Audio Signal based Fault Recognition System

G. Kaleeswaran
PG Student,
Department of EEE,
G.K.M. College of Engineering & Tech,
Chennai.

A. Paulene Lourdu Mary
Assistant Professor,
Department of EEE,
G.K.M. College of Engineering & Tech,
Chennai.

***Abstract*:-This paper considers a multistage hierarchical algorithm of acoustic signal analysis and pattern recognition for identification of moving vehicles in an open environment. The algorithm applies several standalone techniques to enable complex decision-making during event identification. Computationally inexpensive procedures are specifically chosen in order to provide real-time operation capability. The algorithm is tested on pre-recorded audio signals of passing passenger cars and displays promising classification accuracy.**

I. INTRODUCTION

Moving object identification is one of many tasks of environment monitoring systems. The applications of moving motor vehicles identification vary from speed limit control to traffic density analysis and traffic behavior prediction. The most important aspect of such monitoring systems is real-time computation and timely result processing as the nature of the problem most often implies time-critical operation. Most state of the art systems typically rely on single sensor ultrasonic, acoustic, video, infrared, radar, microwave, magnetic, laser, vibration based, etc. signal analysis, otherwise they employ combinational multisensory detectors. The main advantage of acoustic and video methods lies in the ease of data signal interpretability, i.e., the acquired data is perceptual without additional manipulations.

Video based methods of vehicle identification are generally more effective and robust in changing weather conditions if provided sufficient visibility and illumination. However, the large amounts of video data and significantly more complex pattern search algorithms, if compared to algorithms for one-dimensional data streams, put significant constraints on the possibilities of real-time system implementation. Acoustic systems on the other hand do not rely on visibility factors, yet are sensitive to background acoustic noise variation. Thus the accuracy of acoustic system vehicle identification is directly dependent on its ability to distinguish the sound patterns of passing vehicle noise from a limitless amount of noises occurring in the environment.

Acoustic noise analysis provides the possibility to distinguish well separable classes of motor vehicles, such as different passenger cars and trucks. The harmonic nature of the motor noise is, however, seldom present in the vehicle sound pattern due to the fact that motor sounds are well dampened in modern cars. This fact complemented by the Doppler Effect renders the spectral analysis based on fundamental frequency detection ineffective.

This paper considers different methods of digital audio signal analysis, namely the estimation of spectral energy levels and energy envelope, the analysis of several frequency spectrum instantaneous features and spectral pattern matching. The proposed algorithm possesses a hierarchical structure, beginning with the detection of signal perturbation and ending with the classification of the detected vehicle. The algorithm is computationally inexpensive and thus is well implementable on an embedded sensor device. The section of the paper, devoted to system testing, proves that the algorithm is well applicable to the task of identifying motorized vehicles under varying weather conditions. The free field is a region in space where sound may propagate free from any form of obstruction. The near field of a source is the region close to a source where the sound pressure and acoustic particle velocity are not in phase. In this region the sound field does not decrease by 6 dB each time the distance from the source is increased. The near field is limited to a distance from the source equal to about a wavelength of sound or equal to three times the largest dimension of the sound source.

The far field of a source begins where the near field ends and extends to infinity. Note that the transition from near to far field is gradual in the transition region. In the far field, the direct field radiated by most machinery sources will decay at the rate of 6 dB each time the distance from the source is doubled. For line sources such as traffic noise, the decay rate varies between 3 and 4 dB. The direct field of a sound source is defined as that part of the sound field which has not suffered any reflection from any room surfaces or obstacles. The reverberant field of a source is defined as that part of the sound field radiated by a source which has experienced at least one reflection from a boundary of the room or enclosure containing the source.

Frequency analysis may be thought of as a process by which a time varying signal in the time domain is transformed to its frequency components in the frequency domain. It can be used for quantification of a noise problem, as both criteria and proposed controls are frequency dependent. In particular, tonal components which are identified by the analysis may be treated somewhat differently than broadband noise. Sometimes frequency analysis is used for noise source identification and in all cases frequency analysis will allow determination of the effectiveness of 28 Fundamentals of acoustics controls.

There are a number of instruments available for carrying out a frequency analysis of arbitrarily time-varying signals. To facilitate comparison of measurements between instruments, frequency analysis bands have been standardised.

