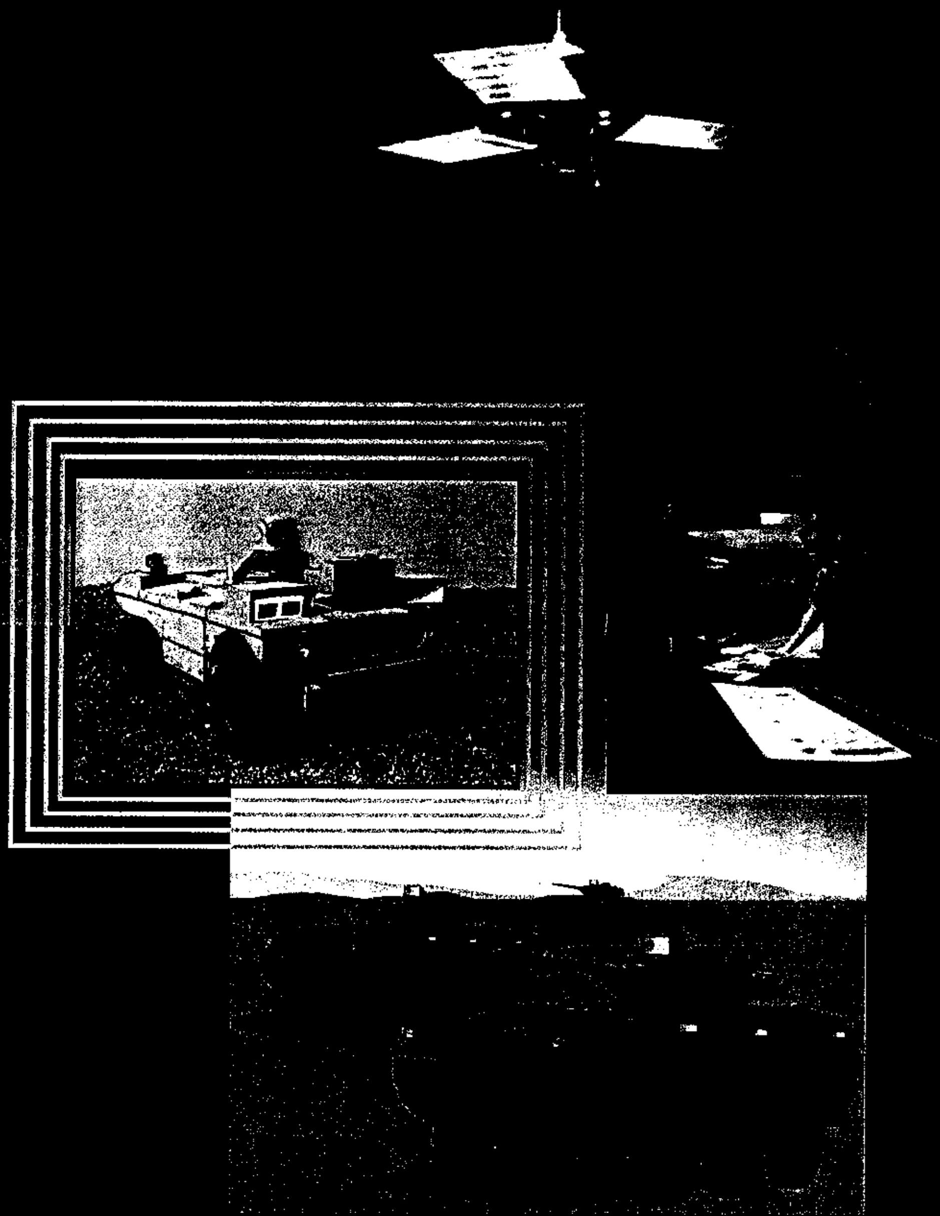

# Advanced Sensors
## CONSORTIUM

ARL Federated Laboratory
4th Annual Symposium

March 21-23, 2000 College Park, MD

PROCEEDINGS

# Vehicle Classification Using Acoustic Data Based on Biology Hearing Models and Multiscale Vector Quantization [1]

**D.A.Depireux**
**S.Varma**
**J.Baras**
Institute for Systems Research
University of Maryland
College Park, MD 20742
didier,varma,baras@isr.umd.edu
http://www.isr.umd.edu/CAAR

