts as a test of similarity. The current piece of signal is identified as SSA if its spectral composition $T(k)$ is close enough to S. As $T(k)$ and S are normalised vectors, θ_s is a dimensionless quantity between 0 and 1.

Finally, the threshold θ_t acts upon the rate of the similarity test, i.e. the signal obtained after detection, and smoothed over time by low-pass filtering. It enables the elimination of short splashes, which frequently occur in turbulent river flows. θ_t is a dimensionless number between 0 and 1.

2.5 Evaluation

To evaluate the efficiency of the SSA detection algorithm, we compared counts obtained by manual and automatic detection. Signal segments associated with SSAs were annotated manually and automatically, and the agreement was reported in a confusion matrix.

As SSAs were only rare events in the case study records, true negative cases were not considered, and three global performance indicators were calculated:

precision: $\mathrm{Pr} = \frac{TP}{TP+FP}$; recall: $\mathrm{Re} = \frac{TP}{TP+FN}$; F-score: $F = 2\frac{\mathrm{Pr}*Re}{\mathrm{Pr}+Re}$, with TP true positive, FP false positive and FN false negative automatic detections of SSAs. The F-score is the harmonic mean of precision and recall and represents a trade-off between false detections and non-detections.

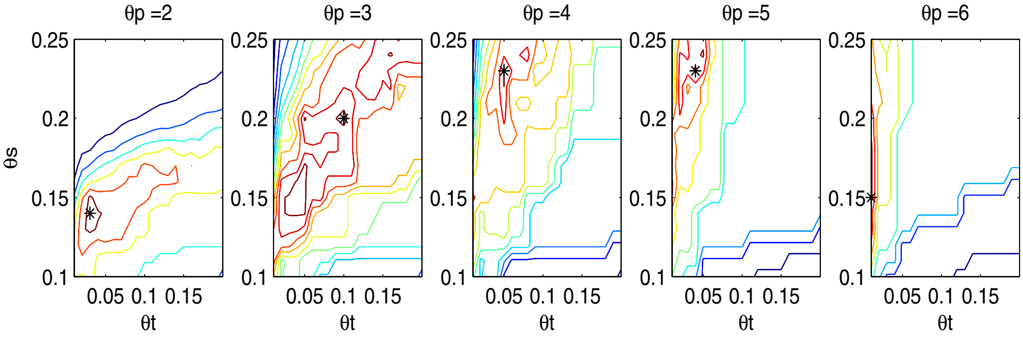


Fig. 5. F-score classification performance results of shad spawning acts recorded in the Vidourle river on 28/5/2015 for different threshold values (θ_p minimum signal power; θ_s maximum spectral distance; θ_t minimum duration of filtered signal). The scale for F-score values ranges from high (red) to low (blue) with maximum indicated by *.

Table 1. Case study information for the Vidourle river, number of shad spawning acts from manual counting and automatic detection, and performance measures for automatic detection compared to manual counting. Threshold values used: $\theta_p = 4$ dB; $\theta_s = 0.23$; $\theta_t = 0.05$ (see text for explanation).

Date	Duration	Manual counting	Automatic counting	True positive	False positive	False negative	Precision	Recall	F-score
2015/05/28	3h55′	22	21	14	7	8	0.667	0.636	0.651
2015/05/30	3h30′	3	2	1	1	2	0.500	0.333	0.400
2015/06/01	3h30′	6	8	5	3	1	0.625	0.833	0.714
2015/06/03	3h47′	5	6	5	1	0	0.838	1.000	0.909
All		36	37	25	12	11	0.676	0.694	0.685

2.6 Case study

The spawning ground for which the test data set was collected is situated in Saint-Laurent-d'Aigouze, on the Vidourle river, a small Mediterranean river. The spawning period usually lasts about eight weeks, from the beginning of April to the beginning of June. Manual counting of shad splashes started in 2008. An average of 10 to 50 spawning acts were recorded per night, with differences between years.

The first tests of the prototype methods began in 2012, but it was only in 2015 that all problems were corrected. The current enhanced smartphone fulfils all basic functionalities such as recording audio signals at predefined times, detecting SSAs and communicating via SMS.

Initially, real-time detection of SSAs with the smartphone was tested using predefined thresholds, which were derived from previous experiments. Unfortunately, the results varied from day to day, depending on conditions. Thus, optimal threshold values seem to depend on the layout of the installation, e.g. position and orientation of the sensor, and site conditions, such as spawning site configuration, hydrological or weather conditions, and external sources of noise.

To improve robustness of the method and define guidelines for choosing the threshold values, we carried out an evaluation of different threshold values using previously recorded data. Figure 5 shows an example of the F-score evaluation results obtained for different threshold values $\{\theta_p, \theta_s, \theta_t\}$. This exercise enabled us to determine a best set of threshold values and to gain insights into the sensitivity of the method to threshold values.

3 Results and discussion

The results for the comparison of manual and automatic SSA detections for the Vidourle river case study are shown in Table 1. Manual counting was carried out through precise marking of SSA signals using Audacity software (available at www.audacityteam.org). The start and end times of each SSA could thus be accurately compared and matched with the automatic detections. The training data used for defining the spectral signature S were recorded on the same spawning ground in 2014 using the same device.

The automatic SSA detection had an F-score of 68%. This score can be viewed as relatively good, considering the noise level and the other sounds present on the records. However, it was regrettable that there were so few SSAs, which prevented us from having statistically significant results.

To obtain more data, we complemented this case study with records coming from other spawning sites, located on the western French rivers Loire and Charente. The results are shown in Table 2, with new tuning of the three thresholds according to the same procedure. Surprisingly, the performance of the automatic detection method was better than for the Vidourle river case study, with an F-score of 88%, although the data were produced using different recording devices, and even the sub-species under study were different (allis shad instead of twaite shad).

Bearing in mind the objective of developing a small, hand-held device such as a smartphone, the detection method was oriented towards a simple algorithm based on a filter bank. A brief comparison with an SVM classifier based on a kernel of radial basis functions (Temko et al. 2006) showed that the

Table 2. Additional validation data from the Loire and Charente rivers, number of shad spawning acts from manual counting and automatic detection, and performance measures for automatic detection compared to manual counting. See text for explanation of thresholds.

Site	Date	Duration	Manual counting	Automatic counting	True positive	FFalse postivie	FFalse negative	Precision	Recall	F-score	Threshold		
											θ_p	θ_s	θ_t
Loire	2012/05/29	38′	19	19	17	2	2	0.895	0.895	0.895	7	0.27	0.04
Charente Crouin	2009/05/08	30′	56	50	47	3	9	0.940	0.839	0.887	5	0.18	0.05
Charente La Baine	2013/05/16	1h30′	172	184	155	29	14	0.842	0.919	0.879	5	0.20	0.08

results are very similar (F-score of 63%). Furthermore, using cross-validation with different partitions between training data and test data, we found that our method seems to be more robust, giving relatively good results with totally new data, contrary to SVM.

Detecting and counting the spawning acts of shads is a challenging issue, which to our knowledge has not yet received satisfactory solutions. The proposed method is based on simple spectral decomposition and filtering. It yielded relatively good results when evaluated offline with recorded data and controlled with a suitable set of parameters. Further evaluations of the equipment with new data will most likely help provide guidelines for tuning these parameters and implement a robust online application for the smartphone.

In conclusion, the rapid evolution of mass electronics and mobile technology is opening up new avenues for monitoring animal species. Combined with signal processing and pattern recognition techniques, the use of such equipment for the quantitative monitoring of shad populations is already feasible. This type of instrument will considerably increase the effectiveness of counting for estimating shad population size, and enable its deployment in new spawning grounds.

Acknowledgements. The research presented in this article was funded by the French Association Migrateurs Rhône Méditerranée (MRM). External data were provided by LOGRAMI and EPTB Charentes.

References

Chanseau M., Castelnaud G., Carry L., Martin-Vandembulcke D., Belaud A., 2004, Essai d'évaluation du stock de géniteurs d'alose *Alosa alosa* du bassin versant Gironde-Garonne-Dordogne sur la période 1987–2001 et comparaison de différents indicateurs d'abondance. Bull. Fr. Pêche Piscic. 374, 1–19.

Coustaux I., Trouilhet J.F., Guilhot J.P. 1994, Reconnaissance de signaux acoustiques à l'aide d'un réseau neuro-mimétique multicouche. J. Phys. IV France 04, 1319–1322.

Diep D., Nonon H., Marc I., Delhom J., Roure, F., 2013a, Mobile Technology Helps Monitor Biodiversity, 24th Int. Bioacoustics Congress, IBAC (abstract).

Diep D., Nonon H., Marc I., Delhom J., Roure F., 2013b, Smartphones for automatic shad counting, 20th Int. Conf. on Systems, Signals and Image Processing, IWSSIP, IEEE, 163–166.

Guyot P., Pinquier J., André-Obrecht R., 2013, Water sound recognition based on physical models. Proc. 38th Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP.IEEE.

Kirchhofer A., Hefti D. (Eds.) 2012, Conservation of Endangered Freshwater Fish in Europe, Birkhäuser Verlag.

Lebel I., Menella J.Y., Le Corre M., 2001, Bilan des actions du plan migrateurs concernant l'alose feinte (*Alosa fallax rhodanensis*) sur le bassin Rhône-Méditerranée-Corse Bull. Fr. Pêche Piscic. 362-363, 1077–1100.

Lebel I., Auphan N., Brosse L., Menella J.Y., 2007, Le Plan Migrateurs Rhône-Méditerranée: actions en faveur de la biodiversité. Cybium 31, 261–273.

Rabiner L., Schafer R., 2010, Theory and applications of digital speech processing. Prentice Hall Press.

Temko A., Nadeu C., 2006, Classification of acoustic events using SVM-based clustering schemes. Pattern Recognition 39, 682–694.

Trouilhet J.F., Coustaux I., Guilhot J.P., 1993, Reconnaissance des formes dans le plan temps fréquence par modélisation paramétrique. In 14° Colloque sur le traitement du signal et des images, FRA, 1993. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images.

Van Den Doel K., 2005, Physically based models for liquid sounds. ACM Trans. Appl. Percept. 2, 534–546.