Applied Acoustics 112 (2016) 171–180

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Predominant environmental noise classification over sound mixing based on source-specific dictionary

María Guadalupe López-Pacheco *, Luis Pastor Sánchez-Fernández, Herón Molina-Lozano, Luis Alejandro Sánchez-Pérez

Centro de Investigación en Computación – Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Nueva Industrial Vallejo, Gustavo A. Madero, México D.F. 07738, Mexico

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 13 May 2016
Accepted 25 May 2016
Available online 2 June 2016

Keywords:
Audio classification
Urban noise
Signal decomposition
Predominant source
Environmental mixture signal

ABSTRACT

This paper presents a methodology to classify predominant urban acoustic sources in real mixed signals. This is based on a source-specific dictionary with atoms in the time–frequency domain using the Orthogonal Matching Pursuit (OMP) algorithm and identifying the class through a proposed selection criterion with a dynamic number of iterations involving a lower algorithm complexity. Several time–frequency atoms were evaluated considering retained energy and relative error to build a source-specific dictionary in the relevant classes. The source-specific dictionary has better results up to 7% in retained energy than to use an individual dictionary such as based on wavelet or Gabor functions, improving classification of predominant sources over sound mixing up to 9% compared to using standard dictionaries. Experimental results on classification are applied to mixture inter-class signals of two or more sources recorded by a real permanent monitoring system in an urban soundscape. The classification performance has successfully achieved identifying a predominant source in real inter-class mixtures of urban soundscapes.

1. Introduction

Nowadays, environmental acoustic analysis has taken great relevance because exposure to high noise levels for long periods of time causes serious damage to the health and comfort of people [1]. In urban areas, high noise levels are mainly generated by human activities related to transport, construction, security, commerce and recreation. There are several investigations related to assessing acoustic level and health damage of these activities in [2–5]. The common control actions for noise reduction are aimed at controlling or removing the individual source that produces noise. A methodology to automatically classify predominant sounds is useful to estimate the acoustic levels of individual sources and it is necessary for the authorities that take decisions to control and reduce noise pollution in urban environments.

There are papers conducting noise analyses about specific urban sources mainly related to transportation activities. A statistical classification of road pavements is conducted in [6] concerning different types of recordings of road pavement segments with the aim of increasing the accuracy of road traffic noise prediction models. Additionally, in [7] analyzed both scooter and motorbike operation patterns to establish how quality of life due to certain specific features of the emitted noise is affected.

In [8] the classification of several airplane classes, based on novel take-off noise signal segmentation in time and parallel Multilayer Perceptron neural networks is shown. The model works with 13 aircraft categories depending on the installed engine type with classification rate above 85% in real environments. Furthermore, in [9] propose a method to estimate the geo-referenced take-off path, using spatio-temporal information extracted from the airplane engine noise signal.

Another application focused on audio classification is presented in [10] where the goal is to identify species of bird based on audio recording using a probabilistic model for audio features and a nearest-neighbor classifier with a new information metric. The classifier achieved over 90% accuracy by identifying 6 species of bird with better results compared to support vector machines. The study of birds and another animal sounds take great relevance in urban soundscapes, due to the animals being affected by the masking effect of high contamination levels in urban zones as shown in [3].

This paper classifies predominant unstructured environmental sounds in recording signals where the sources are mixed,

* Corresponding author.
E-mail addresses: mglp12@hotmail.com (M.G. López-Pacheco), lsanchez@cic.ipn.mx (L.P. Sánchez-Fernández), hmolina@ipn.mx (H. Molina-Lozano), l.alejandro.2011@gmail.com (L.A. Sánchez-Pérez).

http://dx.doi.org/10.1016/j.apacoust.2016.05.020
0003-682X/

improving the classification up to 9% over using individual dictionaries as Gabor or Wavelet, and reducing iterations required for identification. Several sub-dictionaries in the time–frequency domain were tested on all classes to find a combination of them with better results than using a unique dictionary. The classification is based on creating specific dictionaries of each class using the Matching Pursuit algorithm and a proposed selection criterion as a way to discriminate among the various classes. These mixture signals could be actual recordings or as the result of a source separation method proposed in [11–13], for both cases with a predominant sound source.

The most common classes of environmental noise levels are vehicles (including all types of road transportation), vehicle control (whistle sound), emergency sounds (sirens of patrols and ambulances) and the noise generated by crowds performing recreational, commercial or social activities (this category includes sounds of male and female voices, applause, shouts of acclamation, cries for help, and the noise of crowds when walking or running).

This paper is organized as follows: several relevant previous works about signal decomposition with the OMP algorithm are discussed in Section 2. Methodology proposed based on OMP for classification of predominant sounds is described in Section 3. Section 4 contains experimental evaluation and discussion results of the classification performance. Finally, concluding remarks and future research directions are given in Section 5.

