HAL open science

Acoustic counting and monitoring of shad fish populations

Daniel Diep, Hervé Nonon, Isabelle Marc, Jonathan Delhom, Frédéric Roure

To cite this version:

Daniel Diep, Hervé Nonon, Isabelle Marc, Jonathan Delhom, Frédéric Roure. Acoustic counting and monitoring of shad fish populations. International AmiBio Workshop: Recent Progress in Computational Bioacoustics for Assessing Biodiversity, Jun 2013, Bonn, Germany. hal-00838600

HAL Id: hal-00838600

https://hal.science/hal-00838600v1

Submitted on 26 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACOUSTIC COUNTING AND MONITORING OF SHAD FISH POPULATIONS

Daniel Diep[1], Hervé Nonon[2], Isabelle Marc[1], Jonathan Delhom[3], Frédéric Roure[4]

[1]LGI2P, Ecole des Mines d'Alès, France ({daniel.diep, isabelle.marc}@mines-ales.fr)
[2]DIVULCO, Montpellier, France (herve.nonon@divulco.org)
[3]MRM - Migrateurs Rhône Méditerranée, Arles, France (jo.delhom@migrateursrhonemediterranee.org)
[4]GECO Ingénierie, Laudun, France (geco.ingenierie@wanadoo.fr)

ABSTRACT

To estimate the population of migrating shad fish in rivers, a method analysing acoustic signals during spawning is proposed. This method is based on a short-term spectral analysis, combined with a classification with a Gaussian mixture model, it results in detecting and counting shad spawning acts. The application has been implemented on a smartphone, which serves as an automatic counting device and a sound recorder, integrating a number of communication facilities. In terms of detection, the results obtained are very promising, although they show a bias of estimation. The deployment of this device will help assess the population of migrating shad fish and monitor their behaviour from year to year.

Index Terms— acoustic recognition, classification, smartphone application, shad, Alosa fallax rhodanensis

1. INTRODUCTION

Twaite shad (Alosa fallax) is a migratory fish living primarily in the seas which goes up the rivers to breed in spring. In Europe, this species has considerably declined in the midth of the 20th century due to overfishing, pollution and obstacles to migration, and for this reason is now given considerable legal protection [1]. Monitoring the numbers of shads at their reproduction sites is an important indicator for measuring interannual changes in their population. In particular, it enables the evaluation of the effectiveness of structures such as sluices and fish passes, created to facilitate their annual upstream migration [2], [3].

Shad reproduce at night near to the surface of the water, turning quickly and noisily at the time of egg-laying and emitting a characteristic sound lasting a few seconds known as a "spawning splash”. The currently used measuring method is manual counting, from the river bank, by an observer who listens and counts the splashes. This manual counting method is highly restrictive 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 set up; however, they still require considerable intervention on the part of the operator (installation, monitoring of the equipment, deinstallation, listening to the recordings, spawning recognition and counting), and clearly it misses a device that would automatically recognize spawning splashes.

In this paper, we present the design of a new equipment achieved to automatically count the spawning acts of shads (SSAs) at a reproduction site. The central part of this equipment is a smartphone, and this configuration enables to benefit from all the functions attached to it: audio recording capability, high storage capacity, wireless communication, large power autonomy, the whole being integrated in a very small size device.

Automatic detection is a key issue for this application; it has been decomposed in two steps: in the first step, characteristic features are extracted from the acoustic signals in order to provide a representation of the signals in a space of reduced dimension. The second step consists in classifying the signals and detecting spawning splashes, on the basis of a training phase with recorded data. Furthermore, the implementation of the detection method in real time on a simple smartphone obliges to consider fast algorithms necessitating few resources.

The paper is organized as follows: Section 2 describes spawning splashes signals, Section 3 presents the method of automatic detection and Section 4 presents the implementation on a smartphone. Finally, some results are given in Section 5.

