inter.noise HAMBURG 2016

Analysis and automatic detection of anomalous noise events in real recordings of road traffic noise for the LIFE DYNAMAP project

Joan Claudi SOCORÓ; Xavier ALBIOL; Xavier SEVILLANO; Francesc ALÍAS

[1] GTM – Grup de recerca en Tecnologies Mèdia, La Salle – Universitat Ramon Llull

C/Quatre Camins, 30. 08022 Barcelona, Spain

e-mail: jclaudi@salleurl.edu/si21536@salleurl.edu/xavis@salleurl.edu/falias@salleurl.edu

ABSTRACT

The LIFE DYNAMAP project envisions the development of an anomalous noise event detection algorithm that aims at excluding non-road traffic noise sources from the sound pressure levels represented on the dynamic noise maps. With the aim of adapting the algorithm to the real acoustic environment of the project pilot areas, a recording campaign was conducted. As a result, multiple samples of background noise and road traffic noise together with anomalous noise events under different weather conditions were obtained. In this work, we present an in-depth analysis of the recorded database, including a study of the distribution of all the types of collected anomalous noise events, which are also analyzed in terms of their duration and signal-to-noise ratio. Taking into account the conclusions drawn from this real world data analysis, we implement a supervised anomalous noise event detection algorithm, and evaluate its performance in one of the project pilot areas, testing not only its ability to detect anomalous noise events, but also its sensitivity in terms of the acoustic salience of the events with respect the surrounding traffic noise.

Keywords: noise detection, automatic classification, road traffic noise.

1. INTRODUCTION

Traffic noise is one of the main pollutants in urban and suburban areas, which affects the quality of life of their citizens. For instance, the continued exposure to high traffic noise levels has been found to cause harmful health effects, being highly correlated with cardiovascular diseases (1). As cities grow in size and population, the consequent increase in traffic is making this problem more present and evident. In order to address this issue, European authorities have driven several initiatives to study, prevent and reduce the effects of exposure of population to traffic noise. Among them, the European Noise Directive (END 2002/49/EC) is focused on the creation of noise level maps in order to inform citizens about their exposure to noise, besides drawing up appropriate action plans to reduce its negative impact (2). In general terms, these maps are updated every 5 years. On the one hand, this entails a time and cost consuming process that is undertaken by local and regional bodies of government, and on the other hand, the resulting action plans are only evaluated in five-year periods.

As a means to start addressing the aforementioned issues, the DYNAMAP project (Dynamic Acoustic Mapping – Development of low cost sensors networks for real time noise mapping, LIFE13 ENV/IT/001254, see (3)) is aimed at developing a dynamic noise mapping system able to detect and represent the acoustic impact of road infrastructures in real time. This way, the project will develop an approach that will reduce the cost of periodically updating noise maps, as required by the END. To that end, an automatic monitoring system, based on low-cost acoustic sensors and several algorithms to derive reliable acoustic maps, is being designed.

In order to validate the DYNAMAP's project approach, the system will be deployed in two demonstrative pilot areas in the cities of Milan and Rome (Italy). The first one will be located inside the city of Milan, thus allowing to test the system in an urban scenario, while the second one will be located along a major road surrounding the city of Rome (the A90 highway), making it possible to validate the performance of the system in a suburban environment. The specific selection and location of the network has been defined following the process described in (4). Taking into account several environmental and infrastructural factors (e.g. noise levels, population density, number of dwellings, etc.), the candidate areas were ranked based on scores dependent on these factors, selecting district 9 of Milan as the urban pilot area (5), together with a total of 17 critical areas located along the A90 highway in Rome for the suburban setting (4).

As a consequence of automating the road traffic noise mapping, the DYNAMAP system will inevitably have to deal with acoustic events produced by non- traffic sources that could alter the traffic noise levels (e.g. an air-craft flying over, nearby industries or railways, road works, church bells, animals, etc.). In order to increase the road traffic noise mapping robustness, the DYNAMAP system includes an anomalous noise event detection (ANED) algorithm designed to avoid biasing the traffic noise map computation with non road traffic acoustic events. Those events should be detected and eliminated from the noise map computation to provide a reliable picture of the actual road traffic impact.

A first attempt to develop an ANED algorithm was introduced in (6). In that work, a supervised and a semi-supervised machine learning approaches were compared on a database consisting of real road traffic noise (RTN) recordings of the ring road surrounding the city of Barcelona, synthetically mixed with anomalous noise events (ANE) samples extracted from freely available audio databases. Moreover, in order to test the system in different scenarios, synthetic mixtures between road traffic noise and anomalous events were created with two different RTN-to-ANE (i.e., signal-to-noise) ratios: -6 and -12 dB.

