DEVELOPMENT OF NATURAL SOUNDSCAPES –

A PATTERN RECOGNITION APPROACH

Ioannis Paraskevas[1] and Maria Rangoussi[2]

[1] Department of Technology of Informatics and Telecommunications,

Technological Education Institute of Kalamata / Branch of Sparta,

7, Kilkis str., Sparta, GR-23100, GREECE

paraskevas@env.aegean.gr

[2]Department of Electronics,

Technological Education Institute of Piraeus,

250, Thivon str., Aigaleo-Athens, GR-12244, GREECE

mariar@teipir.gr

Abstract

The aim of this work is the mapping of areas of ecological interest focusing on the sound rather than the visual features of this area. Mapping of an area focusing on the visual (morphological) features of the landscape is a mature field with a wide range of applications. On the other hand, 'soundscapes' are maps which convey the sound content of an area at a certain time instant. Sound features encapsulate information which can be combined with the visual features of the landscape, in order to derive useful ecological related conclusions. These include monitoring of the wildlife, or the various human activities in the area. Of great interest is the monitoring of the change in the use or inhabitation of the

area over time – an aim requiring repeated measurements at regular time intervals. In this work, a method is described for the combination of pattern recognition of the sound recordings with the development of the soundscape of the area of interest.

Keywords

Soundscape, sound pattern recognition, feature extraction, classification.

1. Introduction

Research related to the environmental / ecological information of landscapes is usually focused on their visual content e.g. geometric or landscape characteristics of the biotopes. Surveillance engineering methods are typically employed to produce maps at the required resolution. In this research, the sound content of the landscapes is proposed as an additional information stream, aiming to produce useful audio-visual features. Hence, after appropriate processing, sound is used as an additional ecological related indicator of the area of interest.

The term 'acoustic ecology' is first introduced by R. Murray Schafer (Schafer, 1994). Important work in the research area of acoustic ecology is also contributed by the SEKI Group (SEKI Group, 2008). The objective of their study is to examine whether environmental sound recordings convey useful information and consequently, whether features extracted from these recordings can be employed as indicators for the health of biotopes and for the biotopes' dynamic balance. Recently, the interest for the acoustic ecology has increased due to the activities of the World Forum for Acoustic Ecology (WFAE) which was founded in Canada in 1993.

A soundscape (Krause, 1987; Krause, 2002) is the map of a certain region at a given time instant, which is focused on its sound rather than its visual content. Soundscapes form a useful feature of natural conservation (Turner et al., 2001). Indeed, periodic comparison of the soundscapes of a certain area - e.g. regions of the NATURA 2000 network (European Union's network of nature protection areas) - can provide useful ecological related observations. Note that, in this work the terms acoustic / audio are avoided as natural soundscapes may include sounds that are not necessarily audible by the human ear.

For the case of environmental sounds, there are two types of sound pattern recognition: 'coarse' and 'fine'. The first aims to classify sounds that do not belong to the same family, so as to identify the three categories of environmental sounds; namely, human-related, geophysical-related and animal-related (Gage et al., 2001). The second aims to classify sounds that belong to the same family, so as to identify different species of e.g., bats, based on the sounds they produce (Parsons and Jones, 2000). Obviously, the 'fine' type is more demanding, in terms of feature extraction and classification, compared to the 'coarse' type. It is worth mentioning that the majority of existing research work for content based sound pattern recognition belongs to the 'coarse' type (Wold et al., 1996; Foote, 1997; Mingchun and Chunru, 2001; Zhang and Kuo, 2001) rather than to the 'fine' type (Paraskevas and Chilton, 2003; Paraskevas and Chilton, 2004; Paraskevas et al., 2006).

In the proposed method, each sound recording is classified and then placed on the map so as to form the soundscape. The application of pattern recognition to environmental sounds is a hierarchical process. The three main categories of environmental sounds i.e. human-related, geophysical-related and animal-related are initially classified ('coarse' classification). Then follows the classification into subcategories of each of the three aforementioned classes of sounds e.g. classification of different species of bats as a subcategory of the animal-related sounds ('fine' classification). Finally, the soundscape is developed placing each classified recording to the exact area of the map where the sound was recorded. Consequently, the development of a soundscape becomes more complicated as the number of sound classes increases.

