DYNAMAP
DYNAMIC ACOUSTIC MAPPING

inter.noise
HAMBURG 2016

laSalle

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

La Salle – Ramon Llull University

Outline

1. Introduction
2. Audio database
3. Analysis of the recorded database
4. ANED training and results with real data of one pilot area
5. Conclusions and future work
6. Acknowledgements

Outline

1. **Introduction**
2. Audio database
3. Analysis of the recorded database
4. ANED training and results with real data of one pilot area
5. Conclusions and future work
6. Acknowledgements

1 – Introduction

- 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.).
- An Anomalous Noise Event Detection (ANED) algorithm is designed to avoid biasing the traffic noise map computation with non road traffic acoustic events.
- In (6) supervised and a semi-supervised machine learning approaches were compared on a database consisting of real road traffic noise (RTN) recordings (ring road of Barcelona), synthetically mixed with anomalous noise events (ANE) samples extracted from freely available audio databases.

(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.

1 – Introduction

- After the proof of concept in (6), and given the diversity of 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)), collecting nearly 10 hours of audio, which where subsequently labeled and processed.
- The main goal of the campaign was collecting enough representative acoustic data to train, validate and test the ANED algorithm in real conditions. In this work, training and validation is performed using audio data from the Rome pilot area.

(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.

Outline

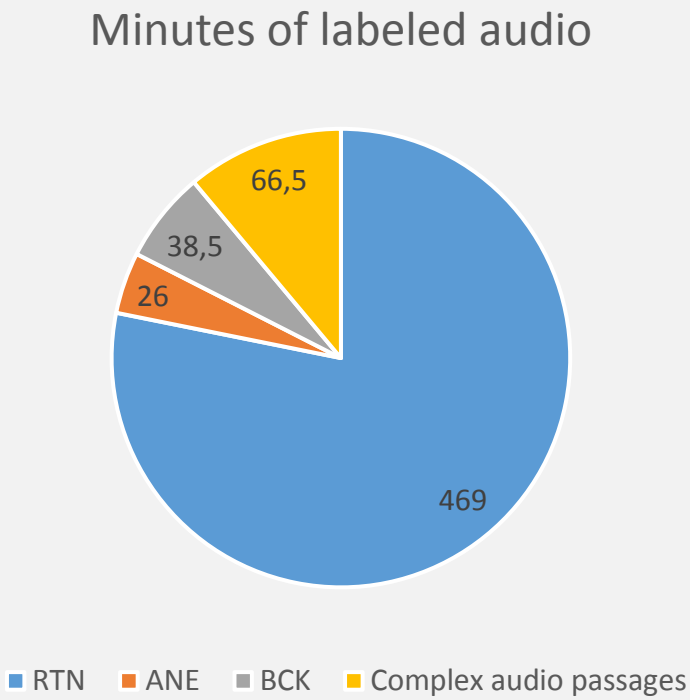
1. Introduction
2. **Audio database**
3. Analysis of the recorded database
4. ANED training and results with real data of one pilot area
5. Conclusions and future work
6. Acknowledgements

2.1 – Recording campaign and audio database generation

Rome

Milan

- Recordings during May 2015 in the two pilot areas in Italy selected for the DYNAMAP project:
  - 6 sites along the A90 highway surrounding the city of Rome (in ANAS S.p.A. portals)
  - 12 roads within the district 9 in the city of Milan
- An exhaustive labeling of all the recorded audio was performed obtaining audio clips of: RTN(road traffic noise), ANE(anomalous events), BCK (background city noise), CMPLX (complex audio passages).
- Labeled audio clips constitute the core for the ANED training, and its performance assessment for the DYNAMAP project.
  - From the perspective of the audio classifier to be designed, the audio database is a highly unbalanced dataset (5% is ANE while 95% belongs to RTN or BCK).



2.1 – Recording campaign and audio database generation

- 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:

1-airp	2-bike	3-bird	4-brak	5-busd	6-chai	7-dog	8-door	9-horn
airplanes	noise of bikes	birdsong	noise of brake or cars' trimming belt	opening bus or tramway door noise	noise of chains	barking of dogs	noise of house or vehicle doors, or other object blows	horn vehicles noise
10-mega	**11-musi**	**12-peop**	**13-sire**	**14-stru**	**15-thun**	**16-tram**	**17-trck**	**18-wind**
noise of people reporting by the public address station	music in car or in the street	People talking	sirens of ambulances, police, etc.	noise of portals structure derived from its vibration, typically caused by the passing-by of very large trucks	thunder storm	stop, start and pass-by of tramways or trains	noise when trucks or vehicles with heavy load passed over a bump	noise of wind, or movement of the leaves of trees