**N.Srour**
**T.Pham**

U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197
nsrour,tpham@arl.mil
http://www.arl.mil/acoustics

**Abstract:**
*The army is interested in using acoustic sensors in the battlefield to perform vehicle identification using passive microphone and seismic arrays. The main advantages of acoustic arrays are that they are non-line of sight, low cost, low power, can be made small and rugged and can provide 360° coverage. Their capability includes target detection, bearing, tracking, classification and identification, and can provide wake-up and cueing for other sensors.*
*Acoustic arrays can be deployed in an expandable tracking system: the outputs of a nework of acoustic arrays can detect, track, and identify ground targets at tactical range by triangulating the reports from several distributed arrays.*
*Here, we present a prototype of vehicle acoustic signal classification. To analyze the signature of the vehicle, we adopt biologically motivated feature extraction models. Several possible representations are used in a classification system. Different vector quantization(VQ) clustering algorithms are implemented and tested for real world vehicle acoustic signal, such as Learning VQ, Tree-Structured VQ and Parallel TSVQ. Experiments on the Acoustic-seismic Classification Identification Data Set (ACIDS) database show that both PTSVQ and LVQ achieve high classification rates.*
*The VQ schemes presented here have the advantage of not having to choose explicitly the features that distinguish the targets. The burden is shifted to having to choose the ``best'' representation for the classifier.*

## Introduction

**A. The problem:** Our application involves an Acoustic Detection System (ADS) which consists in an omni-directional acoustic sensor system that uses an array of microphones and seismic detectors, and a simple processor to detect, track, and classify targets in the battlefield. The array we used consists in three microphones, in a triangle 15 inches apart. When targets are detected, the ADS determines lines of

[1] Prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0001. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

bearing to the targets relative to the position of the array and can cue other systems to approaching targets, or report the results (Detection, Line of bearing and ID) to a central portal.

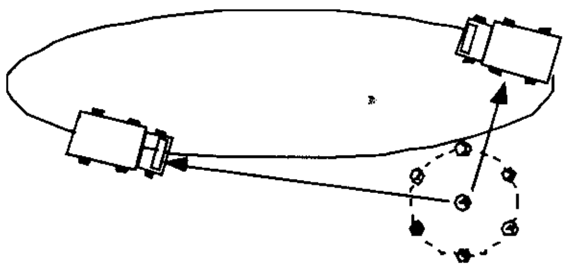

**Figure 1:** Schematic of the situation: two heavy vehicles go around a track while their acoustic signature is recorded by an array of microphones.

**B: Tree Structured Vector Quantizer**:
TSVQ is an example of a classification tree where test vectors are classified stage by stage, each stage giving a better classification than the previous. Each node of the tree is associated with a centroid (a sort of paradigm for a particular class). All test vectors start out at the root node. Then the vector is compared with the centroids of all nodes which are children of the node it belongs to. The vector is classified into the child with the centroid that is closest to it. The vector ends up in a leaf node, and is assigned the class of the leaf node.
Making a vector quantizer in the form of a tree has two advantages. 1), if the tree is balanced, the number of comparisons to be made is $O(\log n)$ with $n$ is the number of vector space partitions. This can be a big factor if is a large number of classes. We can also use parallel TSVQ techniques to further reduce search time. 2), the way the signal vector space is split at each node is indicative of natural partitions in the data-set. The challenge is to preserve fidelity in classification: substituting an optimal partitioning of signal space by a tree structured one reduces optimality. Our goal is to make this difference as small as possible. Proper choice of pre-processing and tree-growing algorithm is crucial.

**C: Multi-Resolution TSVQ (MRTSVQ)**
One special kind of VQ classifying tree is the Multi-Resolution TSVQ (MRTSVQ). Any particular test vector in an MRTSVQ is represented in multiple resolutions or scales. A method of creating such a representation is through affine wavelet transforms. We use auditory cortical filtering as an example of a multi-resolution transform. For a given vector we create a multi-resolution representation. At each level $i$ in the tree, the $i$th resolution vector is compared against the centroids of the nodes in that level. The vector is classified into the node that has the nearest neighbor centroid. At the next level, the next higher resolution of the vector is used for the comparison.
This method offers one advantage over the unembellished TSVQ: at higher levels, where more comparisons have to be made, we use a vector with fewer bits, thus doing many simple computations; progressively finer details are added until satisfactory performance is obtained.
Cutting down on the data presented to the classifier in the early stages does not degrade performance much. In most cases of interest, one does not need all the available data to make simple classification. For example, for speaker ID, the decision whether the speaker is male or female can be made with a rather coarse representation of the sound.
We use a tree-growing algorithm as used by Baras and Wolk (Baras 1993) for classifying radar returns. The tree algorithm uses the Linde-Buzo-Gray (LBG) algorithm (Linde et al, 1980) for VQ at each level of the tree. The algorithm starts with an initially fixed number of centroids. The LBG algorithm is

used to find a distribution of centroids that correspond to a local minimum in the expected squared error distortion. Then an additional centroid is introduced and LBG applied again to find the expected distortion. If the change in distortion from the addition is greater than a fixed fraction of the total distortion, another centroid is introduced and the process repeated. If the change in distortion is lesser than then fixed fraction, the algorithm goes to the next level. The cell in the current level is fixed and the leaf node with the highest value of the distortion is then split at the next lower level (higher resolution). This process goes on until a stopping criterion is satisfied. The stopping criterion can be the final rate of the tree, the number of leaf nodes, the expected distortion of the tree, or any other criterion. This is a greedy method of tree growing, in that the cell with the highest distortion in the current leaf nodes is the one that is split. There are other ways of choosing the split node, among which are, node with largest change in distortion for given rate increase, node with highest entropy and so on.

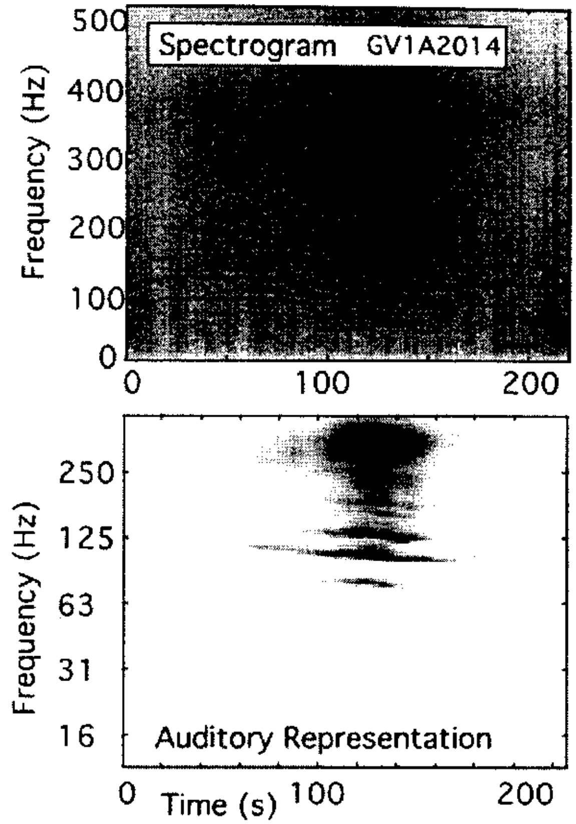


**Figure 2 Top**: spectrogram for a heavy track vehicle of the ACIDS database. When one sees the details, there are two sets of harmonics. **Bottom**: the corresponding auditory representation

### The representations

**A. The auditory representation:** A functional view of the auditory pathway and its attending representations was presented at the previous FedLab meeting (Shamma 99): The cochlea is viewed as a parallel bank of band-pass filters of specific shape and constant Q-factor. The cochlear filter output forms an affine wavelet transform of the stimulus, with the log(freq) spatial axis acting as scale parameter. The cortical representation which is used as a time-frequency decomposition preprocessing stage corresponds roughly to a time-frequency wavelet decomposition of the cochlear representation. Details can be found in Shamma et al 1989, see fig 2.

**B. The pitch extraction:** Once the auditory spectrum is known, we can extract the possible pitches present in it (Shamma and Klein, 00). It might be assumed, for instance, that tanks have two main sources of sound: the tracks and the engine. Hence we would expect to see two pitches. An example of pitch extraction from a heavy track vehicle is shown in fig 3 for the same vehicle as in fig 2. In the bottom of fig 3, the main pitches are at 13.5 Hz, 27 Hz, 35.5 Hz and 54 Hz. Since pitch is often defined with an octave ambiguity, we can assume that the (13.5, 27, 54) Hz values all correspond to the same set of harmonics, whereas the 27 Hz is from a different source.

different scales, from coarsest to finest.

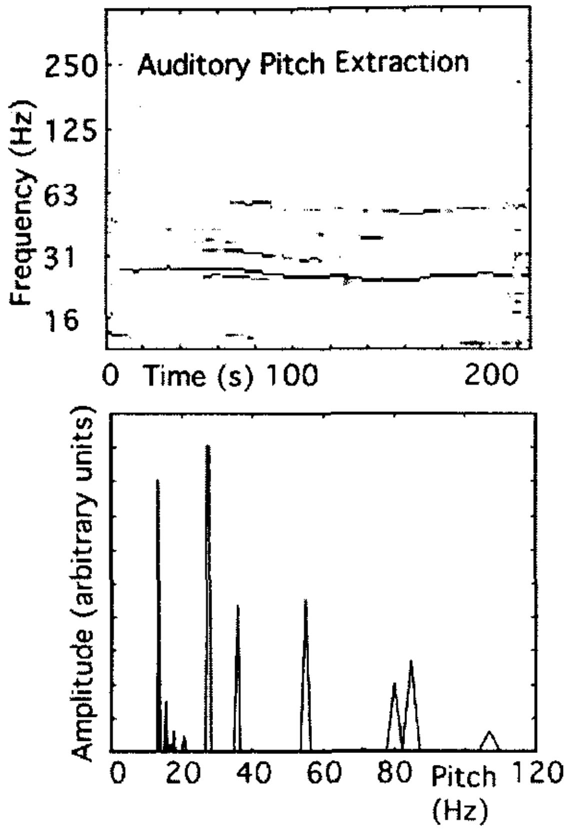


**Figure 3: Top:** possible pitches present in the recording of fig 2. **Bottom:** section of the top graph at around 80 s.

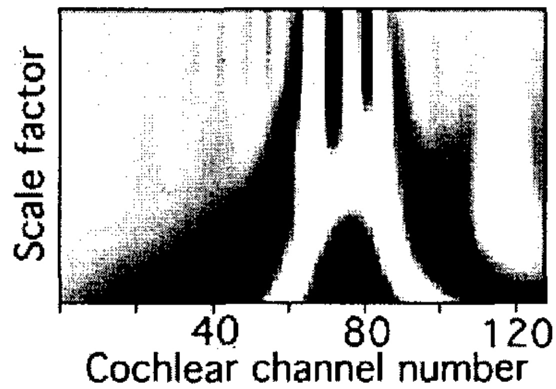


**Figure 4:** Multi-resolution representation from a slice of the cortical representation

Note that because the representation used is based on a log-axis, a change in RPM or velocity (and therefore frequency of the track slap) corresponds to a translation on the log-freq axis. This is important for the training of the algorithm. It might also be possible to classify heavy vehicles according to the ratio of the pitches of the two main pitches present in the recording. This is currently under investigation.

**C The cortical representation:** To use the properties of TSVQ we use a multi-scale representation of the acoustic signal. Such a decomposition is shown in fig 4 for a slice of fig 2. Horizontal slices are a multi-scale decomposition of an instantaneous auditory representation of fig 2, and are used to classify the target's auditory spectrum at different scales, from coarsest to finest.

