# Caracterització de l'impacte dels esdeveniments acústics en els nivells equivalents sonors i en la percepció dels ciutadans per a la confecció de mapes dinàmics de soroll

**Ferran Orga Vidal**

http://hdl.handle.net/10803/674634

# TESI DOCTORAL

| | |
|---|---|
| Títol | Caracterització de l'impacte dels esdeveniments acústics en els nivells equivalents sonors i en la percepció dels ciutadans per a la confecció de mapes dinàmics de soroll |
| Realitzada per | Ferran Orga i Vidal |
| en el Centre | Facultat Internacional de Comerç i Economia Digital La Salle |
| i en el Departament | d'Enginyeria |
| Dirigida per | Dra. Rosa Maria Alsina i Pagès<br>Dr. Francesc Alías i Pujol |

# Resum

La contaminació acústica ha esdevingut un greu problema de salut pública, provocant diversos tipus de malalties i trastorns en les persones. Segons l'Organització Mundial de la Salut, cada any es perden a l'Europa occidental, un milió d'anys de vida saludables per culpa de l'exposició al soroll ambiental.

Per tal d'avaluar i gestionar el soroll ambiental a la Unió Europea, la directiva END 2002/49/CE requereix als estats membres la preparació i publicació de mapes de soroll actualitzats i els plans d'acció relatius, cada cinc anys. Això inclou aglomeracions de més de 100.000 habitants i les principals carreteres, vies de tren i aeroports. Gràcies als avanços tecnològics recents, el paradigma de creació de mapes de soroll ha canviat substancialment, permetent l'automatització de les mesures dels nivells sonors utilitzant xarxes de sensors acústics sense fils per a la generació de mapes de soroll en temps real. Així i tot, aquestes xarxes no poden prevenir una sèrie de situacions que esbiaixarien la mesura real dels nivells equivalents sonors, ocasionant que el mapa no sigui fidel a la realitat que percep el ciutadà, p. ex., el so de les aus, de la indústria, els clàxons, les sirenes, les converses que ocorren prop dels sensors o fenòmens meteorològics com la pluja i el vent.

Aquesta tesi estudia la caracterització dels esdeveniments acústics per a la confecció de mapes dinàmics de soroll de trànsit. L'estudi comença presentant el context de la tesi, el projecte LIFE DYNAMAP, que pretén mesurar els nivells de soroll de trànsit en dues àrees pilot i integrar-los dinàmicament en un mapa de soroll que s'actualitza a temps real. A continuació, es presenta una anàlisi exhaustiva dels esdeveniments que es troben en les dues àrees, la urbana i la suburbana, i s'hi apliquen diverses caracteritzacions. Una de les mesures que es presenta és la de l'impacte en el nivell equivalent sonor ($L_{eq}$), que permet mesurar el biaix que provoca la presència de certs esdeveniments acústics en la confecció dels mapes de soroll de trànsit. També es planteja l'ús de tests perceptius mitjançant mètriques psicoacústiques per tal d'adaptar la caracterització d'aquests esdeveniments a la percepció ciutadana.

L'objectiu principal de la tesi és caracteritzar els esdeveniments d'entorns urbans i suburbans per oferir mapes de soroll més fidels a la realitat percebuda pel ciutadà en relació amb el paisatge sonor on es troba. I durant la tesi es mostra la importància de la detecció de sons en una xarxa de sensors acústics per tal de prevenir errors de mesura en els nivells equivalents i la necessitat d'entrenar el sistema de detecció amb dades obtingudes en els mateixos sensors de la xarxa.

## Agraïments

Aquesta tesi no hauria sigut possible sense els meus dos directors, que m'han ajudat a orientar la recerca i a donar-li forma de tesi: moltes gràcies, Rosa i Francesc. I evidentment, vull agrair l'ajuda del Joan Claudi, que m'ha ajudat a entendre les entranyes del projecte amb una paciència il·limitada.

Aquesta tesi tampoc hauria sigut possible sense els meus companys del GTM, actuals i passats, amb qui hem compartit esmorzars, dinars, molts cafès i una mica de recerca. La vostra amistat i el bon clima que hi ha hagut al nostre grup ha facilitat enormement la meva tasca a la Salle. També m'agradaria agrair la confiança de la Salle i del departament d'enginyeria, que van confiar en mi des d'un principi.

Vull enviar un agraïment molt especial als meus amics, que fan un esforç considerable per no adormir-se cada vegada que em demanen que els expliqui de què va la tesi. Amb una menció especial als amics de Vallvidrera, sense els quals la meva estada al Barcelonès hagués estat insuportable. I una abraçada al Joan Carles, el tipògraf de capçalera, que ha fet que aquesta tesi sigui una mica més llegible.

Finalment, vull agrair el suport de la meva família, que m'ha acompanyat i ajudat al llarg de la tesi, i des de molt abans. Amb una menció especial als meus pares, germana, Aretha i Nina. I, evidentment, a la meva parella, que m'ha acompanyat en aquest trajecte i en els projectes de vida que hem emprès. Us estimo!

Aquesta tesi ha estat possible gràcies a la Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya i als Fons Socials Europeus per l'ajut de contractació de personal investigador novell FI-DGR (2017 FI_B_00243, 2018 FI_B1_00175 i 2019 FI_B2_00168).

**Ferran Orga**
Santa Coloma de Queralt, Baixa Segarra, primavera del 2022.

# Índex

# Índex de figures



# Índex de taules

# Abreviatures

**ANE** *Anomalous Noise Event* (Esdeveniment de soroll anòmal)

**ANED** *Anomalous Noise Event Detector* (Detector d'esdeveniments de soroll anòmal)

**CNN** *Convolutional Neural Network* (Xarxa neuronal convolucional)

**CNOSSOS-EU** *Common Noise Assessment Methods in Europe* (Mètodes comuns d'avaluació del soroll a Europa)

**CRNN** *Convolutional Recurrent Neural Network* (Xarxa neuronal recurrent convolucional)

**DALY** *Disability Adjusted Life Years* (Anys de vida ajustats per discapacitat)

**DES** Detecció d'Esdeveniments Sonors

**DYNAMAP** *Dynamic Noise Mapping* (Mapat dinàmic de sorolls)

**END** *Environmental Noise Directive* (Directiva del soroll ambiental)

**FNN** *Feed-forward Neural Network* (Xarxa neuronal directa)

**GMM** *Gaussian Mixture Model* (Model de barreja gaussiana)

**GPU** *Graphic Processing Unit* (Unitat de processament gràfic)

$\mathbf{L_{eq}}$ Nivell de so continu equivalent (o *Equivalent continuous sound level* en anglès)

**LSTM** *Long Short-Term Memory* (Memòria a curt termini)

**OMS** Organització Mundial de la Salut

**RNN** *Recurrent Neural Network* (Xarxa neuronal recurrent)

**RTN** *Road Traffic Noise* (Soroll de trànsit rodat)

**SBC** *Single-Board Computer* (Ordinador de butxaca)

**SIG** Sistema d'Informació Geogràfica

**SNR** *Signal-to-Noise Ratio* (Relació senyal-soroll)

**UE** Unió Europea

**WASN** *Wireless Acoustic Sensor Network* (Xarxa de sensors acústics sense fils)

Capítol 1

# Introducció

## 1.1 La contaminació acústica i la seva mesura

La contaminació acústica és un dels problemes ambientals més greus que afecten Europa [1, 2]. A part de la molèstia que genera a les persones, també té efectes adversos en la salut, com els acúfens, o els trastorns en l'aprenentatge i el son [3]. A llarg termini, el soroll pot causar estrès [4] i problemes de concentració [5], afectant greument la societat i fins i tot l'economia [6].

Els estudis alerten sobre l'alta exposició dels ciutadans al soroll ambiental en certs entorns. En alguns casos, arribant a nivells crítics com a la ciutat de Tainan (Taiwan), on un 90% dels habitants estan exposats a nivells inacceptables de soroll [7]: 75 dBA, segons la definició del Departament d'Habitatge i Desenvolupament Urbà dels Estats Units [8]. Un altre cas és Curitiba (Brasil), on el 93% dels habitants estan exposats a un nivell de soroll per sobre dels 65 dBA [9] i a Càceres (Espanya), un 90% de la població pateix un soroll per damunt dels 65 dBA durant les hores de feina [10]. Es poden trobar altres casos arreu del món, com a l'Índia [11] o la Xina [12].

El soroll de trànsit és el principal causant de la contaminació acústica en entorns urbans i suburbans [4, 13], i un dels més estudiats per esbrinar els efectes adversos en les persones. S'han trobat relacions entre l'exposició al soroll de trànsit i l'aparició de malalties greus com la diabetis [14] o les malalties cardiovasculars [15]. L'Organització Mundial de la Salut (OMS) quantifica els efectes negatius de l'exposició al soroll de trànsit en un milió d'anys de vida saludables perduts cada any a l'Europa occidental [16].

Per tal d'avaluar i combatre les principals fonts de soroll, l'any 2002, la Unió Europea (UE) va aprovar la directiva 2002/49/CE del Parlament Europeu i del Consell, sobre avaluació i gestió del soroll ambiental: *Environmental Noise Directive* (END) en el seu títol en anglès [17]. I posteriorment, es publicà un marc de treball que pretén millorar la consistència i comparabilitat dels resultats en l'avaluació del soroll en l'àmbit europeu: els mètodes comuns d'avaluació del soroll a Europa (CNOSSOS-EU, de l'anglès *Common Noise Assessment Methods in Europe*) [18]. Els pilars principals de l'END són: *i)* determinar l'exposició al soroll, *ii)* publicar la informació actualitzada referent al soroll per la ciutadania, i *iii)* prevenir i reduir el soroll ambiental on sigui necessari. Per aquest motiu, la UE exigeix que els estats membres publiquin mapes de soroll i els plans d'acció relatius en aglomeracions de més de 100.000 habitants i principals carreteres, vies de tren i aeroports cada cinc anys [17].

Típicament, les mesures de soroll en entorns urbans es duen a terme per tècnics professionals que enregistren el so i en calculen els nivells equivalents de soroll amb sonòmetres certificats [19]. Aquest procediment és costós si es

pretén aplicar a l'àmbit urbà, sobretot si es vol oferir dades de soroll amb certa freqüència. A més a més, les mostres recollides són d'un lloc concret en un període de temps específic, generant així un mapa estàtic que difícilment pot ser fidel a la realitat canviant dels paisatges sonors.

Però gràcies als avanços tecnològics recents, el paradigma de creació de mapes de soroll ha començat a canviar, permetent l'automatització de les mesures dels nivells acústics utilitzant xarxes de sensors acústics sense fils desplegades a les ciutats. Automatitzar el procés de mesurament obre la porta a un mostreig més elevat en l'espai i en el temps, simplificant la caracterització temporal de les mesures, necessària amb el procediment anterior [20]. També es permet un accés més ràpid per part de la ciutadania a la informació referent al soroll, millorant un dels pilars de l'END.