After this proof of concept, and given the diversity of operating scenarios (i.e. urban and suburban), it was necessary to build an acoustic database that could faithfully reflect the characteristics of road traffic noise in real conditions. For this reason, an environmental noise recording campaign was conducted on the two DYNAMAP project pilot areas (see (7) for a detailed description). As a result of the recording campaign, nearly 10 hours of audio were collected, labeled and processed to train acoustic models for subsequent development stages of the ANED algorithm.

The rest of this paper is organized as follows. In section 2, a brief review of the recording campaign and the audio database generation is provided, and a new methodology for automatic SNR ANE labeling is described in order to incorporate perceptual noise equivalent levels in the computation as well as to alleviate the burden of a manual labeling. An in-depth analysis of the recordings gathered in the Rome pilot area is presented in section 3, in which we study the ANE distributions regarding the predominant types of anomalous events, as wells as their durations and signal-to-noise ratios. In section 4 the first results of the ANED trained with the real recordings performed in the Rome pilot area are presented, with the aim of comparing the obtained results with the previous experiments with synthetic mixtures of ANE and RTN that were reported in (6). The choice of audio data from Rome for this study is for ease of comparison, basically due to the similarity of the context that was analyzed in (6). Finally, the concluding remarks and future work are described in section 5.

2. AUDIO DATABASE

In this section, an overview of the audio database constructed from the recording campaign is firstly presented. Secondly, a new methodology to enrich ANE samples with a signal-to-noise ratio (SNR) measure that accounts for the saliency of events with respect the background or road traffic noise is described.

2.1 RECORDING CAMPAIGN AND AUDIO DATABASE GENERATION

An environmental noise recording campaign was performed in May 2015 in the two pilot areas in Italy selected for the LIFE+ DYNAMAP project. Specifically, the recordings were conducted in 6 sites along the A90 highway surrounding the city of Rome (in ANAS S.p.A. portals), and in 12 roads within the district 9 in the city of Milan. The main goal of the campaign was collecting enough representative acoustic data to train, validate and test the ANED algorithm in real conditions. The total amount of recordings was 10 hours of audio, and subsequent labeling and post-processing led to 7 hours, 48 minutes and 38 seconds of road traffic noise samples (labeled as RTN), 38 minutes and 37 seconds of background noise (e.g., quiet noise in a one-way street when no vehicles are present, but some distant traffic noise is perceived, and labeled as BCK), and 25 minutes and 54 seconds of anomalous events (labeled as ANE). The rest of the recorded audio samples were labeled as complex audio passages, since it was not sufficiently clear which was the predominant category from a perceptual point of view.

As described in (7), ANE were labeled by using different subcategories, taking into account the diversity of acoustic phenomena gathered during the environmental recording campaign. These subcategories were defined in order to enrich the description of the occurred acoustic events, and were defined using the following labels (in descendent order of occurrence during the recording sessions): peop (people talking), musi (music in car or in the street), sire (sirens of ambulances, police, etc.), tram (stop, start and pass-by of tramways or trains), stru (noise of portals structure derived from its vibration, typically caused by the passing-by of very large trucks), horn (horn vehicles noise), brak (noise of brake or cars' trimming belt), thun (thunder storm), trck (noise when trucks or vehicles with heavy load passed over a bump), door (noise of house or vehicle doors, or other object blows), bird (birdsong), airp (airplanes), wind (noise of wind, or movement of the leaves of trees), bike (noise of bikes), mega (noise of people reporting by the public address station), busd (opening bus or tramway door noise), chai (noise of chains), and dog (barking of dogs). Further details regarding the recording campaign (e.g., characteristics of the acoustic sensor devices, applied recording methodologies, specific recording locations, etc.) and the post-processing of audio recordings (e.g., audio normalization, labeling and exporting processes) can be found in (7).

It is important to highlight that, besides labeling ANE in terms of their type, another kind of labeling is required to evaluate the ability of the ANED algorithm to detect anomalous events of different intensities. For this reason, we took the labeling process one step further, labeling the ANE in terms of their saliency with respect to the surrounding noise.

In the following section, we describe the approach followed to label the gathered ANE samples in terms of SNR. The main differences of the proposed approach with respect the labeling process explained in (7) are: i) the inclusion of L_{eq} computation based on an A-weighting preprocessing; and ii) an automatic procedure based on selection of specific time regions and averaging for systematize the computation of the SNR attributed to a ANE sample.

2.2 NEW METHODOLOGY FOR SNR ANE LABELING

In (7), ANE samples were manually tagged by computing the relative sound pressure level of road traffic noise or background noise with respect to the ANE level in dBs. However, it is worth noting that in this work, we define the signal to noise ratio or SNR the other way round. That is, it is defined as the relation between ANE and road traffic (or background noise level) in dBs as this definition allows for a better interpretation of the ANE saliency level with respect to the background or road traffic noise levels. This additional information is considered very useful for the subsequent learning stage of the proposed noise classification schemes. The second difference with the approach introduced in (7) is the inclusion of a perceptual-based measure of saliency to consider frequency human sensitivity. Finally, the third difference is moving the approach from a manual to an automatic paradigm, which yields to a dramatic reduction of supervision effort.

