

Dynamic noise map generation using machine learning

Campos, R. R. E¹; Simões, A. B. A. M. R²; Petraglia, M. R.¹ and Torres, J. C. B.¹

¹ Programa de Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil,
{roberto,mariane}@pads.ufrj.br, julio@poli.ufrj.br

² Escola Politécnica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil, anamrsimoes@poli.ufrj.br

Abstract

Urban traffic is one of the most important sources of noise in cities. Noise metrics like LAeq and DNL (day-night level), are based on road traffic characteristics over long averaged periods and do not describe the traffic dynamics in these periods. This paper presents a method to produce dynamic noise maps from flow traffic data obtained by surveillance cameras. The method is implemented in a python application, where the videos are processed by machine learning algorithms. A retrained convolutional neural network is used to identify vehicles in the scene and classify them according to their size, while a tracking algorithm estimates their positions, speeds and accelerations. Such data, extracted from videos, are useful for calculating the sound pressure level at a given location by a standard model. The estimated energy noise level of all vehicles present in the scene are added to produce dynamic noise maps, representing the variation of sound pressure over time, according to the traffic flow. The levels obtained from the data extracted from the video are compared with measurements at control points to validate the model. The results of the proposed model are presented, as well as the its constraints to produce the noise maps.

Keywords: Traffic noise, vehicle information, convolutional neural network .

PACS: please see the instructions in this template.

1. INTRODUCTION

Road traffic noise is a problem in most large cities. It brings several issues for the human health, causing disturbs in the day life, which may vary from annoyance to heart diseases [1–3]. The actual way to prevent high noise exposure is to monitor levels at specific city locations and/or to produce noise maps. Noise mapping is a well established technique to evaluate sound pressure levels (SPL) over the territory. However, it is practically impossible to conduct measurements along a whole city, where a sensor grid of 50×50 m, for instance, would be required to provide a detailed map. Alternatively, simulation is used to generate such maps [4–7], based on traffic data and morphological city characteristics, such as road trajectory, width, slope, building facade reflections and pavement type. Data for noise prediction consists of statistical traffic flow, including number of vehicles per hour in a road, average speed, vehicle category (mo-

torcycle, cars, minivans, buses or heavy trucks) and some street and buildings characteristics, according to the model described in [8]. Even using accurate traffic data and detailed geometry models, such maps present acoustical noise metrics or indexes, which represent averaged energy over long periods of time, such as Equivalent Noise Level L_{eq} , Day-Night Level (L_{DN} or DNL) or their variants (L_D, L_N, L_{DEN}). Therefore, the noise maps present a “static” and averaged view of the noise energy behavior along the time frame analysis, usually hours or day periods.

This work presents an alternative to standard static noise mapping, where a vehicle noise emission prediction model is fed with traffic data extracted from video surveillance cameras. First, a calibration technique is used to map the image to real-world coordinates. Then, machine learning algorithms are used to identify the location of vehicles in each video frame and classify them into some categories. The corresponding class



and average speed data are then used to calculate the noise emission of the vehicle, whose position changes over the video frames, allowing the determination of noise maps.

2. PROPOSED SYSTEM

The proposed system is composed of two parts, the first one related to computer vision (CV) and the second one to a vehicular noise emission model. The CV part is responsible of extracting all the parameters required by the noise emission model, while the other is in charge of generating the corresponding noise levels and their propagation in order to generate the maps. CV is further divided into specific sub-stages. Figure 1 depicts the main stages of the proposed method.

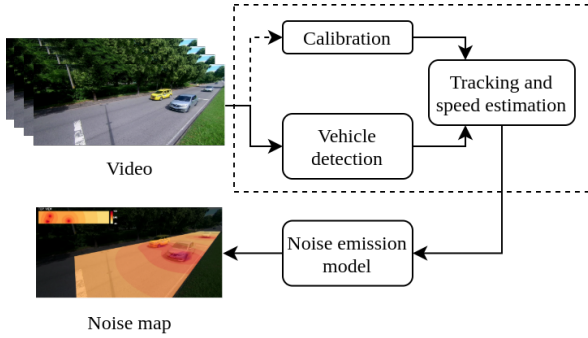


Figure 1: The proposed system flow. Computer vision stages within the dashed rectangle. Arrows indicate data flow, dashed arrow indicates that the data will only be processed once.