Thus the International Standards Organisation has agreed upon "preferred" frequency bands for sound measurement and analysis. The widest band used for frequency analysis is the octave band; that is, the upper frequency limit of the band is approximately twice the lower limit. Each octave band is described by its "centre frequency", which is the geometric mean of the upper and lower frequency limits.

II. BACKGROUND

A. Fundamentals of Digital Audio

Digital audio is the representation of natural sound as a set of digital information. Sound is created when the air is disturbed, usually by a vibrating object. The vibrating object causes ripples of varying air pressure. Very little air actually moves anywhere, as the pressure change is propagated by the collision of air molecules, similar to the way ripples spread across the surface of a pond without causing water currents. These “waves” of varying pressure cause the eardrum to move back and forth. The movement is carried from the eardrum to an organ called the cochlea by a series of tiny bones. The cochlea contains a series of over 10,000 different-sized hairs, which convert these vibrations to nerve impulses. The impulses are carried to the brain and decoded.

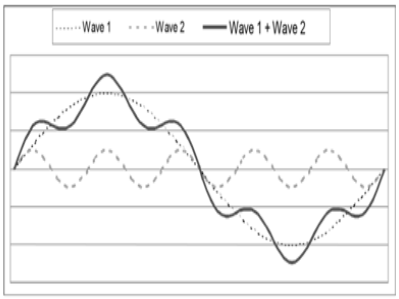

Figure 1: Combining two waves of different frequencies

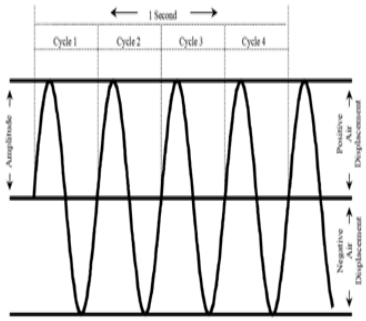

Figure 2: Amplitude and frequency of a simple cyclic wave form

Digital audio signal processing is the analysis and manipulation of audio waveform information, which can become a complex endeavor. Like many computer representations of real-world phenomena, the modeling of sound waves can be performed at any desired level of complexity. The complexity of natural acoustics is mostly due to the complexity of the natural world. As Figure 2-3 illustrates, even a small sample of a human voice waveform is very complex and is not completely periodic. Additionally, the physical properties of objects in a room and the air itself determine what frequencies are reflected or absorbed. Most people can listen to a sound from a familiar source while blindfolded and determine the size of the room. A well-trained ear can even determine the types of materials that comprise the walls.

For a digital computer to process audio, a method of converting audio to and from the domain of digital information is required. The most common format of digitally representing audio information is Pulse Code Modulation. Typically, sound waves are converted to a series of numbers (PCM) as follows: Because an infinite number of data points cannot be recorded to characterize the waveform, a sample is taken at regular intervals. The number of samples taken per second is called the sampling rate. In Figure 3, 43 samples are taken.

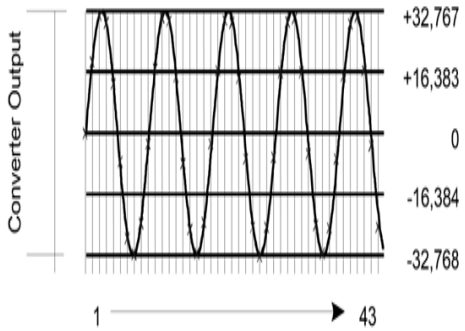

Figure 3: Division into samples

B. METHODS OF SIGNAL ANALYSIS

The audio signal is analysed in frequency domain. The frequency domain representation of the signal is achieved by applying a temporal signal decomposing operation, namely the Fourier Transform (FT). Frequency features are less affected by noise than temporal ones; also most of the temporal features may be approximated in the frequency domain. Furthermore, the frequency spectrum of a temporal signal frame consists of half as many points as there are in the temporal frame, which is relevant in computation complexity critical systems.

The Fast Fourier Transform

The discrete temporal signal is decomposed by the Discrete Fourier Transform (DFT). For a finite duration discrete signal x(m) of length N the DFT function is in this manner the transform is performed along two integer dimensions: m and k, thus having quadratic complexity. In order to reduce its computation, the Fast Fourier Transforms were developed. The proposed system applies a specific implementation of the FFT developed by Frigo and Johnson, called FFTW which is ignored. In order to obtain the absolute amplitude spectrum, the absolute values of this portion of the spectrum are calculated.

