19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

NOIDESc: Incorporating Feature Descriptors into a Novel Railway Noise Evaluation Scheme

PACS: 43.55.Cs

Brian Gygi[1], Werner A. Deutsch[2], Anton Noll[2], Alexander Hirner[2] and Timo Becker[2]
[1]East Bay Institute for Research and Education, 150 Muir Rd. Martinez, CA 94553 USA bgygi@ebire.org
[2]Acoustics Research Institute of the Austrian Academy of Sciences, Reichsratsstrasse 17, Vienna 1010 Austria

ABSTRACT
NOIDESc, a novel framework for the description of noise signals (railway and road) is proposed which has several components: online processing of immission recordings at the train site; automatic generation of metadata for use in the segmentation and annotation of the train signals; incorporation of the recordings into a cumulative collection of calibrated sound recordings; and, the automatic extraction of low-level feature descriptors from the signal. Principal component analysis and cluster analysis on the frequency bands yielded concurrent results, pointing to three main categories of timbre, which can be described as “dark”, “medium” and “bright.” Discriminant analysis validated these categories with an error rate of less than 5.5%. The intensity and duration can similarly each be grouped into three classes, namely “loud” “medium”, “soft”, and “long”, “medium-long”, “brief”, respectively. These findings show the utility of incorporating timbre descriptors into automated noise classification schemes.

INTRODUCTION

Devices for noise measurement are an important factor in urban planning, in situations when noise complaints already exist, or when decisions have to be made how to minimize noise in a community or in parts of a community[1]. Terms such as noise, severe noise, or health-threatening [2]. noise are defined based on medical-psychological data, which often barely correlate with acoustic measures. Studies on the psychological effects of noise are usually conducted using questionnaires, semantic rating scales or other verbal measurements. Multivariate analysis of such data only addresses a subset of causal factors, and often neglects emotional and attitudinal factors. Thus, a classification system for noise is proposed which is based on similarity measurements and low- and high-level descriptors of the noise signal, which can be measured in an automated fashion, without relying on subjects’ verbal assessments[3].

1. DESCRIPTION AND GOALS OF THE NOIDESc PROJECT

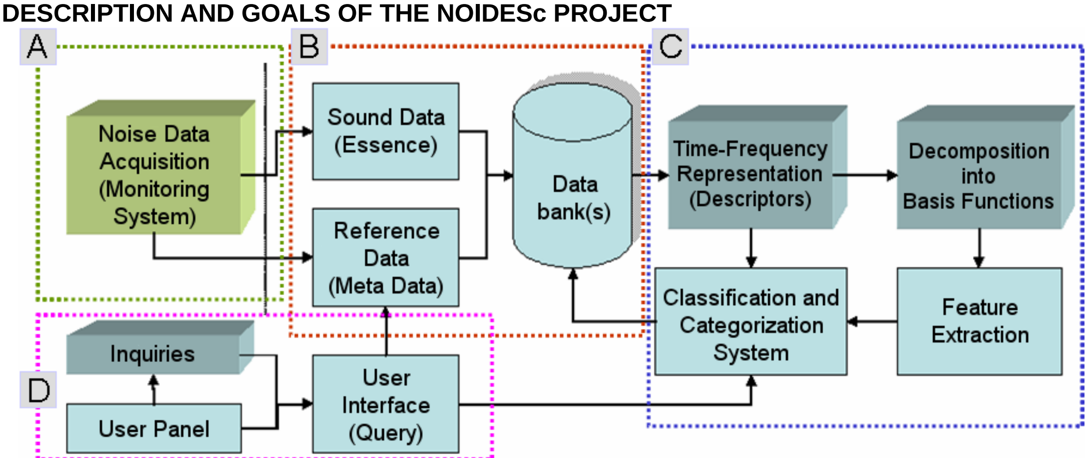

Figure 1: NOIDESc Project overview

This project proposal contains:

1. The development and implementation of tools for both acoustic and (subjective) perceptual relevant descriptors for noise [*9*][4]. This should include a formal model for the automatic calculation of such perceptual descriptors, which are meant to supplement traditional parameters (levels, frequency) captured by existing noise measurement devices.
2. Development of a sufficiently large data base of signal types in order to reliably classify and catalogue different types of noise.
3. The project plan includes the construction of a prototype system in order to evaluate the results of the project.