Cada vegada més projectes de l'àmbit europeu i mundial estan utilitzant aquest nou paradigma de mesurament [21]. A part dels sistemes dissenyats per a mesurar els nivells globals sonors, alguns projectes de monitoratge acústic incorporen la detecció d'esdeveniments acústics per a identificar també la presència de fonts de soroll dins de l'entorn acústic analitzat. El projecte LIFE DYNAMAP [1] pretén monitorar automàticament el soroll de trànsit mitjançant sensors acústics de baix cost i actualitzar els mapes de soroll en temps real [22]. L'acrònim de DYNAMAP prové de *DYNamic Acoustic MAPping*. El sistema es va instal·lar i es testejar a dues àrees pilot diferents a Itàlia, una d'urbana (el districte 9 de Milà) i una de suburbana (la ronda A90 de Roma). Les xarxes de sensors acústics desplegades mesuren el nivell equivalent de so continu ponderat en A ($L_{Aeq}$) a la vegada que detecten si el so capturat és originat pel trànsit (soroll provinent del motor i del soroll de rodament, segons es va acordar en el si del projecte [22]) o no té relació directa amb aquest. D'aquesta manera, el sistema que integra la informació és capaç de discernir si l'$L_{Aeq}$ pertany al soroll de trànsit o no, per tal d'elaborar un mapa acústic que integri solament el soroll provinent d'aquesta font. El sistema actualitza el mapa de so periòdicament amb informació obtinguda de la xarxa de sensors i representa el nivell acústic mitjançant l'índex HARMONICA (informació de soroll harmonitzada pels ciutadans i les autoritats, de l'anglès *HARMOnised Noise Information for Citizens and Authorities*) [23, 24].

En aquesta tesi, s'ha tingut l'oportunitat de treballar amb l'equip tècnic encarregat d'implementar l'algoritme de detecció d'esdeveniments acústics que no són originats pel motor o pel rodament del trànsit (ANED, de l'anglès *Anomalous Noise Event Detector*) i desplegar-lo als entorns reals d'operació: l'urbà i el suburbà. Aquest algoritme s'ha desenvolupat per a detectar si el fragment d'àudio pertany al soroll de trànsit rodat (RTN, de l'anglès *Road Traffic Noise*) o no. La resta d'esdeveniments acústics, com ara sirenes, pluja, trens, tramvies o botzines, s'anomenen en el projecte com a esdeveniments de soroll anòmals (ANE, de l'anglès *Anomalous Noise Event*). S'ha pogut treballar en un projecte real des de la primera fase de la investigació, atorgant una base per a caracteritzar l'impacte dels esdeveniments acústics, en les fases següents.

[1] http://www.life-dynamap.eu/ (visitat: 31/12/2021)

La tesi està dividida en 5 capítols. El capítol 1, la introducció en la qual ens trobem, dona una visió general de la tesi, oferint el context, les preguntes d'investigació i els objectius de la recerca desenvolupada, que van alineats en resoldre les preguntes plantejades. Seguidament, el capítol 2 ofereix l'estudi de l'estat de la qüestió en relació amb els objectius d'aquesta tesi. Primerament, s'analitzen els efectes del soroll en la salut pública; seguidament es tracten les possibilitats de les xarxes de sensors per monitorar el soroll ambiental i, finalment, es descriu l'estat actual dels sistemes de detecció automàtica d'esdeveniments acústics. Després d'això, al capítol 3 es detalla el funcionament de la xarxa de sensors del projecte DYNAMAP, un cas real de monitoratge del soroll de trànsit, que incorpora detecció d'esdeveniments i s'hi explica les contribucions de la tesi en l'ANED. El capítol 4 detalla el procés d'obtenció de les bases de dades i el treball del sistema amb les dades reals. Al capítol 5, s'hi presenta una proposta de mesurament d'esdeveniments acústics que quantifica l'impacte que tenen sobre el nivell de so del trànsit, extraient la contribució de l'esdeveniment en qüestió dels nivells d'$L_{eq}$ mesurat. També s'exploren altres maneres de calcular la molèstia dels esdeveniments acústics a partir d'un estudi perceptiu. Finalment, el capítol 6 recull les conclusions i planteja possibles línies de futur a partir del treball realitzat en aquesta tesi.

## 1.2 Preguntes d'investigació

Tal com s'indica en l'estat de la qüestió, la principal font de soroll en entorns urbans i suburbans és el trànsit. Per aquest motiu, el projecte DYNAMAP ha creat un mapa acústic dels nivells de soroll de trànsit (so de motor i rodament de pneumàtics) que s'actualitza dinàmicament. Ara bé, quan els sensors acústics es col·loquin a l'entorn d'operació real, sembla raonable pensar que captaran altres sons més enllà dels de trànsit, que podrien esbiaixar la mesura dels nivells equivalents de soroll de trànsit. Per això, la primera qüestió que es planteja és la següent:

> P1: *Poden, els esdeveniments acústics no relacionats amb el trànsit, esbiaixar la mesura del nivell equivalent de so en entorns urbans i suburbans? I, doncs, cal tenir-los en compte per confeccionar un mapa acústic que representi el trànsit de forma fiable?*

Aquesta qüestió s'estudiarà en el moment que s'avaluïn els primers resultats del sistema en un entorn d'operació real. S'hi farà front recopilant dades, etiquetant-les i analitzant-les al capítol 5, després de presentar la mesura d'impacte dels esdeveniments.

En una anàlisi detallada dels esdeveniments acústics presents a les àrees pilot del projecte, es planteja que la durada i la prominència d'un esdeveniment condicionen la contribució que aquest tindrà sobre el nivell equivalent mesurat. Per prominència, fem referència a l'elevació del nivell sonor de l'esdeveniment per damunt dels sons que l'envolten, apareix a la literatura citada en anglès com a *saliency*. Per això es planteja la segona pregunta:

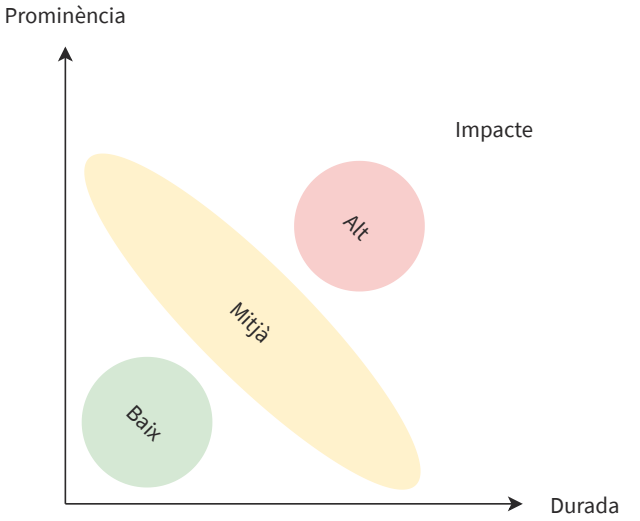



Figura 1.1: Regions d'impacte sobre l'$L_{eq}$ hipotètiques segons la prominència i la durada dels esdeveniments acústics.

> P2: *Provocarien, els esdeveniments amb durada i prominència significatives, un impacte considerable en el càlcul del nivell equivalent? Serà factible classificar-los en regions d'impacte en funció de la prominència i la durada?*

La figura 1.1 mostra les hipotètiques regions d'impacte on poden classificar-se els esdeveniments acústics detectats. En una zona de baix impacte, sembla raonable pensar que s'hi trobaran els esdeveniments que no tenen ni una durada ni una prominència significatives; a la regió d'impacte alt s'hi situarien els esdeveniments que tenen durada i prominència significatives; i a la regió d'impacte mitjà, els esdeveniments que tenen o bé una durada o bé una prominència considerables.

D'aquesta qüestió se'n deriva una altra, que planteja en quina mesura esbiaixaran el càlcul de l'$L_{eq}$ els esdeveniments de les diferents regions d'impacte, de manera individual o agregada dins d'un interval de mesura prèviament definit. Durant el capítol 5, es resoldrà aquesta pregunta i altres qüestions que en deriven amb les dades reals obtingudes de les àrees pilot de DYNAMAP.

També en el moment de presentar la mesura de l'impacte, sorgeix una altra qüestió, que té a veure amb la molèstia percebuda pels ciutadans:

> P3: *És suficient mesurar l'impacte dels esdeveniments en el nivell equivalent de so per tal d'analitzar la molèstia percebuda per les persones?*

A partir d'aquesta pregunta, es plantegen opcions per avaluar la percepció que tenen les persones sobre l'impacte dels esdeveniments acústics d'un determinat entorn sonor. Això s'ha dut a terme mitjançant estudis perceptius que utilitzen dades acústiques reals, per tal d'avaluar les diferents tipologies de sons que hi ha en un paisatge sonor i com són percebudes a nivell psicoacústic.

## 1.3 Objectius

El primer objectiu que es pretén assolir està relacionat amb el funcionament en si de la xarxa de sensors. Cal aconseguir que la xarxa de sensors desplegada a les àrees pilot sigui capaç d'identificar els esdeveniments que no són de trànsit per poder-los eliminar del càlcul del nivell de so equivalent, els anomenats ANEs. Això requereix crear una base de dades de sons de les dues àrees pilot i dur a terme un etiquetatge exhaustiu dels diferents tipus de sons que trobarem en cada cas. I després utilitzar aquestes dades per entrenar un algoritme detector d'esdeveniments acústics capaç de classificar entre soroll de trànsit i esdeveniments anòmals. Finalment, caldrà avaluar la precisió de l'ANED en un entorn d'operació real per tal de comprovar-ne el correcte funcionament.

Una vegada el sistema sigui capaç de funcionar en temps real en un entorn d'operació real, es pretén analitzar la naturalesa dels dos paisatges sonors: el de Roma i el de Milà. Per tal d'aconseguir-ho, es farà un estudi detallat dels esdeveniments recopilats i etiquetats a les dues àrees pilot, per buscar possibles vies d'optimització de l'algoritme en cada zona.

A continuació, a partir dels resultats de l'anàlisi detallada dels esdeveniments, es proposarà una metodologia per a calcular la contribució de cada esdeveniment en el nivell de soroll mesurat. Aquesta mesura, anomenada impacte, es compararà amb altres mètriques dels esdeveniments per tal d'entendre quina és la naturalesa dels esdeveniments que poden esbiaixar la mesura del nivell equivalent del soroll de trànsit en els mapes acústics. D'aquest estudi, en derivarà una anàlisi que pretén avaluar la percepció que tenen els ciutadans dels esdeveniments d'un entorn sonor.

## 1.4 Tesi per compendi de publicacions

El present document és una tesi per compendi de publicacions, on el treball fet durant el període de la tesi ha estat publicat en revistes revisades per experts (*peer-reviewed journals*). A continuació es llisten els articles publicats que formen part del compendi juntament amb un paràgraf detallant-ne la contribució del candidat:

**1** Alsina-Pagès, R. M., Alías, F., Socoró, J. C., Orga, F., Benocci, R. i Zambon, G. **Anomalous events removal for automated traffic noise maps generation**. *Applied Acoustics*, vol. 151 (2019), pàg. 183-192.

