ICSV13 - Vienna
The Thirteenth International Congress on Sound and Vibration
Vienna, Austria, July 2-6, 2006

VECTOR QUANTIZATION TECHNIQUES FOR REMOTE TRAIN DETECTION

Carlos F. Carvalho*[1], Carlos R. Martins[2], and António J. Serralheiro[3]

[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

[2] Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento (INESC-ID) and Escola Náutica Infante D. Henrique

[3] Laboratório de Sistemas de Língua Falada (L2F) do INESC-ID and Centro de Investigação da Academia Militar (CINAMIL)

cfc@isel.ipl.pt

Abstract

In this paper, an automatic system for the detection of moving trains is described. This system is based on an algorithm using vector quantization (VQ) and pattern matching techniques on the rail vibrations. The goal for this system is to remotely detect moving trains, in order to avoid labor accidents unfortunately common among railway workers. This system comprises both hardware and software subsystems. The hardware consists of a receiver having, as input, signals captured by an accelerometer in direct physical contact with the rail. The receiver circuit performs preliminary signal conditioning (amplification and filtering). The software subsystem comprises a pattern-matching process based on the evaluation of distortion measures against pre-stored VQ centroids. These represent in some way the spectra of two different classes of signals: “silence” (or, more appropriately, the absence of a moving train) and “moving train” (approaching or withdrawing). Prior to the detection, a training process (manual or automatic one) must be carried out, so the two VQ centroids can accurately represent the two classes of signals. For both cases, centroids result from a clustering algorithm using reflection coefficients derived from a 14^{th}-order Linear Predictive Coding (LPC) analysis. Incoming signals were sampled at a rate of 16,000 samples per second and windowed into 300ms frames (window displacement is 100ms) and undergo a 14^{th}-order LPC analysis. The recognition and the classification process are thus based on distortion measures to each one of the two centroids. Two distortion measures are used and their performance compared: the Euclidean vector distance [3] and the modified Itakura-Saito distortion [5]. In the former, each vector has its energy normalized to unity while, for the latter, energy also is used in the distance evaluation. For the system assessment, collected data was divided into training and classification corpora (respectively 67% and 33%) and classification results are presented and discussed.

INTRODUCTION

Preliminary work on detection of moving trains has already been addressed in [1] and [2] using an active system based on the time delay of acoustic pulses reflected by the train boggies. The proposed technique was then considered to be insufficient for practical applications and, as such, a more efficient detection process was mandatory. The underlying principle of the detection method presented in this paper is spectral pattern-recognition based on the vector quantization (VQ) of the LPC coefficients of windowed data. Linear Predictive Coding (LPC) is a modeling technique that has been widely used in speech processing.

Several experiments were undertaken to determine which window length should be used. Although in speech processing 30ms (and a 10ms displacement) is generally used, we have found that, for our purpose, a window length of 300ms (with 100ms displacement) was more appropriate for getting reliable detection. We have also found that several well-defined peaks (resonances) were quite evident in the vibration spectra, as depicted in Figure 1 and, in order to account for a good spectral modelling, the prediction order *N*, was chosen to be 14. A sampling rate of 16,000 samples per second was used and each sample was linearly quantized to 16 bits.

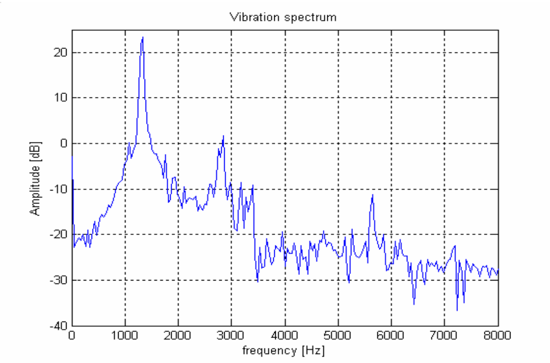


Figure 1 – Rail vibration spectrum due to a moving train, 23 seconds after passing by the sensor

Vector Quantization

It is well known that the excitation source for the LPC model should be iid white noise. In our application, we assume that such a white noise excitation results from the small and random eccentrics that each wheel exhibits. The *N* prediction coefficients, $a(k)$, are computed on a short-term basis over a time frame (window), in which the vibration may be considered approximately stationary.

