2011 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
MODELING & SIMULATION, TESTING AND VALIDATION (MSTV) MINI-SYMPOSIUM
AUGUST 9-11 DEARBORN, MICHIGAN

DETECTION AND IDENTIFICATION OF ACOUSTIC SIGNATURES

Glenn Pietila
Sound Answers Inc.
Canton, MI

Gabriella Cerrato, PhD
Sound Answers Inc.
Canton, MI

Robert E. Smith
RDECOM-TARDEC
Warren, MI

ABSTRACT

Awareness of the surroundings is strongly influenced by acoustic cues. This is of relevance for the implementation of safety strategies on board of electric and hybrid vehicles and for the development of acoustic camouflage of military vehicles. These two areas of research have clearly opposite goals, in that developers of electric vehicles aim at adding the minimum amount of exterior noise that will make the EV acoustically noticeable by a blind or distracted pedestrian, while the developers of military vehicles desire to implement hardware configurations with minimum likelihood of acoustic detectability. The common theme is the understanding of what makes a vehicle noticeable based the noise it generates and the environment in which it is immersed. Traditional approaches based on differences of overall level and/or one-third octave based spectra are too simplistic to represent complex scenarios such as urban scenes with multiple sources in the soundscape and significant amount of reverberation and diffraction effects. This paper will show that the signal processing techniques required to map acoustic perception need to provide more resolution than overall level or one-third octave band based spectra and that the temporal pattern of a sound should be considered.

INTRODUCTION

Acoustic cues have significant contributions to the detection and identification of vehicles, and the understanding of this contribution is extremely important in the development of both civilian and military vehicles. Although both industries have an interest in understanding vehicular detection, the goals for expanding this understanding are significantly different. From the civilian vehicle OEM point-of-view, vehicle detectability, or more appropriately the lack thereof, has become a significant concern in the blind community [1]. In the case of a blind, or distracted, pedestrian the acoustic signature of a vehicle is an important cue warning of an approaching vehicle. Therefore, civilian electric vehicle OEMs are focused on developing warning systems that, in some form, can broadcast a “pleasant” acoustic signature with a high probability of detection. In contrast the goal of a military vehicle OEM is to design a vehicle such that the acoustic signature has a low probability of detection. From this perspective the likelihood of acoustic detection can be minimized by optimizing the vehicle hardware configuration based on the expected background noise and acoustic boundary conditions (reverberations) in the theater of operations. The combination of background noise and boundary conditions is commonly referred to as the soundscape.

In either case it is necessary to understand that the probability of detection is a function of both the vehicle acoustic signature and the acoustic masking component, or soundscape. And that both signals must be defined in greater depth than overall level, or average frequency content, in terms of 1/3rd octave, critical band, or narrow band spectra. The process described in this paper will demonstrate the necessity in providing more than the overall level or average spectral content

Acoustic Detection Models

Acoustic detection has traditionally been evaluated in fairly simplistic terms, using the overall sound pressure level (SPL), or some weighted or adjusted SPL. Improvements on this approach involve using the ratios or differences in average 1/3rd, or critical band spectra between the target sound and the ambient sound (soundscape), but still fall short of fully defining the temporal aspects of the problem. Some common approaches in defining acoustic detection models are as follows.

The aural prediction code ICHIN developed by the U.S. Army to improve the safety and survivability of Army helicopters is discussed briefly by Mueller et. al. [2] and Mueller et. al. [3]. In short the ICHIN code determines the probability of detection based on the difference between the ambient sound level and the target sound level in critical bands. A detailed description is provided in Abrahamson [4]

Ropoza and Fleming [5] discuss a probability of detection model, used in railroad regulatory compliance, based on vector summation of adjusted 1/3rd octave bands in terms of signal-to-noise ratio (SNR).

Miller et. al. [6] discuss audibility of aircraft as the point when the aircraft sound level is “similar” to the ambient sound level.

Hoglund et. al. [7] discuss a study in which subjects were asked to detect helicopter sounds in the context of ambient real-world recordings, including rural, suburban, and urban environments. In this paper the authors conclude that the ambient environment has an impact on the effective SNR level required to detect the presence of a helicopter, and that it is necessary to account for differences beyond treating ambient environments as overall spectrum of a steady-state masker.

Method of Evaluation

To assess the detection and identification of an acoustic target signature immersed in a soundscape a subjective listening study was conducted with a binaural playback system using high fidelity Sennheiser HD 580 headphones. The target sound used in this study is a non-military 6 cylinder diesel engine recorded in a free field acoustic environment. The soundscape is a binaural recording taken in an industrial park with a significant amount of traffic noise. The traffic noise includes passenger cars with gasoline engines, as well as a heavy semi-truck that can be heard applying and releasing its brakes and accelerating. The semi-truck, although powered by a diesel engine, sounds significantly different than the target vehicle to ensure that there will be no false positives based on misidentification. The industrial park can be described as a fairly loud and “busy” environment with significant variation in level and spectral content throughout the recording.

The listening study was designed to assess the probability of detection (PD) for a “stationary” target. The method for determining the PD is a method that could be described as direct elicitation stationary target (DEST), which rather than use an approaching target, uses a stationary target sound immersed in a soundscape [8]. In this method a series of sound files are generated with a varying parameter of choice, such as target level (or distance from the listener), but with a consistent soundscape. The sound files should include target levels that are detectable by all jurors (PD = 1), target levels that are indistinguishable by any jurors (PD = 0) and various levels in between. In this way, rather than study a continuously changing target parameter, the target parameter is varied in discrete steps. For each of the sound files the jurors are asked to indicate if they believe the target is present. The proportion of positive votes is used as an estimate of the PD.

For this study the level of the target sound was reduced by 2.5 dB increments and presented in a random order to the jury. A summary of the results for a 10 jury study is presented in the table in figure 1. In this case one can see that 9 out of 10 jurors correctly identified the target when it was reduced by 15 dB and mixed with the soundscape. The following discussion regarding metric development is presented using the soundscape and target -15dB sound files.

	Votes		
	Yes	No	Total
Target -10 dB	10	0	10
Target -12 dB	10	0	10
Target -15 dB	9	1	10
Target -18 dB	4	6	10
Target -20 dB	4	6	10
Soundscape only	0	10	10

Figure 1: Summary of a DEST listening study for target detection.

Metric development

One approach that can be used to understand the probability of detection problem is to decompose the target sound into its acoustic dimensions, or features. The goal of this approach is to fully describe the target sound in terms of a series of orthogonal, or at least independent, metrics. These metrics can then be correlated to the “measured” probability of detection from the jury study, and used in a linear regression analysis to build a PD model.

One approach is to start by describing all sounds with three general features including amplitude, pitch and timbre. So the goal of metric development is to define each of these features for the target and soundscape. In terms of the target sound the most obvious characteristic is the amplitude, which can be described in physical terms using sound pressure level (SPL) in decibels, or in psychophysical terms using Loudness. The physical term SPL describes the actual amplitude of the sound that can be measured at the listener location, whereas the psychophysical term Loudness [9] describes the perceived amplitude of the sound, so is likely more appropriate for a PD metrics.

Using the example of the stationary diesel engine as the target sound and the industrial park as the soundscape (DEST method) figure 1 compares the Loudness of the two sounds over the 4 second measurement period.

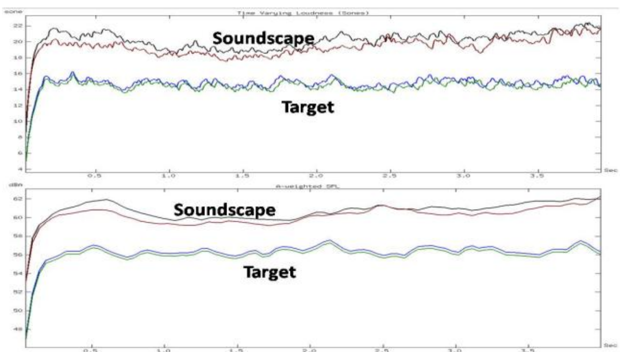

Figure 1: Comparison of the loudness vs time (top) and SPL vs time (bottom) for the Target and Soundscape.

In this case 9 out of 10 jurors were able to identify the target sound as being present in the soundscape despite the fact that the target amplitude is 5 sones (and 4-5 dBA) below the soundscape amplitude. This indicates that although the amplitude of the sound is important it is not the only characteristic that contributes to the PD.

The second subjective characteristic used to describe sound is the pitch, which is essentially the subjective perception of its frequency content. The physical characterization of pitch can be described by the spectral (frequency) content of the sound. In this case if one considers figure 2, where the $1/3^{rd}$ octave spectra from two different signals with the same overall sound pressure level are compared, it is clear that one signal is weighted towards high frequency and the second towards low frequency. In this case it is clear that even though the signals have the same level, a listener could easily discriminate between the two due to the difference in pitch. A metric that could describe the "spectral weight" would be the frequency of the centroid, or the frequency at which half of the energy is above and half the energy is below, termed the 50^{th} percentile frequency [10].

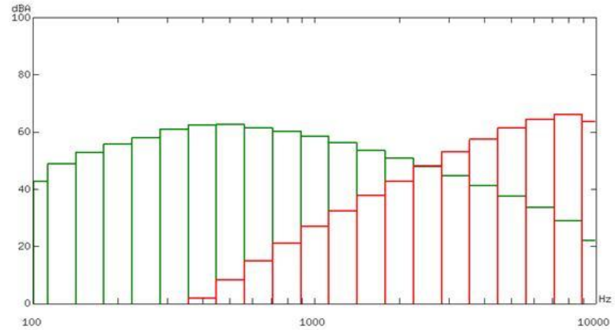

Figure 2: Comparison of two frequency spectra with the same overall sound pressure level, but different spectral content.

This concept applied to the target vehicle is shown in Figure 3, where the average $1/3^{rd}$ octave spectrum of the target vehicle is shown. The average 50^{th} percentile loudness for the target vehicle compared to the soundscape is 533 Hz and 716 Hz respectively.

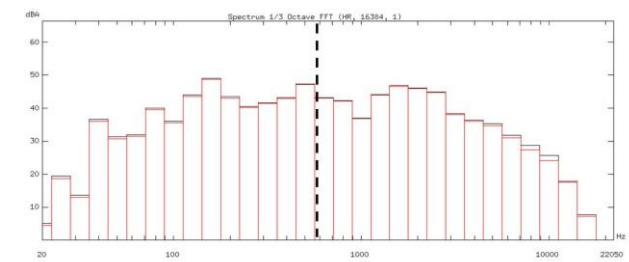

Figure 3: $1/3^{rd}$ octave spectrum for the target sound (diesel engine at idle). The dashed line indicates the 50^{th} percentile frequency, or the frequency at which the area under the loudness spectrum above and below are equal.

The third general feature of sound, often used in music, is timbre. Timbre is loosely defined as the third component of music that is independent of amplitude or pitch. An example often used in music is the ability of a musician to distinguish between the sound from two instruments that are playing at the same amplitude and pitch, and is also referred to as the "color" of the sound. Two aspects of the target and soundscape signals that we are considering that would contribute to the timbre are narrowband tonal content and temporal characteristics. Technically speaking the latter may be classified as rhythm, but that digresses from the intended discussion of this paper. The first characteristic, tonal content, is shown in Figure 4. In this figure one can see that if the soundscape did not contain the higher frequency content (600-1200) then the pitch of the target and soundscape, as estimated by 50^{th} percentile frequency, would likely be the same. In this case there are still discrete peaks that are present in the soundscape but not the target sound, and vice versa. These peaks would also likely allow a listener to discriminate between the two sounds, which would increase the PD. In this case a metric that defines the difference in tonal content, such as a narrowband spectral difference could be used to objectively measure these differences.

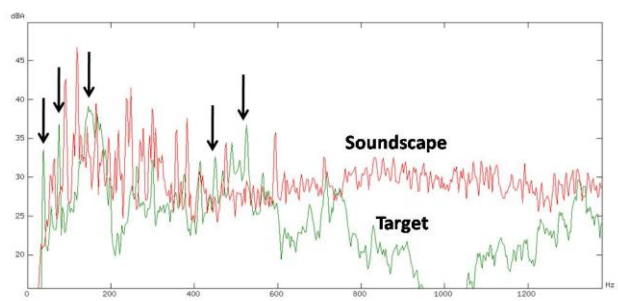

Figure 4: Narrowband frequency spectra comparing the soundscape (red) and the target (green). The arrows indicate the frequencies at which the target exceed the soundscape level.

As part of the detectability study the target shown in Figure 4 was modified such that the peaks that exceed the soundscape were reduced. A narrowband frequency spectra of the modified target and soundscape are shown in Figure 5. In this case one can see the average frequency spectra for the 4 second recording of the target is lower than that of the soundscape spectra. Although the average frequency spectra of the target sound is lower than that of the soundscape a listening study shows that the target sound can still be identified based on the periodic diesel "clatter" sound.

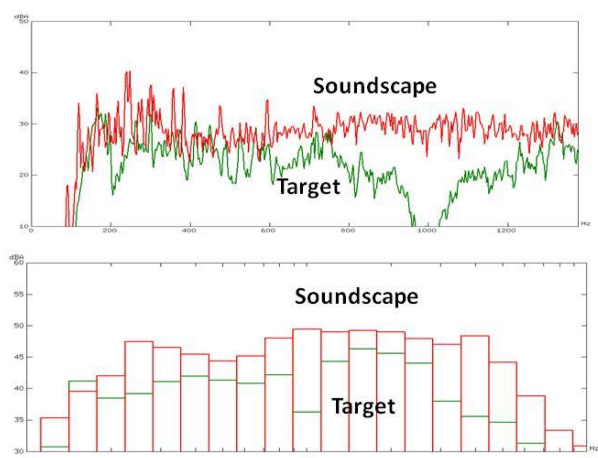

Figure 5: Narrowband and 1/3rd octave spectra of the modified soundscape (red) and target (green).

The temporal nature of the "clatter" demonstrates the second component that would fall within the description of timbre. One metric that could be used to define the temporal characteristic of the sound would be the 50th percentile frequency vs time, shown in Figure 6. This plot essentially describes the pitch variation vs time but temporal characteristics could also include amplitude or tonal variations.

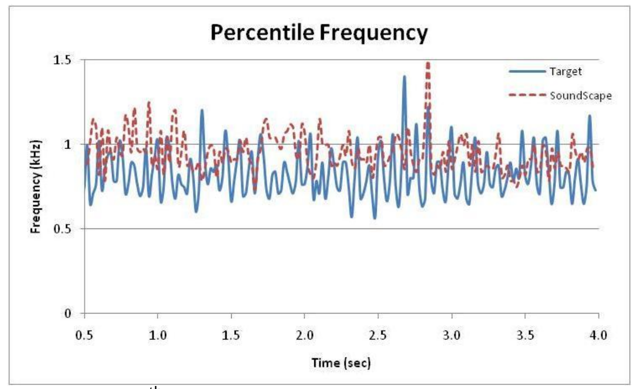

Figure 6: 50th Percentile Frequency vs time for the target and soundscape.

In this figure it appears that although there is a slight offset between the two signals the target signal has a much more regular periodic nature to it as opposed to the soundscape which has much more erratic variations. This regularity in the 50th percentile frequency, or pitch, is a feature that helps in detecting the target sound. In this example the subjective evaluation of the target embedded in the soundscape, shown in Figure 6, confirms that the most identifiable feature of the target in this case is the periodic nature of the diesel clatter. For comparison a white noise signal was shaped with the frequency spectrum of the target sound to create a "steady-state" sound with comparable frequency content. When the two were compared subjectively, the diesel target sound was much more identifiable than the steady sound. This also makes intuitive sense as it would be expected that a target sound that turns on and off in a period manner will be much less likely to "blend-in" with the soundscape.

PD Models

Finally a model of the acoustic detection can be built using the metrics described and developed in the previous section to predict the PD as determined during the jury analysis. This model can be developed in several forms, the simplest of which is a Multiple Linear Regression (MLR) approach, and ranging to more involved solutions such as Artificial

Neural Networks (ANNs) and Radial Basis Function Nets (RBFNs) [11].

In the case of the MLR approach it is recommended to approach the model in a stepwise manner, for example the first metric in the model would likely be a term that describes amplitude. In this case a linear regression model is built using a least squares approach such as:

Where {PD} is vector containing the proportion of votes in favor of detection for each sound file under test, C and C_L are the constants representing a bias and the loudness constant respectively, and $\{\Delta L\}$ is a vector containing the loudness values that correspond to the proportion in {PD}. When this set of equations is solved, the correlation (such as R^2 or the F-statistic) is used to determine if an additional metric is necessary. In the example described in this report it is likely that the next term in the MLR equation would include a metric that gives an indication of pitch, possibly 50^{th} Percentile Frequency. In this case the MLR solution could be written as:

Where ΔPF is a vector containing the difference in the percentile frequency between the target and soundscape, and C_{PF} is a constant representing the percentile frequency weight.

This process is repeated until the correlation between the predicted PD from the MLR model adequately represents the estimated PD from the listening study. This linear model can then be extended in the same way to include the other modalities of interest, or modified slightly to generate a non-linear MR model. A non-linear model would include second order terms such as L^2 or $L*\Delta PF$, such as:

Additional methods of modeling PD could consider intelligent approaches such as ANN and RBFNs. There is a great deal of flexibility in the network architecture, and solution approach in intelligent systems, so a simple network will be presented, but the approach would be optimized for the final system. Figure 8, shows an example of a network that could be used to predict PD.

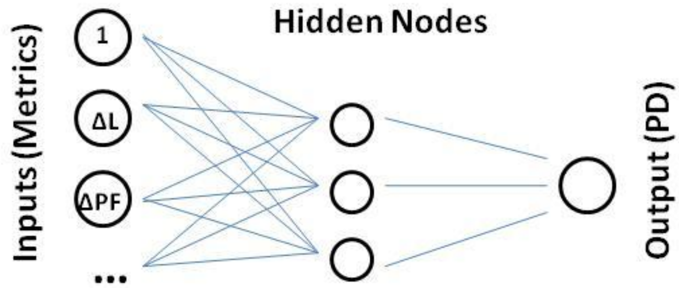

Figure 8: Example of a network architecture that could be used in a multiple perceptron ANN, or RBF network.

In this network the input is a vector containing the metrics that describe the differences between the target and soundscape sounds, and interconnected to all of the hidden nodes. The hidden nodes then sum to generate the network output, or predicted PD. The function of the nodes and method for solving, or "training", the network is dependent on the type of network that is chosen. In the case of a multiple perceptron ANN the nodes would contain a summation and non-linear function, such as a sigmoid function, whose output is then summed with all of the outputs from the other nodes in the hidden layer. The network weights and biases would then be solved using a back propagation algorithm.

If a RBF network is used to model the system the hidden nodes would contain the appropriately designed Basis Functions. In this case the outputs from all of the hidden nodes would be combined as a weighted summation to form the output, which is again the predicted PD.

In either case the network will "learn" to predict the PD value as defined by the proportion of votes that indicate the target vehicle has been detected. The input vector and estimated PD (from a listening study) are taken for each sound file under test.

REFERENCES

[1] P. Goodes, G. Cerrato, E. Meyer, "Investigation into the detection of a quiet vehicle by the blind community", INCE Sound Quality Symposium – Dearborn, MI, July 2008.

[2] A. W. Mueller, C. D. Smith, P. LeMasurier, "Improvement of the Predicted Aural Detection Code ICHIN (I Can Hear It Now)", Noise-Con 93, pp. 403-408, May 1991

[3] A. W. Mueller, C. D. Smith, K. P. Shepherd, B. M. Sullivan, "A New Version of the Helicopter Aural

Detection Program –ICHIN", NASA Technical Memorandum 87745, July 1986

[4] A. Abrahamson, "Correlation of Actual and Analytical Helicopter Aural Detection Criteria (U)", USAAMRDL Technical report 74-102A, Vol. 1, January 1975

[5] A. S. Rapoza, and G. G. Fleming "Determination of a Sound Level for Railroad Horn Regulatory Compliance" U.S Department of Transportation, Federal Railroad Administration Report number DTS-34-RR397-LR1, FRA/RDV-03/28, 2002

[6] N. P. Miller, G. S. Anderson, R. D. Horonjeff, C. W. Mange, J. C. Ross, and M. Newmark, "Aircraft Noise Model Validation Study", Report No. 295860.29 prepared for the National Park Service, 2003

[7] E. Hoglund, D. Brungart, N. Iyer, J. Hamil, F. Mobley, and J Hall, "Auditory acuity for aircraft in real-world ambient environments", J. Acoust. Soc. Am., vol. 128, issue 1, July 2010

[8] M. Meilgaard, G. V. Civille, B. T Carr, "Sensory Evaluation Techniques", 2nd Edition, CRC Press New York, 1991

[9] E. Zwicker, H. Fastl, "Psychoacoustics Facts and Models", Springer-Verlag, Berlin, 1990

[10] R. J. Fridrich, "Percentile Frequency Method for Evaluating Impulsive Sounds", Proceedings of the 1999 SAE Noise and Vibration Conference, SAE Paper No. 1999-01-1851, 1999

[11] L. Fausett, "Fundamentals of Nueral Networks", Prentice-Hall, Upper Saddle River, New Jersey, 1994