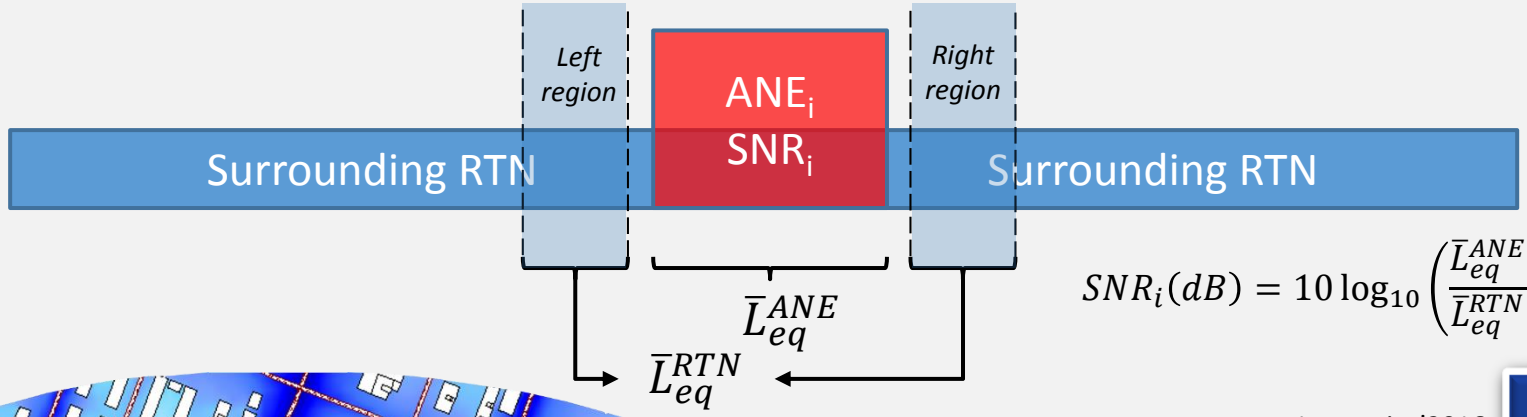
2.2 – New methodology for ANE SNR labelling

- ANE samples were labelled according to their saliency with respect the background traffic or city noise in order to evaluate the ability of the ANED algorithm to detect anomalous events of different intensities.
- The main differences of the proposed approach with respect the labeling process explained in (7) are:
  - $SNR(dB) = 10\log_{10}\left(\frac{\bar{L}_{eq}^{ANE}}{\bar{L}_{eq}^{RTN}}\right)$
  - The inclusion of L_{eq} computation based on an A-weighting preprocessing (including frequency human sensitivity)
  - An *automatic procedure* based on selection of specific time regions and averaging for systematize the computation of the SNR attributed to a ANE sample.

(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.

2.2 – New methodology for ANE SNR labelling

- 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 (T_a).
- An *automatic procedure* is designed to set the time region where the surrounding $\bar{L}_{eq}^{RTN}$ is computed in order to stablish the contextual SNR of each ANE.
- L_{eq} of the RTN during the ANE is approximated by an (naive) estimation based on the closest surrounding RTN regions (*left* and *right*).

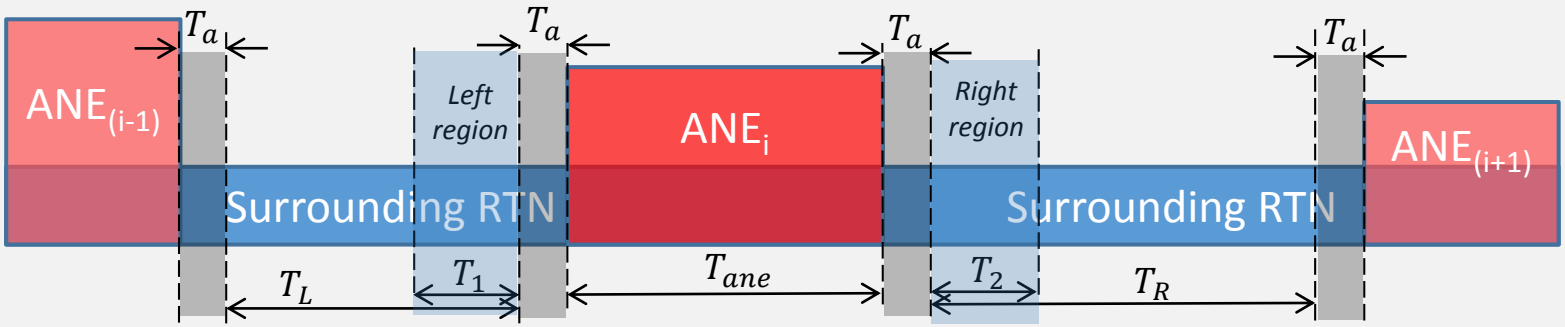


$$SNR_i(dB) = 10\log_{10}\left(\frac{\bar{L}_{eq}^{ANE}}{\bar{L}_{eq}^{RTN}}\right)$$

