Spectral Noise Gate Technique Applied to Birdsong Preprocessing on Embedded Unit

Davi Miara Kiapuchinski
Lab. of Innovation and Technology for Embedded Systems
Federal University of Technology - Paraná
Curitiba, PR, Brazil
davi.miara@ctim.mar.mil.br

Carlos Raimundo Erig Lima
Lab. of Innovation and Technology for Embedded Systems
Federal University of Technology - Paraná
Curitiba, PR, Brazil
erig@utfpr.edu.br

Celso Antônio Alves Kaestner
Academic Department of Computing
Federal University of Technology - Paraná
Curitiba, PR, Brazil
celsokaestner@utfpr.edu.br

Abstract - **This paper proposes an approach for audio preprocessing and noise removal from recordings obtained in natural environments. The method is inspired in the acoustic signature of the audio, and aims to preprocess the recordings of bird songs obtained directly in the field. Using the Spectral Noise Gate technique, the undesired noise is removed on a real application in real time during the recording using an embedded environment. In addition, important statistic features of the audio signal are computed. The main purpose on approach is to eliminate the manual and tedious process of preparing the audio recordings done in the field in order to make them ready to be used as input in other tasks, such as the automatic classification of bird species from recorded bird songs. This is necessary because classification results depend widely from the quality of the input data.**

Keywords - accoustic signature; audio preprocessing; noise removal; birdsong recording.

I. Introduction

Nowadays the ethical concerns with animal welfare suggest the use of non-invasive techniques to study fauna. In the case of birds, ornithologists (bird researchers) defend the use of bioacoustics techniques to identify birds from their audio recorded songs [1]. So, in this case, bioacoustics is the main tool to capture and process birdsong characteristics [2]. In order to do so, several research papers, on the last two decades, propose automatic techniques to classify bird species, using its recorded audio song as input.

The study has evolved such that Fagerlund [3] employs Support Vector Machines (SVM) and decision trees (DT) to produce a classifier that attains an accuracy of 100% in the classification of some bird species. Other example is the work of Vilches, Escobar, Vallejo and Taylor in [4] that obtained accuracy of until 98.39% using several classifiers for three bird species. The same methodology was followed by Lopes, Silla, Koerich, and Kaestner [5][6], where several classifiers and different sets of audio features were analyzed.

However, in practical applications some additional problems must be considered. Bardelli [7] indicates that laboratory tests work pretty well, but in real environments problems like noise caused by winds, other natural causes and also the superposition of audio signals, affect more that 50% of the recordings. In these recordings it is very difficult to classify the bird, and sometimes even to detect the bird song. Agranat [8] also detects this problem and his conclusion is that a manual and tedious preprocessing filtering must be done in order to make the audio signal free of noise and ready to be processed. These drawbacks impose strong restrictions to the application of automatic methods in real environments. These restrictions are, for example, the number of bird species to detect is limited and the frequency range of the audio recorded signals does not encompass the songs from most of the bird species [9].

In this paper we propose a micro-controlled embedded system able to preprocess audio signal and therefore suitable to be applied in real environments. This equipment can be applied not only to classify birds, but also be used as a tool to control biodiversity in sensible areas, natural reserves, parks and also to secure the human life on places where the bird circulation is forbidden, such as near airports or wind plants[10].

II. Audio Processing

The conditions found in real environments involve countless variables that complicate the work of birdsong recording, and in consequence the feature extraction process necessary to subsequent automatic classification. Among the difficulties we emphasize: (1) the wide variety of noises found in real conditions such as the rainfall, wind, density of trees, planes, cars, the noise produced by other animals and so on; and (2) the large bird spectral width of the audio signal, which may vary from 100 Hz to 10,000 Hz; this range differs greatly of the human voice spectrum, focused at approximately 3,000 Hz [8].

The simplest solution to the effectively analysis and the classification process is to apply filters to the signal: the filtering process on digital systems can be done using digital filters [11], eliminating unnecessary frequencies of the sound and attenuating or emphasizing important parts of the signal. Fig. 1 illustrates a signal from a birdsong collected directly in a natural environment. "Blurry" on the spectrum, low frequency random noise produced by several fonts and other unwanted sounds appears in the signal at top. In order to further improve the signal and eliminate unwanted frequencies, additional steps can be executed: digital filters, normalization, approximation, enhancement and cancellation of points can be calculated on the spectrum.