Specifically, the SNR is estimated using an A-weighted L_{eq} computed with the free Matlab "Continuous Sound and Vibration Analysis" toolbox developed by Edward L. Zechman (10), using a 30 ms integration time. An automatic methodology for computing the contextual SNR is subsequently applied to the obtained L_{eq} profiles of the audio signal obtained for each recording session, which is summarized next. The main goal of this process is obtaining two equivalent noise levels in order to compute what we call the *contextual SNR* for each anomalous event present in the recorded masters: a median L_{eq} level within the ANE region and a median L_{eq} level within the surrounding but closer region to the labeled ANE. Once these two equivalent noise levels are obtained the computation of the contextual SNR is straightforward.

Contextual SNR is the proposed solution for obtaining estimations of ANE and road traffic or background noise levels in the time region where the anomalous event occurs. Obviously, this approach has its own limitations. Road traffic noise has a strong a non-stationary behavior and then approximating its level during the ANE time interval using samples from its surroundings is a naïve approach. In addition, during the ANE time interval, the measured L_{eq} is also influenced by the background noise (this is especially true for little salient ANE, i.e. those with low SNR). However, assuming that the estimation of SNR using the aforementioned levels has a limited accuracy we think that the obtained values can be useful enough for the subsequent learning stages of any classification scheme that aims at detecting ANE with the required reliability.

To compute the median L_{eq} level within the closest surroundings of the analyzed ANE, two measurements are made: one before the start of the anomalous event (referred to as *left measurement*), and another after the event has finished (called *right measurement*). When possible, the sum of the lengths of the intervals over which these two measurements are made should equal the duration of the anomalous noise event.

For illustration purposes, let us define T_L as the duration of the closest background or road traffic noise region before the beginning of the anomalous event. Analogously, we define as T_R the duration of the closest background or road traffic noise region after the end of the anomalous event. Let's also define T_1 and T_2 as the time durations of the two background or road traffic noise regions considered to compute the corresponding median L_{eq} (T_1 for the background or traffic noise before the ANE start and T_2 for the background or traffic noise after its end). In all the defined time periods the L_{eq} integration time of 30 ms has been subtracted in order to obtain time regions not affected by transients. Following the previous definitions, it is clear that $T_L \geq T_1$ and $T_R \geq T_2$. The general aim of the proposed approach is obtaining two background or road traffic noise measurement regions such that $T_1 + T_2$ is equal to the anomalous event total duration (T_{ANE}). This way, the obtained averaged L_{eq} measure is computed upon similar conditions. When this condition cannot be accomplished then only the available samples of background and/or road traffic noise are used.

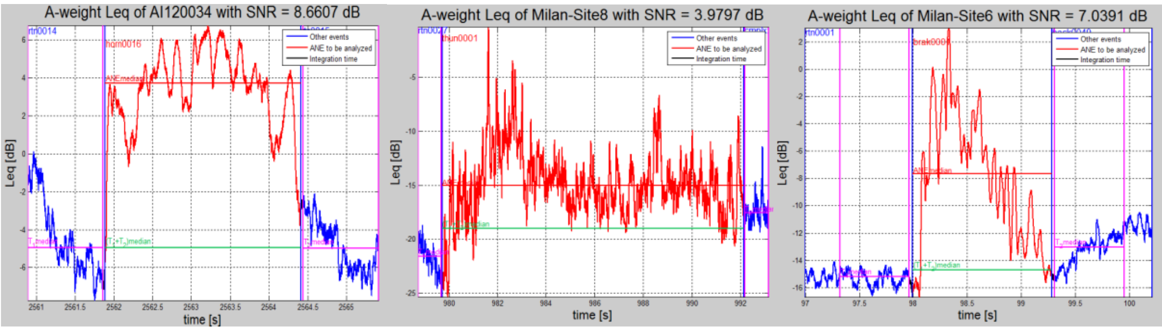

Figure 1 –Examples of anomalous events SNR labeling. From left to right: horn (measured along the A30 motorway in Rome, and with SNR = 8.66 dB), thunder (measured in a Milan road, and with SNR = 3.98 dB) and sound of a brake (also measured in the Milan city, and with SNR = 7.04 dB). The following color palette is used: in red the ANE L_{eq} region and the computed median as a horizontal line, surrounding background or road traffic noise regions are highlighted in blue and its median L_{eq} levels for each side are in magenta, and finally the median L_{eq} of surrounding background or road traffic noise considering both sides is depicted as a green horizontal line within the ANE time region. The SNR is computed as the differences between the median L_{eq} of ANE and RTN. X axis correspond to time in seconds referenced to the start of the recording.