2.2 – New methodology for ANE SNR labelling

- **Example 1:** An anomalous noise event is surrounded by road traffic or background noise

T_a: integration time (30ms)

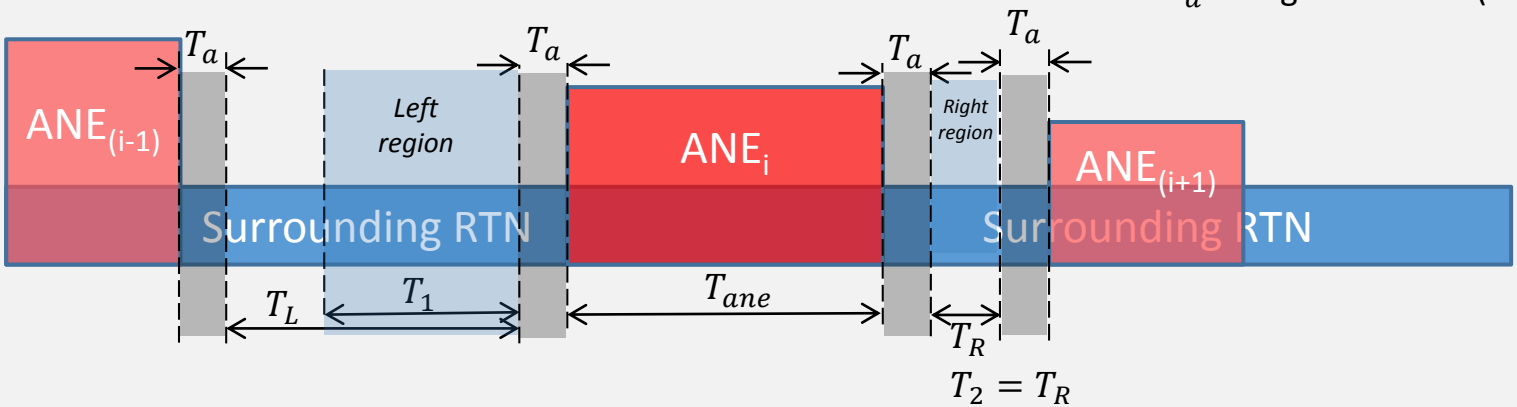


- **General idea**: We search for the closest time regions to the current anomalous noise event (ANE_i) measuring the contextual SNR with a proximity criterion, and trying to obtain as many samples of background or road traffic noise as the samples contained in the anomalous noise event duration ($T_1 + T_2 = T_{ane}$).
- Integration time (used for the L_{eq}) is not considered to avoid transients due to ANE start and end, obtaining more accurate estimations.

2.2 – New methodology for ANE SNR labelling

- **Example 2:** An anomalous noise event is surrounded by road traffic or background noise

T_a: integration time (30ms)