The concept and the execution of the project follow a descriptor pattern similar to the MPEG-7 part 4 audio standard, which have several descriptors enabling a detailed consideration of tone quality, or *timbre*. One of the main hypotheses of this project is that during the classification and evaluation of noise events substantially greater importance should be assigned to timbre than it has been in other noise evaluation schemes to date

2. SOUND DATA COLLECTION AND RECORDING MEASUREMENTS

In order to get a sufficiently large database of train sounds for evaluation, continuous recordings have been made for the past year in Neumarkt/Ybbs, Austria on private property located on a lightly traveled side street. The layout is shown in Figure 2.

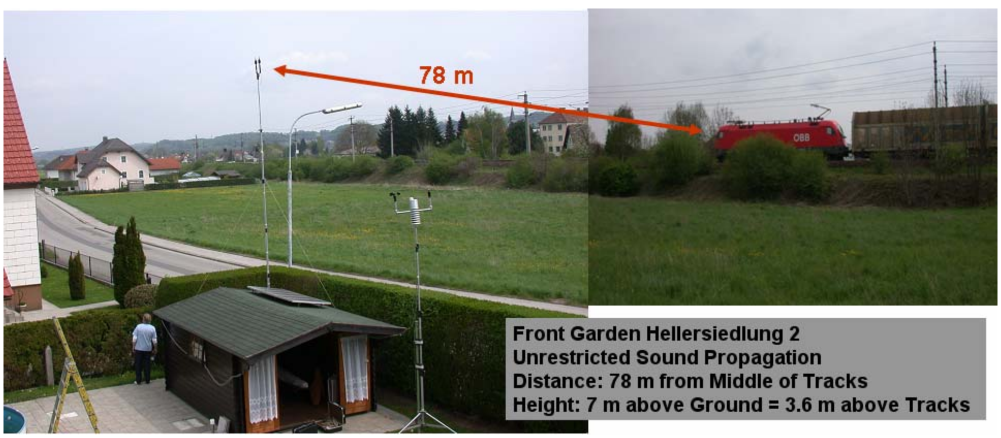

Figure 2: Sound recording location at Neumarkt/Ybbs

The recordings were made with two calibrated ½“ condenser microphones with an isotropic pattern, at a distance of 17 cm parallel to the train tracks. To date there have been a total of 816 hours of recordings processed, corresponding to 485 GBytes of storage.

3. INDIVIDUAL SOUND EVENT SEGMENTATION AND DESCRIPTION

3.1 Segmentation

After initial hand segmentations of the individual sound events were made, standards for automatic segmentation could be decided upon. The short-term and long-term level values, calculated from the FFT of the signal (125 ms frame length, 50% hopsize, Fc = 44.1 kHz) over the frequency range 16-8000 Hz, were compared. For segmenting core segments the threshold value is that level which is reached or exceeded in 1% of the Segment frames ($\mathbf{L}_{01*}$, $\approx \mathbf{L}_{A,1}$), less $\mathbf{K}_{A,\,max*}$ = 10 dB; i.e., $\mathbf{L}_{A,\,max*} = \mathbf{L}_{01*} - \mathbf{K}_{A,\,max*}$. The time intervals (beginning and end of segment) are determined by the maximum points which fall below the $\mathbf{L}_{A,\,max*}$ level.

3.2 Manual annotation of the automatically generated segments

After segmentation and automatic indexing of the segments, the segments were listened to and manually classified according to type of event, as shown in Table 1 below.

a. TrainP: Passenger train	e. Automobile	h. Moped
b. TrainG: Freight train	f. Truck	i. Motorbike
c. Train: Unspecifed train	g. Tractor	j. Hcopter: Helicopter

d. Locomotive alone

Table 1. Manual classification scheme for train types

3.3 Amplitude distribution for the individual sound events

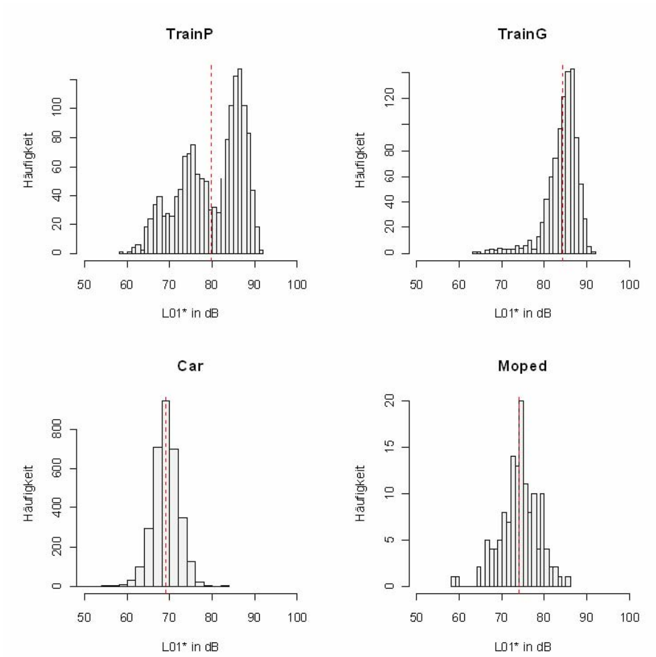

Figure 3. Period histograms showing the typical level distribution in L01* dB(A): The bimodal distribution for passenger trains results from the combination of quickly passing trains and more slowly moving ones starting out *[3]*[5].

3.4 Mean duration for the individual sound events

The data for Neumarkt/Ybbs show the following mean durations/log attack time in sec:

	TrainG	TrainP	PKW	LKW	Motorbike/Moped	Tractor
Mean	19.0 / 5.1	8.8 / 3.2	5.7 / 1.9	5.6 / 3.5	3.8 / 3.7	5.7 / 8.8
SD	6.2 / 4.1	3.2 / 5.7	1.7 / 2.6	1.9 / 3.5	1.7 / 5.5	1.7 / 8.8
N	934	1463	3235	108	189	62

Table 2. Mean duration/log-attack-time for the individual sound events

4. FEATURE EXTRACTION UND CLASSIFICATION

A drastic data reduction can be achieved by extracting the energy in 21 different 1/3-octave frequency bands which would still be sufficiently precise to retain the spectral differences of the individual sound events. From this sparser representation higher order statistical procedures such as Principal Component Analysis and Hierarchical Clustering were performed to reduce the underlying structure to a few main dimensions.

Band 1	63	Band 6	200	Band 10	500	Band 14	1250	Band 18	3150
Band 2	80	Band 7	250	Band 11	630	Band 15	1600	Band 19	4000
Band 3	100	Band 8	315	Band 12	800	Band 16	2000	Band 20	5000
Band 4	125	Band 9	400	Band 13	1000	Band 17	2500	Band 21	6300
Band 5	160								

Table 3. The lower frequency cut off boundaries (in Hz) for the 21 frequency bands used in the ⅓-Octave analysis.

4.1 Principal Component Analysis (PCA)

To decompose the data matrix consisting of the levels in the 21 frequency bands for the individual sound events a Principal Component analysis on the basis of the covariance matrix across the bands was performed, shown in Figure 3.

The first component (PC1) accounted for by far a greater proportion of the total variance than any other component. It shows hardly any frequency-specific information and refers to the energy portion common to all noises across the spectrum. PC2 is predominantly concentrated in the low-frequency region, PC3 lies in the middle frequency range and PC4 is predominantly high, however there are also low frequency portions. PC5 and PC6 are difficult to interpret and account for a very small portion of the total variance.

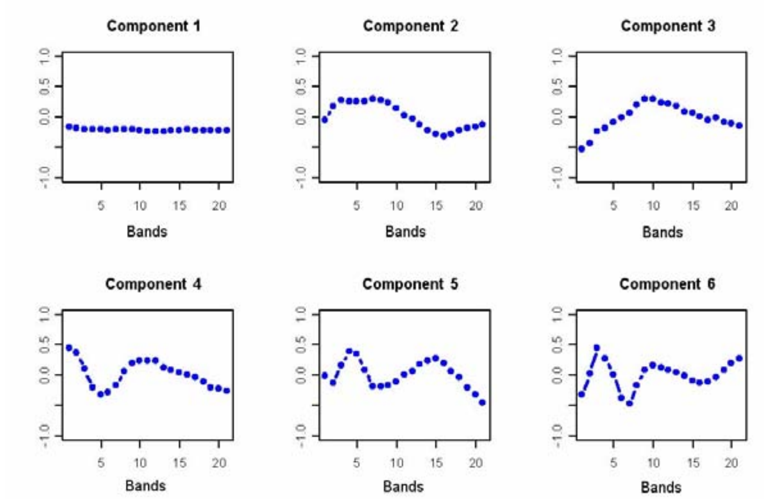

Figure 3. The first six components from the PCA of the segmentwise averaged level values in 21 frequency bands.

4.2 Cluster analysis by Frequency Bands

A hierarchical cluster analysis of the 1/3-octave levels in 21 frequency bands was performed on 954 Train sounds (488 Passenger-, 466 Freight trains, >80 dBL_{01^*}). The results, shown in Figure 4 below, clearly indicate three main clustering of sounds.

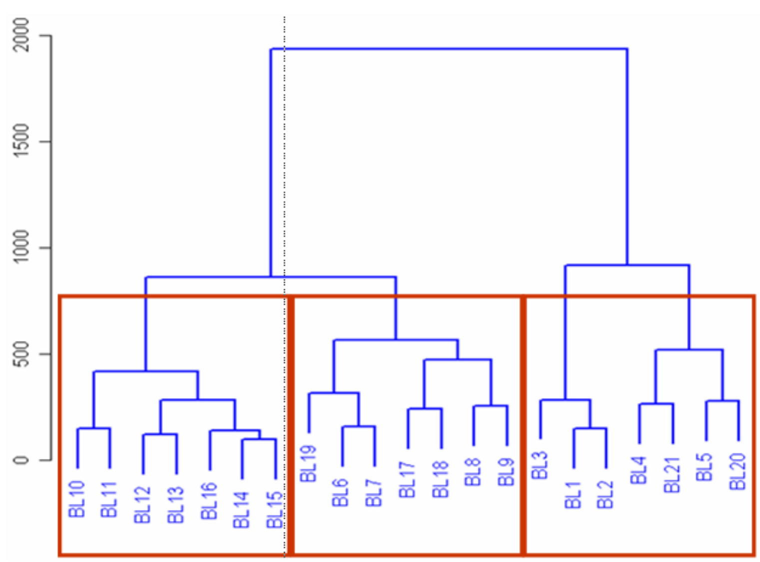

Figure 4. Hierarchical cluster analysis of the 1/3 octave levels in 21 frequency bands

Plotting the mean level values for each grouping by frequency bands reveals that three groups can be divided into into ‘dark’, ‘medium’ and ‘bright’ groups on the basis of their timbre [6, 7, 8], as shown in Figure 5.

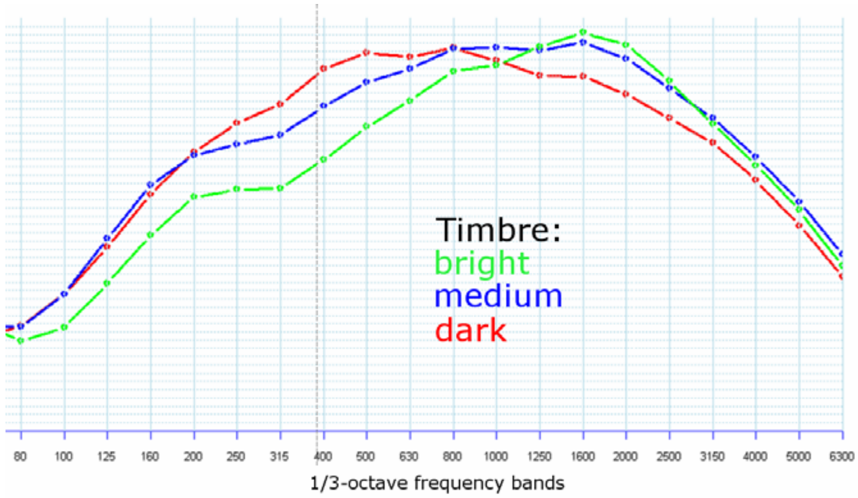

Figure 5. Mean value of the relative 1/3-octave levels in 21 frequency bands according to cluster membership

4.3 Cluster analysis by peak level and duration

Clustering the sounds by peak level (L01*) showed similar tripartite groupings to that found in the clustering by timbre, demonstrated in Figure 6.