In the following, a description of the proposed pattern recognition process is provided along with an example which shows the importance of signal representation for the application of environmental sound pattern recognition.

2. Pattern Recognition for Environmental Sounds

In this Section, the proposed method for the application of sound pattern recognition is presented. As for any pattern recognition application, the process can be divided into two stages: the feature extraction and the classification stage (Duda et al., 2000; Webb, 2002).

In the feature extraction stage, features that encapsulate the information content of the signal are extracted from each sound recording while in the classification stage the sound recordings are classified based on the feature vectors formed in the feature extraction stage (figure 1).

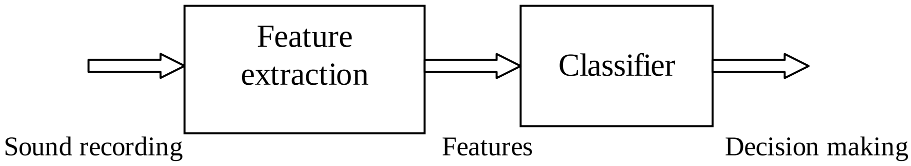

Figure 1: Pattern recognition system model

In the feature extraction stage, class - discriminating features are extracted in order to classify each sound recording to the corresponding class. As in all pattern recognition applications, the classification rate depends on the selection of the appropriate features that form the feature vectors.

The features can be extracted from:

- the time domain signal recordings e.g., zero-crossing rate (ZCR), linear prediction coefficients (Makhoul, 1975),
- the frequency domain e.g. pitch (Rabiner et al., 1976), cepstral coefficients (Childers et al., 1967), bandwidth and
- the time – frequency signal representations (Cohen, 1989; Esmaili et al., 2004) e.g. statistical

features (Lambrou et al., 1998) / coefficients extracted from the magnitude spectrogram (Rabiner and Schafer, 1978; Paraskevas et al., 2006).

The importance of the appropriate signal representation for an efficient feature extraction is illustrated in the following example, where a sound recording consists of two simultaneous sound events. Specifically, the two sound events are: ‘birds croaking’ and ‘waterfall’. Figures 2a, 2b and 2c show the time domain signals of the recordings of:

- ‘sound of birds croaking’,
- ‘sound of waterfall’ and
- the simultaneous occurrence of both sound events i.e. ‘sound of waterfall with birds croaking’, respectively.

Figures 3a, 3b and 3c show the magnitude spectrogram of the sounds presented in figures 2a, 2b and 2c, respectively.