According to [4], let us assume that $\mathbf{x} = [x_1\ x_2\ \ldots\ x_N]^T$ is an *N*-dimensional vector whose *N* components are real valued, continuous-amplitude random variables. What vector quantization does is to map **x** onto **y**, being **y** the quantized value of **x**. **y** is usually called the reconstruction vector or the output vector corresponding to **x**, and

y is assumed to take only one of a finite set of values contained in $\mathbf{Y} = \{\mathbf{y}_i, 1 \le i \le L\}$, where $\mathbf{y}_i = [y_{i1}\ y_{i2}\ \ldots\ y_{iN}]^T$. **Y** is known as the codebook, being L the size of the codebook and $\mathbf{y}_i$ are referred to as the reference patterns or, centroids. So, L represents the number of divisions (cells) in which the N-dimensional space will be divided into, each of them modeled or represented by a centroid contained in **Y**.
Codebook training can be done using appropriate algorithms, such as the K-means algorithm, also referred to as Lloyd algorithm.

It is also important to stress out that if **x** is quantized as **y**, some dissimilarity (distortion) arises due to the quantization process. So, a distortion measure $d(\mathbf{x}, \mathbf{y})$ can be defined to express the "distance" between them. The overall average distortion is

$$D = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} d[\mathbf{x}(n), \mathbf{y}(n)]. \tag{1}$$

TRAINING AND CLASSIFICATION PROCESSES

For training purposes, data was collected in two different situations. Naturally, we collected data from one rail directly excited by a moving train and, also, by trains moving in the neighbor tracks. In this case, the sensed rail was excited through the ballast. Both data were equally considered in our training process.

Data contained in the training corpora were classified as belonging to one of the two already mentioned classes, (i.e. "silence" and "moving train"), thus starting with L = 2. These two classes were individually trained after both automatic and manual segmentations for the initial centroids.

Distortion measures

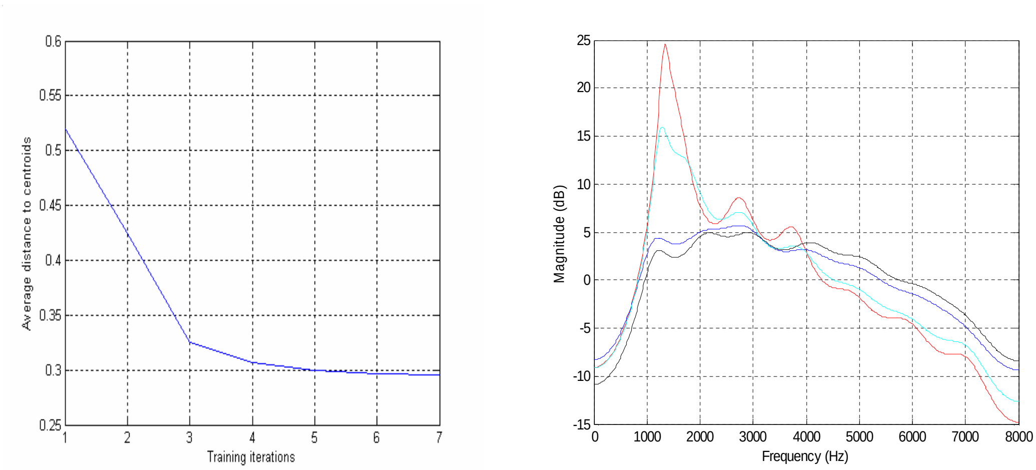
During both training and recognition / classification processes, distortion or distance measures are computed against each of the two centroids. In the training, these centroids are iterated until some decision rule is met, while in the recognition process, centroids are first derived from the training process. Two distortion measures were used: d_E, the Euclidean [3] and $d_{I\text{-}S}$, the modified Itakura-Saito (I-S) [5] distances, defined respectively as

