ICSV14
Cairns • Australia
9-12 July, 2007

14th International Congress on Sound & Vibration

A CEPSTRUM-BASED METHOD FOR REMOTE TRAIN DETECTION

Carlos F. Carvalho[1]

[1]Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa
Departamento de Engenharia de Electrónica e Telecomunicações e de Computadores
Rua Conselheiro Emídio Navarro, 1
P-1950-062 Lisboa, Portugal

cfc@isel.ipl.pt

Abstract

On the subject of remote train detection, previous work has already been made, applying different techniques. The system described in this paper, comprises both hardware (receiver circuit connected to an accelerometer) and software subsystems. The latter subsystem employs a technique based on vector quantization (VQ). Through a training process, two centroids have been derived, resulting from the application of an unsupervised learning algorithm (Lloyd algorithm), using a 20th-order Cepstral analysis. Each centroid represents one of two different classes of signals: “Silence” (or absence of a moving train) and “Moving train” (approaching or withdrawing). What is being presently proposed is an alternative to a previous representation of the centroids, in which they were derived from a 14th-order Linear Predictive Coding (LPC) analysis. Similarly, incoming signals coming from the receiver must also undergo a Cepstral analysis. The input of the software subsystem consists of samples of the input signal, obtained by the hardware subsystem. Sampling rate is 16,000 samples per second, being the samples windowed into 300ms frames, with a window displacement of 100ms. The recognition (or classification) process is based on a distance measure to each one of the two centroids. The Euclidean vector distance measure was used for this purpose. Collected data was divided into training and classification corpora, respectively 67% and 33%. There were considered situations in which one had trains running on the same track where the receiver is connected to, and trains running on neighbour tracks, being the vibrations transmitted through the ground to the receiver circuit. For both situations, classification results are presented and discussed, comparing performances between the Cepstrum-based and the LPC-based pattern matching processes.

1. INTRODUCTION

Regarding the subject of remote train detection, some previous work has already been made. The goal is to detect a moving train as far as possible, concerning a certain location. Preliminary work consisted of an active system which had transmitter and receiver circuits, being addressed in [1] and [2]. The transmitter was responsible for the generation of acoustic pulses (and their transmission to the rail) and the receiver, responsible for the reception of the same pulses, but delayed. Knowing the round trip delay and the velocity of sound in the railway tracks, one could even estimate the distance at which the train was, when there was any. Unfortunately, this system was considered to be inefficient and another approach had to be developed.

An evolution of this system consisted on the use of the receiver circuit only, and on the analysis of the vibrations signal collected from the rail [3]. These vibrations can correspond to the situations of idle activity or of trains running on the rails, approaching or withdrawing. On what concerns to train vibrations, this system made no distinction between situations where there were trains moving on the same track where the receiver was listening, or alternatively, vibrations coming from moving trains running on neighbour tracks. The vibration's signal was modelled using a Linear Predictive Coding (LPC) analysis on each data window. Confronting the coefficients obtained in a given window against pre-stored centroids, the signal contained in a certain window could be classified as representing a moving train, or the absence of a moving train.

The work that was developed, and that led to this paper, can be considered similar to the one based on the LPC analysis, once it is also based on the principle of Vector Quantization (VQ), and had as an inspirational principle the subject of speech recognition.

2. LINEAR PREDICTIVE CODING (LPC) AND CEPSTRAL COEFFICIENTS

When we have a signal that we want to code according to some coding scheme, several alternatives exist to achieve this task. LPC has already been widely used in speech synthesis and recognition systems. The work that was developed used LPC for modelling the data that was being acquired by the receiver circuit. The original signal came from an accelerometer in direct contact with the rail. After performing some experiments, it was chosen a window length of 300ms, with a displacement of 100ms [3]. The sampling rate was 16,000 samples per second, having each sample linearly quantized to 16 bits. LPC has a general model [4] given by,

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^{N} a(k) z^{-k}}, \tag{1}$$

which is assumed to be excited by a source with a flat spectral envelope. Coefficients $a(k)$ are called the predictor coefficients, and represent an optimal estimate to the spectrum of the windowed signal using N poles (the order of the LPC polynomial).

An alternative representation of this information can be obtained using another format, which eventually can be more useful in the sense that can be more suited to the kind of signal we are dealing with, or more physically interpretable.

The LPC cepstral coefficients provide one such alternative representation, and are a very important parameter used in speech recognition. Generally, the order of the set of cepstral coefficients is larger than the one used for LPC coefficients, and it is usually made to be approximately $Q = 3/2 \times N$. Since we have $N = 14$ for the LPC order [3], the order used for the Cepstrum coefficients was $Q = 20$.