En l'article «*Anomalous events removal for automated traffic noise maps generation*» (en català: Eliminació d'esdeveniments anòmals per la generació automàtica de mapes de soroll de trànsit), s'hi explica el paper de l'algoritme ANED per eliminar esdeveniments no relacionats amb el trànsit. Publicat el març de 2019 a l'*Applied Acoustics*, dona una visió del funcionament de l'algoritme i el compara amb un mètode d'etiquetatge manual. L'estudi, on es compara l'etiquetatge automàtic dut a terme per dues versions de l'ANED i el fet per tres etiquetadors experts, conclou que, si bé els errors mínims s'observen en els etiquetatges manuals, també s'hi observa més disparitat en el criteri d'etiquetatge.

El candidat analitza visualment les diferències en els 5 etiquetatges, on s'observa que l'ANED presenta moltes falses alarmes en les seves dues primeres versions i participa activament en la redacció de l'article.

**2** Alsina-Pagès, R. M., Alías, F., Socoró, J. C. i Orga, F. **Detection of anomalous noise events on low-capacity acoustic nodes for dynamic road traffic noise mapping within an hybrid WASN**. *Sensors*, vol. 18, núm. 4 (2018), pàg. 1272.

En un altre estudi, publicat a la revista *Sensors* l'abril de 2018, el candidat participa en el disseny de la xarxa híbrida que combina nodes d'alta i de baixa capacitat per detectar esdeveniments, segons els requeriments de consum de cada sensor. L'article s'anomena «*Detection of anomalous noise events on low-capacity acoustic nodes for dynamic road traffic noise mapping within an hybrid WASN*» (en català: Detecció d'esdeveniments de soroll anòmals en nodes acústics de baixa capacitat pel mapatge dinàmic del soroll de trànsit rodat en una xarxa de sensors híbrida). S'hi presenta un disseny que combina nodes de baixa capacitat en qualitat d'esclaus, que envien els nivells de so als nodes mestre, d'alta capacitat, que al seu torn envien la informació al servidor central.

El candidat realitza l'estimació de la càrrega computacional consumida per l'algoritme i repassa l'oferta d'ordinadors de butxaca que permetrien executar l'algoritme de baixa capacitat, redactant els apartats corresponents en l'article.

**3** Alsina-Pagès, R. M., Orga, F., Alías, F. i Socoró, J. C. **A WASN-based suburban dataset for anomalous noise event detection on dynamic road-traffic noise mapping**. *Sensors*, vol. 19, núm. 11 (2019), pàg. 2480.

Tant el procés d'etiquetatge com l'anàlisi exhaustiu de la base de dades de Roma, obtingudes a través de la WASN, es publiquen a l'article «*A WASN-based suburban dataset for anomalous noise event detection on dynamic road-traffic noise mapping*» (en català: Un conjunt de dades suburbà extret d'una WASN per a la detecció d'esdeveniments de soroll anòmals en el mapatge dinàmic del soroll de trànsit rodat). L'article, publicat el maig de 2019, repassa l'estat de l'art en la detecció d'esdeveniments acústics i les bases de dades emprades a la comunitat científica. A continuació, s'hi explica el disseny de la base de

dades, des de la descripció fins a la caracterització dels esdeveniments, passant pel seu etiquetatge. També s'hi inclou un estudi exhaustiu de la tipologia dels esdeveniments capturats amb la WASN, que formen part del conjunt de dades.

El candidat duu a terme l'anàlisi de les dades i participa en l'etiquetatge de la base de dades; també s'encarrega de la caracterització de les dades i participa activament en la redacció de l'article.

**4** Orga, F., Alías, F. i Alsina-Pagès, R. M. **On the impact of anomalous noise events on road traffic noise mapping in urban and suburban environments**. *International Journal of Environmental Research and Public Health*, vol. 15 núm. 1 (2018), pàg. 13.

El primer càlcul d'impacte individual es presenta a l'*International Journal of Environmental Research and Public Health* (IJERPH) el desembre de 2017. L'article duu per títol *«On the impact of anomalous noise events on road traffic noise mapping in urban and suburban environments»* (en català: Sobre l'impacte dels esdeveniments de soroll anòmals en mapes de soroll de trànsit d'entorns urbans i suburbans). A l'article, s'hi exposa l'estat de l'art en la influència del soroll a la salut pública i l'estat actual dels projectes de monitoratge de soroll a diferents països. A continuació, presenta la metodologia emprada, que consisteix a comparar l'impacte d'un esdeveniment amb la seva durada i el seu SNR. S'analitza l'impacte de tots els esdeveniments recopilats a Roma i a Milà en la primera campanya de gravació, amb tres períodes d'integració diferents, trobant l'impacte en 1, 5 i 15 minuts. Finalment, demostra que tant l'SNR com la durada són clau per determinar l'impacte d'un esdeveniment i s'observa com els esdeveniments que presenten més impacte tenen o bé un SNR o bé una durada significatives.

El candidat descriu la metodologia emprada, detalla els càlculs utilitzats, duu a terme l'anàlisi de dades, la visualització dels resultats i la seva descripció i finalment extreu conclusions conjuntament amb els coautors. Té un paper molt important en la redacció de l'article.

**5** Alías, F., Orga, F., Alsina-Pagès, R. M. i Socoró, J. C. **Aggregate impact of anomalous noise events on the WASN-based computation of road traffic noise levels in urban and suburban environments**. *Sensors*, vol. 20, núm. 3 (2020), pàg. 609.

L'últim treball del compendi, publicat a la revista *Sensors* el gener del 2020, es presenta la mesura de l'impacte agregat i s'avalua també en els dos entorns de DYNAMAP: l'urbà i el suburbà. L'article duu per títol *«Aggregate impact of anomalous noise events on the WASN-based computation of road traffic noise levels in urban and suburban environments»* (en català: L'impacte agregat dels esdeveniments de soroll anòmals en els nivells de soroll de trànsit capturats mitjançant una WASN en els entorns urbà i suburbà). En l'article es presenta una mesura de l'impacte agregat que permet sistematitzar l'anàlisi dels impactes per tots els sensors d'una xarxa de sensors. Mitjançant l'estudi i gràcies al fet de disposar

de més de 300 h de dades, es permet observar com esdeveniments amb un impacte baix o mitjà són capaços d'esbiaixar una mesura de nivell equivalent de manera agregada tant com un esdeveniment d'impacte alt, cosa que no s'havia pogut observar emprant la base de dades inicial de 9 h recopilada manualment. Això ressalta la importància de detectar els esdeveniments no relacionats amb el soroll de trànsit, encara que individualment presentin impactes reduïts, pel seu correcte monitoratge.

El candidat duu a terme l'anàlisi de les dades i participa en l'escriptura dels resultats i la discussió.

## 1.5 Llistat de participacions en congressos en l'àmbit de la tesi

A continuació es llisten les participacions del candidat en congressos durant el període i dins l'àmbit de la tesi doctoral, per ordre cronològic:

1. Zambon, G., Benocci, R., Orga, F., Alsina Pagès, R. M., Alías, F., i Socoró, J. C. **Real-time urban traffic noise maps: the influence of Anomalous Noise Events in Milan Pilot area of DYNAMAP**. INTER-NOISE Congress, Hong Kong, 27-30 d'agost de 2017.
2. Orga, F., Alsina Pagès, R. M., Alías, F., Socoró, J.C., Bellucci, P., i Peruzzi, L. **Anomalous Noise Events Considerations for the Computation of Road Traffic Noise Levels in Suburban Areas: the DYANAMAP's Rome Case Study**. Associazione Italiana di Acustica, 44o Convegno Nazionale Pavia. Pavia, 7-9 de juny de 2017.
3. Socoró, J. C., Alsina-Pagès, R. M., Alías, F. i Orga, F. **Adapting an Anomalous Noise Events Detector for Real-Life Operation in the Rome Suburban Pilot Area of the DYNAMAP's Project**. Euronoise proceedings, Heraklion, Creta, Grècia, 27-31 de maig de 2018.
4. Orga, F., Alías, F. i Alsina-Pagès, R. M. **Impact of Individual Anomalous Noise Events on the Monitoring of Traffic Noise in Urban Areas**. Euronoise proceedings, Heraklion, Creta, Grècia, 27-31 de maig de 2018.
5. Alías, F., Alsina-Pagès, R. M., Socoró, J. C. i Orga, F. **DYNAMAP: A Low-Cost WASN for Real-Time Road Traffic Noise Mapping**. TecniAcústica 2018, Cadis, Espanya, 24-26 d'octubre de 2018.
6. Socoró, J. C., Alsina-Pagès, R. M., Alías, F. i Orga, F. **Analysis of the Acoustic Characteristics of a suburban Multi-Sensor Network for Road Traffic Noise Mapping**. Electronic Conference of Sensors and Applications 2018, en línia, 15-30 de novembre de 2018.
7. Labairu-Trenchs, A., Alsina-Pagès, R. M., Orga, F. i Foraster, M. **Noise Annoyance in Urban Life: The Citizen as a Key Point of the Directives**. 1st International Electronic Conference on Environmental Health Sciences, en línia, 15 de novembre al 7 de desembre de 2018.

8. Alsina Pagès, R. M., Alías, F., Socoró, J.C., i Orga, F. **Performance Analysis of the Acoustic Event Detector in the DYNAMAP's Rome suburban area**. 23rd International Congress on Acoustics, Aachen, Alemanya, 9-13 de setembre de 2019.

9. Alías, F., Socoró, J. C., Orga, F. i Alsina-Pagès, R. M. **Characterization of a WASN-based Urban Acoustic Dataset for the Dynamic Mapping of Road Traffic Noise**. 6th International Electronic Conference on Sensors and Applications (ECSA-6), 15-30 de novembre de 2019.

10. Alsina-Pagès, R. M., Orga, F., Freixes, M., Mallol, R., Aletta, F., Mitchell, A., i Foraster, M. **Urban environment soundscape evaluation: Milan case study of noise events perceptions by citizens**. INTER-NOISE Congress, Hong Kong, 23-26 d'agost de 2020.

11. Alsina-Pagès, R. M., Orga, F., Mallol, R., Freixes, M., Baño, X., i Foraster, M. **Sons al balcó: Soundscape map of the Confinement in Catalonia**. 7th International Electronic Conference on Sensors and Applications (ECSA-7), en línia, 15-30 de novembre de 2020.

12. Alsina-Pagès, R. M., Orga, F., Freixes, M., i Foraster, M. **Citizens' Perceptual evaluation of noise events in an urban environment**. 13th ICBEN Congress on Noise as a Public Health Problem, Estocolm, Suècia, 14-17 juny de 2021.

## Referències

[1] Hänninen, O., Knol, A. B., Jantunen, M., Lim, T.-A., Conrad, A., Rappolder, M., Carrer, P., Fanetti, A.-C., Kim, R., Buekers, J. et al. "Environmental burden of disease in Europe: assessing nine risk factors in six countries". *Environmental health perspectives*, vol. 122, núm. 5 (2014), pàg. 439.

[2] European Commission. *Report from the Commission to the European Parliament and the Council On the Implementation of the Environmental Noise Directive in accordance with Article 11 of Directive 2002/49/EC*. 2017.

[3] Muzet, A. "Environmental noise, sleep and health". *Sleep medicine reviews*, vol. 11, núm. 2 (2007), pàg. 135 - 142.