Two case studies were taken into account:

- **An anomalous noise event is surrounded by road traffic or background noise**. This represents the majority of cases in the recorded databases. Within this case study four possibilities exist: i) when $T_R \geq T_{ane}/2$ and $T_L \geq T_{ane}/2$ then $T_1 = T_2 = T_{ane}/2$ (there are available a half of the ANE duration samples of background or road traffic noise in both sides of the event); ii) when $T_R \geq T_{ane}/2$ and $T_L < T_{ane}/2$ then $T_1 = T_L$ and $T_2 = \max(T_{ane} - T_1, T_R)$ (less samples of background or road traffic noise are available before the anomalous event start than after its end); iii) when $T_R < T_{ane}/2$ and $T_L \geq T_{ane}/2$ then $T_2 = T_R$ and $T_1 = \max(T_{ane} - T_2, T_L)$ (less samples of background or road traffic noise are available after the anomalous event end than before its start); iv) when $T_R < T_{ane}/2$ and $T_L < T_{ane}/2$ then $T_2 = T_R$ and $T_1 = T_L$ (there are less samples of background or road traffic noise than the half of the noise event duration at both sides).
- **Other noise events occur just before and/or after the analyzed anomalous noise event**. In this less frequent scenario, the selection of the time regions where the background or the road traffic noise level is computed is a little trickier. We search for the closest time regions to the current anomalous noise event following a global idea of measuring the contextual SNR with a proximity criterion, and trying to obtain as many samples of background or road traffic noise as the samples contained in the anomalous noise event duration (T_{ane}). In this case, firstly we analyze the background or road traffic noise region that is closer to the analyzed noise event (e.g. let us suppose that this is the road traffic noise time region that is closer but *before* the analyzed noise event start). When this region contains a time interval greater than T_{ANE} in which all the samples are closer to the anomalous noise event than any sample of the opposite side (e.g. *after* the analyzed noise end), or any sample within this region is closest to the analyzed noise event than any sample of the opposite region, then the interval of duration $\min(T_{ANE}, T)$ closest to the analyzed event is selected within this

region (being T the duration of this time region). Otherwise, when it is possible to obtain samples of background and/or road traffic noise from both sides of the anomalous noise event with the general criterion that none of these two time regions are strictly closer than the other, i.e. they contain samples equally distant from the closer ANE sample, then samples from both sides are used to compute the road traffic noise or background noise level.

In figure 1 three examples of the most predominant study case (anomalous noise event surrounded by road traffic or background noise) are shown. The L_{eq} curve is highlighted in different color depending on if the time region is attributed to an anomalous noise event (in red) or to background or road traffic noise (in blue). A period of time equal to the integration time for the L_{eq} computation is located in black at both sides of the anomalous event (hard to see in these examples), in order to avoid transients affect the SNR measurements. The median L_{eq} levels for each time region are shown as magenta horizontal lines (for background and road traffic noise at both sides) and horizontal red lines (for the anomalous event).

3. ANALYSIS OF THE RECORDED DATABASE

The main purpose of this study is analyzing the distribution and durations of the ANE subcategories and their SNR distribution in a real case scenario. To that end, an in-depth analysis of the Rome recordings of our database was performed. These recordings contain 4 hours, 34 minutes and 55 seconds of RTN, and 6 minutes and 51 seconds of ANE, which shows the occasional nature of ANE (i.e., it only represents the 2.5% of the total recorded audio).

3.1.1 ANE distributions

This first part of the study aims at determining which are the predominant types of ANE in the Rome recordings. To that end, the distribution of ANEs has been analyzed by computing the total duration for each ANE subcategory within the recorded database. In figure 2 the distributions of the sum of ANE durations are depicted. As it can be observed, sirens and sound of portals structures (stru) followed by the noise of vehicle horns, people, trucks and car brakes are the most observed subcategories of anomalous events, being the rest significantly less probable. As the list of subcategories were defined in (7) to account for the ANEs recorded both in Rome and in Milan, it can be also observed from figure 2 that some ANE subcategories do not occur in Rome. For instance, no samples of airplanes, bikes, birds, chains, dogs, mega, thunder, tramways or wind were recorded during the Rome recordings. Finally, it is worth noting that some types of anomalous events were collected unexpectedly. For instance, we recorded highway operators talking while doing maintenance works (i.e., peop ANE subcategory), which is a somewhat rare event to collect in the surroundings of a highway.