978-0-7695-4875-3/12 $26.00 © 2012 IEEE
DOI 10.1109/ISM.2012.12

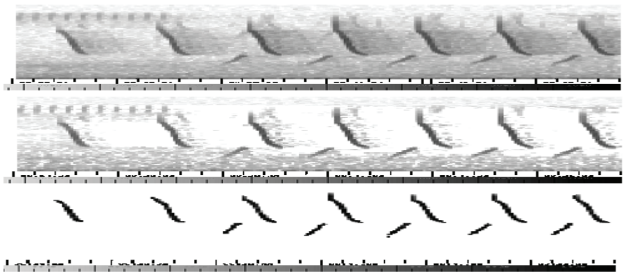

Figure 1. At top a original sample from Northern Cardinal bird, in the middle a filter is applied to reduce the stationary background noise. A filter using Log Frequency Transformation and Power Normalization is applied on the last spectrum, taken from [8]

The technique used in this study is similar and is based on the setting of the values of the signal potency into acceptance limits. Variables like threshold and gain model the filter behavior, letting or not certain signal to 'pass' [12].

As signal preprocessing is a standard work in the focused research area, it is proposed an approach that incorporates this work in the acquisition unit – as embedded software – in order to prepare the samples to be directly used by the automatic classifiers. The main goals are to solve the existing problem of noise elimination from recordings obtained in real environments and also to avoid the any restriction in the bird species that can be target by the systems.

III. Proposed Approach

As seen, there are essential steps to be made before the submission of sound to the automatic classifier. The proposed approach focuses on delivering to the automatic classifier a clean and prepared sound. We follow three steps: (A) a routine of preparation and filtering the sound signal by using the Spectral Noise Gate technique; (B) a feature extraction from the audio signal, in order to compute the sound characteristics to be provided as input to the automatic classifiers and characterize the sound; and (C) the materialization of the two first parts in an embedded environment.

A. Spectral Noise Gate Algorithm

A typical Spectrum Noise Gate is a technique commonly used in sonorous mixing and manipulation that attenuates a signal according to certain threshold [13]. With some variations, it has been used in the general noise removal. In the chosen approach, initially the noise profile or acoustic signature is discovered from the noisy sound previously obtained. After this, some attenuation operations are performed in the sound spectrum. The routine flow is summarized on the Fig. 2.

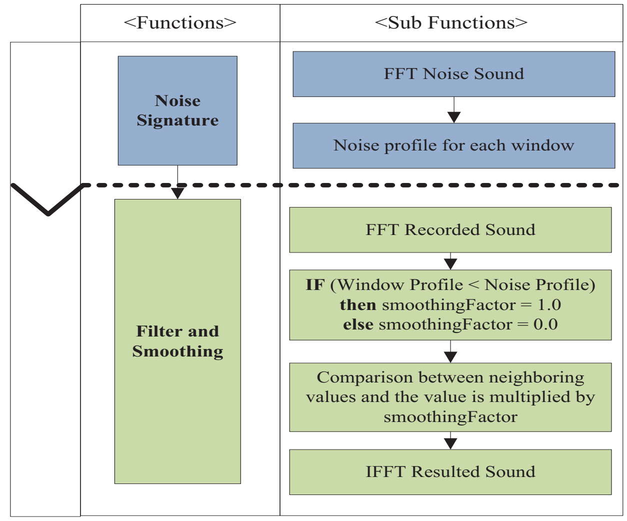

Figure 2. Flow of Activities on Spectral Noise Gate Algorithm

B. Feature Extraction

Several statistical characteristics, on time and frequency domains, can be used to represent the tone.

1) Skewness: indicates the obliquity of the data distribution or how symmetrical the data distribution is [14]. Skewness is computed according to (1), where Mt isa vector with the data distribution, N total of elements, μ is de data average and σ is it standart deviation:

$$St = (\textstyle\sum_{n=1}^{N} Mt[n]) - \mu/\sigma)^3/N \qquad (1)$$

2) Kurtosis: defined by (2) is a statistics employed to measure the "flattening" on the data distribution [14]:

$$Kt = ((\textstyle\sum_{n=1}^{N}(Mt[n]) - \mu/\sigma)^4/N) - 3 \qquad (2)$$

3) Spectral Centroid: compute the spectral shape to informs the balance point of the spectrum, similar to an average of the data [15]; it is computed by (3).

