Proceedings of the Institute of Acoustics

MADRAS, AN ASSISTANT FOR AUTOMATIC NOISE RECOGNITION

D. Dufoumet 01dB, Villeurbanne, France
J. Shelton AcSoft Ltd, Aylesbury, United Kingdom

1. INTRODUCTION

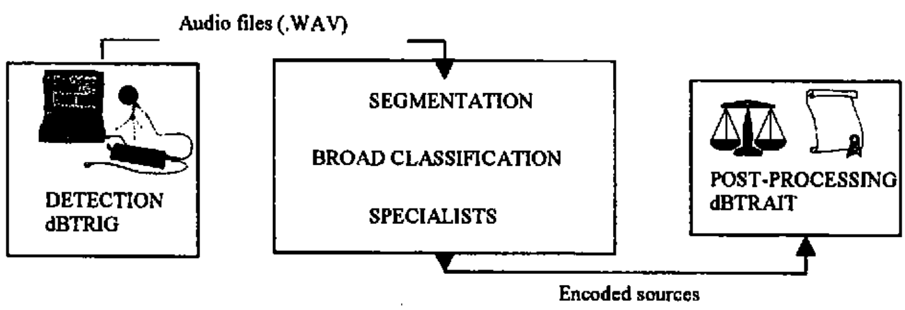
The purpose of the European project MADRAS* (Methods for Automatic Detection and Recognition of Acoustic Sources) was to propose a new generation of intelligent acoustic instrumentation capable of identifying in real-time a wide variety of noise sources, for evaluation of their impact on the noise environment. This multidisciplinary project naturally involves image recognition tools, complex signal processing and artificial intelligence. In this article, we present the architecture of the operational prototype on the Symphonie acquisition platform. Each stage of treatment is detailed with the choices and limitations. Finally, an application of MADRAS is presented, relating to quality control applications, in the study and characterisation of 'Squeak & Rattle' type signals.

2. ARCHITECTURE

2.1 Global assessment

The architecture studied and developed takes account of the different constraints required to realise a system suitable for environmental measurements: low weight, battery power and flexibility. One of the main foci of attention was the reduction of calculation overhead in order to use the techniques available on lightweight platforms (notebook PCs): this resulted in the separation of treatment in the temporal and spectral. Another feature of the resulting method is its adaptable character, not only from the point of view of the integrated software components but also for the user interface : faced with the variability of certain types of signal, the use of on-site training procedures was necessary. Figure 1 shows the final architecture used operationally.

FIGURE 1. MADRAS system



In its prototype version, MADRAS fits between two software components of the 01dB range. The combination of Symphonie & dBTRIG allows the recording of audio signals based on certain criteria; audio files in WAV format are then sent to the MADRAS assistant for simultaneous training and identification, as well as management of the recognition database. The job of MADRAS consists of deciding on the presence of the sources targeted by the user and encoding the measured data. At the end of the measurement/identification, the dBTRAIT software can then treat the coded file automatically, by calculating the individual contributions of the sources and application of standard calculations (nuisance evaluation).

2.2 Detection and Segmentation

The first stage of treatment integrated in MADRAS involves the recording of the signal by the acquisition module: this configures the start of the recording by monitoring the time history of a global indicator (Leq, Fast, Slow, frequency band) using a trigger. MADRAS needs to use an adaptive trigger, based on L90 for example, to take account of the changes in background level as a function of time of day for example, and therefore offering a more adaptable detection. The audio file can also contain contributions from different types of sources. The purpose of the segmentation phase is to build in the ability of the human eye to characterise a time history (global level history) of an event. The principle is to process A-weighted 10ms Short Leq of the recorded event according to a method borrowed from mathematical morphology and usually used in image processing [1], and which also allows access to local (detail) and global (tendency) information at the same time. The first phase of treatment consists of placing two types of markers on the global signal for the peaks (+) and the troughs (o) (fig 2). For each category, a Gaussian smoothing function is applied to attenuate the discontinuities of each group. Two regular curves are obtained describing the peaks and the troughs. The 'trough' curve is essential for describing the tendency during recording. By taking the difference to the original curve, we obtain the patterns. It only rests to compare the average level and the standard deviation of the 'tendency' and 'patterns' curves respectively of the signal during the first second of analysis (pre-delay on recording) to precisely determine the time markers of each component. The example here shows two short events superimposed on a clear tendency. Each of these three components will be treated in a similar way by the various stages of MADRAS.