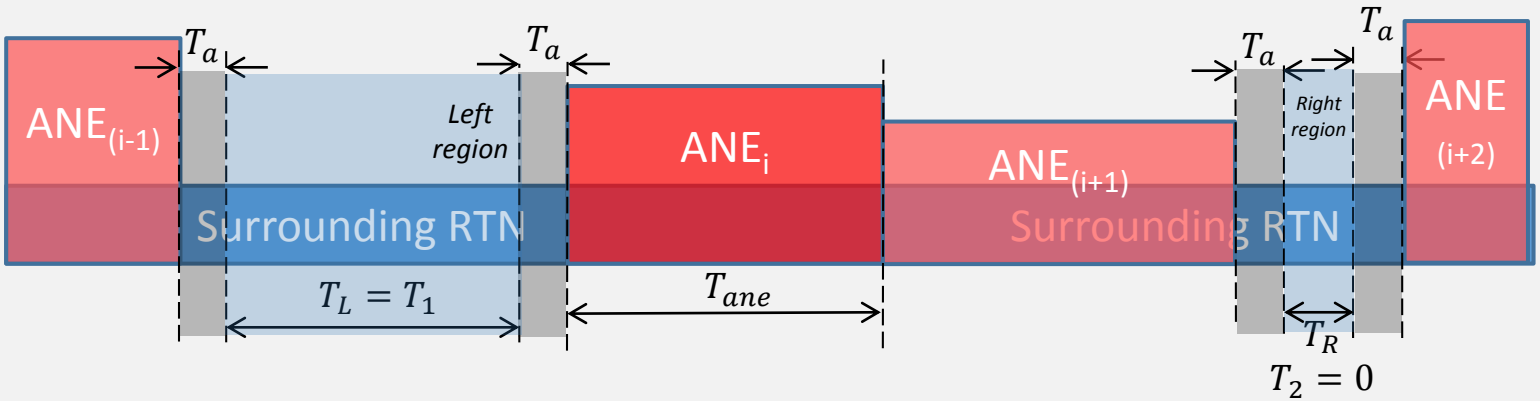


- Contrary to the previous example, the *right region* doesn't contains enough samples to obtain a balanced time region of surrounding RTN, and more samples from the *left region* are here considered for the SNR computation.

2.2 – New methodology for ANE SNR labelling

- **Example 3:** Other noise events occur just before and/or after the analyzed anomalous noise event

T_a: integration time (30ms)



- In this less frequent case, all the samples considered for the SNR computation come from the *left region*, because the RTN samples in the *right region* are further away than the farthest sample of the RTN *left region* (then, $T_2 = 0$).

2.2 – New methodology for ANE SNR labelling

- **Real cases of SNR labeling:**

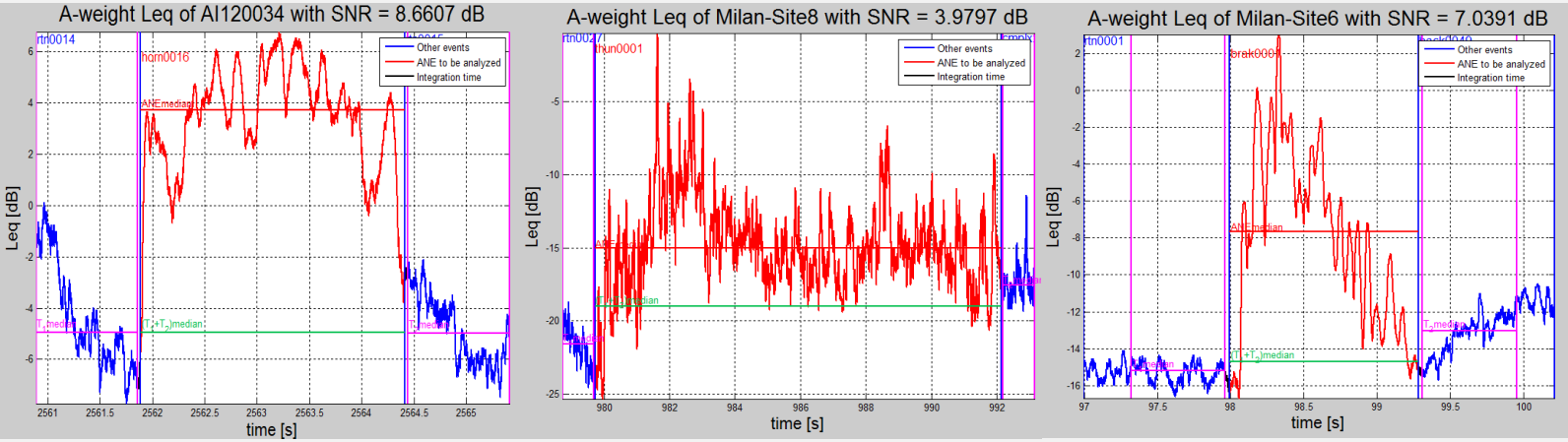


Figure 1 –Examples of anomalous events SNR labeling. From left to right: horn (measured along the A30 motorway in Rome, and with SNR = 8.66 dB), thunder (measured in a Milan road, and with SNR = 3.98 dB) and sound of a brake (also measured in the Milan city, and with SNR = 7.04 dB). The following color 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.

Outline

1. Introduction
2. Audio database
3. **Analysis of the recorded database**
4. ANED training and results with real data of one pilot area
5. Conclusions and future work
6. Acknowledgements

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 – ANE distributions

- Distribution of ANEs has been analyzed by computing the total duration for each ANE subcategory within the recorded database