$$Ct = \textstyle\sum_{n=1}^{N} Mt[n] \times n \,/\, \sum_{n=1}^{N} Mt[n] \qquad (3)$$

4) Spectral Rolloff: Measure the spectral shape that characterizes the data below a certain level; the pattern indicate the cutoff frequency of the noise [15]. Equation (4) represents the Spectral Rolloff with an 85% concentration level:

$$\textstyle\sum_{n=1}^{Rt} Mt[n] = 0.85 \sum_{n=1}^{N} Mt[n] \qquad (4)$$

5) Zero Crossing Rate: is a time domain signal characteristic, indicating the number of times the signal values change from positive to negative; the 'sign' on (5) returns, respectively, 1 or 0 if the result is positive or negative [15]:

$$Zt = \tfrac{1}{2}\textstyle\sum_{n=1}^{N} |sign(x[n] - sign(x[n-1])| \qquad (5)$$

C. Hardware Architecture

In this work we employ the Samsung micro-controller S3C2440A with 289 pins and a 16/32 bit RISC with an ARM920T CPU core [16]. This microcontroller presents features which may be important to Digital Signal Processing (DSP) applications, particularly IO audio interface; direct memory access (DMA) and dedicated data operations.

The embedded Linux OS was used with kernel 2.6.32.2. To compilation it was used the ARM Linux GCC 4.3.2 cross compiler.

IV. Experiments

Using the proposed approach we obtain an acquisition system unit for recording and processing bird song audio signals. The flow of activities in the system starts with the recording of the audio by the microphone in the embedded unit. Next, the sound pass by a spectral analysis and then through the preprocessing steps until finally the corresponding filtered sound and its extracted features are saved on the unit.

1) Recording: It was performed by a mono-channel microphone, the samples are stored in the memory using words of 8 or 16 bits and up to 44,100Hz of sample rate.

2) Spectral Analysis and preprocessing: The spectrum analysis and preprocessing involves the domain transformation of the signal, that is, the convertion of the signal from the time domain to the frequency domain. This process is carried out by the FFT (Fast Fourier Transform). These divisions are subject to a windowing function, that fix the truncations caused by the gaps between the windows. Three different windowing function were computed: the Hamming function, the Hanning function and the Bartllet or triangular windowing function [17].

After this, The Spectrum Noise Gate function is ready to be applied on data. The result, as shown on Fig. 3, is a signal absent from much of the noise and with enhancement of the frequencies that really matter.

3) Feature Extraction: Extraction procedures are applied to represent the signal. Table I presents some features obtained on the time and frequency domain from a sample audio signal recorded with a sampling rate of 8000 Hz and submitted to the FFT using the Bartlett windowing function with 16,384 (2^{14}) samples.

TABLE I. CHARACTERISTICS OF SOUND

Characteristic	***Value***
Total size WAV. file	48,044
Total size fo the data	24,000
Number of bits per sample	16
Recording data and hour	Fri Mar 16 10:36:07 2012
Zero crossing rate	0.0875748
Data Sum	-1.39276
Values not equal zero	23,191
Employed windowing	BARTLETT
Centroid of the data	0.763369
Skewness	0.735164
Data Kurtosis	121.12
Data Rolloff	1,991.21

V. Discussion

The application in natural environments of the proposed approach can be considered as a solution for one important problem in the automatic birdsong classification area.

The target problem is the lack of good results in practical applications, such as the automatic classification of bird songs, in which known solutions are ineffective if the level of noise in the samples is high. The use of the technique of Spectrum Noise Gate sought to remove the noise coming from various and heterogeneous sources, leaving only the bird songs frequencies emphasized in the sound spectrum.

An important point to be considered in this task is the removing of noise coming from "constant" sources, originated by the wind, waterfalls or the rain. These sounds lead to false-positive results in the automatic classifiers when applied in real environments.

To illustrate this point it is present a record of a song of the bird Inhambuguaçu (*Crypturellus obsoletus*), collected in the South of Brazil. During the whole sample it can be heard in the background the sound of a waterfall, which is employed as a noisy sample, using a one second long interval. The Fig. 4 and Fig. 5 compare the spectrum of the sounds before and after the noise removal routine application by the microcontroller. Unwanted frequencies, much of the background noise and the waterfall sound were removed from the signal; the difference is easily perceived when someone listen the sound. So, when the resulting audio signal and the corresponding features are subjected to an automated classifier, we obtain better accuracy results, since the noise characteristics are absent of the signal.

