euronoise

Automatic classification of traffic noise

Manuel A. Sobreira-Seoane, Alfonso Rodríguez Molares and José Luis Alba Castro

University of Vigo, E.T.S.I de Telecomunicación, Rúa Maxwell s/n, 36310 Vigo, Spain
msobre@gts.tsc.uvigo.es

The last review of the international standard "ISO 1996-2:2007, *Determination of Environmental Noise Levels*" [1], in its 6.2 section states that if the Leq of road traffic is measured and the results are going to be used to calculate to other traffic conditions, the number of vehicles and the classification of at least two categories of vehicles: "light" and "heavy" should be registered. In this paper, a first approach to get an automatic classification of vehicles is presented. Some basic classifiers have been tested (k-nearest neighbours, FLD (Fischer Linear Discriminator) and Principal Components. As first approach, the aim of the job was to determine if the different classes (trucks, cars and motorbikes) could be separable using different time and frequency characteristics. The results shows that for some of the characteristics the signals are separable, so a continuous traffic noise signal could be processed to get the information of the number of heavy trucks, cars and motorbikes that passed by during the measurement period. Information of a stereo recording could be used to get information of the direction of the vehicle.

1 Introduction

Time and frequency characteristics of signals provide relevant information thanks to which we could say that a sound contains the individual and unique signature of a certain source. This signature could be considered unique if the right characteristic or characteristics are taken into account. As an example, one could not distinguish between a piano and a violin if the spectral characteristic considered is just the fundamental frequency of the note they are playing. If a piano note is recorded and reversed in time (played backwards), then, although the spectral contain is the same, the time envelop of the sound and the time envelop of every harmonic has changed in such a way that the sound is not far away from the one a bowed string. Therefore both, time and frequency characteristics, are quite important to distinguish or classify different sound sources.

If the complexity of the problem increases (classification of sources of the same kind) the number of time and frequency characteristics to consider the sound signature as unique will increase. The noise emitted by a diesel engine of a heavy truck and the one of a light vehicle are not so different. Anyway, most of us can distinguish between the sound of a truck and the sound of a car. So the characteristic or the set of time and frequency characteristics that makes this sound different should be found to proceed with an automatic classification of these sources. Once the set of characteristics are stated, different classification algorithms could be used to determine if a new sound belongs to one of the classes that have been modeled with the previous characteristics analysis. It is quite clear that the final result will depend on the combination the set of features chosen and the classification method selected. With some experience and knowledge on classification techniques, some of the methods can be selected and some others just rejected. Anyway, the process to get good results and to improve them is a kind of trial and error test.

The process of the classification of noise sources includes several stages: first the sound should be preprocessed (background noise suppression, segmentation of continuous signal into single events, etc). Once preprocessed the signal features will be extracted. A vector of characteristics (signature of the source) is then sent to the classification algorithm which be then report the class (or set) the signal belongs to. In a previous stage, the classes should be defined and the model trained with a set of known signals. The figure (1) shows the basic structure of a classification system.

Noise sources signature recognition in general and vehicle noise classification in particular has been studied very little compared to speech recognition or music genre classification, although some related literature can be found. The feature extraction techniques and the classification algorithms used can be found in the common literature on the topic [5, 6, 7].

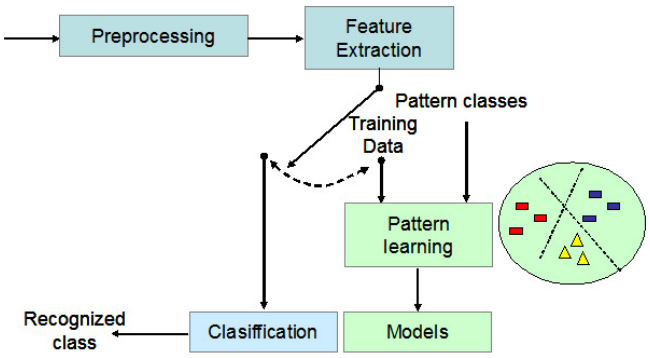

Figure 1: Basic structure of a classification system

To develop this work on automatic classification, a database of vehicles pass-by signals has been recorded: signals of 100 different motorbikes, 100 cars and 100 heavy trucks have been recorded. A flat road with mid-density traffic, shown in figure (2) was selected to get a set of clean signals. Any recording with high background noise or wind was rejected. As first approach, the possibility of simultaneous vehicles passing by is not considered and is left for next future research. Two microphone have been used, so the speed and sense of and sense of circulation of the vehicle can be also estimated.

Figure 2: Road selected to record the database.

2 Vehicle Detection

In this section a brief description of the vehicle detection stage is described. This is the critical stage, whose role is to detect whether a vehicle has passed by and send the segment of signal to the feature extraction block. The vehicle detector just says if traffic noise is present, extracting the traffic noise signal from the background signal. The traffic signal could be a single vehicle (light or heavy) or a combination of vehicles (simultaneous pass by).The kind (or class) of event will be decided by the classification stage. A basic algorithm, to separate the traffic signal has been used. The equation (1) defines the Short Time Energy for the N-sample of a frame t.

$$STE_t = \sum_{n=0}^{N-1} |x_t[n])| = \frac{1}{N} \sum_{k=0}^{N-1} |X_t[k]|^2 \qquad (1)$$

Any given frame will be cataloged as environmental noise frame or traffic noise frame depending on the vaule of the STE compared to a given threshold. The best approach tested to fix the values chosen for the thresholds, TH, is besed on the statistical noise levels,L_N , indicating the sound level that is exceeded a certain fraction $N\%$ of the time over a given interval (e.g., 15 minutes). The L_{90} level could be considered as the background noise level, although the time percentage L_N will have to be adjusted for our particular case depending on the location's traffic flow average. Consequently, the appropriate L_N value as silence TH will be used and a multiple of this as $TH_{traffic}$ ($TH_{silence} + 3dB$ and $TH_{silence} + 6dB$ depending on the traffic conditions). The figure (3) shows an example of segmentation of the traffic noise signal using the STE.

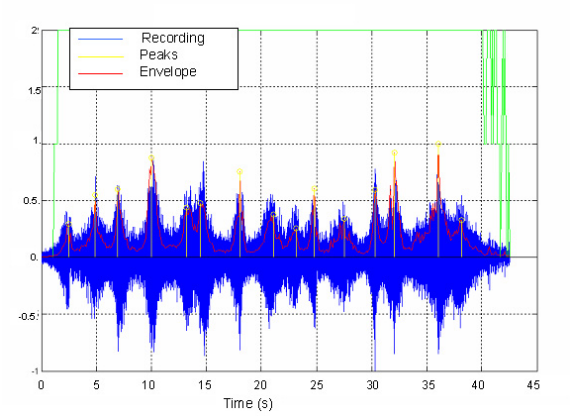

Figure 3: Example of vehicle detection using STE with the continuous traffic signal used for test of the classification methods.

Once the traffic noise intervals are detected, the next step is try to isolate each of the individual events (a traffic event could contain two or more simultaneous vehicles). Once this objective is achieved, we will be ready to proceed to the next stage: classification of samples. The simplest way of detecting whether a vehicle is passing or not is by analysing the temporal evolution of the envelope signal, looking for maximum value peaks. As we are dealing with blocks of a certain length N for the analysis, a rough scaled estimation of the envelope can be easily determined via the STE of every individual frame. For the purpose of this job, this procedure will give us an accurate enough estimation as long as N is short enough. The smaller the value of N , the closer possible vehicles will be detected. The figure (4) shows the detection of different vehicles with a high degree of overlapping. The traffic noise is then cleaned, removing the background noise, using a estimation of the background signal taken in the silence periods [2, 3].

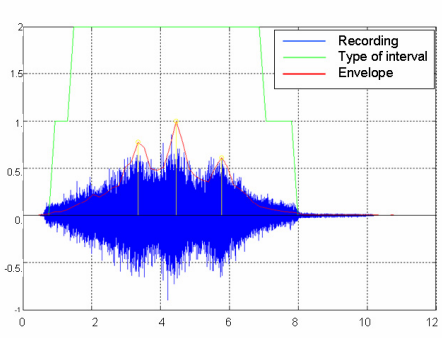

Figure 4: Example of traffic segmentation with high overlapping.

3 Features extraction

The choice of a feature set is the crucial step in building a pattern classification system, for its results will determine the classifier's final response. These features will constitute a new feature space that will replace the original sample space for classification. Therefore, in order to get high accuracy for classification, a good set of representative characteristics should be selected. Thus these parameters can be grouped into two categories according to the domain in which they are calculated. These categories are spectral features (frequency-domain) and temporal features (time-domain). In the next subsections both categories and the features tested are described. The definition of these magnitudes and the signal analysis procedures are described in the classic bibliography on signal processing, as [5]. The use of these features with in pattern recongnition is described in [6].

3.1 Temporal features

Zero Crossing Rate–ZCR: this parameter is defined as the number of time-domain zero crossings within a processing frame and, although it is calculated in the time-domain, it gives an idea of the frequency content of the signal, showing its noisiness. It can be calculated with the following expression:

$$ZCR_t = \frac{1}{2} \sum_{n=0}^{N-1} |sign(x_t[n]) - sign(x_t[n-1])| \qquad (2)$$

where sign() represents the sign function, with value equal to 1 for positive arguments (including zero) and -1 for negative ones.

3.2 Spectral Features

Spectral Centroid: it represents the the centre of gravity of the spectral power distribution. It is related

to the brightness of a sound (more high-frequency than middle or low-frequency content), and so the higher the centroid, the brighter the sound. The spectral centroid for a processig frame t can be calculated as:

$$Centroid_t = \frac{\sum_{k=0}^{N-1} |X_t[k]| .k}{\sum_{k=0}^{N-1} |X_t[k]|} \quad (3)$$

Spectral Rolloff Point: [8]: this feature measures the frequency below which a specific amount of the spectrum magnitude resides. It measures the"skewness" of the spectral shape. The rolloff point is calculated as:

$$SR = max_m \left\{ \sum_{k=0}^{m} |X_t[k]| \leq TH \cdot \sum_{k=0}^{N-1} |X_t[k]| \right\} \quad (4)$$

where the threshold, TH, takes values between 0.85 and 0.99.

Subband Energy Ratio–SBER: the ratio of the energy in a certain frequency band to the total energy. Its expression is, being S_i the i-th sub-band:

$$SBER_t = \frac{\sum_{k \in S_i} |X_t[k]|^2}{\sum_{k=0}^{N-1} |X_t[k]|^2}\} \quad (5)$$

The spectra are divided into non-uniform intervals, typically 4 full octave sub-bands:

$$\begin{aligned} S_1 &= [0, f_0/8] \\ S_2 &= [f_0/8, f_0/4] \\ S_3 &= [f_0/4, f_0/2] \\ S_4 &= [f_0/2, f_0] \end{aligned}$$

where f_0 is half of the sampling frequency. The figure 5 shows the SBER for the 4th subband. It can be seen how there are clear differences between three classes: motorbikes, cars and heavy trucks, so this is one of the main features to be considered to solve the problem of automatic classificacion of traffic noise.

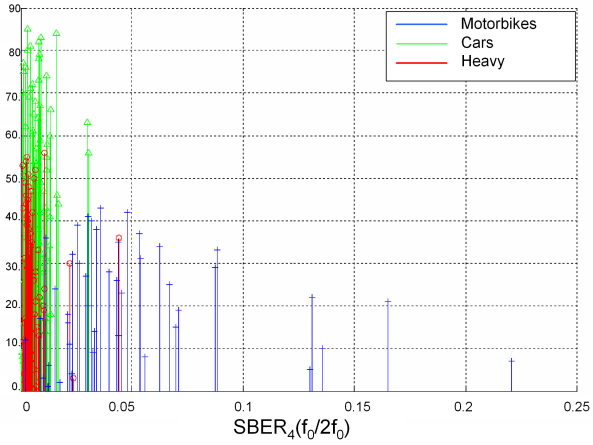

Figure 5: Energy ratios for the 4th subband.

3.3 Perceptual features: Mel parametrization

Mel-Frequency Cepstral Coefficients (MFCC) are a perceptual parameter that can be used to characterize our traffic noise signals. The sense of "perceptual" lies in the fact that they are meant to approximate the response of the human auditory system: that is, if a person is able to recognize whether a given noise belongs to either a conventional car or a motorcycle, it might be possible to reproduce, or at least approximate, those subjective features upon to which the human ear is dependent. For instance, 13 MFCC coefficients are usually employed to represent voice, although for classification purposes 5 of them have been proved to be just enough [6]. Their performance when applied to our concrete theme will be discussed later.

To obtain the MFCC, the signal is filtered in frequency domain with a Mel scale filter bank. Then, the inverse Fourier Transform of the logarithm of the Spectrum is obtained.

4 Classificacion algorithms

4.1 k-Nearest Neighbour – k-NN

The k-NN classifier places the points of the training set in the feature space and picks the k points nearest to the test point. Thus, a given point in the space will be assigned to a concrete class if this is the most frequent class label among the k nearest training samples. If just one feature is used, the Euclidean distance can be used as measure, but this can distort the calsisfication for an N-dimension space, where N features can be used. To avoid this, the Mahalanobis distance defined in Eq (6) is used.

$$d_M(x, y) = \sqrt{(x-y)^T C^(-1)(x-y)} \quad (6)$$

where C is the covariance matrix of the training set of data. The use of this measure has two main advantages over the Euclidean distance:

- It decorrelates the different features, though this decorrelation is done to the whole set of training samples as one entity, and not for every class separately. This relies on the assumption that the covariance matrix is the same for all classes, which is not true for a majority of the practical cases.
- The Mahalanobis metric is scale-invariant, i.e., it does not dependent on the scale of measurements, which means it automatically scales the coordinate axes of the feature space.

The choice of the number of neighbours to be considered, k, it depends on the data. High values of k will reduce the effect of noise in the classification, but the borders between classes becomes more complex.

4.2 Fischer Linear Discriminant – FLD

Classifiers based on Linear Discriminant Analysis are supervised methods that employ the label information

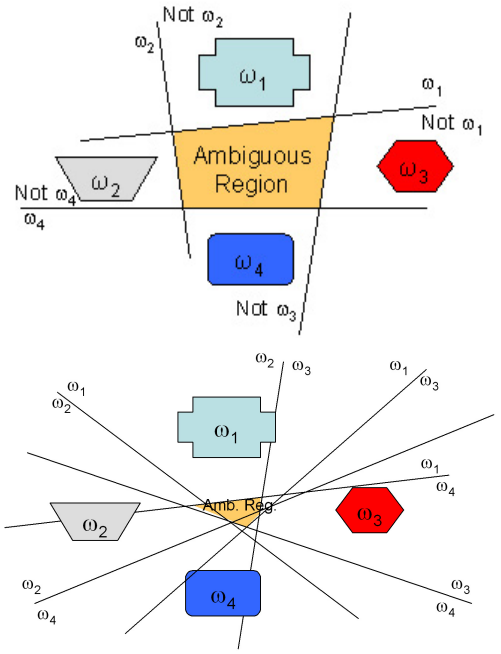

Figure 6: Linear boundaries stablished by One versus All (top) and One versus one (bottom)

of the training data to establish a linear boundary between the classes. With this purpose, the analysis seeks to project the data from a d-dimensional space onto a line, the discriminant direction. If this is interpreted geometrically, the surface of decision is a hyperplane H_s, and the discriminant direction is orthogonal to this hyperplane that separates the zones of decision. This method only works, consequently, for two separable e categories (C1 and C2), although this can be extended to an arbitrary number of classes.

The discriminant direction will be the solution of minimizing/maximizing a criterion function. Fisher Linear Discriminant (FLD) analysis proposes the projection onto the vector w that maximizes the separation of the data in a least-squares sense (Least Mean Square, LMS), weighted by the total within-class scatter [9], which means the criterion is the Mahalanobis distance. A complete description of the FLD can be found at [9].

FLD analysis is only valid for two category classification. Of more classes are implicated the analysis should be extended. The natural generalization of FLD to c classes ($c > 2$) is called *Multiple Discriminant Analysis* and involves c-1 discriminant functions. Another solution to the classification of multiple classes is to divide the problem in several two-class classification. This approach can be fulfilled following two different strategies:

1. **One-Versus All Classification.** This method suggests the training of c classifiers (one class is the positive and the others constitute the negative). So, each of these classifiers will make a class estimation, so that at the end the assigned class will be the one that achieves a higher margin (in case more than one positive class is estimated).

2. **One-Versus One Classification.** This other strategy proposes, instead, the implementation of $c(c-1)/2$ two-category classifiers, such that all the possible combinations are covered. Then, a voting strategy is adopted: each binary classifier generates a"vote", and the estimated class will be that with larger number of "votes".

As can be inferred from figure(6), the One versus One classification has become more popular since it offers a more accurate performance(the ambiguous region is smaller).

5 Results

In oder to test the automatic classification possibility for traffic noise sources (motorbikes, cars and heavy trucks), a database with 100 items of each class was recorded. The signals were recorded in PCM format, with a sampling frequency $fs = 44100bps$ and 16 bits per sample. For purpouse of classification, each signal was downsampled to 11025 bps, so the effective bandwith for the analysis (feature estraction) is $f_s/2 = 5512Hz$. 40 signals of ewach class were selected as set of training and the other signals were used to test the performance of the classifiers.

The ZCR showed good behaviour to discriminate between heavy vehicles and motorbikes, but it was not the best discriminant feature between cars and heavy trucks. Similar results have been obtained for the Spectral Centroid.

The sub-band energy ratio showed a good behaviour: the heavy trucks present higher energy concentration at low frequencies while the power density is higher at high frequencies for the motorbikes. The bands with more discriminant power were:

$$S_3 = \quad [f_0/4, f_0/2] \cong [1.4kHz2.8kHz]$$
$$S_4 = \quad [f_0/2, f_0] \cong [2.8kHz5.5kHzkHz]$$

The SBER for the 4th subband has been shown in the figure (5).

Other spectral feature showing good discriminant properties for this case is the *Spectal Rolloff* with threshold values between 0.55 and 0.70.The last feature showing good discriminant results was the MFCC.

As the standard ISO 1996-2 [1] states that the number of vehicles during the measurement period in "at least" two classes, heavy and light vehicles, should be reported. The first approach was to consider the possibility to distinguish between those two classes; the table (1) shows the error probability using single features. A $k_N N$ with k=3 and a Fisher Linear Discriminant were used. It can be observed how the SBER showed the best result.

The table (2) shows the result of the extension of the previous job to three classes. Both SBER and MFCC showed a good behaviour with the 3-NN classifier. The table (3) shows the results when MFCC, SBER and the Spectral Rolloff are used simultaneously. It can be observed how a simple 3-NN or a FLD with a One versus One strategie shows good results.

It must be considered that the purpose of the classification of vehicles when measuring pass-by noise is to extrapolate the results of the measurement to other traffic conditions. The traffic noise emited by a road is funtion of the $10log(N)$, where N is the number of vehicles. So an error of a 10% leads to an error around 0.5

Parameters	Error probability (%)	
	3-NN	**FLD**
ZCR	30.34	35.45
Spec. Centroid	51.49	38.25
Spectral Rolloff	28.68	26.12
SBER	12.18	10.17
MFCC	15.73	13.57

Table 1: Error probabilities for two classes using single features (heavy and light vehicles)

dB in the estimation of the sound pressure level. The expected error in tha calculation of traffic noise is even larger, mainly to the weather conditions, so a error of 10 % in the estimation of the number oh vehicles of each class could be asumed although further improvements are needed to get a lower error probability.

Parameters	Error probability (%)		
	3-NN	**FLD**	
		one vs all	**one vs one**
ZCR	39.98	-	43.97
Spec. Centroid	38.43	-	27.04
Spectral Rolloff	31.94	-	22.01
SBER	16.24	36.21	27.59
MFCC	18.42	17.19	15.76

Table 2: Error probabilities for three classes using single features

6 Conclusions

The paper showed a 1st approach to the problem of automatic classification of traffic noise signals. It has been identified the Subband Energy Ratio as the feature with higher discriminant performance. This spectral characteristic together with the MFCC and the Spectral Rolloff leads to good results using a 3-NN classifier.

The results presented in the paper are good enough to be promising, which means that it should be worth further research to improve the results: the database should be extended and the training sets should be bigger. It should be considered the possibility to extend the number of classes to deal with the problem of joint signals (simlutaneous pass by of different vehicles), and the use of different classification techniques as neural networks could be considered.

Parameters	Error probability (%)		
	3-NN	**FLD**	
		one vs all	**one vs one**
MFCC, SBER Spec. Rolloff	10.07	13.36	11.82

Table 3: Error probabilities using the best combination of three joint features

Acknowledgments

This work has been partially financed by the Spanish MEC, ref. TEC2006-13883-C04-02, under the project An-ClaS3 “Sound source separation for acoustic measurements´´.

References

[1] ISO 1996-2:2007.”Acoustics - Description, measurement and assessment of environmental noise. Part 2: Determination of environmental noise level”. 2nd Edition, (2007)..

[2] Vary, P. ”Noise suppression by spectral magnitude estimation – mechanism and theoretical limits”. *Signal Processing* 8(4), 387-400 (1985)

[3] Kamath, S. and P. Loizou (2002). ”A multi-band spectral subtraction method for enhancing speech corrupted by colored noise”. *In Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing* (ICASSP’02) (2002)

[4] Harb et al.,”Voice-Based Gender Identification in Multimedia”, *Applications Journal of Intelligent Information Systems*, 24:2/3, 179-198 (2005).

[5] John G. Proakis,Dimitris Manolakis. *Digital Signal Processing. Principles, Algorithms and Applications*. Prentice Hall, Febrero 2004. ISBN 3-528-35558-1.

[6] Enrique A. Cortizo, Manuel Rosa-Zurera and F. López Ferreras.”Application of Fischer Linear Analysis to Speech/Music Classification”. *Proceedings of EUROCON*, Belgrado 2005, pp. 1666-1669.

[7] Dietrich W. R. Paulus, J. Hornegger.*Applied Pattern Recognition*. Fourth Edition. Ed. Vieweg, Febrero 2004. ISBN 3-528-35558-1.

[8] V. Peltonen. *Computational Auditory Scene Recognition*. Master of Science Thesis, Tampere University of Technology.

[9] Max Welling. *Fischer Linear Discriminant Analysis*. At `http://www.cs.huji.ac.il/~csip/Fisher-LDA.pdf`