[4] Jakovljevic, B., Paunovic, K. i Belojevic, G. "Road-traffic noise and factors influencing noise annoyance in an urban population". *Environment international*, vol. 35, núm. 3 (2009), pàg. 552 - 556.

[5] Stansfeld, S. A. i Matheson, M. P. "Noise pollution: non-auditory effects on health". *British medical bulletin*, vol. 68, núm. 1 (2003), pàg. 243 - 257.

[6] Goines, L. i Hagler, L. "Noise Pollution: a Modern Plague". *Southern Medical Journal-Birmingham Alabama-*, vol. 100, núm. 3 (2007), pàg. 287 - 294.

[7] Tsai, K.-T., Lin, M.-D. i Chen, Y.-H. "Noise mapping in urban environments: A Taiwan study". *Applied Acoustics*, vol. 70, núm. 7 (2009), pàg. 964 - 972.

[8] Fund, P. H. O. "Department of Housing and Urban Development" (2006).

[9] Zannin, P. H. T., Diniz, F. B. i Barbosa, W. A. "Environmental noise pollution in the city of Curitiba, Brazil". *Applied Acoustics*, vol. 63, núm. 4 (2002), pàg. 351 - 358.

[10] Morillas, J. B., Escobar, V. G., Sierra, J. M., Gómez, R. V. i Carmona, J. T. "An environmental noise study in the city of Cáceres, Spain". *Applied Acoustics*, vol. 63, núm. 10 (2002), pàg. 1061 - 1070.

[11] Chakrabarty, D., Chandra Santra, S., Mukherjee, A., Roy, B. i Das, P. "Status of road traffic noise in Calcutta metropolis, India". *The Journal of the Acoustical Society of America*, vol. 101, núm. 2 (1997), pàg. 943 - 949.

[12] Brown, A. i Lam, K. "Levels of ambient noise in Hong Kong". *Applied Acoustics*, vol. 20, núm. 2 (1987), pàg. 85 - 100.

[13] Botteldooren, D., Dekoninck, L. i Gillis, D. "The influence of traffic noise on appreciation of the living quality of a neighborhood". *International journal of environmental research and public health*, vol. 8, núm. 3 (2011), pàg. 777 - 798.

[14] Sørensen, M., Andersen, Z. J., Nordsborg, R. B., Becker, T., Tjønneland, A., Overvad, K. i Raaschou-Nielsen, O. "Long-term exposure to road traffic noise and incident diabetes: a cohort study". *Environmental health perspectives*, vol. 121, núm. 2 (2013), pàg. 217.

[15] Dratva, J., Phuleria, H. C., Foraster, M., Gaspoz, J.-M., Keidel, D., Künzli, N., Liu, L.-J. S., Pons, M., Zemp, E., Gerbase, M. W. et al. "Transportation noise and blood pressure in a population-based sample of adults". *Environmental health perspectives*, vol. 120, núm. 1 (2012), pàg. 50.

[16] World Health Organization i the Joint Research Centre of the European Commission. *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. Inf. tèc. World Health Organization. Regional Office for Europe, 2011, pàg. 1 - 126.

[17] EU. "Directive 2002/49/EC of the European Parliament and the Council of 25 Juny 2002 relating to the assessment and management of environmental noise". *Off. Journal of the European Communities*, vol. L189/12 (2002).

[18] Kephalopoulos, S., Paviotti, M. i Anfosso-Lédée, F. "Common Noise Assessment Methods in Europe (CNOSSOS-EU)". *Publications Office of the European Union*, vol. Report EUR 25379 EN (2002), pàg. 1 - 180.

[19] Ripoll, A. i Bäckman, A. "State of the art of noise mapping in Europe". *European Topic Centre Terrestrial* (2005).

[20] Barham, R., Chan, M. i Cand, M. "Practical experience in noise mapping with a MEMS microphone based distributed noise measurement system". *$39^{th}$ International Congress and Exposition on Noise Control Engineering (Internoise 2010)*. Vol. 6. 2010, pàg. 4725 - 4733.

[21] Alías, F. i Alsina-Pagès, R. M. "Review of Wireless Acoustic Sensor Networks for Environmental Noise Monitoring in Smart Cities". *Journal of sensors*, vol. 2019 (2019).

[22] Sevillano, X., Socoró, J. C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., Coppi, P., Nencini, L., Cerniglia, A., Bisceglie, A., Benocci, R. i Zambon, G. "DYNAMAP – Development of low cost sensors networks for real time noise mapping". *Noise Mapping*, vol. 3 (1 Maig de 2016), pàg. 172 - 189.

[23] Mietlicki, C., Mietlicki, F., Ribeiro, C., Gaudibert, P. i Vincent, B. "The HARMONICA project, new tools to assess environmental noise and better inform the public". *Proceedings of the Forum Acusticum*. Krákow, Poland, 2014, pàg. 7 - 12.

[24] Bellucci, P., Peruzzi, L. i Cruciani, F. R. "Implementing the Dynamap system in the suburban area of Rome". *Proceedings of INTER-NOISE 2016*. Institute of Noise Control Engineering. Hamburg, Germany, 21 - 24 Agost de 2016, pàg. 6396 - 6407.

Capítol 2

# Estat de la qüestió

En aquest capítol, es llisten els estudis més rellevants en el camp de la tesi, començant pels projectes principals que utilitzen xarxes de sensors acústics per monitorar els nivells de soroll. A continuació, es repassen els diferents sistemes de detecció d'esdeveniments que s'utilitzen a les xarxes de sensors, que permetrien d'identificar els esdeveniments acústics que hi ocorren i, finalment, es detallen els efectes del soroll a la salut dels ciutadans.

## 2.1 Monitoratge acústic mitjançant xarxes de sensors sense fils

El monitoratge acústic en entorns urbans s'ha dut a terme habitualment a través de mesuraments duts a terme per experts emprant dispositius certificats. Això presenta diversos inconvenients, com el cost de realitzar-los, el mostreig d'hores i dies arbitrari i el fet de no poder gaudir de la informació al moment d'obtenir-la. A partir de la popularització de les xarxes de sensors acústics sense fils (o WASNs, de l'anglès *Wireless Acoustic Sensor Network*) s'estan duent a terme projectes de monitoratge automàtic dels ambients acústics mitjançant xarxes de sensors. Per exemple: el DREAMsys, que monitora diverses àrees del Regne Unit utilitzant una xarxa de sensors distribuïts [1]; l'UrbanSense, que monitora soroll urbà, diòxid de carboni i monòxid de carboni en temps real al Canadà [2]; el Senseable, que mesura el nivell de so equivalent en temps real a diferents punts de la ciutat de Pisa [3]; el LIFE Monza, que posa èmfasi en els components de baix cost per monitorar el soroll a la ciutat de Monza, a Itàlia; el SONYC, que ha desplegat 56 sensors de baix cost a la ciutat de Nova York per monitorar el soroll urbà i classificar els sons que hi ocorren en temps real basant-se en queixes rebudes al telèfon d'emergències [4], la xarxa de Màlaga, on s'han desplegat 8 sensors de baix cost capaços de mesurar fins a 86 paràmetres diferents [5] o la de Madrid, que incorpora 31 sensors equipats amb sonòmetres de classe 1 [6]. A Barcelona també s'ha desplegat una xarxa amb 130 sensors acústics per mesurar la qualitat acústica de la ciutat [7, 8]. I, en alguns països, també s'estan col·locant sensors acústics a les autovies, com a Burdwan, on monitoren els nivells de soroll i en fan anàlisi estadística. Alguns autors estudien quina ha de ser la localització dels sensors per avaluar múltiples fonts de soroll [9, 10].

A la literatura es poden trobar diversos estudis que tenen en compte la tipologia de les dades, en lloc d'avaluar només el nivell equivalent del soroll. Com per exemple el projecte *Smart Sound Monitoring*, que creua la informació acústica recollida amb enquestes de percepció [11]. D'altres propostes utilitzen sistemes de reconeixement de so per a obtenir informació dels sons i establir relacions entre els esdeveniments identificats i les enquestes recollides. I

amb l'objectiu de monitorar el soroll de trànsit, el projecte LIFE DYNAMAP identifica els esdeveniments no relacionats directament amb el trànsit rodat i eliminar-los de la computació del nivell sonor recollit amb sensors de baix cost [12].

Estudis més recents, pretenen utilitzar dades sobre el trànsit als carrers com a predictor de les emissions acústiques a la xarxa viària [13], obrin la possibilitat de reduir la quantitat de sensors acústics necessaris per monitorar una zona concreta.

## 2.2 Detecció d'esdeveniments sonors

La Detecció d'Esdeveniments Sonors (DES) és un camp de recerca en evolució constant i s'aplica amb diferents finalitats, com ara la vigilància [14, 15], el monitoratge de fauna [16], la salut [17] o el monitoratge d'entorns urbans [11, 18, 19]. Amb una terminologia que varia en la literatura, la detecció d'esdeveniments sonors o acústics consisteix a detectar i classificar diferents esdeveniments que ocorren en un flux de dades acústiques. Sovint, aquesta tasca implica proveir un descriptor textual sobre l'esdeveniment acústic que es detecta, com ara una etiqueta o classe [20]. En funció del nombre d'esdeveniments simultanis que es pretenen detectar, la DES es pot aplicar en un domini monofònic o polifònic. Els classificadors més utilitzats en la literatura són les màquines de vector de suport [21], els models de mescla de Gaussianes [22], els arbres de decisió [23] i els models ocults de Màrkov [24].

Tot i que la base teòrica de les xarxes neuronals data del segle XIX [25, 26] i el primer model computacional va ser creat el 1843 [27], no ha estat fins a la primera dècada del s. XXI que s'ha popularitzat el seu ús. Aquest augment en l'ús de les xarxes neuronals a la literatura ha estat provocat per la tendència a l'augment de potència computacional i l'ús de les unitats de processament gràfic (o GPUs, de l'anglès *Graphic Processing Unit*) com a eines de computació de propòsit general [28]. L'arquitectura més simple és la de les xarxes neuronals directes (o FNNs, de l'anglès *Feed-forward Neural Network*), on la informació només es mou en una direcció [29], des dels nodes que formen la capa d'entrada cap als de sortida, passant per les capes ocultes, si hi són presents. En canvi, les xarxes neuronals recurrents (o RNNs, de l'anglès *Recurrent Neural Network*) presenten realimentacions basades en connexions entre els nodes de la mateixa capa, de capes anteriors o amb si mateixos. Una de les implementacions de les RNNs és la de les xarxes de memòria a curt termini (o LSTM, de l'anglès *Long Short-Term Memory*), construïda amb unitats formades per una cèl·lula de memòria, una porta d'entrada, una porta de sortida i una porta d'oblit [30]. Una altra xarxa d'interès és la xarxa neuronal convolucional (o CNN, de l'anglès *Convolutional Neural Network*), que implementa la funció matemàtica de la convolució en alguna de les seves capes. Mitjançant els filtres convolucionals, la xarxa és capaç d'entrenar matrius amb una gran capacitat d'abstracció, cosa que la fa ideal en el camp del reconeixement d'imatge i vídeo. La primera CNN va ser la

implementada per Yann LeCun el 1989 per reconèixer els codis postals escrits a mà del servei postal dels Estats Units [31].