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

and

$$d_{I-S}(\mathbf{u}, \mathbf{v}) = u_1 v_1 + 2 \sum_{i=2}^{N} u_i v_i \ , \tag{3}$$

where the **u** and **v** vectors refer to the autocorrelation of the LPC coefficients for the windowed data and for the centroid. The set of coefficients that describes the centroids is obtained by averaging all corresponding windows within each class. This class partition was, as mentioned, performed both manually and automatically. The underlying idea in the manual segmentation was to provide each centroid with suitable data. In fact, human decision was based on listening tests in order to decide which data would indicate a moving train or background noise. The automatic procedure did receive all available data and the algorithm was free to decide how to allocate data to both centroids. Initially, the resulting centroid from averaging all data was partitioned in order to have a new additional centroid. Data was now reclassified with these two centroids, and a new clustering within each class was performed using the Lloyd algorithm. This process was then iterated and just seven iterations were enough to generate the two final centroids. However, it should be noted that this average was not taken directly from the LPC coefficients. Instead, the corresponding reflection coefficients were derived from the LPC polynomial to ensure that the models (centroids) through the training process would always be stable. Figure 2a) depicts the average distance, taken from equation (1), as a function of the number of iterations. Figure 2b) depicts the Bode amplitude diagram of the two 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.



a) *b)*

Figure 2 – a) Average distance to the centroids as a function of the number of iterations and b) Bode amplitude diagram of the centroids, both manual and automatic (see text).

Classification is made upon a minimum-distortion or nearest neighbor selection rule, according to equation (4):

$$q(\mathbf{x}) = \mathbf{y}_i, \qquad \text{iif } d(\mathbf{x},\mathbf{y}_i) \le d(\mathbf{x},\mathbf{y}_j), \qquad j \ne i, \qquad 1 \le i, j \le L \tag{4}$$

This overall process (training and recognition) is summarized in Figure 3.

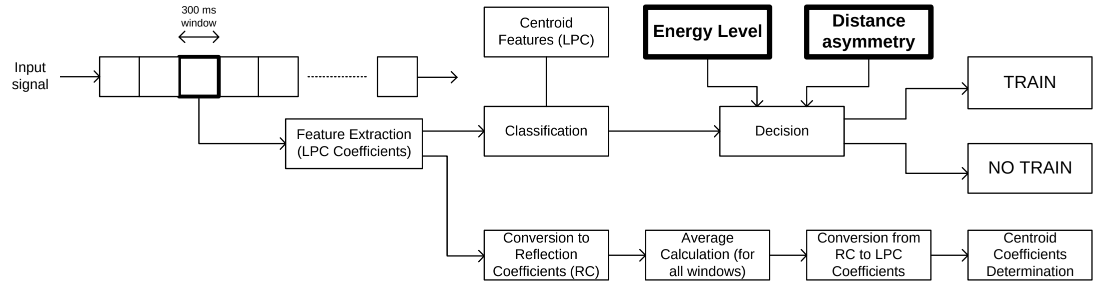


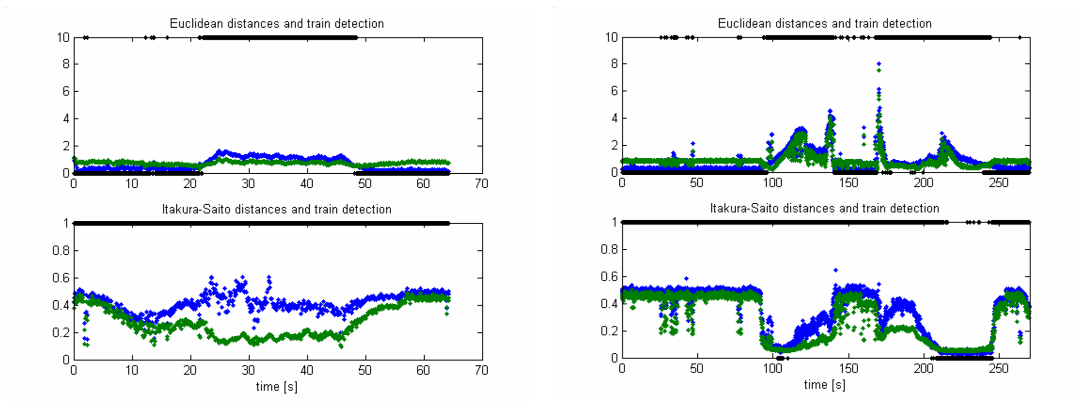
Figure 3 –Block diagram of the detection process.