2.1 Camera calibration

The main purpose of camera calibration in our application is to be able to measure distances, between two arbitrary points, on the road plane in real-world coordinates. Calibration is achieved by computing three vanishing points (VP) which define the vehicle motion [9] and assuming that the observed road segment is straight. The first VP corresponds to the lines on the road which are parallel to the traffic flow. The second VP is perpendicular to the first VP and parallel to the road plane. The third VP corresponds to the lines perpendicular to the road plane. VP_1 and VP_2 are calculated as the intersection points of manually extracted straight lines from the road plane, these lines are illustrated on Figure 2. Finally, considering that the camera principal point is coincident with the image center, VP_3 and cam-

era parameters can be calculated [9].

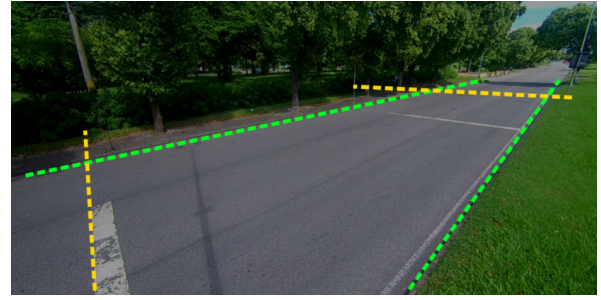


Figure 2: Calibration lines, VP_1 is generated from the intersection of green lines and VP_2 from the intersection of the yellow lines. VP_3 (normal to the plane road) is calculated from VP_1 and VP_2 . Therefore, it is not necessary to annotate them manually.

2.2 Vehicle detection

To be able to detect vehicles on the images, a Mask R-CNN network [10] was used. This network was trained on COCO dataset [11] and is able to detect up to 80 object classes, eight of which are vehicles (bicycle, car, motorcycle, bus, truck, train, boat and airplane). For each detected object, a bounding box is provided along with the object class and its binary mask.

2.3 Tracking and speed estimation

Detection is carried out on a frame-by-frame basis and then a tracking stage is added to provide a temporal relationship between them, allowing to extract the trajectory of each vehicle and the subsequent speed estimation. We use the SORT tracking algorithm [12], which is updated every frame by assigning new detections to the tracks. This algorithm has a small memory that allows the continuation of the trajectory when no vehicles are detected in a certain number of consecutive frames, thus preventing the separation of one trajectory into several. Then, the vehicle class corresponding to a completed trajectory is assigned as the class with repetitions greater than fifty percent of the number of detections contained by the trajectory, since sometimes the detector may assign the wrong class to a vehicle. When this limit is not reached, the entire trajectory is discarded as there is no certainty of the type of vehicle detected. Also, trajectories with small numbers of detections are discarded.

For speed estimation, we take as reference the

approach described in [9]. A 3D box of the vehicle is aligned with the real scene from tangential lines to the vehicle's contour, obtained from the binary mask of the previous stage, and passing through all VP (please refer to the original paper for detailed explanation). The original algorithm was modified, such that only the 3D bounding box horizontal projection (parallel to the road) is used. Additionally, we set the centroid of the horizontal projection as a vehicle position point to map from image to world coordinates, different from the reference that uses the midpoint of the lower front edge of the 3D box. This allows us greater stability of the reference point against the misalignment produced when computing the masking tangential lines. Next, the vehicle position, in image coordinates, is projected and scaled to world coordinates using the camera calibration parameters and a reference vehicle length/width. Finally, the vehicle speed (v) is estimated as $v = l(t_{0-1})/t_{0-1}$, where $l(t_{0,-})$ is the distance travelled by the vehicle from the beginning (t_0) to the end (t_1) of the tracking.