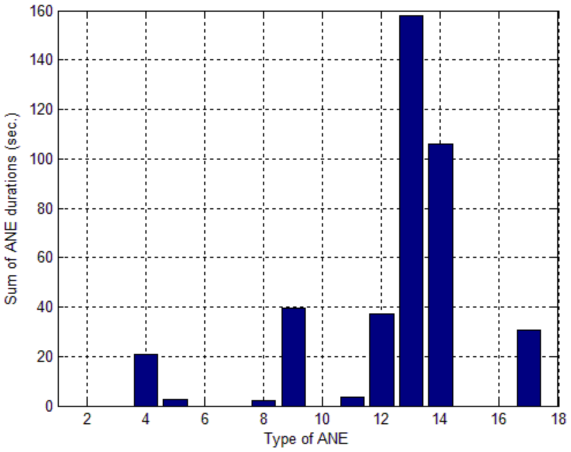

Figure 2 –Sum of total ANE durations for each type of ANE for the Rome locations. The X axis show the type of ANE following the next correspondence: 1 (airp), 2 (bike), 3 (bird), 4 (brak), 5 (busd), 6 (chai), 7 (dog), 8 (door), 9 (horn)10 (mega), 11 (musi),12 (peop), 13 (sire), 14 (stru), 15(thun),16 (tram), 17 (trck), 18 (wind). Left: global (two locations).

The recording campaign was long enough to observe the main possible types of anomalous noise events but not all of them (which, by definition are any of those that are different from the road traffic noise). For instance, in the context of a highway like the one studied in the Rome ring it could be difficult that a birdsong distorted the traffic noise during audio measurements, but there is also the possibility that a bird approaches to the sensor and its noise exceeds the background traffic noise. Of course, other type of anomalous noise events which can eventually be more intense than the background traffic noise could be observed during longer recording campaigns (e.g. airplanes, not reported during Rome recordings).

3.1.2 ANE durations

In this section, we study the durations of the observed ANE subcategories in Rome. In figure 3 the boxplots of ANE durations are shown for each type of ANE subcategory (X axis, following the same numeration as in figure 2). As it can be observed, the sounds of sirens constitute the longest ANE, while the shortest ones are very short and impulsive-like noises (which were assigned to “door” label). However, also brake noises, people, sounds of trucks and noise of portal’s structure have significant durations.

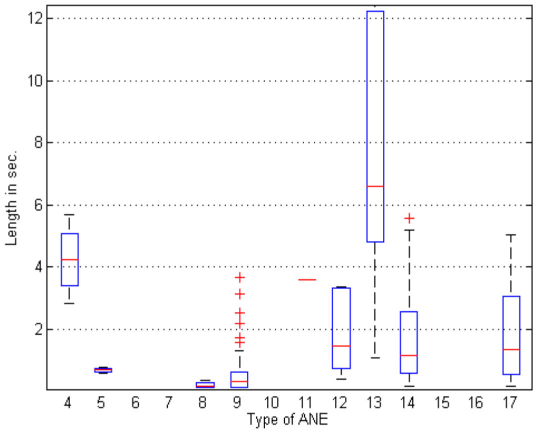

Figure 3 –Boxplots of the durations of each ANE type in Rome (see figure 2 for X axis details).

3.1.3 SNR distributions

In this section, the analysis of the contextual SNR (see section 2.2) is described with the aim of obtaining a more accurate description of the saliency of anomalous noise events with respect to the surrounding road traffic noise in real scenarios. The boxplots of the measured SNR for each type of ANE collected in Rome are depicted in figure 4. As it can be observed, the median SNR values are located in the range between 0 and 5 dB. Hence, anomalous events observed in Rome show a lower saliency than the synthetic mixtures studied in (6). The traffic density during the recordings in Rome was generally high or very high, which made difficult to obtain audio passages where other types of noises surpassed significantly the background traffic noise. The mean value of SNR for the observed anomalous events is 1.5 dB.

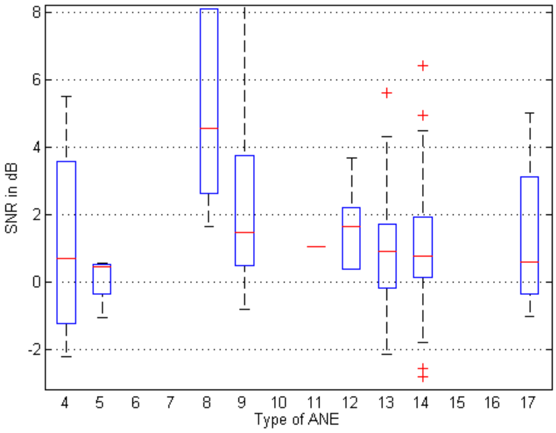

Figure 4 –Boxplots of contextual SNR for each ANE type in Rome (see figure 2 for X axis details).