AUTOMATIC REMOTE TRAIN DETECTION

Preliminary results

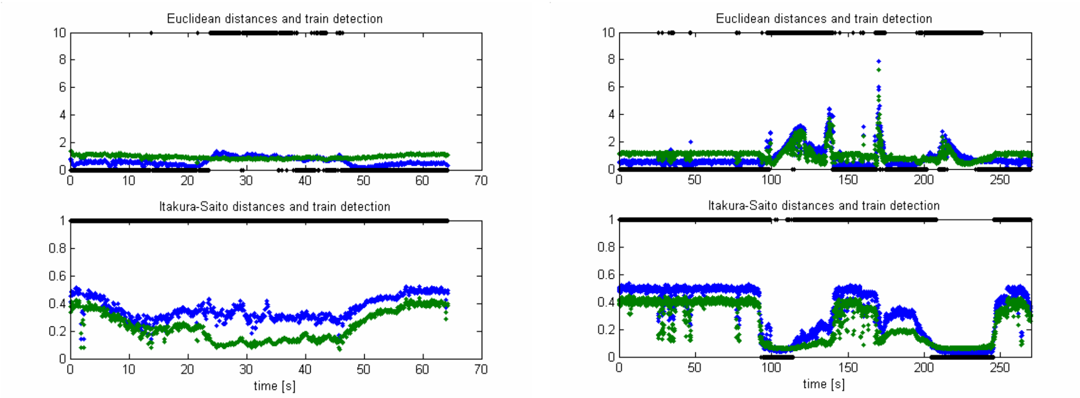
Figures 4 and 5 depict, respectively, the distances of the test data to the two centroids, using both Euclidean (top plots) and modified I-S (bottom plots) distances, running on data not used in the training process (test data). In both figures, a) depicts classification of a moving train on the contiguous rail while, in b), moving on the same rail as the sensor's. In Figures 4 and 5 there are also presented detection results. In these Figures, black dots represent the detection decision of a moving train (amplitude 10 for Euclidean distance; 1, for I-S distance) or no train (amplitude 0).

For all graphics, the green dots indicate the distance to the "train" centroid and the blue dots indicate the distance to "silence" centroid.



a) *b)*

Figure 4 – Classification results for centroids resulting from manually segmented data.



a) *b)*

Figure 5 – Classification results for centroids resulting from automatic clustering data.

As it can be seen, the results present a considerable number of detection errors in the form of *false alarms*. These false alarms must be eliminated as much as possible, in order to increase the confidence in this system. These preliminary results also show that the Euclidean distance exhibits a more reliable indication of train movement while, for these examples, the I-S distance indicates train movement generally all the time, when this is not the case.

Modification of the recognition approach

To account for detection errors, we decided to include the short-time energy and its changes (delta-energy) in the decision algorithm, as shown in Table 1. Furthermore, the decision thresholds should encompass different scenarios. Those two thresholds (low and high) dealing with both the low-medium and the medium-high segments were heuristically established and were set to a value of 5% and 60% of their energy and relative distances values to the centroids, respectively. These changes have been highlighted in the classification process diagram in Figure 3. Also, to suppress isolated detection "peaks", further improving the elimination of false detections of moving trains, a non-linear filter (majority voting) is used over a 5s window.