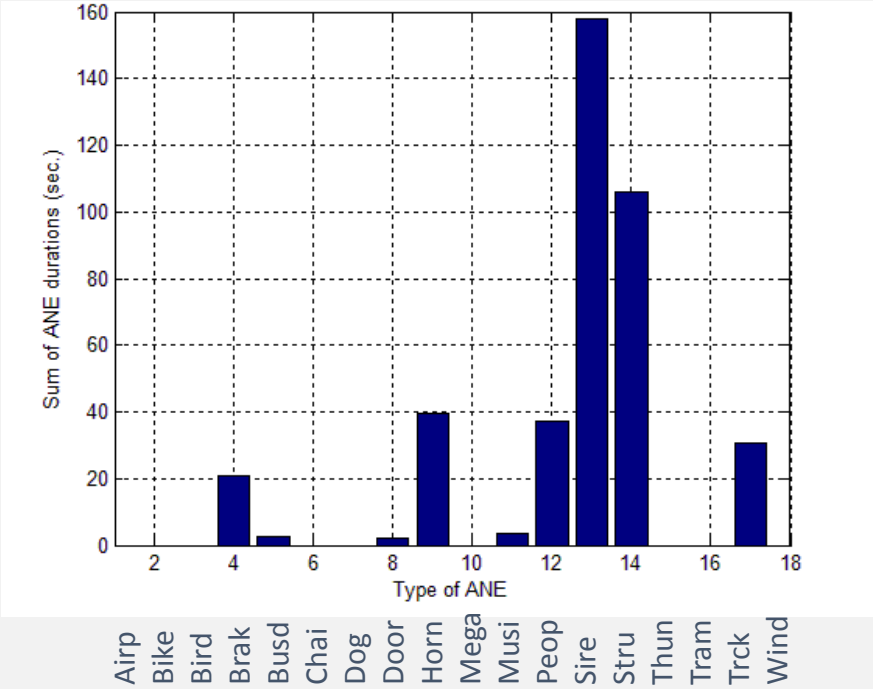


Figure 2 –Sum of total ANE durations for each type of ANE for the Rome locations.

- Sirens, “stru” are more probable followed by horns, “peop”, trucks and car brakes.
- Some ANE observed in Milan (e.g. dogs’ barks) were inexistent in Rome highway.
- Some unlikely ANEs were observed (e.g. peop) while others not (e.g. bird approaching the sensor)

3.2 – ANE durations

- Boxplots of ANE durations are shown for each type of ANE subcategory

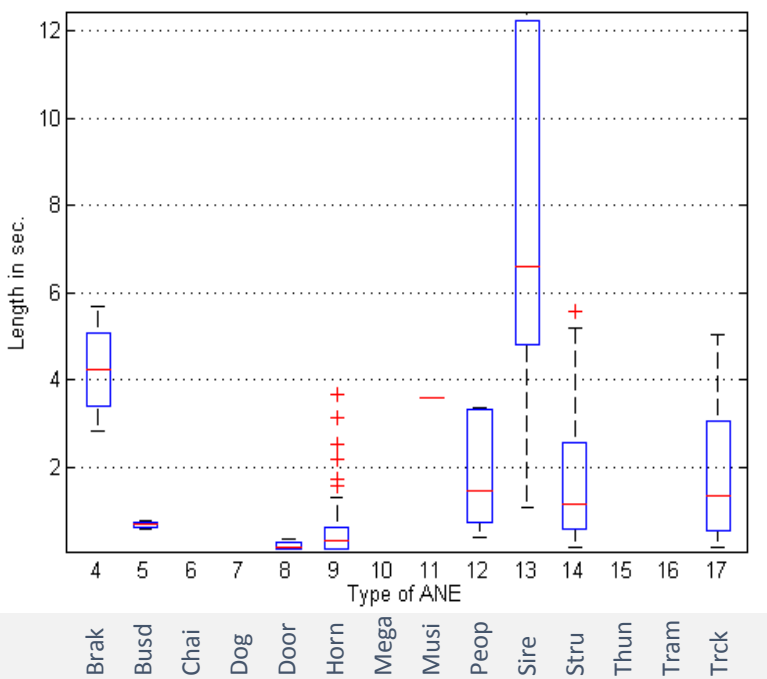


Figure 3 –Boxplots of the durations of each ANE type in Rome.

- Longest ANEs are sirens while shortest ones are “door”-like sounds.
- However, brake noises, people, sounds of trucks and noise of portal’s structure have significant durations

3.3 – SNR distributions

- Distributions accounting for the saliency of ANE in real road traffic scenarios were also studied. The boxplots of the measured SNR for each type of ANE collected in Rome are depicted in figure 4.