4. ANED TRAINING AND RESULTS WITH REAL DATA OF ONE PILOT AREA

The ANED algorithm is designed to detect the presence of noise events other than road traffic noise following a "detection-by-classification" approach (6), which performs classification of sequential audio segments based on signal windowing plus a feature extraction module followed by a binary classification scheme. To that effect, both road traffic noise and background city noise are considered to belong to the same RTN class for classifier learning purposes. Thus, the system has to decide if an input audio segment belongs to the ANE or the RTN category.

It is important to note that the ANED algorithm is asked to detect non-traffic noise events that could potentially distort the mapped noise levels in a broad range of scenarios, ranging from detecting ANEs in quiet situations – e.g., in single way isolated roads with very low traffic conditions – to highly noisy environments – e.g., in a very dense highway.

Following the same approach described in (6), in this work we consider two machine learning strategies: i) a supervised approach, where two acoustic models are trained from the labeled database (one for RTN class and another for the ANE class), and ii) a semi-supervised approach, where only one acoustic model is built (the one that represents the RTN class) although a binary decision threshold is optimized based on examples from both acoustic noise classes (the reader is referred to (6) for further details about both approaches). On the one hand, this strategy was proposed to alleviate the difficulty of obtaining a representative collection of anomalous audio events, taking into account that they can be highly local, occasional, and with a very diverse and unpredictable nature. On the other hand, road traffic noise was supposed to present more stable patterns, which in turn it could be modeled with datasets gathered from reasonably short recording campaigns.

4.1 Audio signal parameterization and classification

For comparison purposes, we parameterize the signals from the Rome pilot area with the same two types of audio features considered in (6): the biologically-inspired Gammatone Cesptral Coefficients (GTCC) (8), and the Mel-Frequency Cepstral Coefficients (MFCC). In both cases, 13 coefficients have been extracted to parameterize the audio input from a Hanning window of 30 ms with 50% of overlap.

Moreover, in (6) two machine learning techniques were evaluated as core techniques for the ANED algorithm implementation: K-Nearest Neighbor (KNN) and Fisher's Linear Discriminant (FLD). The choice of these two classification techniques was motivated by the fact that they both provide a certain distance measure that can be interpreted as a measure of similarity (or threshold) between the dominant class (i.e., RTN) and the input noise frame. At classification runtime, this distance measure is compared with the optimized threshold to provide the final binary decision in the semi-supervised approach. In this work, only FLD has been evaluated using the complete audio database collected from the Rome pilot area. On the one hand, FLD showed the best performance in (6), and on the other hand, the KNN demands for huge memory resources on real life data, making running simulations on a standard PC almost unfeasible. This is mainly because KNN requires managing all the database at classification runtime.

As explained in (6), the measure used for classifying audio frames in the semi-supervised approach with the FLD classifier corresponds to an estimation of the log probability that road traffic noise is the source of the input analyzed signal frame (thus, a value close to 0 shows high similarity to this class while high negative values show that the input could be classified as an anomalous event). The optimal threshold used in the semi-supervised approach was based on obtaining an equally minimum value of both type I and type II errors (false positives and false negatives) (9).

The classifier is trained within a binary approach using always the following two labels: ANE for those frames that belong to any of the aforementioned anomalous events subcategories; and RTN for frames that belong to road traffic noise. In the Rome pilot area no background city noise (BCK) was measured because only recordings during the day with dense or highly dense traffic conditions were collected within the recording campaign. However, in future studies concerning also recordings from the city of Milan, background city noise will be attributed to the RTN label during the classifier learning stage.

4.2 Results and discussion

The evaluation process is performed following a 4-fold cross validation scheme following (6). In each repetition, training + validation and test subsets are changed so as to obtain statistically reliable results. As regards the supervised version of the ANED algorithm, training plus validation data (75% of the total available data) contains both classes (RTN and ANE). In contrast, in the semi-supervised version, training data (37.5% of the total available data) contains only RTN class, while the validation set used for the threshold optimization (37.5% of the data) contains both RTN and ANE samples. All the assessments are obtained at a frame level, i.e. every 30 ms the test data is assigned to a specific class label (ANE or RTN) by each version of the ANED classifier, and the evaluation measures are computed with respect to the manual labels obtained from the labeling process (i.e., the so called ground truth).

With the aim of providing the classifier with the proper diversity of training data, the ANE frames are randomly selected and distributed into the 4-cross validation scheme, assuring that all type of ANEs are present in each fold for the learning+validation and the test partitions in a similar proportion. This process is particularly important to guarantee a fair ANED evaluation in this work, considering the unbalanced distribution of ANE and RTN within the collected database in Rome (see section 3) with respect to the balanced nature of the artificially generated database in (6).