A la literatura s'hi pot trobar molts exemples de xarxes neuronals utilitzades per la detecció d'esdeveniments sonors. Per exemple, una FNN després d'un enfinestrat dels esdeveniments acústics [32] o les CNNs, que tenen la capacitat d'aprendre en temps i en freqüència mitjançant els filtres convolucionals [33]. També les RNN, que poden proveir el context d'un temps anterior utilitzant les connexions retroalimentades [34, 35]. Una barreja entre filtres convolucionals i xarxes recurrents, dona lloc a les xarxes neuronals recurrents convolucionals (o CRNN, de l'anglès *Convolutional Recurrent Neural Network*). Aquestes xarxes es beneficien de l'aprenentatge de les dimensions de freqüència i temps i, a més a més, aprenen sobre el context temporal dels esdeveniments, que permet detectar certes repeticions o intermitències que poden ajudar a la detecció [36].

Un estudi més recent, aplica la xarxa convolucional de CapsNet [37], destinada a millorar la detecció d'imatges canviades de mida o perspectiva i superposades, a la detecció d'esdeveniments acústics, millorant els resultats de les xarxes convolucionals existents [38]. Tanmateix, aquestes tècniques estan restringides als conjunts de dades que tingui la base de dades utilitzada per entrenar-les, que haurà d'oferir fragments de gravació llargs. I la capacitat de computació pot condicionar també l'arquitectura del model utilitzat.

## 2.3 Influència del soroll a la salut pública

La molèstia és l'efecte més conegut del soroll ambiental [39], però els estudis demostren que també pot tenir altres efectes adversos en la salut humana.

Els efectes més rellevants en són els trastorns en el son [40], les pèrdues auditives i els *tinnitus* (o acúfens) [41]. Però el soroll ambiental també és causant de trastorns en l'aprenentatge d'infants [42], especialment, per la presència del soroll d'aeronaus [43]. I s'ha arribat a utilitzar el nivell de soroll ambient com a indicador en el rendiment dels estudiants a classe [44]. Alguns estudis també han trobat una relació entre la contaminació acústica i l'aparició de problemes cardiovasculars, per exemple l'infart de miocardi [45], la hipertensió [46], així com alteracions en la pressió sanguínia [47, 48]. Segons l'OMS, cada any sorgeixen 48.000 casos nous de cardiopatia isquèmica deguts al soroll ambiental que acaben provocant 12.000 morts. En altres investigacions, se'n destaquen efectes com la diabetis [49] o trastorns en la salut mental [50]. Un estudi fins i tot destaca que hi ha una relació entre el soroll ambiental generat per les aeronaus i el naixement de nadons prematurs [51]. I, a més a més, s'ha demostrat que degrada els espais residencials, socials i de treball, i els centres d'aprenentatge, de manera que també genera efectes negatius en el desenvolupament socioeconòmic dels llocs afectats [52].

L'OMS quantifica els anys de vida saludables perduts a Europa en termes d'«anys de vida ajustats per discapacitat» (o DALYs, de l'anglès *Disability Adjusted Life Years*) [53] i conclou que les malalties relacionades amb l'exposició al

soroll produeixen, cada any, una pèrdua d'un milió d'anys de vida saludables a l'Europa occidental.

Jakovljevic *et al.*, en un estudi on utilitza indicadors com l'estrès, l'edat i la localització de l'habitació dins la llar, determina que la majoria d'adults entrevistats troben que el trànsit és la font de soroll més molesta [54]. Estudis com aquest, demostren que, de totes les fonts de soroll ambiental, el soroll de trànsit n'és la més rellevant en entorns urbans, i per això, és també la font més estudiada. Botteldooren *et al.* compara un conjunt d'indicadors relacionats amb l'exposició al soroll i remarca la rellevància que té el soroll de trànsit en la qualitat de vida dels veïnats [55]. També un estudi de Van Renterghem *et al.* ressalta la importància de tenir una façana silenciosa en els habitatges [56]. Öhrström conclou que el soroll de trànsit augmenta el cansament en les persones i destorba el son [57], en un estudi on també destaca que tenir accés a una part silenciosa de l'habitatge contribueix al benestar fisiològic i psicològic de les persones que hi viuen. Per més exemples, el lector pot consultar l'estudi de l'OMS sobre les intervencions en el soroll de transport i els seus impactes en salut a Europa [58].

## Referències

[1] Barham, R., Chan, M. i Cand, M. "Practical experience in noise mapping with a MEMS microphone based distributed noise measurement system". *39$^{th}$ International Congress and Exposition on Noise Control Engineering (Internoise 2010)*. Vol. 6. 2010, pàg. 4725 - 4733.

[2] Rainham, D. "A wireless sensor network for urban environmental health monitoring: UrbanSense". *IOP Conference Series: Earth and Environmental Science*. Vol. 34(1). IOP Publishing. 2016, pàg. 012028.

[3] Nencini, L., De Rosa, P., Ascari, E., Vinci, B. i Alexeeva, N. "SENSEable Pisa: A wireless sensor network for real-time noise mapping". *Proceedings of the EuroNoise 2012*. Prague, Czech Republic, oct. de 2012, pàg. 10 - 13.

[4] Bello, J. P., Silva, C., Nov, O., Dubois, R. L., Arora, A., Salamon, J., Mydlarz, C. i Doraiswamy, H. "SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution". *Communications of the ACM*, vol. 62, núm. 2 (2019), pàg. 68 - 77.

[5] López, J. M., Alonso, J., Asensio, C., Pavón, I., Gascó, L. i Arcas, G. de. "A Digital Signal Processor Based Acoustic Sensor for Outdoor Noise Monitoring in Smart Cities". *Sensors*, vol. 20, núm. 3 (2020), pàg. 605.

[6] Asensio, C., Pavón, I. i De Arcas, G. "Changes in noise levels in the city of Madrid during COVID-19 lockdown in 2020". *The Journal of the Acoustical Society of America*, vol. 148, núm. 3 (2020), pàg. 1748 - 1755.

[7] Camps, J. "Barcelona noise monitoring network". *Proceedings of EuroNoise 2015*. Maastrich, Netherlands: EAA-NAG-ABAV, 31 Maig - 3 Juny de 2015, pàg. 218 - 220.

[8] Camps-Farrés, J. i Casado-Novas, J. "Issues and challenges to improve the Barcelona Noise Monitoring Network". *Proceedings of EuroNoise 2018*. Heraklion, Crete – Greece: EAA –- HELINA, 27 - 31 Maig de 2018, pàg. 693 - 698.

[9] Licitra, G., Gallo, P., Rossi, E. i Brambilla, G. "A novel method to determine multiexposure priority indices tested for Pisa action plan". *Applied Acoustics*, vol. 72, núm. 8 (2011), pàg. 505 - 510.

[10] Miedema, H. M. "Relationship between exposure to multiple noise sources and noise annoyance". *The Journal of the Acoustical Society of America*, vol. 116, núm. 2 (2004), pàg. 949 - 957.

[11] De Coensel, B. i Botteldooren, D. "Smart sound monitoring for sound event detection and characterization". *Proceedings of the 43rd International Congress on Noise Control Engineering (Inter-Noise 2014)*. Melbourne, Australia, 16 - 19 Novembre de 2014, pàg. 1 - 10.

[12] Sevillano, X., Socoró, J. C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., Coppi, P., Nencini, L., Cerniglia, A., Bisceglie, A., Benocci, R. i Zambon, G. "DYNAMAP – Development of low cost sensors networks for real time noise mapping". *Noise Mapping*, vol. 3 (1 Maig de 2016), pàg. 172 - 189.

[13] Lan, Z. i Cai, M. "Dynamic traffic noise maps based on noise monitoring and traffic speed data". *Transportation Research Part D: Transport and Environment*, vol. 94 (2021), pàg. 102796.

[14] Crocco, M., Cristani, M., Trucco, A. i Murino, V. "Audio surveillance: a systematic review". *ACM Computing Surveys (CSUR)*, vol. 48, núm. 4 (2016), pàg. 52.

[15] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N. i Vento, M. "Audio Surveillance of Roads: A System for Detecting Anomalous Sounds". *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, núm. 1 (Gener de 2016), pàg. 279 - 288.

[16] Cakir, E., Adavanne, S., Parascandolo, G., Drossos, K. i Virtanen, T. "Convolutional recurrent neural networks for bird audio detection". *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE. 2017, pàg. 1744 - 1748.

[17] Goetze, S., Schroder, J., Gerlach, S., Hollosi, D., Appell, J.-E. i Wallhoff, F. "Acoustic monitoring and localization for social care". *Journal of Computing Science and Engineering*, vol. 6, núm. 1 (2012), pàg. 40 - 50.

[18] Alsina-Pagès, R. M., Alías, F., Socoró, J. C., Orga, F., Benocci, R. i Zambon, G. "Anomalous events removal for automated traffic noise maps generation". *Applied Acoustics*, vol. 151 (2019), pàg. 183 - 192.

[19] Salamon, J. i Bello, J. P. "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification". *IEEE Signal Processing Letters*, vol. 24, núm. 3 (març de 2017), pàg. 279 - 283.

[20] Mesaros, A., Heittola, T. i Virtanen, T. "Metrics for polyphonic sound event detection". *Applied Sciences*, vol. 6, núm. 6 (2016), pàg. 162.

[21] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J. i Scholkopf, B. "Support vector machines". *IEEE Intelligent Systems and their applications*, vol. 13, núm. 4 (1998), pàg. 18 - 28.

[22] Reynolds, D. A. "Gaussian Mixture Models." *Encyclopedia of biometrics*, vol. 741 (2009).

[23] Elizalde, B., Kumar, A., Shah, A., Badlani, R., Vincent, E., Raj, B. i Lane, I. "Experimentation on the dcase challenge 2016: Task 1-acoustic scene classification and task 3-sound event detection in real life audio". *Detection and Classification of Acoustic Scenes and Events*, vol. 2016 (2016).

[24] Eddy, S. R. "Hidden markov models". *Current opinion in structural biology*, vol. 6, núm. 3 (1996), pàg. 361 - 365.

[25] Bain, A. *Mind and body: The theories of their relation*. Vol. 4. Henry S. King, 1873.

[26] James, W. *The principles of psychology, Vol. 2. Henry Holt and Company*. 1890.

[27] McCulloch, W. S. i Pitts, W. "A logical calculus of the ideas immanent in nervous activity". *The bulletin of mathematical biophysics*, vol. 5, núm. 4 (1943), pàg. 115 - 133.

[28] Oh, K.-S. i Jung, K. "GPU implementation of neural networks". *Pattern Recognition*, vol. 37, núm. 6 (2004), pàg. 1311 - 1314.

[29] Schmidhuber, J. "Deep learning in neural networks: An overview". *Neural networks*, vol. 61 (2015), pàg. 85 - 117.

[30] Hochreiter, S. i Schmidhuber, J. "Long short-term memory". *Neural computation*, vol. 9, núm. 8 (1997), pàg. 1735 - 1780.