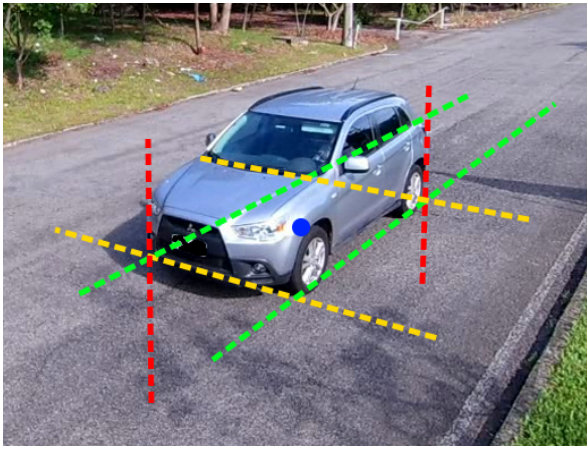


Figure 3: Vehicle horizontal projection plane build from tangent lines between the contour to VP_1 (green), VP_2 (yellow) and VP_3 (red). Centroid in blue.

2.4 Dynamic Noise map

A static noise map based only on traffic flow depends on several parameters such as vehicle category, speed, tyre and pavement types, rain conditions and so on. A noise emission model for a single vehicle was presented in [13], based on engine, gear state, exhaust systems, tires type, among other contributions. The model also considers the emission predominance of some systems over others. For example, that engine noise

is preponderant over tyre noise at low velocity and high rpm, while for a vehicle running at constant high speed, the engine noise is negligible when compared to tyre rolling. Therefore, based on vehicle's category, its acceleration state and speed, its sound power L_W can be estimated using Ianoni's method. Given individual vehicle sound power and considering it as a point source with spherical propagation, the sound pressure level can be calculated at any point (or receiver) over the horizontal grid or in the 3D space, as illustrated in Figure 4. The distance $d(i, j)$ between the vehicle location i and the receiver position j is used to calculate the sound pressure level L_I on each grid element, according to Equation 1, which also considers the vehicle sound power L_W .

$$L_I(i, j) = L_W - 20 \log d(i, j) - 11 \quad (1)$$

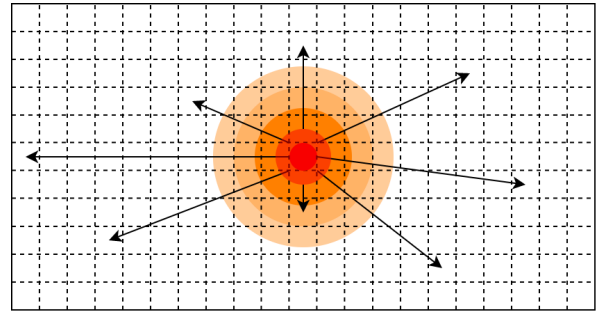


Figure 4: Discretized region. Sound pressure level (red circles) is predicted for every grid position (receiver).

This data is stored in a matrix which represents the distribution in space of noise emission from a single source. If more than one vehicle is present at the scene, individual noise matrices are energetically added to represent the sound level received by each grid element. A static map is then produced for each frame by superposing the sound pressure level contribution of all vehicles present in the scene. A matrix corresponding to the energetic grid values is used to overlay color levels to the original frame or visualize it in different ways. The temporal sequence of all video frames provide us a dynamic noise map. The dynamic map is obtained by creating a video sequence of the static ones.



3. EVALUATION

In order to evaluate the proposed system, measurements in a controlled environment were carried out. It consisted of a car (*light vehicle*) passing in front of a microphone and a video camera emulating a video surveillance system. The camera was mounted on a pole at 2.6m height and is capable of recording video at 1080p and 30 fps. The microphone, previously calibrated with a 94 dB and 1 kHz signal, was located on the edge of the track, at a 1.5m height. Audio was recorded at 48 kHz and SPL was computed in 1/30 second windows, in order to match the video frame rate. The speed reference was obtained from a GPS device located inside the car. A total of 20 pass-by trials were performed with the car at constant speed, with ranges between 30 and 80 km/h in an environment with about 50 dBA background noise.