Instantaneous Feature Extraction

In order to acquire the specific signal properties several features are extracted from the amplitude frequency spectrum. These are referred to as instantaneous features due to the fact that they are extracted from every single spectral frame independently, not relying on previous information. The list of features is signal-specific and is formed during the process of sample signal analysis in order to distinguish well

separable, desirably weakly correlated features, which best indicate the nature of signal fluctuations corresponding to the concerned events. The six spectral features considered in this paper are extracted from the absolute magnitude spectrum frame X (k) of length K.

Root Mean Square (RMS) Energy of the power spectrum conveys the general spectral energy level: The Mel-scale is chosen for its increasing spread towards the higher frequencies which ultimately means that the bands of lower frequencies, where most of the spectral energy resides, are shorter than the bands of low-energy higher frequencies. This allows for better distribution of spectral energy by bands.

The **spectral centroid** represents the first central moment of the magnitude spectrum. It is calculated as the frequency averaged over the absolute magnitude spectrum:
Spectral roll-off measures the frequency below which a certain amount of spectral energy resides. This amount denoted by TH = [0,1] is the threshold, which in this case is chosen to be equal to TH = 0. In the proposed algorithm the RMS energy is used independently. On the other hand the rest of the mentioned features are concatenated into a feature vector, which is analyzed during the later stages of classification. The application of instantaneous features is further described.

Attack Sustain Release Envelope

The process of a vehicle passing the measurement point at any given velocity consists of three stages: approach (spectral energy increases), passing (spectral energy remains stable), retreat (spectral energy decreases). This pattern is detected by applying the Attack Sustain Release (ASR) envelope estimation. It is conducted by analyzing the RMS spectral energy as denoted. The amount of deviation of RMS energy of the present frame X_{RMS} (i) is estimated by the difference between it and the mean value of M previous RMS energy readings to account for noise. The parameter $\delta \in [0,1]$ is the lower threshold of energy deviation. Such, RMS energy deviation is coded to three states by the following principle: where 1 denotes energy increase, 0 denotes stable energy levels and -1 denotes energy decrease. Therefore, the transitions $1 \rightarrow 0 \rightarrow -1$ and $1 \rightarrow -1$ are suspected for car passing event occurrence and the quantities of -1, 0, and 1 coded frame denote the lengths of attack, sustain, and release components respectively.

III.PROPOSED WORK

The proposed hierarchical algorithm, presented in Figure 3.9., consists of two independent stages. The hierarchical decision making scheme (on the left), firstly differentiates relatively loud sounds from mild background noise, secondly distinguishes vehicle-produced sounds from heavy background noise and lastly estimates the vehicle type from a set of predefined types. This part of the algorithm operates in a frame-by-frame manner, computing a single class label per signal frame. The ASR envelope estimating procedure runs parallel to the decision-making one and complements the past frame classifications with reassurance of positive vehicle passing event detection. The hierarchy of the algorithm is conditioned by the supremacy of vehicle detection priority over vehicle classification priority, i.e., differentiation between vehicle-produced sound and other types of noise is more important than correct vehicle type estimation. The real time constraints for the algorithm consist of limiting the processing time of every iteration to the time duration of a single frame.

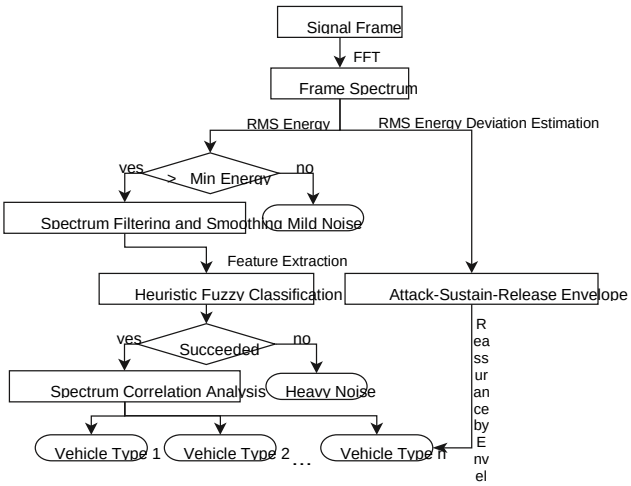

Figure 4: Block diagram of the proposed hierarchical algorithm for vehicle detection and classification