The coarse scale (lower part of figure) captures the broad and skewed distribution of energy in the auditory spectrum, while the finer scale (upper part of figure) captures the detailed harmonics structure. In the other intermediate cortical scales, the dominant harmonics are highlighted while the weaker ones are suppressed. Thus intermediate scales emphasize the most valuable perceptual features within the signal. The cortical filter is a redundant representation, not all the scales are necessary for the classification algorithm.

## Results[2]

In the ACIDS database, most vehicle are recorded for dozens of runs, corresponding to different speed and gear, different terrain (desert, arctic, normal roadway, and etc), and different recording systems. This database represents an ideal opportunity for classification research.

[2] The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

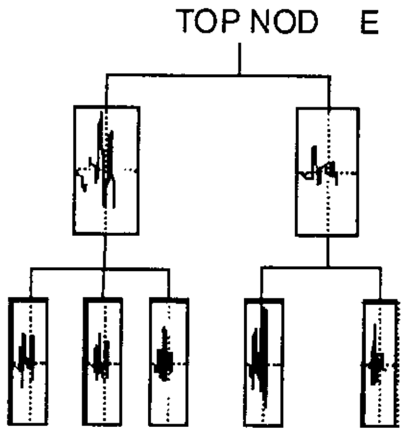


**Figure 5:** First two levels of the VQ tree for the cortical representation of fig 4

| | |
|---|---|
| Type 1: | heavy track vehicle |
| Type 2: | heavy track vehicle |
| Type 3: | heavy wheel vehicle |
| Type 4: | light track vehicle |
| Type 5: | heavy wheel vehicle |
| Type 6: | light wheel vehicle |
| Type 7: | light wheel vehicle |
| Type 8: | heavy track |
| Type 9: | heavy track |

**Table 1**: Different vehicles in the ACIDS database.

We have trained the TSVQ algorithm on the type of representation shown in fig 4 and have obtained a classification tree, the top of which is shown in fig 5. Preliminary results with this specific representation is shown in table 2. We will present results for other types of pre-processing and representations during the talk.


**References**

1. J.S. Baras and S.I. Wolk, (1993) Hierarchical Wavelet Representations of Ship Radar Returns, Technical report T.R. 93-100, Institute for Systems Research, U. of Maryland at College Park
2. Y. Linde, A. Buzo and R. Gray, An algorithm for Vector Quantizer Design, IEEE Trans. Comm., Vol COM-28, No. 1, pp 84-95, Jan 1980
3. S.A.Shamma, D.A.Depireux, C.P.Brown, N. Srour and T. Pham, Signal Processing in Battlefield Acoustic Sensor Arrays, in the FedLab 99 proceedings
4. S.A.Shamma and D.J.Klein, to appear in J.Acoust.Soc.Amer 2000.


| Predicted\True | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | **93.077** | 37.5 | 18.958 | 2.2917 | 0 | 0 | 11.563 | 14.167 | 4.25 |
| 2 | 5.9615 | **36.25** | 2.7083 | 1.0417 | 0 | 0 | 3.125 | 0.2083 | 19 |
| 3 | 0.1923 | 0 | **43.958** | 0 | 0 | 0 | 5.3125 | 0 | 2.25 |
| 4 | 0 | 1.25 | 1.875 | **78.542** | 0 | 0 | 38.438 | 0.625 | 1.75 |
| 5 | 0 | 25 | 0.8333 | 0.625 | **0** | 0 | 1.25 | 0.4167 | 4.75 |
| 6 | 0 | 0 | 0.625 | 0 | 0 | **0** | 6.875 | 4.1667 | 6 |
| 7 | 0 | 0 | 4.5833 | 3.9583 | 0 | 0 | **6.25** | 0.2083 | 1.75 |
| 8 | 0.1923 | 0 | 3.125 | 3.5417 | 0 | 0 | 20.938 | **79.583** | 8.75 |
| 9 | 0.5769 | 0 | 23.333 | 10 | 0 | 0 | 6.25 | 0.625 | **51.5** |
| Total % | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

[illegible]

**Table 2: Classification results with the representation of fig 4**