In figure 5 the results obtained with FLD classifier using the two machine learning strategies (supervised and semi-supervised) and the two types of audio parameterizations (GTCC and MFCC) are shown. For comparison purposes to (6) the same two types of evaluation measures are considered: i) the global accuracy, which accounts for the number of correct classifications with respect to the total evaluations; and ii) the macro-averaged F1 measure, which is defined as the harmonic mean of the macro-averaged recall and precision (both values are computed as a mean of the recall and precision of both categories – ANE and RTN). Contrary to the work presented in (6), the F1 measure is a macro-averaged version across the two categories, which suits better for unbalanced datasets as the one is being evaluated in this work. In (6) the F1 measure value of the ANE class was computed, which was appropriate in the context of balanced datasets.

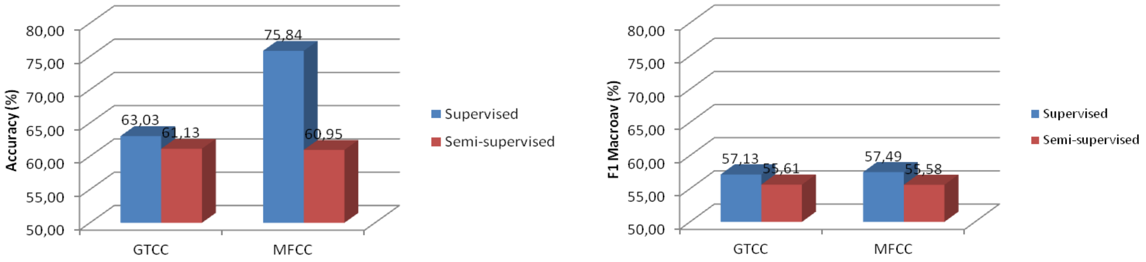

Figure 5 – Results of ANED algorithm considering the Rome pilot area real recordings and with the FLD classifier. Global accuracy is depicted at left while at right the macro-averaged F1 measure is shown, both in %.

As it can be observed from figure 5, the supervised version of the ANED algorithm shows the best results both in terms of accuracy and F1 macro-averaged, as opposite to what was observed in (6) when considering synthetic mixtures of ANE and RTN with 6 and 12 SNRs (ANE-to-RTN). Additionally, the averaged results show better performances when using MFCC than GTCC, which also draws a different picture that the results obtained with synthetic mixtures.

In particular, it can be observed that the accuracies obtained with MFCC are higher than the ones obtained with GTCC for the supervised approach (75.84% with MFCC compared to 63.03% with GTCC). However, this improvement is significantly reduced when comparing both F1 macro-averaged values. This can be explained because the recall of RTN improves in 21% when using MFCC instead of GTCC, while the ANE recall is reduced in 20%. Moreover, the ANE precision improves in 25% when using MFCC rather than GTCC, while the RTN precision remains nearly constant. As RTN is the dominant class (i.e., it represents the 97.5% of the total number of samples), the improvement of RTN recall is significantly biasing the computed accuracy. In contrast, as the F1 macro-averaged measure is designed to better represent the response of a classifier when dealing with significantly unbalanced classes shows a different behavior. Specifically, no clear preference between both types of audio features is observed, since the differences in F1 macro-averaged values are lower than 0.5% (i.e., no statistically significant values are obtained).

Therefore, while the accuracy is a standard performance measure for evaluating classification schemes when train and test datasets are balanced (there are similar proportion of samples for each class), the F1 macro-averaged is best suited for evaluating classification schemes on unbalanced datasets as the one at hand, that is, the Rome audio database, where the proportion between ANE and RTN constitutes a highly unbalanced dataset (2.5% vs. 97.5%, respectively).

Additionally, it is worth noting that the obtained F1 values are significantly lower to the ones obtained with the artificial database (6) (around 25% lower in average), besides being far from the desired values for a competent classifier. This shows the complexity of real life audio data and it makes necessary to develop further studies to address the observed challenges behind this type of data, such as the unbalanced nature of the database and the high diversity of SNR values.

For instance, if the experiments conducted in (6) are repeated by setting the SNR to the mean value observed in the Rome database, which is 1.5 dB (see section 3.1.3), the obtained performance of the classifiers decreases a 15% in terms of accuracy in comparison with results reported with 6 dB artificially fixed SNR. However, in these artificially mixed experiments the datasets are better balanced in terms of class labels, and then both F1 and accuracy measures achieve similar values (with a mean value of approximately 70% across different audio parameters and learning strategies).

5. CONCLUSIONS AND FUTURE WORK

This paper constitutes a step forward from our initial research towards developing the ANED algorithm for the DYNAMAP project by considering an audio database collected in a real life scenario. From our initial work based on artificially mixing ANE data with real RTN, the labeling of audio samples has been improved in regard the automatic computation of the saliency of anomalous events with respect to the surrounding background or road traffic noise (SNR) besides attaining for perceptual L_{eq} measures. As a result, we improve the reliability of the training and the subsequent evaluation processes of any classification scheme that can be adopted to develop the ANED algorithm. Moreover, an in-depth study of real life audio data gathered in the recording campaign of the DYNAMAP project has been presented for the Rome pilot area. The conducted analysis of the recorded database has been performed considering ANE distributions, ANE durations and SNR distributions, showing a dramatically different scenario as the one defined in (6).