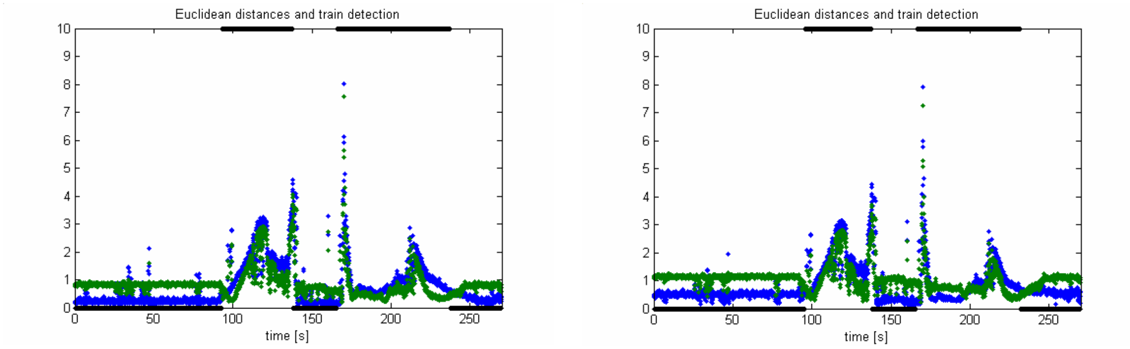
Energy level	*Difference of the Distances to the Centroids*	*Decision*
High	High	Trust classification
High	Medium	Consider TRAIN
High	Low	Consider TRAIN
Medium	High	Trust classification
Medium	Medium	Trust classification
Medium	Low	Consider TRAIN
Low	High	Trust classification
Low	Medium	Trust classification
Low	Low	Consider NO TRAIN

Table 1 – Combination of energy level and distance asymmetry to join to the classification information taken by the distortion measures.

EXPERIMENTAL RESULTS

Detection of moving trains

Figure 6 depicts detection results (black line) for both manually segmented data (left-hand plots) and for automatic data clustering (right-hand plots). Only the Euclidean distance was used as by our preliminary results, and for a train moving (approaching, stopping and withdrawing) in the same tracks as the sensor's.



a) *b)*

Figure 6 – Classification by the Euclidean distance and information on energy level and distance asymmetry, with centroids a) from manual segmentation and b) from automatic training.

Train distance assessment

A detection system should also provide prospective users with an estimation of the approaching train(s). This task was also carried manually through the registration of the time instants were a train could be perceived (approaching phase), or just to faint to be sensed (withdrawing phase). Previous work [2] has shown that 25dB/km is a typical value for energy attenuation as a function of distance. Therefore, distance could be easily assessed by combining this attenuation with the energy level of the vibrations, as shown in Table 2. It is quite interesting to note that human detection can be more subtle, detecting approaching trains much farther away. However, automatic detection of approaching trains does happen when they are significantly distant (more than 1km away).

	Human detection distance [m]	*Automatic detection distance [m]*
Approach phase	1580	1240
Withdrawal phase	1320	830

Table 2 – Estimated train detection distance by human perception and by automatic system.

CONCLUSIONS

In this paper, an automatic remote train detection system was described. This (passive) system is based in VQ techniques over data collected from accelerometers placed on the train tracks. An LPC feature extraction pre-processing stage allows 300ms windowed signal to be tested against pre-stored centroids in order to check if they are "silence" or "moving train". This preliminary detection is complemented with energy and (time) delta-energy to further improve detection. Both automatic and manual generated centroids were used in the evaluation of this system, yielding similar results.

In these experiments, only suburban (electric) trains were considered and we intend to pursuit this work over other kind of trains, namely those with large diesel locomotives, since they do "excite" the tracks in quite a different way and, as such, extra centroids may be needed.

ACKNOWLEDGEMENTS

The authors would like to thank to REFER (Rede Ferroviária Nacional), in particular to Eng. João Alves, for allowing and providing the necessary means for the signal acquisitions to take place, as well as to Mr. Victor Areias, and to all others that gave their contribution for making the acquisition campaigns as fruitful as possible.

REFERENCES

[1] Carvalho Carlos F., Martins Carlos R., Nunes Mário S., "Remote Train Detection System", Proc. ICSV12 – Twelfth International Congress on Sound and Vibration, Lisbon, Portugal, July 2005.

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

[3] Deller John R., Hansen John H. L., Proakis John G., *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, 1993.

[4] Makhoul J., Roucos S., Gish H., "Vector Quantization in Speech Coding", Proc. IEEE, vol. **73**, no. 11, pp. 1551-1588, Nov. 1985.

[5] Tsao C., Gray R., "Matrix Quantizer Design for LPC Speech Using the Generalized Lloyd Algorithm", IEEE Trans. Acoust., Speech, Signal Processing, vol. **ASSP-33**, pp. 537-545, June 1985.