[31] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. i Jackel, L. D. "Backpropagation applied to handwritten zip code recognition". *Neural computation*, vol. 1, núm. 4 (1989), pàg. 541 - 551.

[32] Cakir, E., Heittola, T., Huttunen, H. i Virtanen, T. "Polyphonic sound event detection using multi label deep neural networks". *2015 international joint conference on neural networks (IJCNN)*. IEEE. 2015, pàg. 1 - 7.

[33] Piczak, K. J. "ESC: Dataset for Environmental Sound Classification". *Proceedings of the 23rd ACM International Conference on Multimedia*. MM '15. Brisbane, Australia: ACM, 26 - 30 Octubre de 2015, pàg. 1015 - 1018.

[34] Parascandolo, G., Huttunen, H. i Virtanen, T. "Recurrent neural networks for polyphonic sound event detection in real life recordings". *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2016, pàg. 6440 - 6444.

[35] Adavanne, S. i Virtanen, T. "Sound event detection using weakly labeled dataset with stacked convolutional and recurrent neural network". *arXiv preprint arXiv:1710.02998* (2017).

[36] Çakır, E., Parascandolo, G., Heittola, T., Huttunen, H. i Virtanen, T. "Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, núm. 6 (juny de 2017), pàg. 1291 - 1303.

[37] Mukhometzianov, R. i Carrillo, J. "CapsNet comparative performance evaluation for image classification". *arXiv preprint arXiv:1805.11195* (2018).

[38] Vesperini, F., Gabrielli, L., Principi, E. i Squartini, S. "Polyphonic Sound Event Detection by Using Capsule Neural Networks". *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, núm. 2 (2019), pàg. 310 - 322.

[39] Miedema, H. i Oudshoorn, C. "Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals." *Environmental health perspectives*, vol. 109, núm. 4 (2001), pàg. 409.

[40] Muzet, A. "Environmental noise, sleep and health". *Sleep medicine reviews*, vol. 11, núm. 2 (2007), pàg. 135 - 142.

[41] Śliwińska-Kowalska, M. i Zaborowski, K. "WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Permanent Hearing Loss and Tinnitus". *International Journal of Environmental Research and Public Health*, vol. 14, núm. 10 (2017), pàg. 1139.

[42] Lercher, P., Evans, G. W. i Meis, M. "Ambient noise and cognitive processes among primary schoolchildren". *Environment and Behavior*, vol. 35, núm. 6 (2003), pàg. 725 - 735.

[43] Hygge, S., Evans, G. W. i Bullinger, M. "A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren". *Psychological science*, vol. 13, núm. 5 (2002), pàg. 469 - 474.

[44] Chetoni, M., Ascari, E., Bianco, F., Fredianelli, L., Licitra, G. i Cori, L. "Global noise score indicator for classroom evaluation of acoustic performances in LIFE GIOCONDA project". *Noise Mapping*, vol. 3, núm. 1 (2016).

[45] Babisch, W., Beule, B., Schust, M., Kersten, N. i Ising, H. "Traffic noise and risk of myocardial infarction". *Epidemiology*, vol. 16, núm. 1 (2005), pàg. 33 - 40.

[46] Bluhm, G. L., Berglind, N., Nordling, E. i Rosenlund, M. "Road traffic noise and hypertension". *Occupational and environmental medicine*, vol. 64, núm. 2 (2007), pàg. 122 - 126.

[47] Dratva, J., Phuleria, H. C., Foraster, M., Gaspoz, J.-M., Keidel, D., Künzli, N., Liu, L.-J. S., Pons, M., Zemp, E., Gerbase, M. W. et al. "Transportation noise and blood pressure in a population-based sample of adults". *Environmental health perspectives*, vol. 120, núm. 1 (2012), pàg. 50.

[48] Babisch, W., Swart, W., Houthuijs, D., Selander, J., Bluhm, G., Pershagen, G., Dimakopoulou, K., Haralabidis, A. S., Katsouyanni, K., Davou, E. et al. "Exposure modifiers of the relationships of transportation noise with high blood pressure and noise annoyance". *The Journal of the Acoustical Society of America*, vol. 132, núm. 6 (2012), pàg. 3788 - 3808.

[49] Sørensen, M., Andersen, Z. J., Nordsborg, R. B., Becker, T., Tjønneland, A., Overvad, K. i Raaschou-Nielsen, O. "Long-term exposure to road traffic noise and incident diabetes: a cohort study". *Environmental health perspectives*, vol. 121, núm. 2 (2013), pàg. 217.

[50] Guite, H., Clark, C. i Ackrill, G. "The impact of the physical and urban environment on mental well-being". *Public health*, vol. 120, núm. 12 (2006), pàg. 1117 - 1126.

[51] Nieuwenhuijsen, M. J., Ristovska, G. i Dadvand, P. "WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Adverse Birth Outcomes". *International Journal of Environmental Research and Public Health*, vol. 14, núm. 10 (2017), pàg. 1252.

[52] Goines, L. i Hagler, L. "Noise Pollution: a Modern Plague". *Southern Medical Journal-Birmingham Alabama-*, vol. 100, núm. 3 (2007), pàg. 287 - 294.

[53] World Health Organization i the Joint Research Centre of the European Commission. *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. Inf. tèc. World Health Organization. Regional Office for Europe, 2011, pàg. 1 - 126.

[54] Jakovljevic, B., Paunovic, K. i Belojevic, G. "Road-traffic noise and factors influencing noise annoyance in an urban population". *Environment international*, vol. 35, núm. 3 (2009), pàg. 552 - 556.

[55] Botteldooren, D., Dekoninck, L. i Gillis, D. "The influence of traffic noise on appreciation of the living quality of a neighborhood". *International journal of environmental research and public health*, vol. 8, núm. 3 (2011), pàg. 777 - 798.

[56] Van Renterghem, T. i Botteldooren, D. "Focused study on the quiet side effect in dwellings highly exposed to road traffic noise". *International journal of environmental research and public health*, vol. 9, núm. 12 (2012), pàg. 4292 - 4310.

[57] Öhrström, E., Skånberg, A., Svensson, H. i Gidlöf-Gunnarsson, A. "Effects of road traffic noise and the benefit of access to quietness". *Journal of Sound and Vibration*, vol. 295, núm. 1 (2006), pàg. 40 - 59.

[58] Brown, A. L. i Kamp, I. van. "WHO environmental noise guidelines for the European region: A systematic review of transport noise interventions and their impacts on health". *International Journal of Environmental Research and Public Health*, vol. 14, núm. 8 (2017), pàg. 873.

Capítol 3

# Un cas real de monitoratge del soroll de trànsit: el projecte DYNAMAP

El soroll ambiental es pot definir com l'acumulació de la contaminació acústica generada per activitats humanes a l'exterior [1]. Les fonts que generen soroll són moltes, com l'activitat industrial o l'oci, però la principal font de soroll és el transport: el trànsit rodat, marítim, ferroviari i aeri [2, 3, 4].

Per aquest motiu, el projecte LIFE DYNAMAP[1] se centra en el monitoratge del soroll de trànsit rodat per generar un mapa acústic que s'actualitza automàticament mostrant-ne els nivells sonors de forma dinàmica. El projecte desplega dues xarxes pilot per verificar el funcionament del sistema dissenyat i implementat; una desplegada al nucli de Milà i, l'altra, al llarg de l'autovia A90 de Roma [5]. A la figura 3.1, s'hi pot observar l'esquema general de DYNAMAP, que inclou la xarxa de sensors, el servidor i la generació del mapa de soroll. Les dades acústiques captades amb els sensors de baix cost s'analitzen contínuament en el mateix sensor per tal d'extreure'n el nivell equivalent de so ($L_{A1s}$) i una etiqueta que indica cada segon si el fragment pertany a RTN o ANE ($RTN_{1s}/ANE_{1s}$) [6] mitjançant l'ANED, que utilitza tècniques d'aprenentatge automàtic per a classificar els fragments d'àudio segons les seves característiques acústiques [7]. El servidor del projecte rep les dades dels diferents nodes i en calcula els nivells sonors del soroll de trànsit ($L_{A1s,RTN}$), filtrant-ne els esdeveniments classificats per l'ANED com a anòmals. Finalment, s'actualitza el mapa de soroll precalculat agregant els nivells de soroll de trànsit extrets dels sensors un cop filtrats els nivells etiquetats com a ANEs [8] sobre un Sistema d'Informació Geogràfica (SIG). L'actualització dinàmica del mapa es fa en un interval que varia des de 5 minuts a 1 hora, en funció de l'hora del dia i l'àrea pilot [9]. Per la visualització dels nivells, es calcula i mostra l'índex HARMONICA [10, 11], que representa el nivell de soroll amb una escala de 0 a 10, calculat amb dos descriptors: el nivell de soroll de fons i els pics de soroll que n'emergeixen.

A la figura 3.4 s'hi pot observar el diagrama de blocs d'un node de la xarxa de sensors. Els nodes capten el senyal acústic amb sensors de baix cost i engeguen dos processos en paral·lel; per una banda, en calculen el nivell de so equivalent amb la ponderació A [12], és a dir, l'$L_{Aeq}$ i, per altra banda, executen l'ANED, que classifica el senyal entrant en soroll de trànsit o en esdeveniment anòmal. Aquests dos processos tenen una sortida per cada segon, que s'enviarà al servidor encarregat d'integrar la informació rebuda de cada node i actualitzar el mapa de soroll de manera dinàmica. El detall de l'algoritme ANED s'explica a la secció 3.2.

[1] http://www.life-dynamap.eu/ (visitat: 31/12/2021)

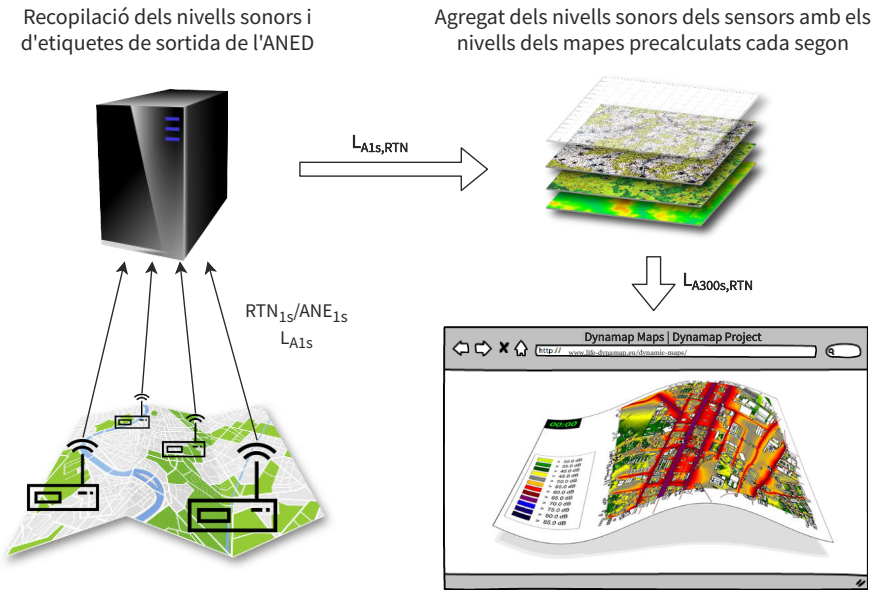



Figura 3.1: Esquema general de LIFE DYNAMAP

Els resultats d'integrar els nivells de soroll de trànsit al mapa de soroll dinàmicament es poden consultar a la pàgina web del projecte. A continuació es presenten dues captures del visor en línia, on s'observen els nivells de soroll en l'índex HARMONICA dels espais oberts (els edificis estan exclosos de la representació). A la figura 3.2 s'hi veu una porció del districte 9 de Milà, on s'aprecia un canvi en els nivells de soroll substancial entre les zones prop de les carreteres principals i els espais envoltats d'edificis. I a la figura 3.3 s'hi observa una porció del mapa dinàmic de Roma a l'autovia A90, on es pot contemplar efecte que produeixen els edificis prop de l'autovia, reduint dràsticament el nivell de soroll a l'altra banda d'aquests. A [13] s'hi detalla la precisió i la fiabilitat de les mesures acústiques integrades al mapa dinàmic.