Moreover, we have analyzed the results obtained by the two versions of the ANED algorithm (supervised and semi-supervised) when training the FLD classifier with real audio data. The high diversity of the observed ANE SNR levels in the real environment draws a new picture of the addressed problem and poses an even more challenging scenario as the one envisioned in (6). The results reported in this paper with real recordings are significantly lower than the ones presented in (6), since, for instance, extreme SNR=+12 dB values have not been observed in Rome. Moreover, due to dramatically unbalanced nature of the gathered database (ANE only represents the 2.5% of the data), the conclusions drawn from the previous work are not longer applicable if working with the whole database. As a consequence, we have observed that: i) the KNN approach is not able to run on standard PC resources, which makes its implementation unfeasible on the remote sensors designed for the project, ii) the semi-supervised version of the ANED algorithm, which was able to improve some of the results with the supervised technique in a synthetically controlled experiment, is not yet capable to outperform the supervised classifier in a more realistic scenario, iii) the classifier tends to better recognize the major class while the minority class (i.e., ANE class) tends to be disregarded, which is the opposite way the ANED algorithm is asked to perform, and iv) it dramatically affects the performance measures that are designed for evaluating balanced problems such as the accuracy. As an example, if the classifier would only respond with the RTN label the accuracy would be around 97%, while in reality it would not working properly as any anomalous event would not be detected. As a means to deal with this issue, in this work we have considered the F1 macro-averaged measure in order to better evaluate the classification performance considering the relative importance of both classes.

In future works, we plan to study different approaches to address the unbalanced nature of the problem at hand, as it dramatically affects the performance of both ANED algorithm approaches, yielding significantly different conclusions from the ones obtained in our previous work. Moreover, we will keep working on studying the ANE database by including the Milan pilot area data in order to achieve more general conclusions as the ones derived from the analysis of Rome pilot area with respect to the artificially added ANE samples to the RTN recorded in Barcelona ring.

ACKNOWLEDGEMENTS

The research presented in this work has been partially supported by the LIFE DYNAMAP project (LIFE13 ENV/IT/001254) and the Secretaria d'Universitats i Recerca del Departament d'Economia i

Coneixement (Generalitat de Catalunya) under grant refs. 2014-SGR-0590 and 2015-URL-Proj-046.

REFERENCES

1. W. Babisch; “Transportation noise and cardiovascular risk”, Noise&Health, 10, pp. 27–33, 2008.
2. EU Directive 2002/49/EC of the European parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise, Official Journal of the European Communities, L189/12, July 2002.
3. Dynamic Acoustic Mapping – Development of low cost sensors networks for real time noise mapping, web: http://www.life-dynamap.eu/es/
4. S. Radaelli, P. Coppi, A. Giovanetti, R. Grecco; “The LIFE DYNAMAP project: automating the process for pilot areas location”, Proc. 22nd International Congress on Sound and Vibration (ICSV22), Florence (Italy), 12-16 July 2015.
5. G. Zambon, R. Benocci, A. Bisceglie; “Development of optimized algorithms for the classification of networks of road stretches into homogeneous clusters in urban areas”, Proc. 22nd International Congress on Sound and Vibration (ICSV22), Florence (Italy), 12-16 July 2015.
6. J.C. Socoró, G. Ribera, X. Sevillano, F. Alías; "Development of an Anomalous Noise Event Detection Algorithm for dynamic road traffic noise mapping", Proc. 22nd International Congress on Sound and Vibration (ICSV22), Florence (Italy), 12-16 July 2015.
7. F. Alías, J.C. Socoró, X. Sevillano, L. Nencini; "Training an Anomalous Noise Event Detection Algorithm for Dynamic Road Traffic Noise Mapping: Environmental Noise Recording Campaign", Proc. TecniAcústica 2015, Valencia (Spain), p. 345-352, 21-23 October 2015.
8. X. Valero, F. Alías; "Gammatone Cepstral Coefficients- Biologically Inspired Features for Non-Speech Audio Classification", IEEE Transactions on Multimedia, vol. 14, n. 6, pp. 1684-1689, 2012.
9. S. Furui; "Cepstral analysis technique for automatic speaker verification", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 29, n. 2, pp. 254-272, 1981.
10. E. L. Zechman; "Matlab Continuous Sound and Vibration Analysis toolbox", http://www.mathworks.com/matlabcentral/fileexchange/21384-continuou