Related work

Although a number of studies exist related to the monitoring of twaite shad and its relative species allis shad (Alosa alosa) using acoustic records [4], to our knowledge very few authors have been working on the automatic detection of shad spawning acts. Only [5] and [6] have conducted some experiments with these acoustic signals, extracting spectra

estimates with autoregressive parameters using the Levinson-Durbin method, and classifying shad splashes using a Multi-Layer Perceptron neural network. However, the results showed an important sensitivity to environment noise, yielding a high rate of false detections. Furthermore, such experiments have not been followed by actual field exploitation.

2. ACOUSTIC SIGNALS

Acoustic signals generated by shad spawning acts (SSA) are particularly difficult to analyze and discriminate: they occur at random in a noisy environment populated by a lot of nocturnal animals; their statistics are non-stationary and the sounds may notably differ from one to the other. Physically, their modeling is complex, although some numerical simulation models [7] are able to reproduce them.
Non-stationarity of the phenomenon is clearly visible on the spectrogram of Figure 1b, which represents the evolution of power spectrum versus time.

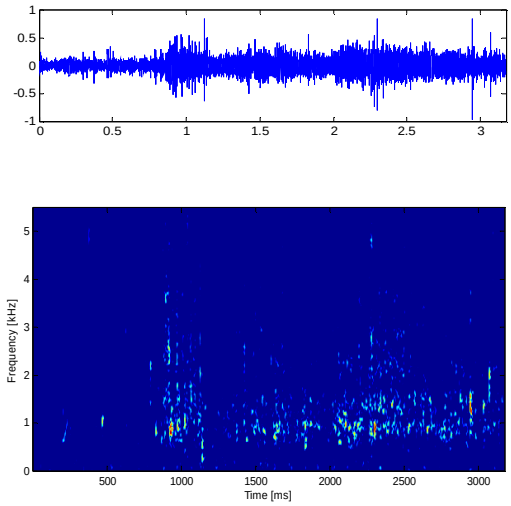

Fig. 1: a SSA signal a) vs. time, b) time - frequency spectrogram

Sensors

An important part of the study has been devoted to the selection of the acoustic sensor. We started our recordings with a microphone equipped with a parabolic reflector. Like for optic or electromagnetic waves, the parabolic reflector concentrates at its focus point the power of sound waves parallel to the central axis [8]. The Sony ECM-PB1C microphone, which integrates a 17 cm diameter parabolic reflector, was found to realize a good compromise between size, price and quality.
Unfortunately Sony stopped its production and we were conducted to create a substitute with a tailor-made parabolic dish associated with Olympus ME52W microphone (see Fig.2). Measurements revealed that both sensors have very close characteristics in sensitivity, range and directivity.

Fig. 2 : The Sony ECM-PB1C microphone (right) and its substitute (left)

3. AUTOMATIC DETECTION

The method of automatic detection of shad spawning acts comprises an extraction of signal features, followed by an algorithm of classification.

3.1. Sound database

Our database was built from 272 records segmented manually, and was composed of sounds sampled at 44.1 KHz, having a duration between 3 s and 10 s, and variable amplitudes (table 1).

data set	SSA	non SSA	total
training	56	80	136
test	56	80	136

Table 1: Sound database

Non-SSA signals are all types of environmental sounds encountered during the recordings, such as frog calls, dog barkings, wind, train passings or human voices. Some sounds overlapping with SSAs were also included in the test set.

3.2. Feature extraction

Signals were analyzed in the time-frequency domain in order to extract pertinent features. In this work we considered spectrum coefficients calculated on successive time frames of 4096 samples, corresponding to 93 ms. For each frame, the energy of the signal has been computed using FFT (Fast Fourier Transform) along 10 spectral bands covering a frequency range from 100 Hz to 5000 Hz on a logarithmic scale. The 10 triangular filters are centered on 100, 274, 485, 742, 1055, 1435, 1899, 2426, 3149, 3984 Hz.

The calculation of spectrum coefficients is straightforward, operations are schematized in Figure 2a.

In comparison, MFCC (Mel Frequency Cepstrum Coefficients) need more computation steps. MFCCs are features commonly used to process acoustic signals such as speech or music. In association with HMM (Hidden Markov Models), they have proven to constitute an efficient tool in the domain of speech recognition [9].
Figure 2b summarizes the sequence of operations composing a MFCC calculation. After signal conditioning with a pre-emphasis filter and a Hamming window, the FFT spectral transform is filtered by a bank of 42 filters distributed along a Mel scale, filter centers ranging from 133 to 6854 Hz. Finally, 13 MFCC parameters are extracted after a log transform and a DCT (Discrete Cosine Transform). We used for MFCC the same time slices of 93 ms as previously.