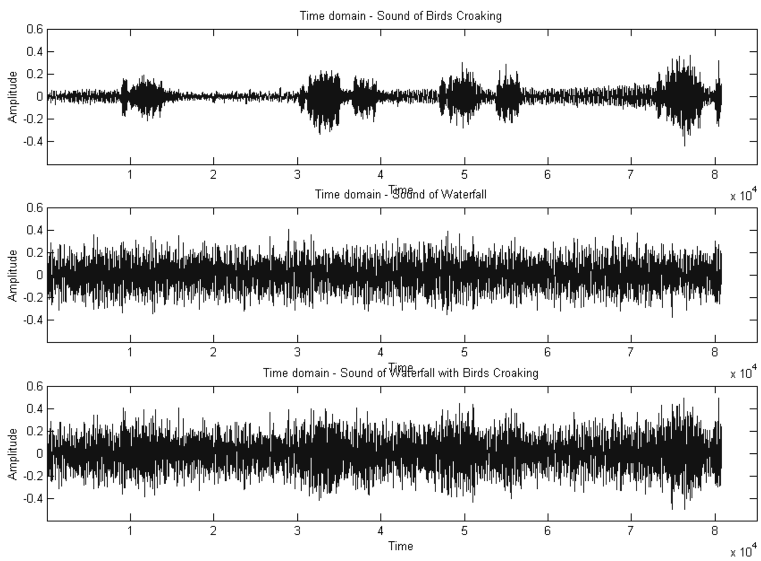

Figure 2a: Time Domain – Sound of Birds Croaking

Figure 2b: Time Domain – Sound of Waterfall

Figure 2c: Time Domain – Sound of Waterfall with Birds Croaking

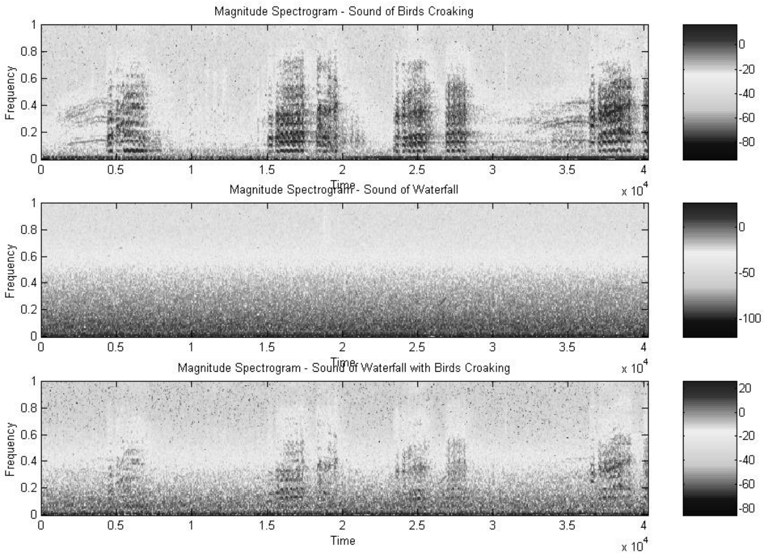

Figure 3a: Magnitude Spectrogram – Sound of Birds Croaking

Figure 3b: Magnitude Spectrogram – Sound of Waterfall

Figure 3c: Magnitude Spectrogram – Sound of Waterfall with Birds Croaking

From figure 2c, it is observed that in the case of the simultaneous occurrence of the two sound events i.e. 'simultaneous sounds of waterfall and of birds croaking', it is not possible to distinguish from the time domain these two events, whereas, these events are distinguishable from the corresponding magnitude spectrogram (figure 3c). Consequently, for this case, the feature vector should be formed from features extracted from the frequency (magnitude spectrogram) rather than from the time domain of the signal.

It is known that, for efficient feature extraction, mathematical transforms may be applied to the signal e.g. Fourier transform (Proakis and Manolakis, 1992), Hartley transform (Bracewell, 1986). The feature vectors formed from the features extracted from each recording should present its information content to the classifier in a compact manner. Different kinds of classifiers e.g. distance metric classifiers, neural networks (Kohonen, 2001) etc. can be employed depending on the intrinsic characteristics of the sound classes aiming to classify.

The environmental sounds recorded from microphones that are spread to different places of an area, form the sound database. Then, each sound recording is classified / identified via pattern recognition methods. The classified sound recordings are placed to the corresponding location on the sound map thus, forming the soundscape. Therefore, the proposed method may be put in steps as follows:

i. Development of the database of environmental sound recordings.
ii. Features are extracted from the aforementioned sound recordings (sound signals) which will be employed for the classification stage of the sound pattern recognition.
iii. Classification (coarse / fine) of the environmental sound recordings based on the features extracted from the sound signals in step (ii).
iv. Development of the soundscape employing the classified sound recordings (sound signals).

3. Conclusions

The aim of this work is to present a research method which combines the application of sound pattern recognition with the development of soundscapes for areas of ecological interest. The periodic development of soundscapes for the same area is a useful tool for the detection of changes in the ecosystem. Soundscapes can be combined with other sources of information (e.g. territorial morphology), via Geographic Information Systems (GIS) software, in order to develop complete models of areas of interest.