Lower Energy Threshold

The first stage of the hierarchical procedure is the estimation of sufficient signal energy. The energy level of a signal frame is calculated and compared to the lower energy threshold, if the threshold is not exceeded, the procedure terminates and the frame is marked as mild noise. The estimation of the lower energy threshold occurs during algorithm parameter estimation by means of test signal analysis. The initial threshold is chosen as the minimal value of RMS energy of all the frames that correspond to vehicle passing instances.

Heuristic Fuzzy Classification

The sound pattern of a moving object passing a measuring device is not consistent. Due to the variance of signal energy and complex spectral shape alternations caused by engine sounds, moving object velocity and the Doppler Effect, and also the influence of background noise, the spectral features of the signal also vary to some extent. Thus the variance of a feature vector of length L produces an L-dimensional feature space, to which all the feature vectors corresponding to an event of the same class label must belong. The extent of this belonging is estimated using a fuzzy inference algorithm derived by a heuristic training procedure. This fuzzy algorithm operates by applying fuzzy inference to an input feature vector. The algorithm is relatively lightweight if L does not exceed 20-30, that being the reason of applying spectral features instead of the whole spectral vector.

Reassurance by ASR Envelope

As it was said earlier, the detection of the ASR dynamic of signal energy complements the past identification results. If the ASR pattern is detected, an additional notification is generated and the labels assigned to the frames

belonging to the time interval of this pattern are searched for the most frequent one. Additional restrictions may also be applied to the ASR envelope detection. Such if the potential velocity of the moving object is known, the lower and upper bounds for the attack, sustain or release components may be specified, so the detection is invalid if these restrictions are not met.

IV.RESULTS

A.First Test Signal

The first test audio signal was measured using a Shure SM58 microphone and digitized using a Roland Edirol UA25EX audio signal processor at 44.1 kHz sampling rate in mono channel mode and saved in a 16-bit Waveform Audio File (WAV) format. For the acquisition of the test signal a microphone was placed at an empty parking lot and two cars (Mercedes S320 and Mazda MX-5) were in turn passing the microphone stand at a speed of 35 – 45 km/h at the passing point, starting to accelerate from a distance of approximately 40 meters. Each car has overall passed the microphone three times: the Mercedes first three passes and the Mazda second three passes. The sounds were acquired during summer time in mild weather conditions.

For testing, the frame length of 2^{14} = 16384 samples is chosen, which corresponds to 0.3715 seconds at a sampling rate of 44.1 kHz. The signal feature vector comprises of eight features: four band energy features (four bands of 1-824, 8242616, 2616-6514, 6514-15000 Hz), spectral centroid, spectral roll-off and spectral slope. In total 2 class labels are used: 1 for Mercedes and 2 for Mazda. The reference spectral vectors used in correlation analysis are estimated by averaging several spectra of sounds produced by vehicles of the same class, in total one reference vector per class is applied.

The heuristic fuzzy algorithm, trained to identify the general vehicle feature space, succeeds in doing so for the majority of signal frames thus allowing the correlation coefficient calculation procedure to analyze only the frames corresponding to vehicle pass time intervals. The fourth subplot of Fig. 2. shows, that the correlation coefficient values are unreliable during the periods be instances they become more separate, indicating an obvious leader.

B. Second Test Signal

The second test signal was acquired using a miniature condenser microphone Sennheiser KE 4-211-2 and an embedded computing device GumstixOvero Water (600MHz, 256MB RAM, 4GB microSD). The signal was also sampled at 44.1 kHz mono channel mode and saved to a 16-bit WAV file.Signal acquisition was conducted at a lively two-lane highway during dense traffic in late fall under heavy wind and light rain. The frame length was chosen the same as for the first test signal. For the vehicle classes two were chosen: 1 for passenger cars and 2 for trucks and busses. Feature vectors comprise of eight features, which are the same as for the first signal, except the bands for the band energy features are less spread: 220-818, 818-2592, 2592-6438, 6438-14780 Hz. For the acquisition of reference spectral vectors, the same technique as before is used.