Les tasques del projecte LIFE DYNAMAP es duen a terme entre set socis: l'administració de carreteres d'Itàlia (ANAS), les autoritats locals (Ajuntament de Milà i AMAT), les empreses privades ACCON i BlueWave i les Universitats de Milà "Bicoccaï La Salle (Universitat Ramon Llull). L'administració de carreteres i les autoritats s'encarreguen de proveir la logística i els recursos per desplegar la xarxa de sensors, la Universitat de Bicocca s'encarrega de la ubicació dels sensors de la xarxa, la Universitat de La Salle desenvolupa l'algoritme detector d'esdeveniments, ACCON implementa el mapa dinàmic i BlueWave s'encarrega del maquinari i manteniment dels dispositius, incloent-hi el calibratge [14]. El calibratge consisteix en col·locar un senyal de 94 dB i 1 kHz durant 10 segons


Figura 3.2: Porció del mapa de soroll dinàmic del districte 9 de Milà (es pot consultar en temps real a https://milano.noisemote.com/noisemap).

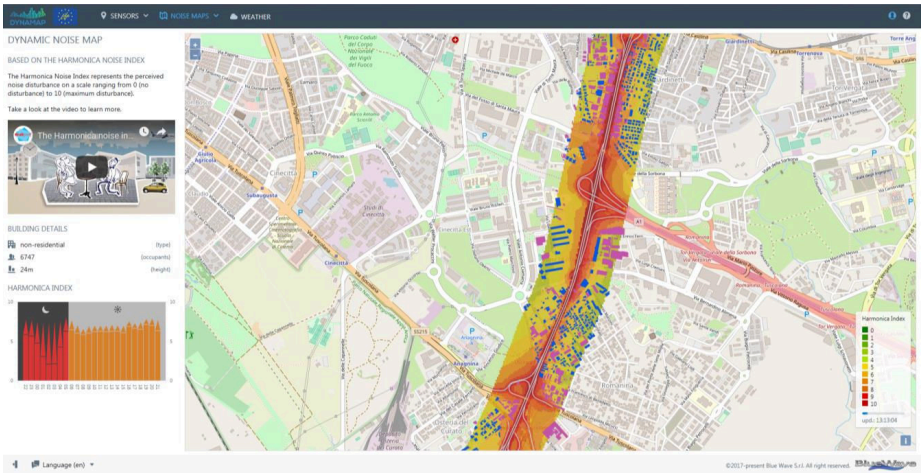



Figura 3.3: Porció del mapa de soroll dinàmic al llarg de l'autovia A90 que encercla Roma (es pot consultar en temps real a https://roma.noisemote.com/noisemap).

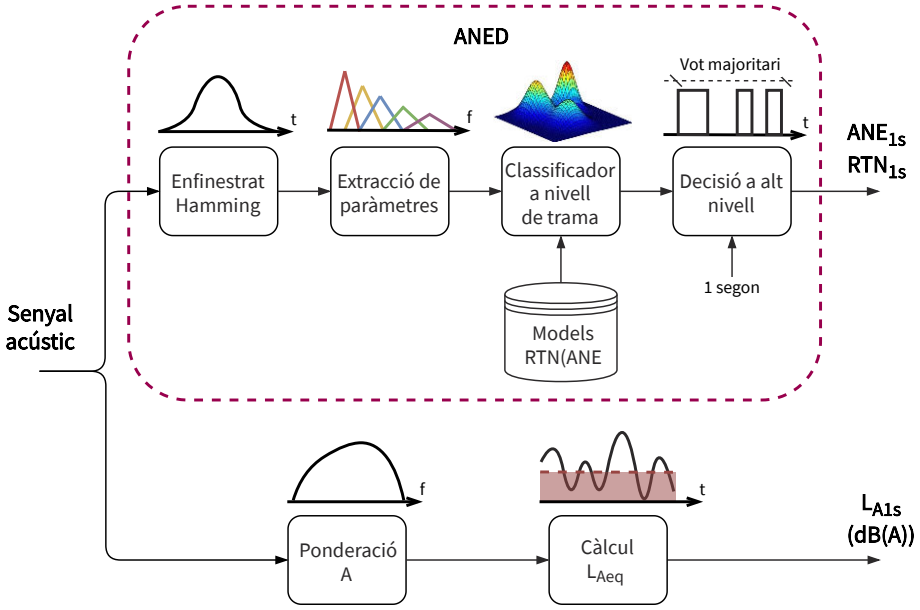



Figura 3.4: Diagrama de blocs d'un node de la xarxa de sensors de DYNAMAP

per detectar biaixos en la mesura del nivell sonor equivalent, aquesta tasca es duu a terme periòdicament segons el desgast del sensor acústic.

## 3.1 La xarxa de sensors DYNAMAP

El projecte LIFE DYNAMAP té dues xarxes desplegades: una al districte 9 de Milà [15] i l'altra a l'autovia A90 de Roma [16]. Cada una d'aquestes xarxes consta de nodes esclaus que incorporen l'algoritme detector d'esdeveniments anòmals i envien el nivell equivalent de soroll i el resultat de la classificació a un node mestre. Aquest node mestre aglutina les sortides dels altres nodes esclaus i envia els valors d'$L_{Aeq}$ i l'etiqueta RTN/ANE al servidor, que integra la informació amb el mapa acústic precalculat i l'actualitza de forma dinàmica. De nodes esclaus n'hi ha de dos tipus, els d'alta capacitat i els de baixa capacitat. Els primers, anomenats *Hi-Cap*, disposen de connexió a la xarxa elèctrica i els segons, *Lo-Cap*, requereixen abastament autònom per poder funcionar (per exemple, una placa solar) [17].

Finalment, l'algoritme s'ha desplegat satisfactòriament en les dues àrees pilot, i els ciutadans en poden observar el mapa acústic actualitzat al lloc web del projecte DYNAMAP [2]. Les localitzacions exactes dels punts de mesura a Milà es poden veure a la figura 3.5, i les localitzacions de Roma, a la figura 3.6.

[2] http://www.life-dynamap.eu/dynamic-maps/ (visitat: 31/12/2021)

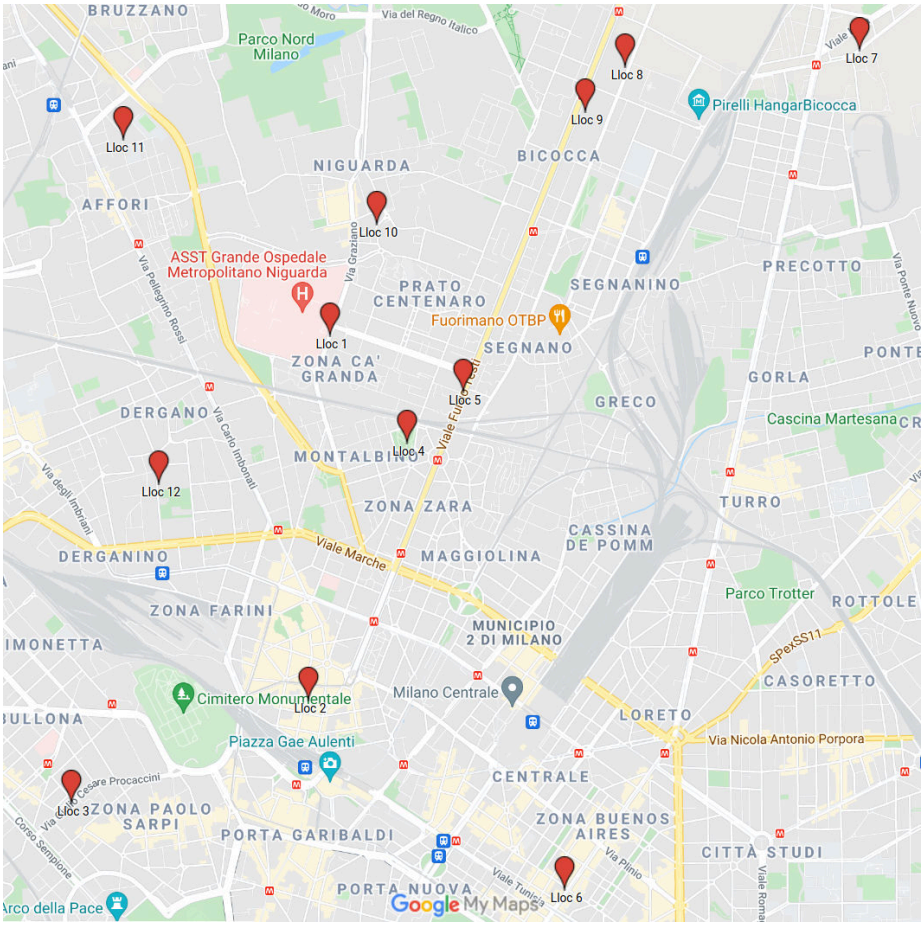



Figura 3.5: Punts de recollida de dades de Milà (imatge extreta de Google Maps).

## 3.2 Algoritme de detecció d'esdeveniments de soroll anòmals (ANED)

L'ANED és l'algoritme encarregat de classificar els esdeveniments en soroll de trànsit rodat o esdeveniments de soroll anòmals, que no estan directament relacionats amb el soroll de trànsit. La classificació entre els dos tipus de sons permet al sistema de discernir els nivells acústics que no provenen directament del trànsit i prescindir-ne en l'actualització dinàmica del mapa acústic [18].