2. Review of previous work

In [14], a Support Matching Pursuit method was introduced to resolve individual echoes in measuring the thickness of layered structures where the resulting echoes from two successive interfaces overlap in time. Another study about sparse coding is the one conducted in [15] to improve the noise robustness of a speech recognition system. Furthermore, the MP technique has been used as a similarity measure based only on the dictionary, coefficients, and residual information, for fuzzy clustering and classification of signals in [16].

There are several areas of non-speech audio classification applications besides environmental acoustics, such as music genre classification [17] achieving 97.1% accuracy in real-word audio collections. Sparse representation of musical signals has been reported in [18], using source-specific dictionaries that efficiently capture music signal characteristics, tested by real piano recordings. Another study that reports sparse approximations is the one conducted in [19] for drum sound classification. The decomposition signal is given using MP and the dictionary is a set of atomic functions learned in an unsupervised manner for mixtures of percussions sounds. The classification accuracy in 3-class database is 87.8%.

The task of recognizing environmental sounds to identify a scene or context recording by audio sensor has been reported in [20] where it is proposed using the MP algorithm to obtain effective time–frequency features. The MP-based method is supplemented with Mel-Frequency Cepstral Coefficients (MFCC) to yield an averaged accuracy rate of 83.9% in discriminating fourteen classes. Moreover, the present paper is focused on the recognition of sound events predominant in a mixture of several sounds, and not on characterizing the general acoustic environment types as a whole.

In [21], the authors work on classifying environmental audio signals constructing the time–frequency matrix using the MP time–frequency distribution (MP-TFD) technique and then apply the non-negative matrix decomposition to decompose the TFM into its significant components. The results show 10% accuracy-rate improvement compared to the MFCC features over environmental classes. However, the classes used (airplane, helicopter, drum, flute, piano, male, female, animal, bird and insect) differ to those studied in the present work, where the main feature is that there is a predominant source (vehicle, whistle, siren, car horn and crowd) with a high acoustic level within a mixture of several sources inter-class for urban soundscapes as shown in [11].

A system using MP for audio classification was presented in [22], this work is very similar to our proposed technique in the use of MP decomposition for signal classification. Their approach classified time-varying warning signals from an actual acoustic monitoring system using modified MP algorithm. The time–frequency decomposition is applied over classes of warning signals from real-world faulty structures with a misclassification rate of 1%. However, the type and variance of the database sound classes is unclear since only a sample spectrogram of each class is shown. The dictionary elements are TF-shifted versions of the learning signals from each class. When the modified MP algorithm stops, the net contribution of the correlation coefficients from each class is used, and the class is selected by a test statistic with a fixed number of iterations.

Research conducted in [22] differs from the present paper mainly in the dictionary used and the selection class criterion. This paper proposes a dictionary based on several families of wavelets that model the characteristics in time–frequency domain and the selection of class does not require a fixed number of iterations, but the class is selected by evaluating criteria at each iteration decreasing the fixed iteration number commonly used in MP.

3. Methodology proposed to classify based on Orthogonal Matching Pursuit

The Orthogonal Matching Pursuit (OMP) is an iterative process for decomposing signals in a redundant dictionary as it is analyzed in [23]. This technique provides a sparse linear expansion of waveforms where convergence is guaranteed due to the residual signal approaching zero energy. The dictionary D is a collection of parameterized waveforms ϕ_γ of length n called atoms. The parameter γ represents the index for time–frequency jointly, as in the case of a time–frequency dictionary described in [24].

The performance of classification using OMP algorithm depends mainly on the dictionary used. Section 3.1 shows a background review about several dictionary atoms in time–frequency domain; these atoms were tested by experiments to build the source-specific acoustic dictionary over predominant classes (see Table 2). Fig. 1 shows a block diagram of the proposed methodology to classify predominant sources from a mixture signal in time–frequency domain which consists of two stages: creation of source-specific dictionary shown in Section 3.2 and classification based on OMP algorithm evaluating the maximum energy reached at each iteration and stops when the maximum value is unique. The algorithm steps are listed in Table 1.

3.1. Selection of atoms for decomposition

It is relevant for classification performances to choose the most suitable atoms that represent the classes using several types of sub-dictionaries as described in [25]. The different classes are better adjusted to a kind of sub-dictionary than any other, thus, dissimilarity generates among the other classes. Several types of atom could be used in a dictionary due to the convergence OMP decomposition being not dependent on the atom type. Dictionaries widely used in frequency (DCT, COS, SIN) and time–frequency domain (Gabor, DB, SYM, COIF) are given as follows.

(A) Discrete Cosine Transform (DCT) orthonormal basis is defined as: