NOIDESC: CONTENT BASED DESCRIPTION OF TRAIN NOISE

Werner A. Deutsch[1], Holger Waubke[1], Brian Gygi[1], Anton Noll[1] and Matthias Stani[2]

Acoustics Research Institute[1]
Reichsratsstrasse 17, A-1010 Vienna, Austria
Werner.Deutsch@oeaw.ac.at (e-mail address of lead author)
Federal Institute of Heat and Sound Technology[2]

Abstract

Measures for describing noise data are usually selected from national and international standards or guidelines. These standards almost exclusively employ integrated levels with varying temporal and spectral weighting, primarily A-weighting and temporal exponential smoothing with a 1s time constant. More specific descriptions use octave- and third-octave band spectra. Because of the averaging methods used, these parameters are not suitable to reproduce transients (isolated short time events) or to represent perceptual relevance sufficiently. A direct psychoacoustic evaluation using only annoyance estimates remains questionable, because of the large variability of annoyance estimates both between and among individuals. The temporal interpersonal variability is based (among other factors) on differential sleep patterns, attitudes towards the cause of the noise and short-time trends in the social environment.

A solution that is independent from psychological variability, but characterizes noise events in more detail, uses multiple features that are derived from the waveform yet provide perceptual relevance. The international ISO standard MPEG7-4 was defined during the past years for semi-automatic description of audio in multimedia. The descriptors outlined in the MPEG7-4 standard are here tested in detail on a large number of recorded train segments. A method is proposed to reduce the large number of MPEG-7 descriptors to a smaller set that is relevant for noise events.

Using MPEG-7 descriptors and some related acoustic measures, the similarity space for the train sounds is developed, and the features that best predict the structure of the space are isolated. This method singles out the features that best allow detection of acoustic events and also to identify general classes of trains with the highest risk of annoyance. Among these features are complex timbral and envelope descriptors. The project includes the installation of a monitoring system at a railway track to prove the relevance of the used feature set.

INTRODUCTION

From the earliest attempts to measure environmental noise, researchers have tried to find descriptors that would be objective (i.e. independently measurable), reliable, would reflect acoustic features of the sounds and which would predict individuals' responses to the noise, usually expressed as annoyance ratings. One measure which has been widely adopted as a standard is the day-night average sound level (DNL) first described in [14]. This uses the average A-weighted sound level with a relatively slow time constant of about 1 second. Other variants have been employed, such as the L_{eq}, but nearly all of them have employed the A-weighted sound level with temporal smoothing.

While these measures have performed adequately for predicting long-term responses to non-specific background noise of a fairly constant level, the temporal and spectral resolution are insufficient for capturing short-term transients or spectral peaks that have a great deal of perceptual relevance both from an informative standpoint [4] and an annoyance standpoint [12].

Further, the use of annoyance itself as a dependent measure has been questioned. Annoyance ratings are highly variable [3] due to several non-acoustic factors such as different measurement techniques [1] demographics [11] and the attitude of the listener to the sound source: for instance, railway noises are consistently found to be less annoying than road or airplane noises of the same level [9, 13]. It has also been suggested that annoyance be replaced as a dependent variable by other measures, such as number of community complaints or sleep disturbance [7].

The above issues also apply to automatic noise assessment methods. Automatic devices for noise measurement are an important factor in urban planning, or when deciding whether an area should become a suburb or an industrial zone. They are also helpful 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. Measuring noise helps to prioritize attempts to protect the community from noise, and they aid in the computer simulation of planned steps for noise reduction. Thus, we propose a classification system for environmental noise 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 the subject's verbal assessment. This will also allow communities flexibility in determining specific noise standards, since all judgments can be scaled to a baseline measure.

In the past several years, quite a bit of work has been done on content-based retrieval for music and audio [2], which aims to recover features of music and audio (such as musical genre) from the waveforms and using this information as metadata for further classification and search. To standardize this procedure, the International Standards Organization (ISO) developed the MPEG-7/4 standards for Audio [6]. Some examples of these descriptors are AudioSegmentType, SoundClassificationModel, and SpectralCentroid. A fair bit of research has investigated the capabilities of MPEG-7 descriptors in handling music and general audio [8]. In this paper content-based descriptors will be applied to a particular type

of environmental sound: train noises. It is expected that this approach will enable classification and similarity ratings of train sounds without the established problems of the previous measures listed.

DATABASE OF TRAIN SOUNDS

The main database of trains sounds used in these studies was from field recordings supplied by TGM. The recordings were from the same location on three different dates in 1997. Details For each date the recordings for that date were incorporated into a single .wav file. These recordings yielded 180 usable segments of train sounds, with 70 being from fast trains (Schnellbahn), 61 from freight trains, 22 from passenger trains, 19 from Tractor trains and the remainder were unspecified. The durations of the segments ranged from 6 s to 61.8 s, with a mean of 24.34 s. The Maximum A-weighted amplitudes of the segments (L_{max}) ranged from 70 dBA to 101 dBA with a mean of 92.12 dbA. For comparison purposes, a second, smaller database of recordings from the Nordbahn was obtained (details of the recording conditions are also in the Appendix). The recording was from a single date in 1995 and again all the recordings were included in a single .wav file. This recording contained 21 useable segments of train sounds, of which 16 were passenger trains, 3 tractor trains and only 2 freight trains. The durations of these segments ranged from 7.8 s to 60.55 s with a mean of 23.9 s. The L_{max} levels were between 78. dBA to 98 dBA with a mean of 89.62 dBA. The .wav files were all read into STx by the suppliers of the recordings who also provided the annotations of each segment boundary and metadata about each segment, such as the type of train, length, number of cars and any unusual features about the recording (such as a car passing by or the presence of a strong wind).

ACOUSTIC ANALYSIS OF TRAIN SEGMENTS

Statistics of the Steady State Portions of Train Segments

The steady state portion of each train segment was extracted using STx. This was defined as the section that ranged in amplitude between $L_{.01}$ (the level corresponding to the 99th percentile of amplitude values) and 10 dB below $L_{.01}$. Figure 1 shows a train segment with the outer segment boundaries as annotated by TGM and the inner boundaries as selected by STx. The STx-generated segment boundaries were used for further analysis

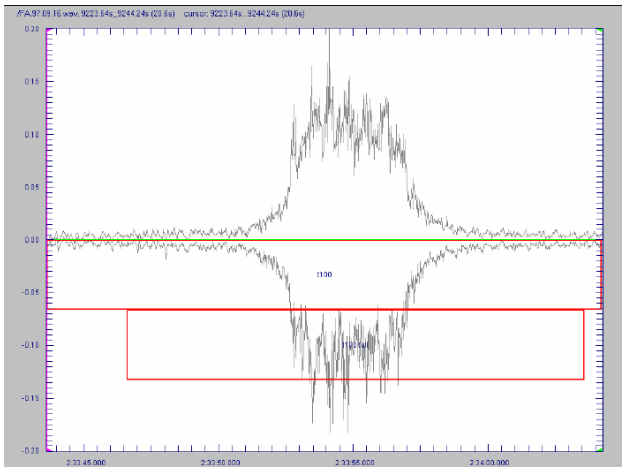

Figure 1. Train segment with annotated and automatic segment boundaries

The distribution of durations for the steady state segments was heavily positively skewed because of the large number of fast trains. The distribution of L_{Max} values was almost normal. Histograms for both distributions are shown in Figure 2.

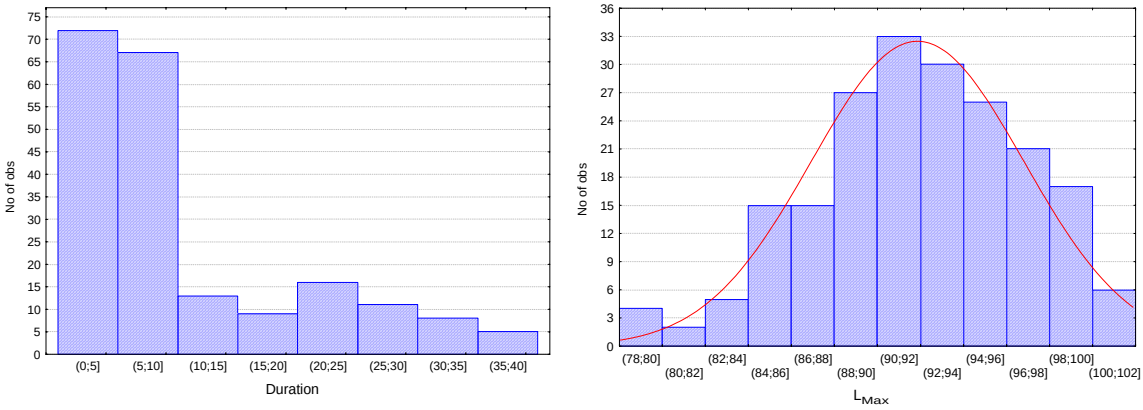

Figure 2. Duration (seconds) and LMax (dBA) distributions for the steady portions of all segments

MPEG-7 Features (As Defined in ISO/IEC FDIS 15938-4) Used for Analysis

TemporalCentroid Descriptor

The temporal centroid is defined in section 5.3.20 of the ISO standard as the time averaged over the energy envelope, and is a fairly common representation of the envelope of a signal.

AudioSpectrumCentroid Descriptor

The AudioSpectrumCentroid descriptor is defined in section 5.3.8 as the center of gravity of the log-frequency power spectrum, calculated using a sliding FFT window of 50 ms with 75% overlap.

AudioSpectrumSpread Descriptor

AudioSpectrumSpread (section 5.3.9) describes the second moment of the log-frequency power spectrum, or the RMS deviation of the log-frequency power spectrum with respect to its center of gravity.

Measures Related to MPEG-7 Features Used for Analysis

Spectrum Skew

The Spectrum Skew is closely related to the AudioSpectrumCentroid and Spread descriptors, in that it is the third moment of the log-frequency power spectrum.

Spectrum Slope

The slope of the Spectrum was extracting by calculating the best-fitting linear regression equation to the power spectrum.

Spectral Flux

Spectral flux is another measure of the change in the spectrum over time. As described by [10], it is the running correlation of spectra in short (50 ms) time windows.

Additional Features

Modulation Spectrum Peaks

The modulation spectrum, first suggested by [5], reveals periodic temporal fluctuations in the envelope of a sound. The algorithm used here divides the signal into frequency bands approximately a critical band wide, extracts the envelope in each band, filters the envelope with low-frequency bandpass filters (upper F_c ranging from 1 to 32 Hz), and determines the power at that frequency. The result, shown in Figure 3, is a plot of the depth of modulation by modulation frequency. Peaks in the modulation spectrum of train sounds correspond to significant events, such as flat wheels and the individual cars, which are notated in Figure 3. The train segments examined, most had a very low peak at about 2 Hz, representing the individual cars of the trains. Some trains had noticeable peaks at about 5 Hz; listening to those segments suggested those peaks are caused by flat wheels on the trains. Te modulation spectrogram is also useful for distinguishing train from non-train sounds, also shown in Figure 3.

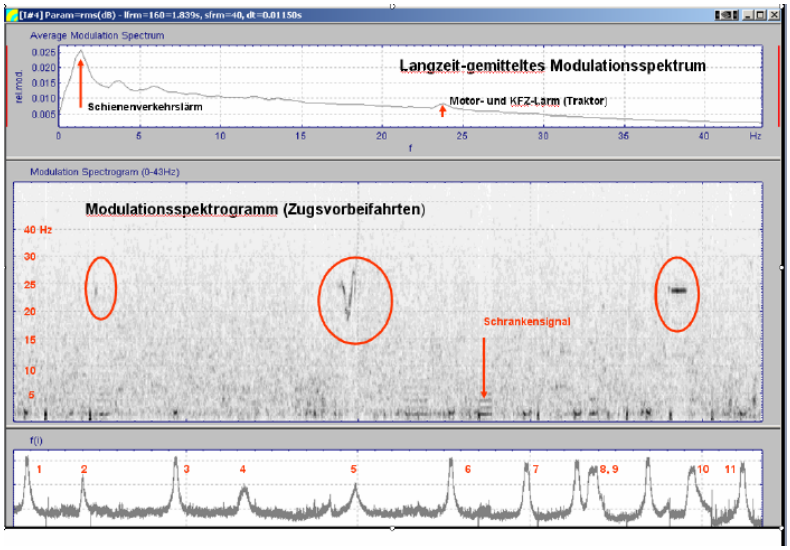

Figure 3. (Top panel to bottom) Modulation spectrum, spectrogram and wave form of a train sound. The marked peaks in the modulation spectrum correspond to specific train sounds, as noted in text. The circled sounds do not belong to trains.

ANALYSES

Intercorrelations

Several of the MPEG-7 and MPEG-7 related descriptors were highly correlated, as shown in Figure 4, which plots the short-term RMS, Spectrum Centroid, Spread, Skew, Flux and Slope as a function of time for the steady state portion of a single train sound. It is clear the RMS, Centroid and Spread are highly correlated (with a negative correlation for the Skew), as are the Slope and the Flux. These correlations across all train segments are shown in Table 1.

Table 1. Intercorrelations of acoustic parameters

	Spread	Skew	Slope	Flux
Centroid	**-.86**	**-.98**	-.63	-.51
Spread		**.88**	.62	.50
Skew			.70	.60
Slope				**.97**

Of these parameters, the most parsimonious and least susceptible to different recording conditions (including RMS) is the spectral centroid. The distribution of spectral centroids across train segments is nearly normal, ranging from 1966 to 3619 Hz, with a mean of 2818 Hz, and *SD* of 342.9 Hz. Moreover, the spectral centroid is quite good at predicting the train class. Predictions as to whether a train was a fast train, passenger train or freight train were 65% correct using only the spectral centroid. Moreover, since the spectral centroid correlates highly with Terhardt's measure of roughness, there is a clear association with a perceptual feature. This analysis shows that spectral centroid can be a useful parameter for assessing similarity of train sounds.

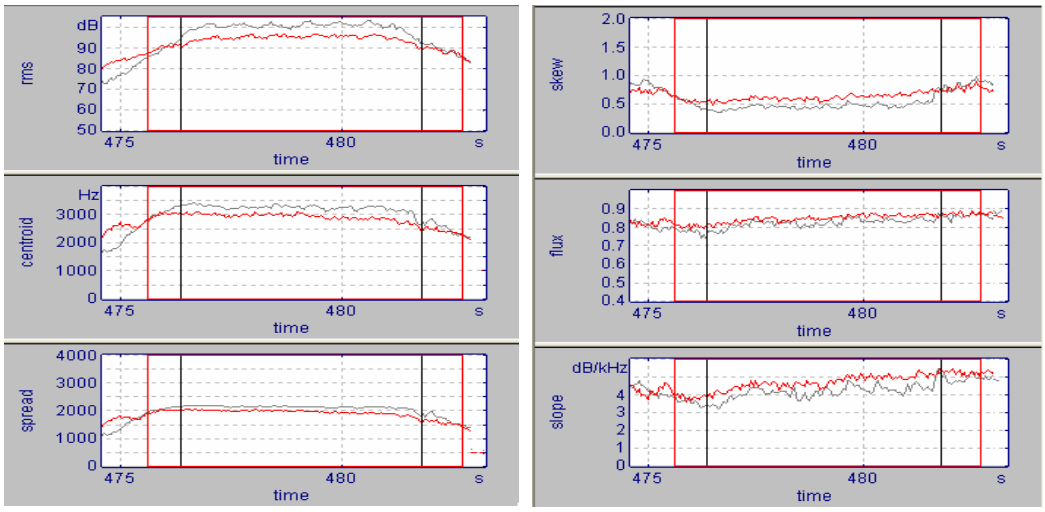

Figure 4. Plots of parameters in short-time windows for one train sound

Hierarchical Clustering

In an effort to quantify timbral differences in the sounds, the long-term spectra of all 201 sound segments were subjected to a hierarchical clustering analysis. The clustering method was based on distance, using pairwise agglomeration and the Ward Clustering Method. A five-cluster solution seemed to group aurally different train segments. A plot of the mean spectra for each cluster is shown in Figure 5.

These clusters captured complex timbral features of the sounds, incorporating many of the variables noted above. Using the variables L_{Max}, Centroid, Spread, Skew, Slope and Flux ,a discriminant analysis classification model was able to predict the clustering of a particular segment with 86% accuracy. Although there is a relationship between spectrum and train class (the class of a segment predicted its cluster with 56% accuracy) each cluster had trains from each of the different classes, so there all commonalities in the spectra that transcend train class.

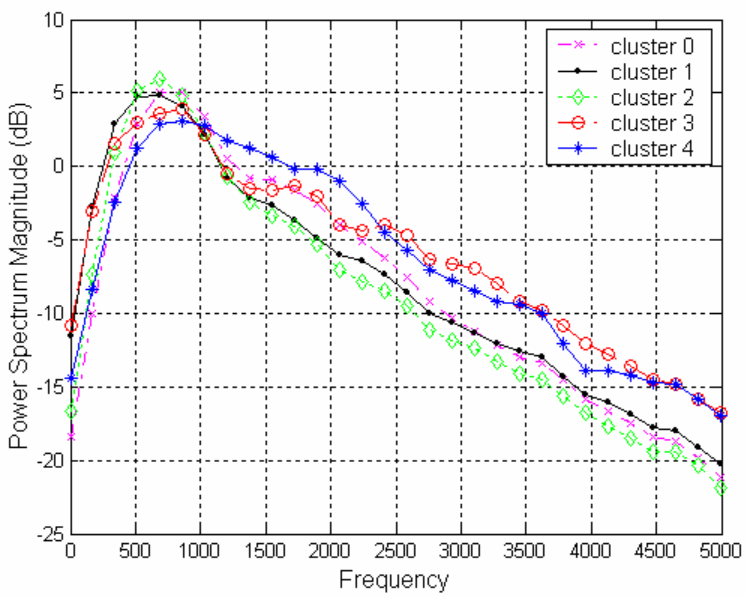

Figure 5. Mean spectra for each cluster

SUMMARY

The analyses in this paper have pointed the way to easily derived and understandable descriptors of train sounds which are independent of subjective variability and are not as easily affected by different recording conditions as the traditional measures of Loudness and Duration are. One MPEG-7 feature in particular, AudioSpectrumCentroid, was found to distribute normally among the train segments used in this database and was highly predictive of train class (e.g. fast train, passenger train or freight train). The peaks in the modulation spectrum correspond to particular

features of the trains such as the individual cars, engine noise and flat wheels. Finally, a clustering based on the spectra of the segments revealed some basic timbral classes for the train sounds which seem to include several complex features of the sounds. Taken together, these descriptors can enable an automatic noise assessment system to make sophisticated judgments about the nature of a particular train sound and its similarity to other known examples of train sounds.

REFERENCES

[1] Fields, J.M., "Railway Noise Annoyance in Residential Areas: Current Findings and Suggestions For Future Research". J. Sound Vib., **51**, 343-351 (1977).

[2] Foote, J., "Content-based retrieval of music and audio". Multimed. Storage Archiv. Syst., **2**, 138-147 (1997).

[3] Green, D., M. and S. Fidell, "Variability in the criterion for reporting annoyance in community noise surveys". J. Acoust. Soc. Am., **89**(1), 234-243 (1991).

[4] Gygi, B., G.R. Kidd, and C.S. Watson, "Spectral-temporal factors in the identification of environmental sounds". J. Acoust. Soc. Am., **115**(3), 1252-65 (2004).

[5] Houtgast, T. and H.J.M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria". J. Acoust. Soc. Am., **77**(3), 1069-1077 (1985).

[6] ISO, *MPEG-7: Multimedia Content Description Interface, Part 4: Audio*. 2001, ISO.

[7] Job, R.F.S., "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction". J. Acoust. Soc. Am., **83**(3), 991-1001 (1988).

[8] Kim, H.-G., N. Moreau, and T. Sikora, "Audio Classification Based on MPEG-7 Spectral Basis Representations". IEEE Transactions On Circuits And Systems For Video Technology, **14**(5), 716-725 (2004).

[9] Knall, V. and R. Schümer, "The Differing Annoyance Levels of Rail and Road Traffic Noise". J. Sound Vib., **87**(2), 321-326 (1983).

[10] Lakatos, S., P.C. Cook, and G.P. Scavone, "Selective attention to the parameters of a physically informed sonic model". J. Acoust. Soc. Am., **107**(5, Pt.1), L31-L36 (2000).

[11] Miedema, H.M.E. and H. Vos, "Demographic and attitudinal factors that modify annoyance from transportation noise". J. Acoust. Soc. Am., **105**(6), 3336-3344 (1999).

[12] Schomer, P.D., "On a theoretical interpretation of the prevalence rate of noise-induced annoyance in residential populations---High-amplitude impulse-noise environments". J. Acoust. Soc. Am., **86**(2), 835-836 (1989).

[13] Schreckenberg, D., et al., "An interdisciplinary study on railway and road traffic noise: Annoyance differences". J. Acoust. Soc. Am., **105**(2), 1219 (1999).

[14] Schultz, T.J., "Synthesis of social surveys on noise annoyance". J. Acoust. Soc. Am., **64**(2), 377-405 (1978).