The results of signal analysis are presented in Fig. 3. As the time intervals between car passes are very short and often non-existent altogether, reference class labels, which are also used during fuzzy algorithm training, are introduced in the first subplot. The results are as follows: out of 46 instances of class 1 vehicles, 37 were successfully detected and classified, 5 were undetected and 4 were confused with class 2; for 11 instances of class 2 vehicles, 9 were correctly classified, 1 was not detected and 1 confused with class 1. Thus the classification accuracy for class 1 vehicles is 80.43% and for class 2 – 81.82%. The vague differences between passenger and heavy cars of some vehicles lower the classification quality. Furthermore, a heavy truck can emit a noise loud enough for it to mask the sound of a nearer lighter car thus making it undetectable.

General Testing Results

The algorithm operates well in both the cases of motor vehicles passing with a certain time interval between the passes and heavy traffic. Though, if the flow of vehicles is consistent and very dense, the decrease of identification quality is witnessed. The influence of background noise, such as wind, is reduced due to the algorithm's multistage decision making logic. Thus the algorithm is applicable under different weather conditions.

Providing a variety of tunable parameters, the sensitivity of the algorithm is proven to be well adjustable to the needed extent. This provides the opportunity to apply the algorithm for classification of various types of moving objects not limited to motorized vehicles. Results are given in below from figure 5 to 16.

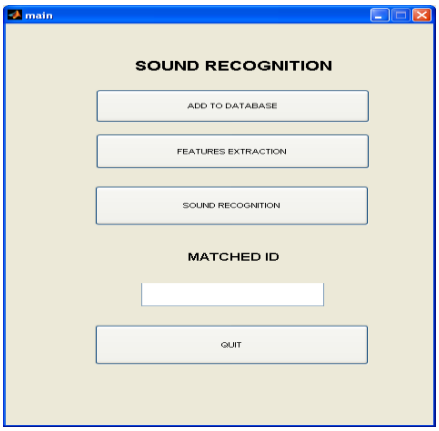

Figure 5: Sound recognitions

Figure 6: Menu to select option

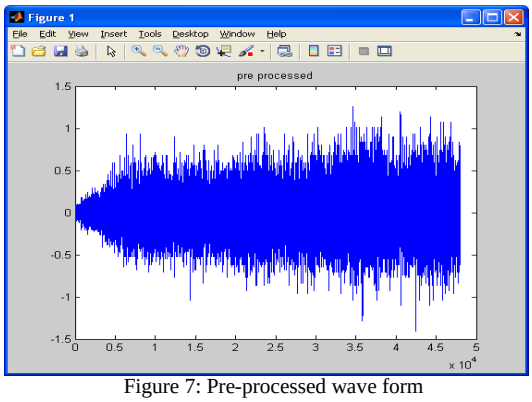

Figure 7: Pre-processed wave form

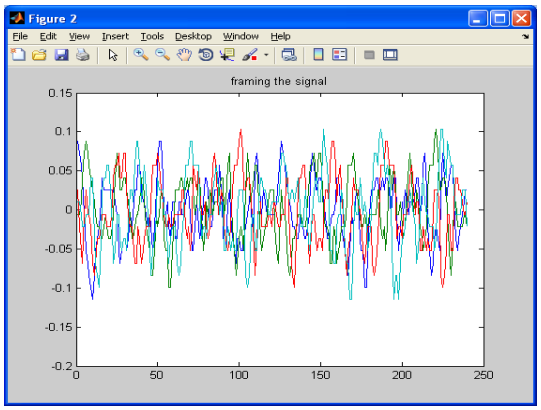

Figure 8: Framing the signal

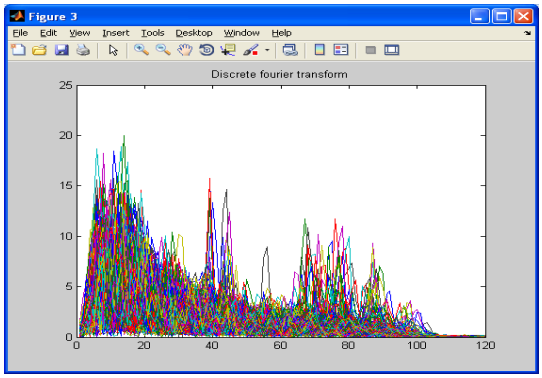

Figure 9: Discrete Fourier transform

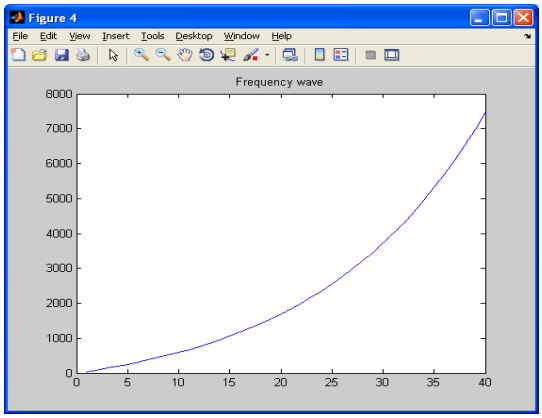

Figure 10: Frequency wave

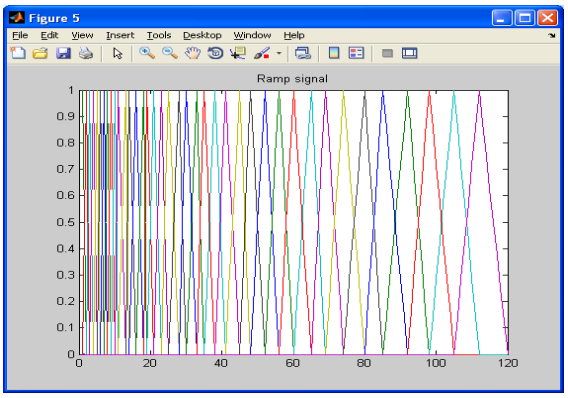

Figure 11: Ramp signal

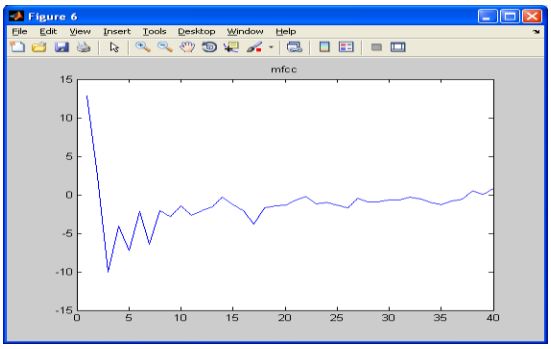

Figure 12: MFCC waveform

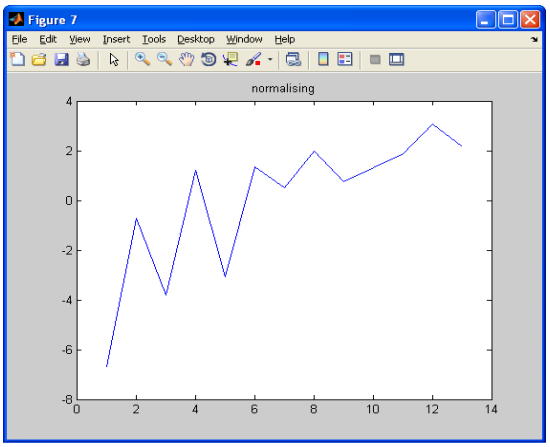

Figure 13: Normalizing waveform

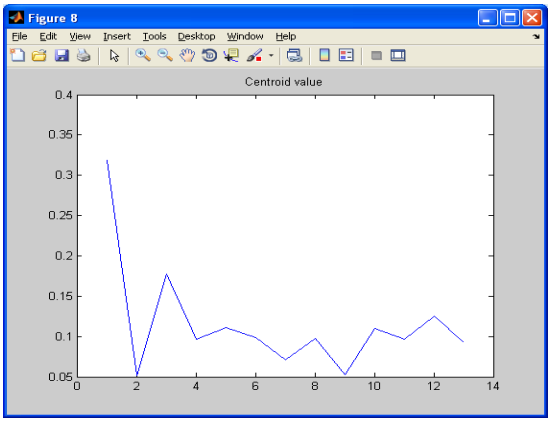

Figure 14: Centroid value

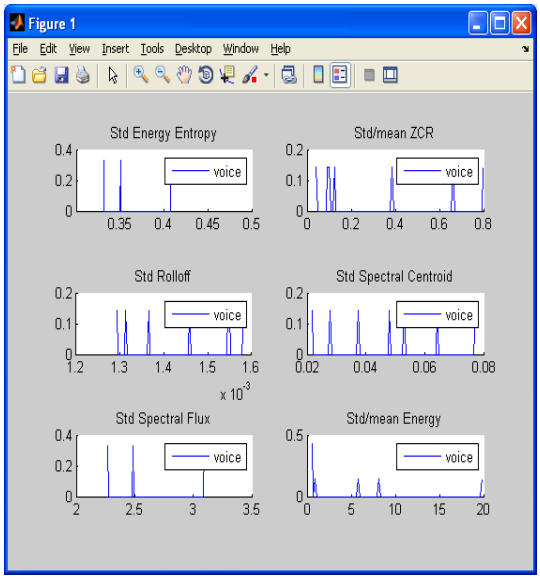

Figure 15: Various standard and mean values

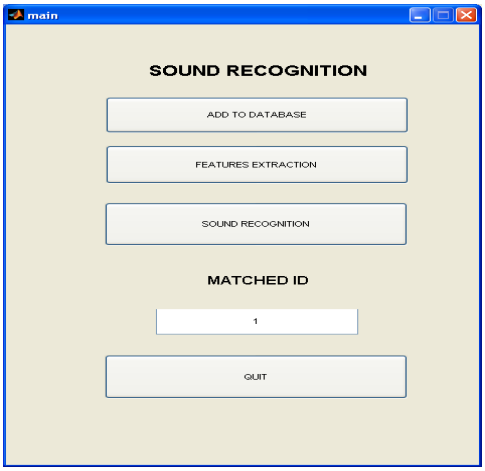

Figure 16: Final result based on voice matching

V.CONCLUSION

In this work we have introduced a hierarchical algorithm for moving vehicle identification by means of acoustic noise analysis. The algorithm is developed specifically for real-time application and is therefore computationally inexpensive and simple in computation. Algorithm testing results indicate the algorithms potency in the task of detecting and classifying motor vehicles.

For future developments algorithm robustness may be increased by applying soft discretization to the transitions of the algorithm decision-making path thus transforming its appearance to a fuzzy tree. The final class label therefore may be decided based on degrees of membership.

REFERENCES

[1] T. Takechi, K. Sugimoto, T. Mandono and H. Sawada, "Automobile identification based on the measurement of car sounds," Proc. 30th Annual Conference of IEEE. IECON 2004, vol. 2, pp. 1784- 1789, November 2004.

[2] A. Starzacher and B. Rinner, "Single Sensor Acoustic Feature Extraction for Embedded Realtime Vehicle Classification," Proc. International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 378-383, December 2009.

[3] N. A. Rahim, M. P. Paulraj, A. H. Adom, and S. Sundararaj, "Moving vehicle noise classification using backpropagation algorithm," Proc. 6th International Colloquium on Signal Processing and Its Applications (CSPA) 2010, pp. 1-6, May 2010.

[4] S. Maithani and R. Tyagi, "Noise Characterization and Classification for Background Estimation," Proc. International Conference on Signal Processing, Communications and Networking, pp. 208-213, January 2008.

[5] S. S. Yang, Y. G. Kim and H. Choi, "Vehicle identification using wireless sensor networks," Proc. IEEE SoutheastCon, 2007, pp. 41-46, March 2007.

[6] M. Frigo and S. G. Johnson, "FFTW: an adaptive software architecture for the FFT," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 3, pp. 1381-1384, May 1998.

[7] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," CUIDADO I.S.T. Project Report, 2004.

[8] A. Riid and E. Rustern, "An integrated approach for the identification of compact, interpretable and accurate fuzzy rule-based classifiers from data," Proc. 15th IEEE International Conference on Intelligent Engineering Systems (INES), pp. 101-107, June 2011.

[9] Y. Peng and P. Flach, "Soft Discretization to Enhance the Continuous Decision Tree Induction," Proc. Integrating Aspects of Data Mining, Decision Support and Meta-Learning, pp. 109–118, September 2001.