Author's pre-trained (COCO) Mask-RCNN network uses ResNet-50+FPN backbone with standard convolutional and fully connected heads for mask and box prediction, respectively.

The noise emission model [13] corresponding to the *light vehicle* category at constant speed is expressed by the Equation 2.

$$L_W(v) = \begin{cases} 53.6 + 26.8 \log v & , \text{for } v > 11.5 \\ 82 & , \text{for } v \leq 11.5 \end{cases} \quad (2)$$

Where speed v is expressed in km/h.

4. RESULTS

Video camera recordings were processed to extract vehicle speed and its distance to the microphone location. The mean absolute error (MAE) between the estimated and measured speed is around 1.36 ± 1.30 km/h, which is significantly lower than the MAE obtained using the original algorithm (4.98 ± 1.92 km/h) [9]. Figure 5 shows a comparison between estimated and measured speeds. As can be seen, a good agreement between estimated and reference speed was achieved. Only few cases presented larger deviations, specially around 60 km/h, which might be caused by the vehicle passing on the side of the road, furthest away from the camera, region where the calibration error increases.

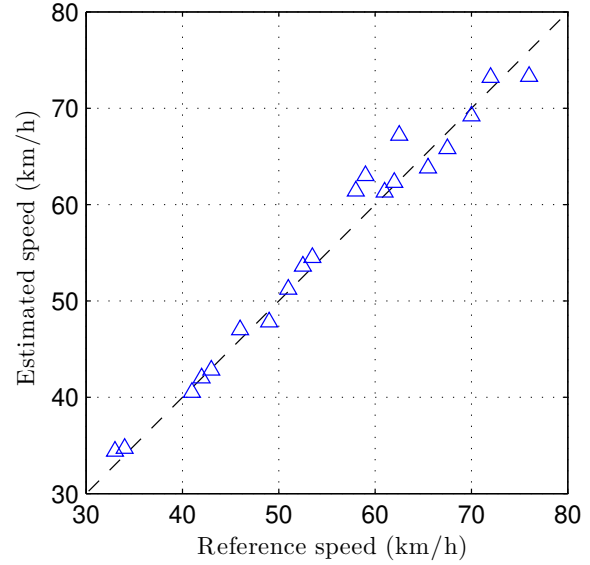


Figure 5: Reference vs estimated speed.

A comparison between estimated and measured SPL was also carried out. Estimated SPL was calculated using the vehicle category, estimated speed and distance when the car passes in front of the virtual receiver location, according to Equations 2 and 1. This SPL was then normalized to a distance of 3.75m from the receiver, which corresponded to the average vehicle pass-by distance. The measured background noise level was added to the predicted SPL in order to match with evaluation scenario. Measured SPL was calculated using a time frame of

Figure 6 shows the approximation of the measured SPL to the predicted SPL curve for all trials. The MAE when the vehicle is in front of the microphone is 1.15 ± 1.02 dBA with larger deviation for speeds between 30 and 50 km/h. The maximum variation reached 2.88 dBA, a value that is accordance with noise mapping acceptable deviation. Additionally, the speed estimation error corresponds to a 0.5 dBA variation in the predicted SPL at the receiver. Variation which was considered not significant to affect the interpretation of the results.

From Figures 5 and 6 and considering the estimated MAEs, the proposed system was considered quantitatively validated to proceed with the dynamic maps elaboration. Figure 7 shows some frames taken from a dynamics noise map video sequence. The top left sub-image corresponds to