The proposed approach was also evaluated according to the technical efficiency on the embedded environment. Fig. 6 presents some of the runtimes of the algorithm Spectrum Noise Gate, according to the selected sampling frequency and the type of windowing function. To obtain these results we use samples collected in the embedded environment with a length of 3 seconds and 16 bits of precision per word, and a noisy acquisition with 1 second long and the same precision. The FFT times for the selected frequencies range from 0.30 to 1.90 seconds, for 8000 Hz and 44100 Hz frequencies

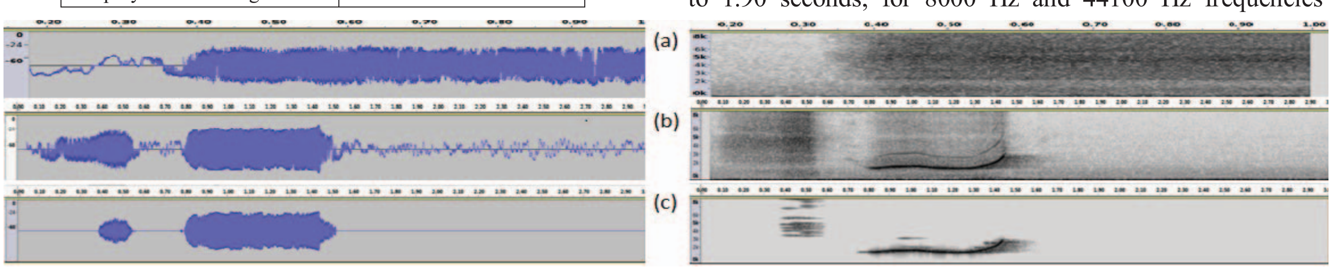

Figure 3. Spectrum Noise Gate process. (a) represents a signal with 1 second of duration (x axis) with a common noise present on omni-directional microphones; (b) shows the original sound, recorded by the presented approach, with 3 seconds. In (c) is shown the result of Spectrum Noise Gate algorithm emphasizing only the important part of spectrum. The left figures show the signal in time domain (time X dB levels) and the graphs on the right illustrate the signal on the frequency domain (time X frequency).

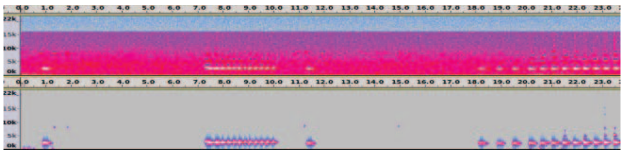

Figure 4. Spectrum of bird call before, with the water noise, and after the Spectrum Noise Gate Technique (time X frequency)

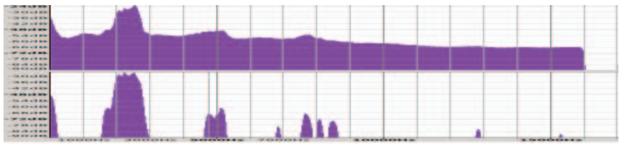

Figure 5. Signal Power Levels before and after the Technique (frequency X dB levels)

respectively. A sampling frequency that is sufficiently precise for the analysis of sounds of nature is 22,050 Hz, which would take on average one second in the proposed method. In terms of execution of the noise removal routine the Bartlett windowing function stands out over others two windowing functions.

VI. Final Remarks

In this paper we propose an approach to audio signal preprocessing and noise removal, which intends to be performed directly by the sound acquisition unit in real environments. The proposal uses a preprocessing routine, which applies specific filters in a recorded audio signal, in order to eliminate noise and spurious sounds aiming reduce the continuous manual work done up to now of prepare the samples to be input of the automatic classifiers.

The approach, through the sound extracted features, aims to provide solid inputs to automatic birdsong classifiers, maintaining the possibility of using the acquisition unit in real and dynamic situations. In order to do so, we design a specific system, functional and portable, similar to a smart recorder.

The performance of the equipment is adequate to the target problem: the total time elapsed in the process of noise removal is about 24 seconds; this execution time is satisfactory, however, depending on the situation, it is necessary parsimony in the parameterization of the system to avoid long delays in the process on a real condition. Some solutions for example, are to choose a not so high sampling rate or still parameterize a short sample to be analyzed.

As future work, we intend to improve the embedded software to reduce the processing time. Also, to include the classification procedure itself in the embedded software, obtaining a complete autonomous classification device.