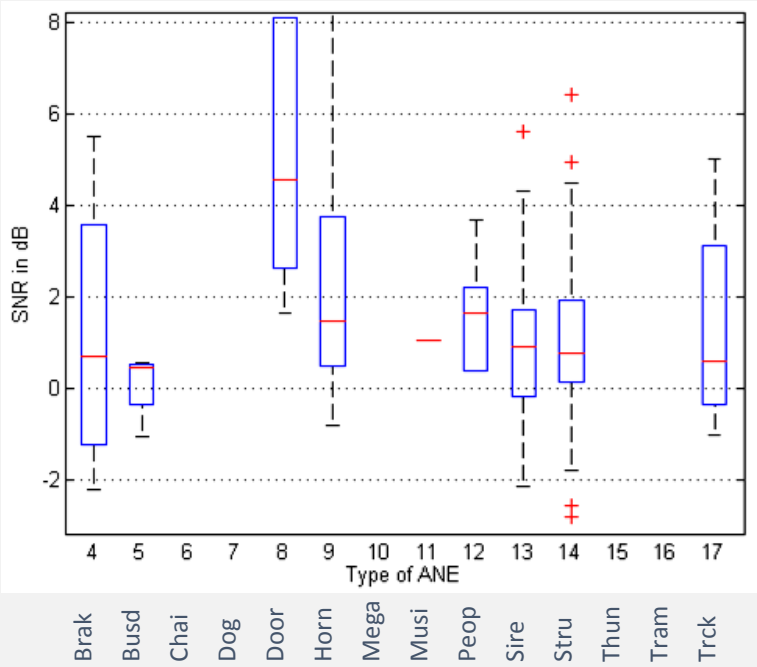


Figure 4 –Boxplots of contextual SNR.

- Median SNR values are located in the range between 0 and 5 dB
- Anomalous events observed in Rome show a lower saliency than the synthetic mixtures studied in (6) (SNR used values were 6 and 12 dB).
- The mean value of SNR for the observed anomalous events is 1.5 dB.

(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.

Outline

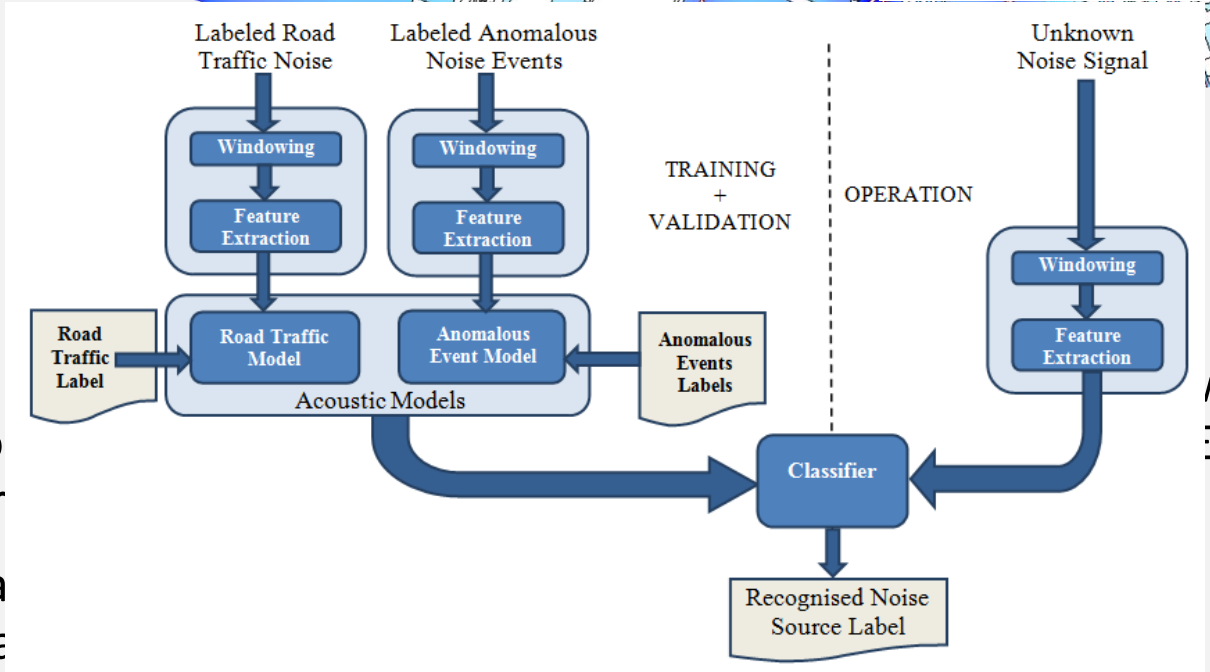
1. Introduction
2. Audio database
3. Analysis of the recorded database
4. **ANED training and results with real data of one pilot area**
5. Conclusions and future work
6. Acknowledgements

4 – ANED training and results with real data of one pilot area

- ANED performs **binary classification** of each audio frame (30 ms window length) in the following two categories: RTN (road traffic noise) and ANE (any non road traffic noise event).
- Following the same approach described in (6), in this work we consider two machine learning strategies:

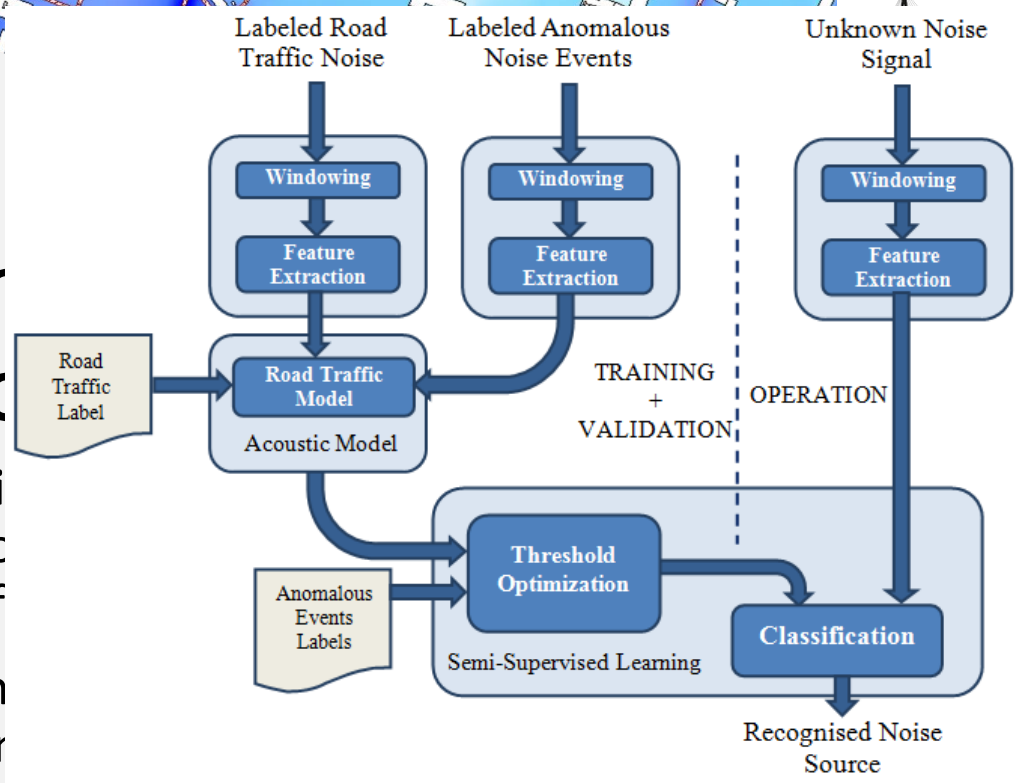
(6) J.C. Socoró, G. Ribera, X. Sevillano, F. Alías; "Development of an Anomalous Noise Event Detection Algorithm for dynamic road traffic noise mapping", Proc. 22nd International Congress on Sound and Vibration (ICSV22), Florence (Italy), 12-16 July 2015.

4 – ANED tra
of one pi



- ANED performs N
  length) in the fo
  (any non road tr
- Following the sa
  two machine lea
  - a **supervised approach**, where two acoustic models are trained from the labeled database (one for RTN class and another for the ANE class)

4 – ANED trair ta of one pilc



- ANED performs **bi** vindow length) in the follc d ANE (any non road traf
- Following the sam nsider two machine learr
  - a **supervised approach**, where two acoustic models are trained from the labeled database (one for RTN class and another for the ANE class)
  - 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 (see (6) for further details about both approaches). This strategy was proposed to alleviate the difficulty of obtaining a representative collection of anomalous audio events (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.

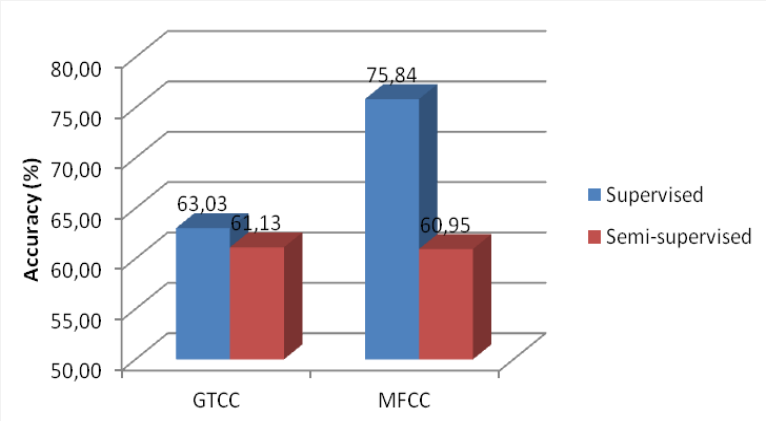
(6) J.C. Socoró, G. Ribera, X. Sevillano, F. Alías; "Development of an Anomalous Noise Event Detection Algorithm for dynamic road traffic noise mapping", Proc. 22nd International Congress on Sound and Vibration (ICSV22), Florence (Italy), 12-16 July 2015.
(8) X. Valero, F. Alías; "Gammatone Cepstral Coefficients- Biologically Inspired Features for Non-Speech Audio Classification", IEEE Transactions on Multimedia, vol. 14, n. 6, pp. 1684-1689, 2012