The cepstral coefficients, $c(k)$ can be derived directly from the set of previously determined LPC coefficients $a(k)$, using the following recursion.

$$c(k) = a(k) + \sum_{i=1}^{k-1} \frac{i}{k}\, c(i)\, a(k-i)\,. \tag{2}$$

This is performed over all $a(k)$ coefficients, for k = 1 to k = N, computing one coefficient $c(k)$ in each iteration. However, this expression only gives us the coefficients $c(k)$ up to k = N, but we are still left the computation of coefficients running from N + 1 to Q. These coefficients are given by the following recursion, where k starts at N + 1 and goes up until Q, computing one coefficient $c(k)$ in each iteration:

$$c(k) = \sum_{i=k-N}^{k-1} \frac{i}{k}\, c(i)\, a(k-i)\,. \tag{3}$$

The two recursions stated above came from [5]. Cepstral coefficients, being the coefficients of the Fourier transform representation of the log magnitude of the spectrum, have already been shown, and are considered to be, more robust for speech recognition than the LPC coefficients. The goal of this work is to verify in what measure this robustness is also extendible to the class of signals we are dealing with, taking advantage of this benefit.

3. THE TRAINING PROCESS

Using an acquisition apparatus, date were collected from the rail. In some situations, we had signals concerning to trains running on the same track where the receiver was connected to. However, in some other situations, signals were originated by the rail excitation of moving trains on adjacent tracks, which could also reach the receiver circuit indirectly through the ground. What would be most natural to consider was that these additional signals would have no interest, nevertheless both data were equally considered in the training process, because the goal is to classify the vibrations to belonging to a moving train or not, regardless of their source. Consequently, data will be grouped into two classes, “Silence” and “Moving train”, leading us to have a codebook with a size L = 2.

There is a training corpora, consisting of about 67% of the whole set of data, leaving the rest 33% for recognition. The two classes were individually trained according to an automatic procedure which received all available data and the algorithm was free to decide how to allocate data to both centroids. Each centroid is made up of coefficients concerning a 14th-order LPC analysis. In a starting phase, we had a centroid resulting from the average of all data. Afterwards, this was partitioned in order to have a new additional centroid, since the codebook will have two. Data was now reclassified with these two centroids, and a new clustering within each class was performed using the Lloyd algorithm [4]. This process was then iterated to generate the two final centroids. Within each iteration, classification is made upon a minimum-distortion or nearest neighbour selection rule.

However, it should be noted that the mentioned average was not taken directly from the LPC coefficients. Instead, the corresponding reflection coefficients (RC) were derived from the LPC polynomial for stability reasons. When the iterations are finished, the RC coefficients are converted back to LPC coefficients.

Training was also made using a manual segmentation of the training data, where the main idea is to provide each centroid with the suitable data. This accounts for human decision, based on listening tests in order to decide which data indicates a moving train or merely background noise or other kind of noises (but no train).

Figure 1 depicts the Bode amplitude diagram of the two automatically resulting centroids (red and blue plots) and the manually clustered centroids (cyan and black plots).

Blue and black plots correspond to the "Silence" centroids while red and cyan are for the "Moving train" centroids.

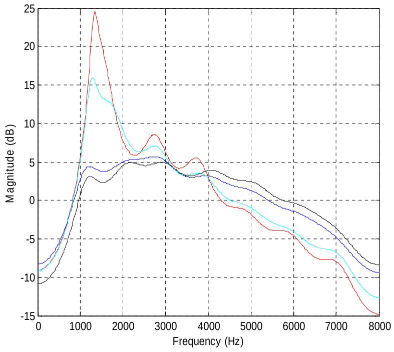


Figure 1 - Bode amplitude diagram of the centroids, both manual and automatic.

To conclude the training process, the set of LPC coefficients is converted into a set of 20th order cepstral coefficients using the recursions of equations (2) and (3), giving the final coefficients determination.

The overall training process is summarized in Figure 2.

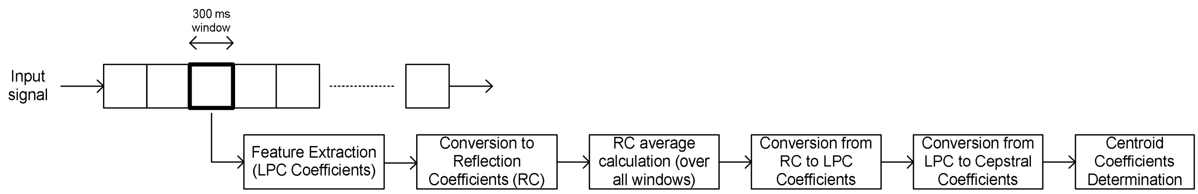


Figure 2. Block diagram of the training process.