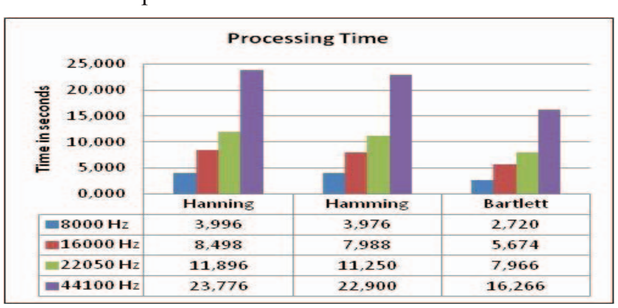

Figure 6. Processing time alternating sampling rate

Acknowledgment

This research project is supported by the Brazilian research agencies CNPq and Araucaria Foundation.

References

[1] F.C. Straube, "Newsletter BOCEV (Bird Observers Club of the European Valley," (in portuguese) Oct. 2005. Available at http://maicon-mohr.blogspot.com.br/2010/04/entrevista-com-fernando-straube.html. Accessed on July 2010.

[2] A.G. M. Betts, G.D. Forbes, A.M. Villard and J. Gunn, "The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence," Ecological Modelling, vol. 191, issue 2, jan. 2006, pp. 197 – 224, doi 10.1016/j.ecolmodel .2005.04.027

[3] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines". EURASIP Journal on Advances in Signal Processing**,** vol 2007, Mar. 2007, pp. 64 -72, doi: 10.1155/2007/38637.

[4] E. Vilches, I.A. Escobar, E.E. Vallejo and C.E. Taylor, "Data mining applied to acoustic bird species recognition," in Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06), vol. 3, pp. 400–403, Hong Kong, August 2006.

[5] M. T. Lopes, C. N. Silla Jr., A. L. Koerich, C. A. A.Kaestner, Feature set comparison for automatic bird species identification. IEEE Internmational Conference on Systems, Man and Cybernetics, Anchorage, AK, USA, October 2011: 965-970

[6] M. T. Lopes, L. L. Gioppo, T. T. Higushi, Celso A. A. Kaestner, C. N. Silla Jr., A. L. Koerich: Automatic Bird Species Identification for Large Number of Species. IEEE International Symposium on Multimídia, Dana Point, CA, USA, December 2011: 117-122

[7] R. Bardelli, et. al. "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," Pattern Recognition Letters, vol 31, issue 12, Set. 2010, pp. 1524–1534, doi 10.1016/j.patrec.2009.09.014.

[8] I. Agranat, "Automatically identifying animal species from their vocalizations," 2009 Acoustics. Inc**.,** Concord, Massachusetts, p. 22, 2009.

[9] A. L.McIlraith and H. C. Card, "Birdsong recognition using backpropagation and multivariate statistics," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2740–2748, Nov. 1997.

[10] J. Bell, N. H. K. Burton, A. J. Musgrove, M. M. Rehfish, "A model for assessing bird strike risk at proposed new airports," Bird Strike Committee Proceedings, 5th Joint Annual Meeting, Toronto, Canada, pp. 1-10, Aug. 2003.

[11] C. H. Yehia and N. M. Viera, "Digital signal processing in audio engineering," (in portuguese) Fisrt Brazilian Seminar in Audio Engineering, Belo Horizonte, Brazil, pp. 9, 2002.

[12] M. J. Terrell and J. Reiss, "Automatic Noise Gate Settings for Multitrack Drum Recordings," Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09), pp. 1-4, Set. 2009.

[13] J. Hodgson, Understanding Records: A Field Guide to Recording Practice, Continuum International Publishing Group, p. 272, 2010

[14] D. Downing and J. Clark, "Business Statistics," 3th. Edition, Barron Educational Series, 1997.

[15] C. Silla, C. A. A. Kaestner e A. L. Koerich, "Automatic Music Genre Classification using Ensemble of Classifiers," Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'07), Montreal, Canada, pp. 1687-1692, Oct. 2007.

[16] Sansung Eletronics. "S3C2440A, 32-bit CMOS microcontroller user's manual, ". Rev. 1, Yongin-City, pp. 596, 2004.

[17] B. Champagne and F. Labeau, Discrete Time Signal Processing, Class Notes for the Course ECSE-412, Department of Electrical & Computer Engineering, McGill University, 2004, p. 190-194.