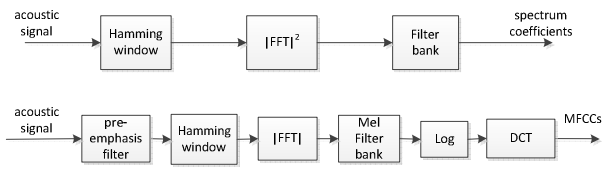

Fig. 2: Feature extraction of a) spectrum coefficients b) MFCCs

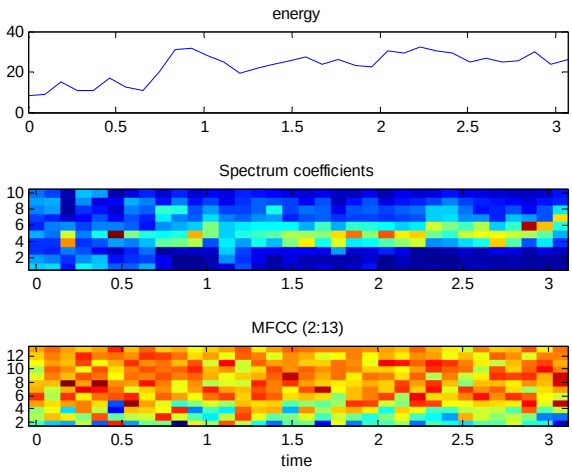

Fig. 3: Features extracted from a SSA signal, a) energy, b) spectrum coefficients, c) MFCCs

Figure 3 shows the features extracted from the SSA signal presented in Figure 1. Energy per frame has been calculated using short-time FFT, spectral coefficients and MFCCs are represented versus time on a colored scale of amplitudes.

3.3 Classification

Classification of signals in the acoustic domain is covered by a set of statistical methods such as HMM (hidden Markov models), NN (artificial neural networks) or SVM (support vector machines), which are supervised learning models (see e.g. [10] or [11] for an overview).
However, the non-stationarity of SSA signals indicates that extracted features evolve in a discontinuous way, feature vectors of a given shape seeming to occur intermittently. Therefore, instant features corresponding to time slices cannot be all tagged as being parts of a SSA signal. This particularity led us to orientate the detection issue toward unsupervised learning, i.e. clustering methods. In this domain, the GMM is a well understood statistical model, and presents the advantage of computational efficiency.

GMM (Gaussian Mixture Model)

This clustering method assumes that the data are generated by a mixture of probability distributions in which each component represents a different cluster.
Consider a set of N points (x_1, …, x_N) in $\mathbb{R}d$ to be clustered into K groups. The data is seen as N observations of a d-dimensional random vector with density:

$$\Phi_\theta(x) = \sum_{k=1}^{K} \alpha_k \Phi_k(x)$$

where Φ_k is the density of the normal distribution $N(\mu_k,\Sigma_k)$ and α_k the weight of this component in the mixture $\left(\sum \alpha_k = 1\right)$. The parameter $\theta=(\alpha,\mu,\Sigma)$ defines the model, with $\alpha=(\alpha_1,...,\alpha_K)$, $\mu=(\mu_1,...,\mu_K)$ and $\Sigma=(\Sigma_1,...,\Sigma_K)$.
The Expectation-Maximization or EM algorithm gives a means to estimate the parameters of the model. If we denote $p_\theta(\omega_k \mid x_i)$ the posterior probability that given x_i the point belongs to cluster ω_k, the EM algorithm maximizes the likelihood function of θ :

$$L(\theta) = \prod_{k=1}^{K} p_\theta(\omega_k \mid x_i)$$

Once clusters have been determined using the EM algorithm, it is simple to obtain a classifier. We associate to each class (SSA /non SSA) the combination of clusters, i.e. the mixture of Gaussian densities that provides the highest score, the score being the sum of true positives and false negatives in the confusion matrix.

Results of this classification with a varying number of clusters Nc are reported in Table 2, where the score has been expressed as a percentage.
We observe that the scores do not progress when Nc > 5. Spectrum coefficients give generally better results than MFCCs.

	spectrum coef.		MFCCs	
nb. clusters	training	test	training	test
2	67.6	60.3	75.0	70.6
3	86.8	77.9	71.3	69.9
4	89.7	82.4	83.8	77.9
5	91.2	85.3	88.2	89.7
6	86.8	77.9	83.8	72.1
7	88.2	84.6	89.0	79.4
8	89.0	82.4	87.5	77.9
9	90.4	88.2	85.3	75.0

Table 2: percentages of correct classification

As an example, Table 3 shows the confusion matrix obtained with Nc=5 and spectrum coefficients on the test set. For this classification, we obtained a True Positive Rate (TPR) of 78.6% and a False Positive Rate (FPR) of 10.0%.

	detected as SSA	detected as non SSA
actual SSA	44	12
actual non SSA	8	72

Table 3: A confusion matrix

4. A SMARTPHONE IMPLEMENTATION

The target device of our application is a Samsung GT-B7350 Omnia Pro4, a cell phone running Windows Mobile 6.5. To connect the microphone with its parabolic antenna to the audio input of the smartphone, we inserted an additional preamp in order to adapt the input level.

According to our needs, we developed the following software functions:

Handheld recorder: this application aims at recording and storing all sounds of interest, and to build a database for further analyses. The recorder makes use of the audio functionality of the cell phone and its facility of storage on a SDHC memory card. To reduce the volume of data, a suppression of blanks has also been included.

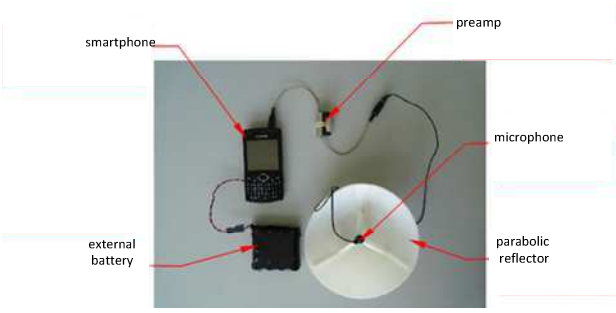

Fig. 4: A view of the device setup

Splash detection and counting: this function implements the automatic SSA detection presented in Section 3, based on spectrum coefficients and GMM classification. To meet real-time constraints, a particular attention has been given to the programming code, e.g. by using an integer-FFT algorithm. Furthermore, some thresholds have been added on signal energy and detection trigger in order to adapt the method to on-line detection.

Task scheduler: this component gives the ability to launch applications at pre-defined times. It allows the smartphone to remain on standby during day and to start recorder and detection at night.

Remote control via GSM and SMS: this communication tool has been designed to drive the smartphone and monitor it remotely, in case of difficulty to access the device. When solicited, the smartphone sends reports about its state, SSA counts, etc.

Wireless LAN (WiFi): this module can be possibly used to download data towards a collecting device, as long as transfer times are not excessive.

Autonomy: a 32 GB SD memory card and a 9600 mAh battery enable the device to work in autonomy during one week. To improve energetic autonomy, a photovoltaic sensor has been added to the equipment. Besides, data storage capability can be much improved by using a sound compression technique such as MP3 or Ogg-Vorbis (option not yet implemented).

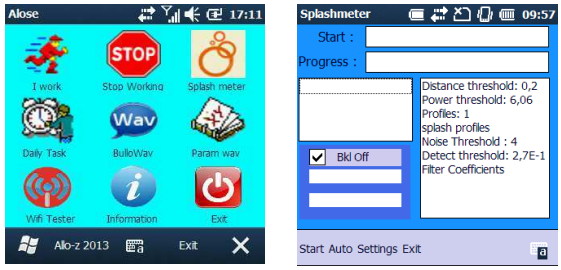

Fig. 5: Displays of the smartphone application

5. RESULTS

A test version of the smartphone was installed in spring 2012 on a shad spawning ground located on river Cèze, in the south of France. The smartphone equipped with a parabolic microphone and a battery was mounted on a tripod placed on the bank, at a height of 3 meters.