3.1 Distance measure

During the training process, a distance measure is computed against each of the two centroids, which are iterated until some decision rule is met. The distance measure that was used was the Euclidean distance [6], stated by

$$d_E(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{N} |x_i - y_i|^2} = \sqrt{[\mathbf{x}-\mathbf{y}]^T[\mathbf{x}-\mathbf{y}]}\,. \tag{4}$$

This distance measure is the most suited when using cepstral coefficients.

4. THE RECOGNITION PROCESS

The classification or recognition process is accomplished by individually analysing each incoming frame containing the signal listened from the rail. For each frame, a 14th order LPC analysis is made, followed by a conversion to a set of 20th order cepstral coefficients. According to the Euclidean distance and following the nearest neighbour rule, determination is made upon the class to which the data contained in the frame belongs to, by using the previously determined centroids, based on cepstral coefficients.

However, as was seen in [3] by taking preliminary results, the classification exhibited a great deal of false alarms, foreseeing the need for using something else to aid in the decision process. It was seen that the distance asymmetry (the difference between the distances to each centroid) and the energy level could be useful for reducing these detection errors. After performing some tests, a heuristic was established to decide which decision to take based on the information contained in these two parameters [3]. The same heuristic was used in this work in order to put both approaches evenly with the purpose of having a better performance comparison.

This recognition process can be summarized according Figure 3.

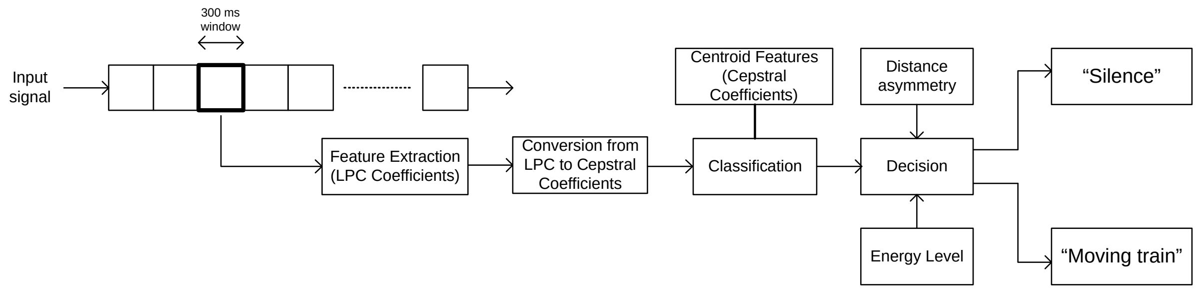


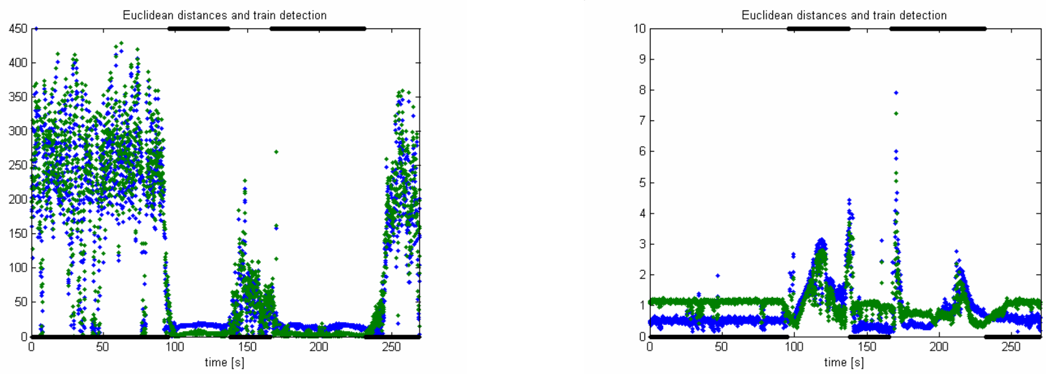
Figure 3. Block diagram of the recognition process.

5. EXPERIMENTAL RESULTS

Taking the classification corpora, data was subjected to classification in order to assess the performance of the system. In the following figures, for each frame of the signal, three informations are given in the form of dots. The green dots indicate the distance to the “Moving train” centroid, the blue dots indicate the distance to the “Silence” centroid and the black dots give an indication about the output of the system. This indication, when high, tells us that the frame was classified as “Moving train”, and as “Silence”, otherwise.

The magnitude of the distances is different when using cepstral or LPC coefficients because the metric is also different, according to each approach.

In Figures 4 and 5 there are depicted situations for trains running on the same track as the sensor’s, or on adjacent tracks, respectively. The obtained result is confronted with the result obtained for the same situation when using an LPC-based system. In order to have a better level of comparison, results for both cepstral and LPC-based algorithms come from an unsupervised training algorithm, leaving aside the manual training.



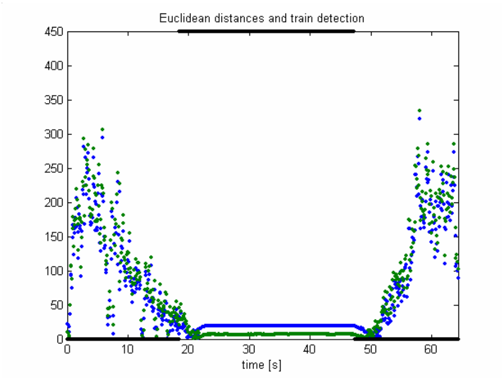
a) Using cepstral coefficients

b) Using LPC coefficients.

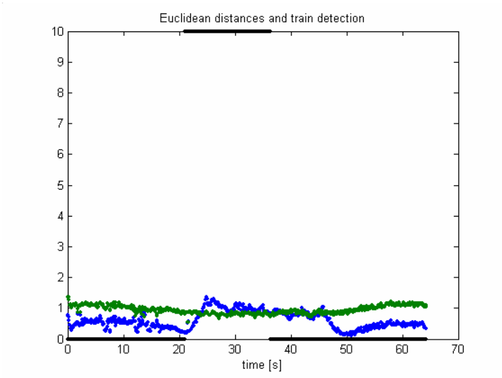
Figure 4 - Classification results for centroids resulting from automatic clustering data, for a train moving (approaching a train station, stopping and withdrawing) on the same track as the sensor’s.

At a first glance, in Figure 4, the train indications using one or the other type of coding representation, yield very similar results.

In Figure 5, it can be seen that in the cepstrum-based method, the train begins to be detected before the moment it does by the LPC-based method and that in the withdrawing phase, it can be sensed up to some later instant using the first method when compared with the second.



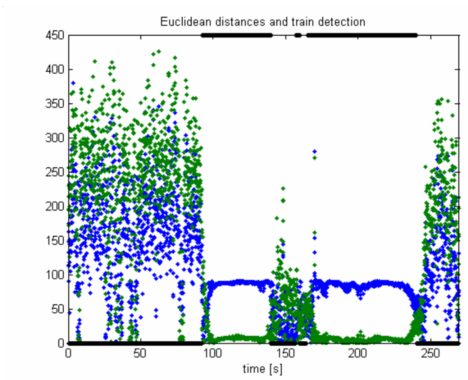
a) Using cepstral coefficients



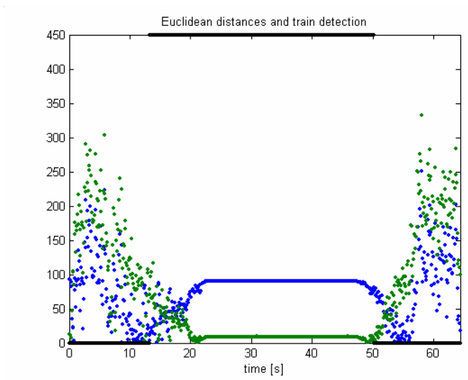
b) Using LPC coefficients.

Figure 5 - Classification results for centroids resulting from automatic clustering data, for a moving train (non-stopping) on a contiguous railway line.

It should be emphasised that it is preferable to have a fully automated system in which human intervention is reduced to the indispensable. However it is interesting to have an idea about the performance of the cepstral-based system when training is accomplished through a manual segmentation of the training data. In Figure 6, classification for both of the previously tested situations is shown, but the underlying training procedure is based on a manual partition of the training corpora. It can be noted that for the situation of the train on the same track as the sensor's, there are false alarm errors, so the results obtained by the system where the training was automatically done, show to be better. On the other hand, the indication of moving train vibration on a contiguous track stands for a longer period.



a) Moving train (approaching a train station, stopping and withdrawing) on the same track as the sensor's.



b) Moving train (non-stopping) on a contiguous railway line.

Figure 6 - Classification results for centroids resulting from manual clustering data, using cepstral coefficients.

5.1 Train distance assessment

The sooner the train is detected, more time we have to take any necessary measures. According to a study done in [1], by knowing the velocity of sound in the railway tracks, and the time when the train begins and ends to be sensed, one can estimate the distance at which the train begins and ends to be detected (in the beginning of the approaching phase and in the end of the withdrawing phase, respectively).

In order to have a clear comparison on the performance of this automatic detection system, using one type of coefficients or the other, Table 1 shows the average results according to all the experiments that were made, showing the distances at which the train begins and ends to be detected. These results include distances using LPC coefficients, already presented in [3], and using cepstral coefficients.

	Automatic detection distance using LPC coefficients [m]	*Automatic detection distance using Cepstral coefficients [m]*
Approach phase	1240	1040
Withdrawal phase	830	890

Table 1 – Estimated train detection distance by the automatic system, using LPC coefficients and cepstral coefficients.

It can be noted that, in the approaching phase, the detection distance by using cepstral coefficients is poorer than the one by using LPC coefficients. However, in the withdrawing phase, the use of cepstral coefficients lets us track the train up to a longer distance. The tests underlying these results used data with trains moving on the same track as the sensor's, because only in this situation it makes sense to have an estimate of the distance to the detected train.

In situations where we have trains running on neighbour tracks, whose vibrations get to the sensor through the ground, we cannot have an estimate of the distance to the train. This is because we have not taken into account for the transfer function of the terrain between the tracks where the train is moving, and the track where the accelerometer is attached to. We can only take into account the time indication for the detection. Based on tests performed in these situations, it was seen that one could, in average, detect a train 6.5 seconds before the same detection could be achieved by using LPC coefficients, and that the train could be perceived (in the withdrawing phase) for more 10.4 seconds than by the LPC-based algorithm.

6. CONCLUSIONS

In this paper, an automatic remote train detection system has been described. The underlying principle for this system is pattern recognition using vector quantization, taking as a base, the same methods used for speech recognition. Through this, one can classify the frames of data collected from an accelerometer placed on the railway tracks, to belonging to one of two classes: "Moving train" or "Silence".

Prior to classification, this system is trained in order to get the coefficient values that model each of the two centroids. The model is established using a 20^{th} order cepstral analysis. These coefficients are obtained by firstly performing a 14^{th} order LPC analysis and then converting this set of coefficients to the set of cepstral coefficients just mentioned. These are only an alternative representation of the LPC coefficients, but are more suited for speech recognition than the first. In this work it was tested how this could also apply to vibrations coming from the railway line, pursuing the same objective as in [3].

It was seen that the obtained results, when considering trains moving on the same track where the sensor (accelerometer) was connected, were roughly the same, though slightly inferior to those obtained in [3], where LPC analysis was used.

On the other hand, if data indicating moving trains, coming from neighbour tracks, reach the sensor, when comparing the moments when the train begins and ends to be detected, it could be seen that the cepstrum based system was able to keep track of the train's movement in a wider amount of time. So, in this particular aspect, the cepstrum approach has shown to have a better performance than the LPC-based system.

In the future, one could consider the study of the frequency response of the ground to understand why the filtering performed by the soil is favourable to this cepstrum based method.

On the other hand, if one wants to think of this system as being only sensitive to vibrations present in the same track as the sensor's, there could considered an adaptive noise cancellation [7], in order to avoid interference from neighbour tracks. Some applications can be developed in order to make the acquisition in a certain railway track immune to the vibrations resulting from traffic of trains on neighbour tracks.

REFERENCES

[1] C. F. Carvalho, *Sistema de Detecção Remota de Comboios*, MSc thesis, Instituto Superior Técnico (Universidade Técnica de Lisboa), Lisboa, Maio 2004. (in Portuguese)

[2] Carlos F. Carvalho, Carlos R. Martins and Mário S. Nunes, "Remote Train Detection System", *Proceedings of the Twelfth International Congress on Sound and Vibration* (ICSV12), 11-14 July 2005, Lisbon, Portugal.

[3] Carlos F. Carvalho, Carlos R. Martins and António J. Serralheiro, "Vector Quantization Techniques for Remote Train Detection", *Proceedings of the Thirteenth International Congress on Sound and Vibration* (ICSV13), 02-06 July 2006, Vienna, Austria.

[4] John Makhoul, Salim Roucos and Herbert Gish, "Vector Quantization in Speech Coding", *Proceedings of the IEEE*, vol. **73**, no. 11, pp. 1551-1588, Nov. 1985.

[5] Panos Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall, 1987.

[6] John Deller, John Hansen, John Proakis, *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, 1993.

[7] Bernard Widrow and Samuel Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.