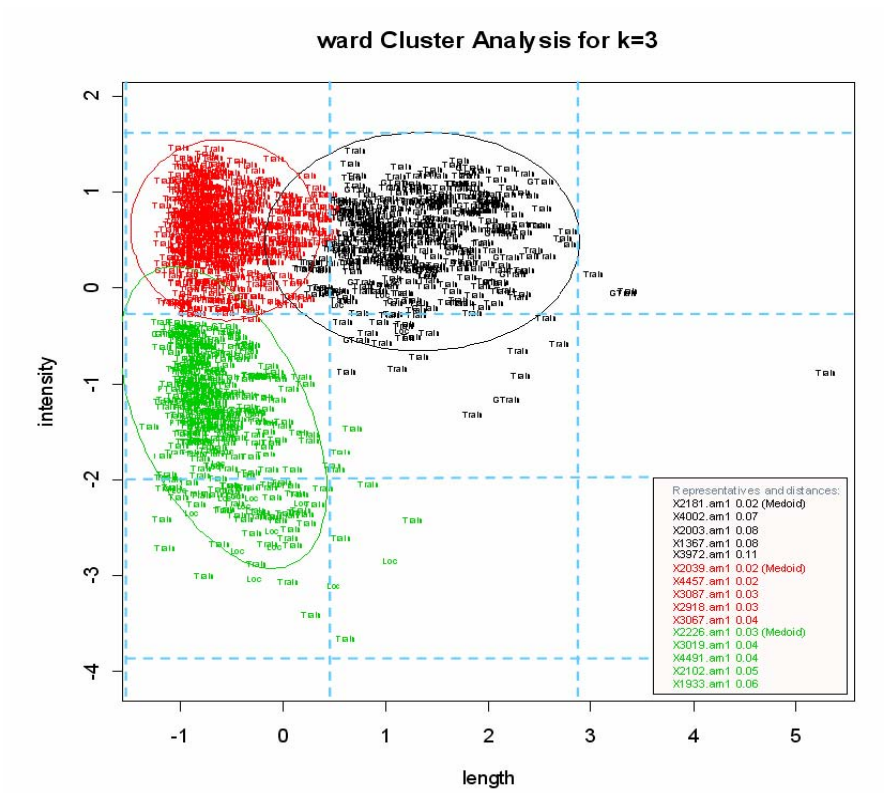

Figure 6. Hierarchical cluster analysis based on segmentation level (L01*, ≈ LA,1) and duration

Cluster analyses based on segmentation level and duration reveal three main groupings of sounds, which can be described as “short-quiet”, “short-loud” and “medium-loud”. Some areas in the matrix are almost totally unoccupied, e.g. “quiet-long.”

4.4 Discriminant Analysis

In order to compare of the results from the PCA (PC1 to PC5) with the allocations from the cluster analysis, a discriminant analysis was performed. Input data are subsets of the first 5 main components, from which 32 forecast models were derived. The regressand was the cluster affiliation from the cluster analysis. The results show that by including the first five main components the classification error rate for grouping sounds into the three tone quality classes achieves is 5.45%.

PC_1	PC_2	PC_3	PC_4	PC_5	False classification rate in %
+	+	+	+	+	5.45
+	+	+	-	+	7.65
+	+	+	+	-	7.74
-	+	+	+	+	7.90
+	+	+	-	-	9.70
-	+	+	+	-	10.08
-	+	+	-	+	10.72
-	+	+	-	-	12.05
+	+	-	+	-	25.98
+	+	-	+	+	26.09
+	+	-	-	-	27.15
+	+	-	-	+	27.46

Table 4. Results of the discriminate analysis predicting cluster membership listed by inclusion of principal components. For the sake of space not all 32 possible configurations are shown.

5. CONCLUSIONS

1. The reduction of the spectral resolution on 21 frequency bands (third octaves, 63 - 8,000 hertz) for the classification of timbre is precise enough to describe these sounds, with minimal loss of information about the spectral variance (timbre).
2. The timbre can be divided in 3 classes, described here as "dark", "medium" and "bright".
3. Sound pressure level and train duration can be likewise grouped in 3 classes e.g., “very loud”, “medium loud“, “soft“, and “long“, “medium long“and”brief“, respectively. Finer gradations and their dimensioning are freely selectable.
4. Cluster analysis and PCA provide converging evidence for the existence of the timbre categories, which was validated by discriminant analysis results showing an error rate for classification of 5.5%.

It is hoped that with the suggested methodology and the obtained results an improved signal-based and objective classification utilizing timbre information can be developed for train noises.

References:

[1] Zwicker, E. & H. Fastl (1990): Psychoacoustics. Springer-Verlag, Berlin.
[2] Guski, Rainer (2000): Stellungnahme zu den medizinischen Gutachten M8 und M9 bezüglich des Ausbaus des Flughafens Schönefeld; Seite 9. http://www.bvbb-ev.de/Guski.pdf
[3] Deutsch W.A. (1996): Visualisierung von Bahnlärm, In: Tagungsband, Fachausschuss für Akustik, Österreichische Physikalische Gesellschaft.
[4] Stevens, Stanley Smith (1975): Psychophysics: introduction to its perceptual, neural, and social prospects. John Wiley & Sons, Inc.
[5] Deutsch W.A (1994): Schallspektrographie. In: Psychoakustik, Tagungs-band ÖAL u. BM Umwelt, Jugend u. Familie, 153-170.