Unitary tests of the prototype have allowed to validate the choices of configuration. More specifically, audio files have been recorded and used to evaluate the method of SSA detection. They cover 10 days during which sounds have

been continuously recorded between 11 pm and 3 am, giving 40 hours of recordings.

At first, a training set has been selected with 40 files of 5 minutes, each of them containing at least a manually identified SSA. The classifier was trained by extracting the spectrum coefficients and applying GMM; the on-line detection algorithm of the smartphone was then applied to the data, giving the following results.

Number of identified SSAs : 48
Correctly detected: 39 (81%)
Not detected: 9 (19%)
False detections: 4 (8%)

Contrary to the results presented in Section 3, the on-line version of the detector does not perform a complete segmentation of signals, and the True Negative case of the confusion matrix is therefore irrelevant (we do not detect non-SSA signals).
The results obtained seem fully consistent with the TPR and FPR obtained previously.

Another evaluation consisted in applying the automatic detection on the whole data set, and comparing the results with those recorded by manual counting. Over the total period, 74 SSAs were counted automatically and 55 manually, giving an excess of 35%.

6. CONCLUSION

The work presented in this paper describes a new application which aims at estimating and monitoring the populations of shads on their reproduction sites.
Indeed, mass market devices like smartphones offer today a real opportunity to integrate signal processing and pattern recognition techniques. They help reduce the development time by providing increasingly integrated audio solutions, and provide in the same time new means of communication.
The first results obtained in spring 2012 led us to plan the deployment of counting devices at a larger scale in spring 2013.
Current work will try to reduce the bias of estimation by implementing more efficient detection methods, and optimize some parameters such as sampling frequency or data compression.

ACKNOWLEDGEMENT

This study was supported by the French PLAGEPOMI (*Plan de Gestion des Poissons Migrateurs*, Management Plan of Migratory Fish)

REFERENCES

[1] LIFE-Projekt Maifisch - *The re-introduction of allis shad (Alosa alosa) in the Rhine System* (2011). Available at http://ec.europa.eu/environment/life/project/Projects/index.cfm?fuseaction=search.dspPage&n_proj_id=3121

[2] Baglinière JL and P. Elie (eds) (2000). *Les Aloses (Alosa alosa et Alosa fallax spp.)*. Cemagref Editions. INRA, Paris.

[3] I. Lebel, J.Y. Ménella, M. Le Corre , 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.

[4] M.L. Acolasa, V. Véron, H. Jourdana, M.L. Bégoutc, M.R. Sabatié and J.L. Baglinière, "Upstream migration and reproductive patterns of a population of allis shad in a small river (L'Aulne, Brittany, France)". *ICES Journal of Marine Science*, April 2006; 63(3) : 476-484.

[5] J.F. Trouilhet, I. Coustaux, J.P. Guilhot : "Reconnaissance des formes dans le plan temps-fréquence par modélisation paramétrique". *14e colloque GRETSI*, 1993

[6] I. Coustaux, J.F. Trouilhet, J.P. Guilhot : " Reconnaissance de signaux acoustiques à l'aide d'un réseau neuro-mimétique multicouche", 3rd French conference on acoustics, *J. Phys. IV France 04* (1994)

[7] K. Van Den Doel, "Physically-based models for liquid sounds", Proc. ICAD 04, *10th Int. Conf. Auditory Display*, Sydney, 2004.

[8] Z. Chen and R.C. Maher, "Parabolic Dish Microphone Principles and Characteristics", progress report prepared for *Advanced Acoustic Concepts*, Inc., Hauppauge, NY, 10 pages, 2005.

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

[10] Temko, Andrey, and Climent Nadeu. "Classification of acoustic events using SVM-based clustering schemes." *Pattern Recognition* 39.4 (2006): 682-694.

[11] A. Dufaux, et al. "Automatic sound detection and recognition for noisy environment." *Proc. of the X European Signal Processing Conference*. 2000.