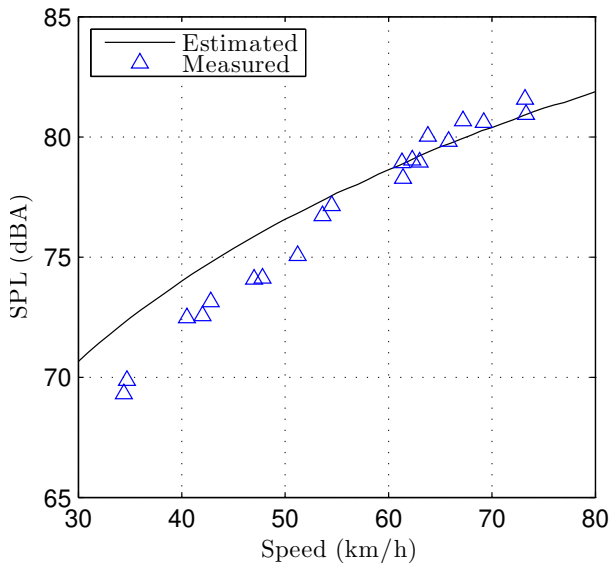


Figure 6: Reference speed vs estimated speed.

the generated map, which is projected onto the road using the calibration parameters calculated at the beginning. The projected and blended map allows us to perceive a coherence between the real and estimated position of the vehicles as well as the SPL distribution.

5. CONCLUSION

In this work, a vehicular noise emission prediction model was proposed as an alternative to standard static noise mapping. The proposed model is based on traffic data extracted from video surveillance cameras. A mean absolute error of 1.15 ± 1.02 dBA was observed between the estimated and measured SPL, validating the noise emission model proposed for the generation of dynamic noise maps. It is worth mentioning that the results presented in this work correspond only to the *light vehicle* category. Future work consists of evaluating real street scenarios with all vehicle categories, to consider acceleration profiles and different camera viewpoints.

6. ACKNOWLEDGEMENTS

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code PROBRAL-2018918853P.

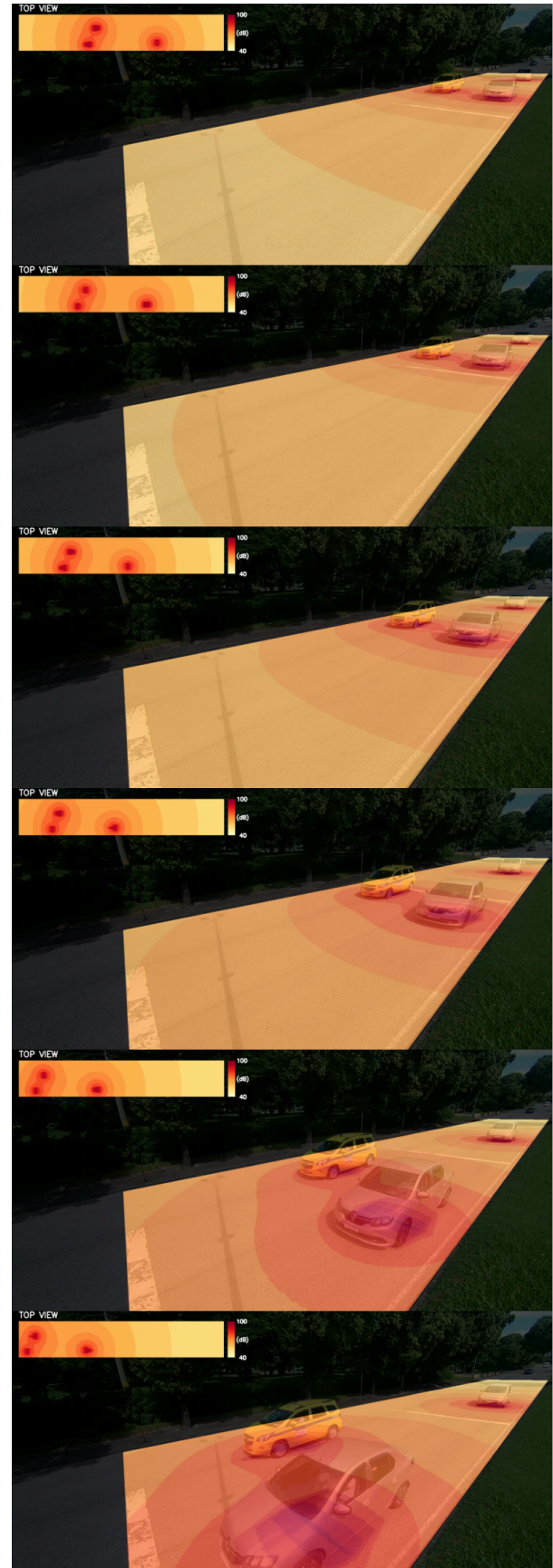


Figure 7: Dynamic noise map for a video sequence.