This « visual » comparison obviously doesn't allow resolution of the cocktail-party effect, but this simplified approach, easy to implement, gives satisfactory results in a wide variety of situations.

Complex signals from mixed sources cannot normally be resolved by MADRAS.

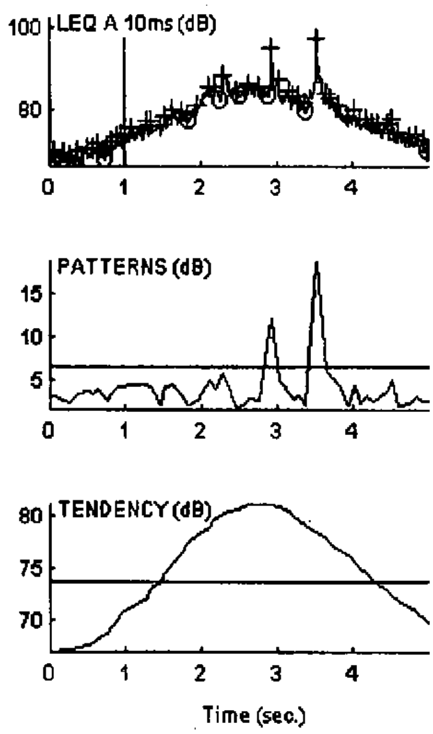


FIGURE 2. A car pass-by with 2 gun-shots

2.3 Broad class classification

This stage consists of the main original component of MADRAS. Wishing to take as many decisions as possible in the temporal domain in order to reduce the processing, this phase transforms MADRAS into a branch classifier. Each of the events output from the segmentation will be classified here into the following five metaclasses: impulsive signals, vehicle pass-bys, mass-transport pass-bys (trains, planes), stationary noises and bursts of energy (shouts, intermittent noises).

Once more, the idea is to reproduce the visual interpretation of a level time history of Short Leq (50ms) of the event. For this, five geometric measures will be calculated for each event : duration (seconds), dynamic range (dB), slope (dB/sec.), surface area of the curve and the activity (number of appearances above the average level). Each event is now represented in a five dimensional space. A database is then built up from a high number of recordings on site (several thousand events). To interpret the separation of each individual in this new space, we are going to reduce the number of dimensions by using a multivariable analysis of data, in particular breaking them down into principal components. To avoid the disagreements of the classical ACP methods (linear methods, empirical choice of principal directions), we use an auto-associative memory neural network (fig. 3). The teaching phase of the network consists of presenting the same vector (geometric values) at the inputs and outputs of the network. The network (multi-layer perceptron with gradient backwards-propagation) will adjust its weightings (connections) to minimise the quadratic error between the desired output and that obtained in response to the input vector. We do this by compressing the intermediate information into only two dimensions (hidden layer), which are going to define the new axes of the representation space. At the end of the training phase (stable connections), the different weightings v1 and v2 indicate the contributions of each measurement variable to the new components (fig. 4a). The outputs of the hidden layer x1 and x2 for each individual in the database indicates the partition in the new space (fig. 4b).

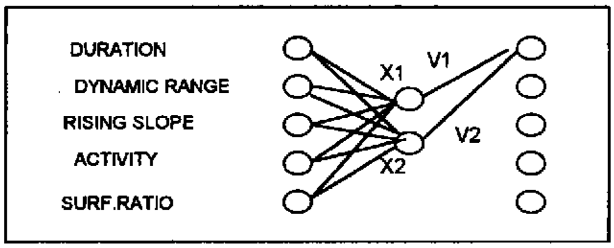


FIG 3.
Clustering memory

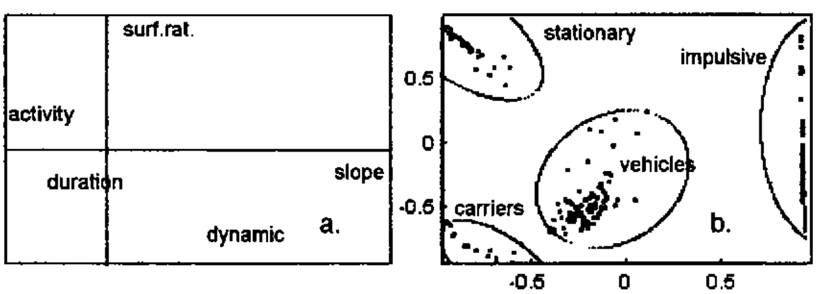


Fig. 4. Factorial analysis plan
a. Variables
b. Patterns

The results clearly show a decision map indexed by a power axis (slope against activity) and a shape axis (dynamic range against surface ratio). We can state with confidence that the map of individuals shows the natural grouping into separate classes. Thanks to the quadratic definition of surfaces, we can define five areas in the map: impulsive noises, stationary noises, vehicle pass-bys, mass-transit pass-bys. The exterior of these zones indicates energy bursts. The recognition procedure is thus very simple: it consists of measuring the geometrical characteristics of each new event, and then projecting these on to a decision map, to decide the allocation of one or other metaclass. Within each metaclass, a specific treatment is then applied.

2.4 Specialists

The function of each specialist is to offer a more detailed classification within each metaclass. Each carries a training overhead adapted to the type of signal to be classified. The main ones are presented here:

Isolated vehicles: a neural network (classical multi-layer perceptron) was trained in the laboratory on several hundred vehicle signatures. Using the average 1/3-octave spectra, from 50 Hz to 5 kHz, it can distinguish the following classes: cars, motorcycles, mopeds, vans and heavy lorries. The results of successful recognition are of the order of 85% depending on the training database used.
Stationary noises: faced with the wide range of such sources (fans, air conditioners, mowers, etc..), an on-site training protocol was defined. The user must construct a model for each source to be identified. Each time a « stationary » event is detected, a mask (1/3 octave L90) is presented to the operator: this can then be adjusted for each frequency band (levels, tolerances) to define an individual average for each source. Once the user is satisfied with the stability of the parameters, the mask is written to the database. The specialist is then run in recognition mode, which consists of comparing the new measured tolerances to those in the database (least-squares fit). A recall of these different sources in a post-processing situation is then possible for new identifications or for use in training.
Impulsive noises: The procedure used is the same as that for stationary sources. Only the analysis method is different: here we use an orthogonal wavelet analysis of the Daubechies type [2]. This analysis is particularly well suited to the characterisation of transient events and achieves a stable tolerance mask after only around 10 training examples.

2.5 Performance

The actual prototype was tested in several industrial environments in which the various noise sources were rarely mixed. Concentrating on one or other particular source in a multi-source context gave recognition results of the order of 80%.
The great weakness of the system is in the detection of events. In effect, an approximate detection (truncated audio recording) leads to an erroneous segmentation (event time markers) and uncertain analyses. Additional effort is required to propose more robust and « intelligent » detection structures.
In the case of a correct detection (fine control of the detector), percentages of the order of 90% were obtained. This percentage could be improved by adapting the networks to each situation, after a brief description of the situation by the user, at the beginning of setting up the measurement, for example (loading of dynamic databases).
MADRAS has the benefit of a modular structure, in which each block (detection, segmentation, specialist, etc...) can be adapted to any situation. Although several environmental applications could be envisaged (monitoring of airports and leisure areas), specific needs in terms of industrial quality control are also excellent examples of applications, as described in the following.

3. SQUEAK-AND-RATTLE APPLICATION

3.1 Introduction

The appearance of « squeaks and rattles » in car cabins can lead to serious economic consequences for the car makers, as the image of low-quality will affect consumer perception. Following demands of the car makers and their suppliers to better understand the signals and their appearance, the MADRAS methodology has been adapted and applied to this problem.

3.2 Detection improvement

The first essential modification of the architecture presented earlier concerns the structure of the selected detection. Faced with the strong transient character of such events (some milliseconds), the detection based on short-term level is not sufficiently reactive or dynamic. A new structure must therefore be developed and integrated in MADRAS, with the following principles: The signal is passed simultaneously through two sliding analysis windows, one (the shortest) ahead of the other in time. The detection consists of evaluating a contrast function between the two estimations made in each window. The long estimation (over 500ms) is used to characterise the background level. We then impose a pseudostationarity on the background noise, of period 500ms, which is large enough to cover environmental variations on a test bed or on the road where the system is to be used. At the same time, a short-term estimation (40ms) gives the changing character of the signal. The resulting estimation from each window can be narrow band spectra, partial band spectra or partial octave spectra. Actual tests can be used to evaluate the relative performance of these spectral estimators as a function of the signal to be detected. Once the two estimations are made, a contrast function based by example on a band-to-band spectral comparison decides or otherwise the selection of an event of interest, which can then be treated by the MADRAS network. Depending on this state (detection or not), simple rules decide whether to update or block the noise estimation. This detection structure can also be adapted according to the variations in background noise and the dynamic or transient characteristics of the required noises.

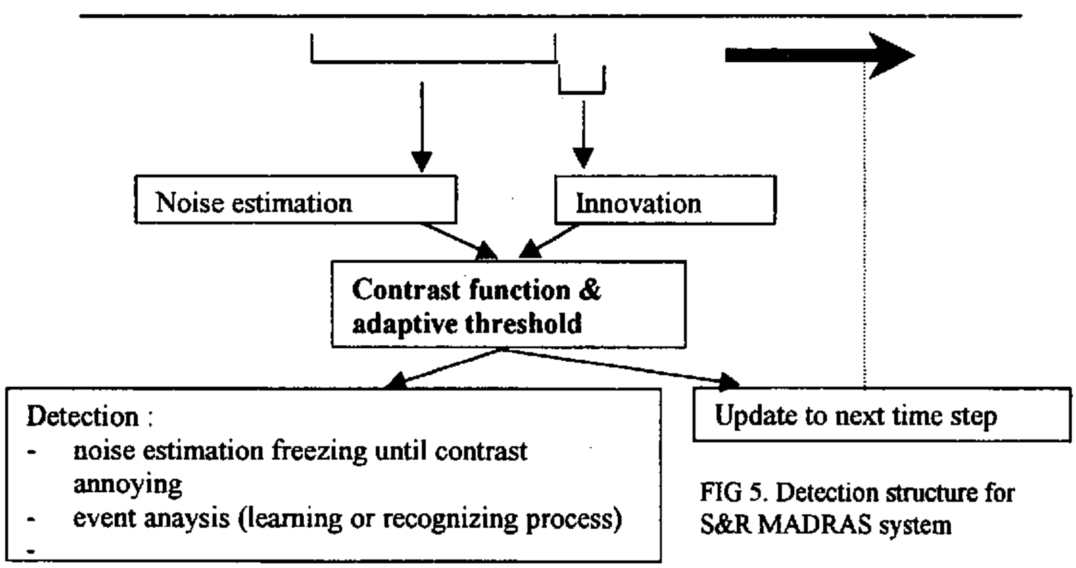


FIG 5. Detection structure for S&R MADRAS system

3.3 Learning process

The second adaptation of the MADRAS system is the specific analysis method of transient signals. One point of interest in Squeak and Rattle is in the characterisation of the annoyance of the vehicle passengers. We need to find therefore a method, which is sufficiently discriminatory to allow a detailed training, and recognition of signals, as well as taking into account the « annoying » character of the signal. For this, the analysis method used consists of a hearing model for « transients », which as well as allowing the construction of a tolerance for each signal to identify, also includes an intrinsic sensitivity to the characterisation of annoyance [3].
The method used, a veritable « cochleogram », is based on a bank of filters in specific barks [4]. In addition to classical frequency masking [5], each filter includes a time response that depends on the band and level of the signal. Two RC-type filters integrate the signal on the rise and fall to reproduce the characteristics of human perception as faithfully as possible.
Tests are continuing to balance the performance and ease of use (building of tolerances) between this specialist and the generic one developed in MADRAS (Mallat wavelets). As with all wavelets,

this method concentrates on the acoustic signature in those bands of interest that contribute to the building of a stable tolerance.

4. CONCLUSION

This article has presented the basic building blocks which make up the MADRAS methodology, which, when used in a dynamic and adaptive way, allow its use in many different situations where automatic identification of signals is of great interest.
An example of adaptation to the problem of Squeak & Rattle has been presented, in which specialised tools for detection and analyse have allowed the MADRAS methodology to be customised to the problem. Validations of the system are currently under way in collaboration with the car makers and sub-suppliers, to finalise the best adapted choices of parameters.

*with financial support of the European project CEEDG12, in collaboration with CSTB, CEDIA and ENTPE

References

[1] J. Serra, « Mathematical morphology and image analysis », Academic Press, London, 1982
[2] I. Daubechies, S. Mallat, A.S. Willsky, « Wavelets transforms and multi-resolution signal analysis », IEEE Transctions on information theory, vol. 38, 2, 1992
[3] S. Veste, « Maîtrise des paramètres acoustiques pour le test dichotique », Congrès de neuropsychologie de langue française, Caen, France, 1992
[4] M. Stanley, « A efficient implementation of the Patterson-Holdsworth auditory filter bank », Apple Computer Technical report, Perception Group, 1993
[5] Zwicker and Fastl, « Psychoacoustics, Facts and Models », Ed. Springer Verlag, 1999.