References

Bracewell R.N. (1986). The Fourier Transform and Its Applications. 2nd edition McGraw-Hill.

Childers D.G., Skinner D.P., and Kemerait R. (1967). The cepstrum: a guide to processing. *Proceedings of IEEE*, pp. 1428 – 1443.

Cohen L. (1989). Time-frequency distributions-a review. *Proceedings of the IEEE*, volume:

77, issue: 7, pp. 941 – 981.

Duda R.O., Hart P.E., and Stork D.G. (2000). Pattern Classification. 2nd edition John Wiley & Sons, Ltd.

Esmaili S., Krishnan S., and Raahemifar K. (2004). Content based audio classification and retrieval using joint time-frequency analysis. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume: 5, pp. V- 665-8.

Foote J. (1997). Content-based retrieval of music and audio. *SPIE Multimedia Storage and Archiving Systems (II)*, vol.3229, pp. 138-147.

Gage S.H., Napoletano B.M., and Cooper M.C. (2001). Assessment of ecosystem biodiversity by acoustic diversity indices. *American Institute of Physics*, 109, 2430.

Kohonen T. (2001). Self-organizing maps. 3rd Edition Springer-Verlag Berlin-Heidelberg-New York.

Krause B. (1987). Bioacoustics, habitat ambience in ecological balance. *Whole Earth Review*, 57.

Krause B. (2002). Wild soundscapes: Discovering the voice of the natural world. Berkeley, California: Wilderness Press.

Lambrou T., Kudumakis P., Speller R., Sandler M., and Linney A. (1998). Classification of audio signals using statistical features on time and wavelet transform domains. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume: 6, pp. 3621 - 3624.

Makhoul J. (1975). Linear Prediction: A tutorial review. *Proc. IEEE*, pp. 561-580.

Mingchun L., and Chunru W. (2001). A study on content-based classification and retrieval of audio database. *International Symposium on Database Engineering & Applications*, pp. 339 – 345.

Paraskevas, I., and Chilton, E. (2003). Audio classification for retrieval from multimedia databases. *Proceedings of the EC-VIP-MC, 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications*, Zagreb, Croatia, pp.187–192.

Paraskevas, I., and Chilton, E. (2004). Combination of Magnitude and Phase Statistical Features for Audio Classification. *Acoustics Research Letters Online*, 5 (3), 111–117.

Paraskevas, I., Chilton, E., and Rangoussi, M. (2006). Audio classification using features derived from the Hartley transform. *Proceedings of the 13th Intl. Workshop on Systems, Signals and Image Processing (IWSSIP'2006)*, Budapest, Hungary, pp.309–312.

Parsons S., and Jones G. (2000). Acoustic Identification of twelve species of echolocation bat by discriminant function analysis and artificial neural networks. *The Journal of Experimental Biology*, 203, 2641-2656.

Proakis J.G., and Manolakis D.G. (1992). Digital Signal Processing Principles, Algorithms, and Applications. Macmillan Publishing Company.

Rabiner L., Cheng M., Rosenberg A., and McGonegal C. (1976). A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, volume: 24, issue: 5, pp. 399 – 418.

Rabiner L., and Schafer R.W. (1978). Digital processing of speech signals. Prentice-Hall.

Schafer R. M. (1994). The soundscape: Our sonic environment and the tuning of the world. Rochester, Vermont: Destiny Books.

SEKI Group (2008). www site: http://envirosonic.cevl.msu.edu/seki

Turner M.G., Gardner R.H., and O'Neill R.V. (2001). Landscape ecology in theory and practice: Pattern and process. New York, New York: Springer-Verlag New York, Inc.

Webb A.R. (2002). Statistical Pattern Recognition. 2nd edition John Wiley & Sons.

Wold E., Blum T., Keislar D., and Wheaton J. (1996). Content-Based classification, search and retrieval of audio. *IEEE Multimedia Fall 1996*, pp. 27-36.

Zhang T., and Kuo C.C.J. (2001). Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4.