REFERENCES

- [1] S. Fidell, T. Schultz, and D. M. Green. A theoretical interpretation of the prevalence rate of noise-induced annoyance in residential populations. *The Journal of the Acoustical Society of America*, 84(6):2109–2113, 1988. doi:[10.1121/1.397056](https://doi.org/10.1121/1.397056).
- [2] SA Stansfeld, B Berglund, C Clark, I Lopez Barrio, P Fischer, E Ohrstrom, MM Haines, J Head, S Hygge, I van Kamp, and BF Berry. Aircraft and road traffic noise and children's cognition and health: a cross-national study. *The Lancet*, 365(9475):1942 – 1949, 2005. ISSN 0140-6736. doi:[http://doi.org/10.1016/S0140-6736\(05\)66660-3](https://doi.org/10.1016/S0140-6736(05)66660-3).
- [3] Yvonne de Kluizenaar, Sabine A. Janssen, Frank J. van Lenthe, Henk M. E. Miedema, and Johan P. Mackenbach. Long-term road traffic noise exposure is associated with an increase in morning tiredness. *The Journal of the Acoustical Society of America*, 126(2):626–633, 2009. doi:[10.1121/1.3158834](https://doi.org/10.1121/1.3158834). URL <http://dx.doi.org/10.1121/1.3158834>.
- [4] Shinichi Sakamoto. Road traffic noise prediction model asj rtn-model 2018: Report of the research committee on road traffic noise. *Acoustical Science and Technology*, 41(3):529–589, 2020. doi:[10.1250/ast.41.529](https://doi.org/10.1250/ast.41.529).
- [5] Sh. Givargis and M. Mahmoodi. Converting the uk calculation of road traffic noise (cortn) to a model capable of calculating laeq,1h for the tehran's roads. *Applied Acoustics*, 69(11):1108–1113, 2008. ISSN 0003-682X. doi:<https://doi.org/10.1016/j.apacoust.2007.08.003>.
- [6] Diego Murillo Gomez, Juan Gil Carvajal, Valentina Zapata Rodriguez, and José Téllez García. Assessment of the rls 90 calculation method for predicting road traffic noise in colombian conditions. *Revista Facultad de Ingeniería Universidad de Antioquia*, 1, 04 2015. doi:[10.17533/udea.redin.n75a17](https://doi.org/10.17533/udea.redin.n75a17).
- [7] F Anfosso Ledee, M Paviotti, and S Kephelopoulos. *Common noise assessment methods in Europe (CNOSSOS-EU) : to be used by the EU Member States for strategic noise mapping following adoption as specified in the Environmental Noise Directive 2002/49/EC*. Publications Office, 2012. doi:[doi/10.2788/32029](https://doi.org/10.2788/32029).
- [8] J. Quartieri, Nikos Mastorakis, Gerardo Iannone, Claudio Guarnaccia, Salvatore D'Ambrosio, Antonio Troisi, and T Lenza. A review of traffic noise predictive noise models. pages 72–80, 01 2009.
- [9] Marketa Dubska, Adam Herout, and Jakub Sochor. Automatic camera calibration for traffic understanding. *Proceedings of the British Machine Vision Conference*, 2014.
- [10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017. doi:[10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322).
- [11] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2014. URL <http://arxiv.org/abs/1405.0312>. cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list.
- [12] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468, 2016. doi:[10.1109/ICIP.2016.7533003](https://doi.org/10.1109/ICIP.2016.7533003).
- [13] J. Quartieri, Gerardo Iannone, and Claudio Guarnaccia. On the improvement of statistical traffic noise prediction tools. *Proceedings of the 11th WSEAS International Conference on Acoustics and Music: Theory and Applications, AMTA '10*, 06 2010.