4.1 – Audio signal parameterization and classification

- Signals coming from the Rome pilot area were parametrized with the same two types of audio features considered in (6):
  - GTCC (8) and MFCC
  - 13 coefficients , Hanning window of 30 ms, 50% of overlap
- Fisher's Linear Discriminant (FLD) was used as classifier, using the log probability that road traffic noise is the source of the input analyzed signal frame as the measure that must be compared to an optimized threshold in the semi-supervised approach.
- Contrary to the work in (6) KNN was not included in this study because KNN demands for huge memory resources on real life data (4.5 hours of parameterized audio), making running simulations on a standard PC almost unfeasible.
- 4-fold cross-validation scheme, ensuring that in each learning+validation and test partitions there are similar distributions of ANE subcategories

4.2 – Results and discussion

- Global accuracy and also F1 macro-averaged measures were computed for each configuration (audio parameters vs. learning approach)
- Contrary to the work in (6), 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.



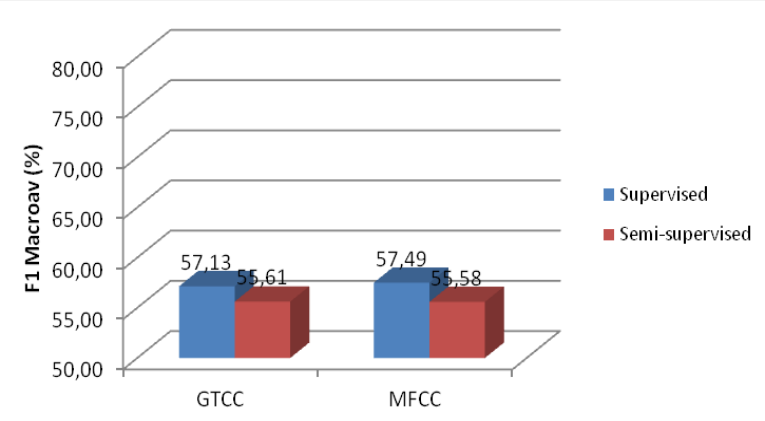


Figure 5 – Results of ANED algorithm considering the Rome pilot area real recordings and with the FLD classifier. Global accuracy is depicted at left while at right the macro-averaged F1 measure is shown, both in %.

4.2 – Results and discussion

- **Supervised version shows better performance than the semi-supervised approach**, as opposite to what was observed in (6) with synthetic mixtures of RTN and ANE with 6 and 12 dB of SNR.
- 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 (6)).
- The greater improvement of accuracy when comparing results of MFCC with regard GTCC is due to the fact that the dataset is highly unbalanced (RTN>>ANE), while macro-averaged F1 measure reflects a less significant improvement.
- F1 values are **significantly lower** to the ones obtained with in (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**.

(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.

Outline

1. Introduction
2. Audio database
3. Analysis of the recorded database
4. ANED training and results with real data of one pilot area
5. **Conclusions and future work**
6. Acknowledgements

5 – Conclusions and future work

- Improvement of labeling of audio recordings with an **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.
- Audio database analysis of one pilot area (Rome) reflects the **diversity of real life recordings** (ANE subcategories distributions, ANE durations and ANE saliencies – SNR –).
- ANE SNR distributions of highway Rome recordings show a significantly different scenario, and **results of ANED trained with real data are consistently far from the ones reported in (6) with synthetic mixtures** of RTN and ANE:
  - Mean ANE SNR in real recordings was 1.5 dB, and median values for each ANE subcategory were between 0 and 5 dB, while it was fixed to 6 and 12 dB in (6).

(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.

5 – Conclusions and future work

- Dramatically unbalanced nature of the gathered database (ANE only represents the 2.5% of the data):
  - **KNN unfeasible** to work due to too high demand of memory and computational resources
  - 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**
  - 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 this dramatically affects the performance measures that are designed for evaluating balanced problems such as the accuracy (→use of macro-averaged F1 measure).
- Future work:
  - address the unbalanced nature of the problem, and improve recognition performance
  - studying the ANE database by including the Milan pilot area

(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.

Outline

1. Introduction
2. Audio database
3. Analysis of the recorded database
4. ANED training and results with real data of one pilot area
5. Conclusions and future work
6. **Acknowledgements**

6 – 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.