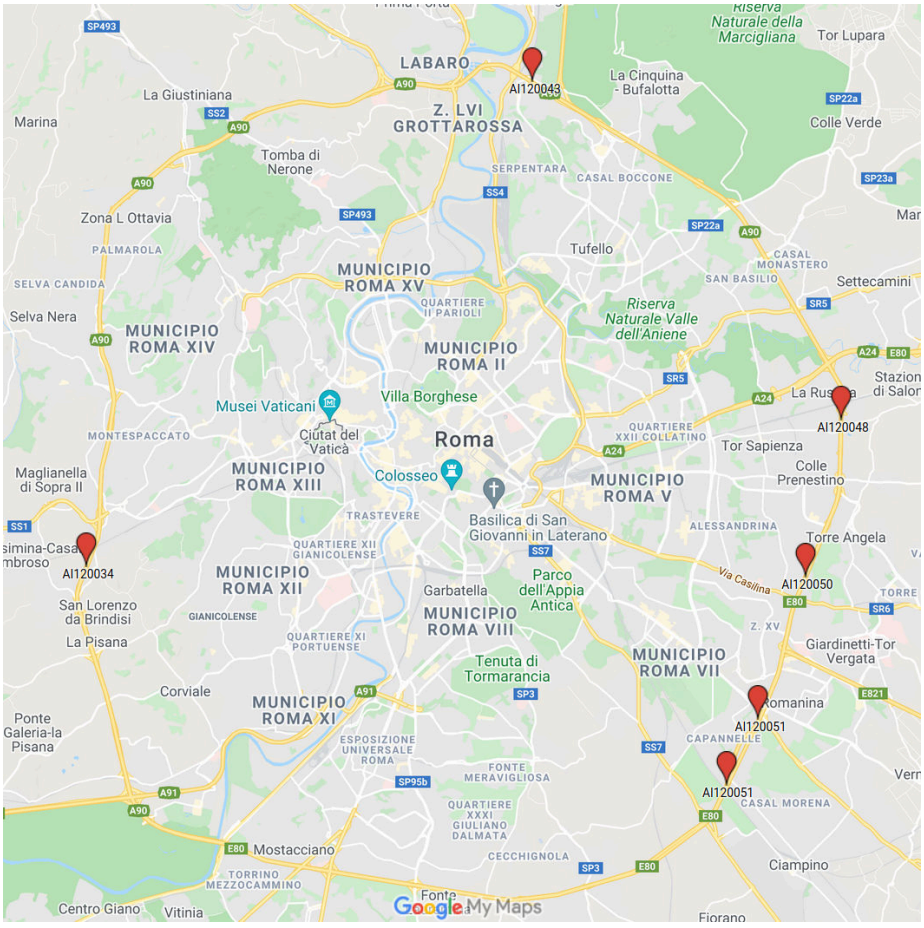



Figura 3.6: Punts de recollida de dades de Roma (imatge extreta de Google Maps).

### 3.2.1 Fases de disseny i implementació de l'ANED

L'algoritme de detecció d'esdeveniments anòmals es desenvolupa en diverses fases, les quals se sintetitzen en els punts següents:

1. Recollida manual de dades en els llocs on es desplegarà la xarxa.

2. Etiquetatge de les dades acústiques recollides, aproximadament 4,5 h d'àudio per cada localització [19].

3. Entrenament de l'ANED fora de línia (en una plataforma de proves programada en MATLAB®).

4. Testatge de l'ANED i validació dels resultats amb les dades recollides manualment.

5. Migració de l'ANED al maquinari dels nodes de la xarxa DYNAMAP, programat en C++ [16].

6. Recollida de dades reals a través dels sensors desplegats.

7. Entrenament de la segona versió de l'algoritme que utilitza dades recollides amb la xarxa de sensors de cada àrea pilot.

8. Validació de l'ANED executat als sensors de la xarxa amb dades recopilades als sensors.

Les fases presentades s'han executat amb l'ordre definit, però han requerit un conjunt d'iteracions derivades dels processos de testatge i validació.

### 3.2.2 Funcionament de l'ANED

L'algoritme de detecció d'esdeveniments anòmals té per finalitat detectar si el fragment a classificar no pertany directament al trànsit rodat, perquè el servidor l'elimini de la computació del mapa dinàmic de soroll. Aquesta classificació la duu a terme cada segon, marcant el fragment amb l'etiqueta RTN si pertany al trànsit rodat o l'etiqueta ANE si és un esdeveniment anòmal.

L'algoritme segueix quatre passos que es recullen a continuació i que es poden observar al diagrama de blocs de la figura 3.4 dins el requadre amb el nom d'ANED:

1. Enfinestrat de l'àudio a nivell de trama: la primera classificació de l'ANED es fa a nivell de trama de 30 ms. Per això s'utilitza una finestra Hamming de la longitud de la trama aplicant una superposició del 50% per compensar les parts de l'àudio atenuades per la finestra.

2. Extracció de paràmetres: s'extreuen els coeficients cepstrals en les freqüències de Mel (o MFCC, de l'anglès *Mel-Frequency Cepstral Coefficients*) [20].

3. Classificador a nivell de trama: s'utilitza un classificador binari basat en models de mescla gaussiana (o GMM, de l'anglès *Gaussian Mixture Model*), entrenat prèviament amb el subconjunt de dades d'entrenament, per classificar entre RTN i ANE a nivell de trama de 30 ms.

4. Decisió del classificador: s'aplica un vot majoritari entre les decisions preses en cada trama per donar una etiqueta de sortida per cada segon d'àudio d'entrada.

### 3.2.3 ANED de baixa capacitat

Com ja s'ha mencionat, la xarxa de sensors de DYNAMAP és híbrida, i utilitza sensors d'alta capacitat (*Hi-Cap*) i de baixa capacitat (*Lo-Cap*). Els nodes que tenen connexió a la xarxa elèctrica poden disposar de més potència computacional, cosa que els permet d'executar l'ANED. Però els nodes que s'alimenten de manera alternativa, no poden garantir un subministrament elèctric constant i, per tant, disposen de menys potència computacional.

Per resoldre això, s'ha dissenyat i avaluat una simplificació de l'algoritme menys exigent computacionalment [17]. Això es fa utilitzant finestres que seleccionen subbandes de l'espectre Mel per classificar directament els esdeveniments, marcant un llindar per cada subbanda d'energia. Així, es redueix la complexitat del detector d'esdeveniments, que es redueix a $\frac{1}{6}$ de la càrrega computacional de l'ANED original, però com a conseqüència se'n redueix la precisió de la detecció un 10%.

## 3.3 Publicacions destacades

El candidat ha participat activament en algunes de les tasques de desenvolupament de l'algoritme ANED, cosa que li ha aportat un aprenentatge en profunditat en el camp dels mapes sonors i els algoritmes de detecció d'esdeveniments. També ha tingut una implicació important en la tasca d'etiquetatge, sobretot en el conjunt de dades reals per a l'entrenament i la validació de l'algoritme. A continuació, se'n detallen les contribucions.

En l'article «*Anomalous events removal for automated traffic noise maps generation*» (en català: Eliminació d'esdeveniments anòmals per la generació automàtica de mapes de soroll de trànsit), s'hi explica el paper de l'algoritme ANED per eliminar esdeveniments no relacionats amb el trànsit. Publicat el març de 2019 a l'*Applied Acoustics*, dona una visió del funcionament de l'algoritme i el compara amb un mètode d'etiquetatge manual. L'estudi, on es compara l'etiquetatge automàtic dut a terme per dues versions de l'ANED i el fet per tres etiquetadors experts, conclou que, si bé els errors mínims s'observen en els etiquetatges manuals, també s'hi observa més disparitat en el criteri d'etiquetatge. S'analitza visualment les diferències en els 5 etiquetatges, on s'observa que l'ANED presenta moltes falses alarmes en les seves dues primeres versions.

En un altre estudi, publicat a la revista *Sensors* l'abril de 2018, es participa en el disseny de la xarxa híbrida que combina nodes d'alta i de baixa capacitat per detectar esdeveniments, segons els requeriments de consum de cada sensor. L'article s'anomena «*Detection of anomalous noise events on low-capacity acoustic nodes for dynamic road traffic noise mapping within an hybrid WASN*» (en català: Detecció d'esdeveniments de soroll anòmals en nodes acústics de baixa capacitat pel mapatge dinàmic del soroll de trànsit rodat en una xarxa de sensors híbrida). S'hi presenta un disseny que combina nodes de baixa capacitat en qualitat d'esclaus, que envien els nivells de so als nodes mestre, d'alta capacitat, que al seu torn envien la informació al servidor central. Es realitza l'estimació de la càrrega computacional consumida per

l'algoritme i repassa l'oferta d'ordinadors de butxaca (o SBC, de l'anglès *Single-Board Computer*) que permetrien executar l'algoritme de baixa capacitat.

## Referències


[1] European Commission. *The green paper on Future Noise Policy*. Inf. tèc. 1996.

[2] EU. "Directive 2002/49/EC of the European Parliament and the Council of 25 Juny 2002 relating to the assessment and management of environmental noise". *Off. Journal of the European Communities*, vol. L189/12 (2002).

[3] Botteldooren, D., Dekoninck, L. i Gillis, D. "The influence of traffic noise on appreciation of the living quality of a neighborhood". *International journal of environmental research and public health*, vol. 8, núm. 3 (2011), pàg. 777 - 798.

[4] Alberts, W. i Roebben, M. "Road Traffic Noise Exposure in Europe in 2012 based on END data". *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Vol. 253. 4. Institute of Noise Control Engineering. 2016, pàg. 3944 - 3955.

[5] Sevillano, X., Socoró, J. C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., Coppi, P., Nencini, L., Cerniglia, A., Bisceglie, A., Benocci, R. i Zambon, G. "DYNAMAP – Development of low cost sensors networks for real time noise mapping". *Noise Mapping*, vol. 3 (1 Maig de 2016), pàg. 172 - 189.

[6] Alías, F., Alsina-Pagès, R. M., Orga, F. i Socoró, J. C. "Detection of Anomalous Noise Events for Real-Time Road-Traffic Noise Mapping: The Dynamap's project case study". *Noise Mapping*, vol. 5, núm. 1 (2018), pàg. 71 - 85.

[7] Socoró, J. C., Alías, F. i Alsina-Pagès, R. M. "An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments". *Sensors*, vol. 17, núm. 10 (2017), pàg. 2323.

[8] Zambon, G., Benocci, R., Bisceglie, A., Roman, H. E. i Bellucci, P. "The LIFE DYNAMAP project: Towards a procedure for dynamic noise mapping in urban areas". *Applied Acoustics*, vol. 124 (2017), pàg. 52 - 60.

[9] Bellucci, P., Peruzzi, L. i Zambon, G. "LIFE DYNAMAP: making dynamic noise maps a reality". *Proceedings of EuroNoise 2018*. Heraklion, Crete – Greece: EAA –- HELINA, maig de 2018, pàg. 1181 - 1188.

[10] Mietlicki, C., Mietlicki, F., Ribeiro, C., Gaudibert, P. i Vincent, B. "The HARMONICA project, new tools to assess environmental noise and better inform the public". *Proceedings of the Forum Acusticum*. Krákow, Poland, 2014, pàg. 7 - 12.

[11] Bellucci, P., Peruzzi, L. i Cruciani, F. R. "Implementing the Dynamap system in the suburban area of Rome". *Proceedings of INTER-NOISE 2016*. Institute of Noise Control Engineering. Hamburg, Germany, 21 - 24 Agost de 2016, pàg. 6396 - 6407.

[12] Pierre, R. L. S. i Maguire, D. J. "The impact of A-weighting sound pressure level measurements during the evaluation of noise exposure". *NOISE-CON 2004*. Baltimore, Maryland, des. de 2004, pàg. 1 - 8.

[13] Bellucci, P., Peruzzi, L. i Nencini, L. "LIFE DYNAMAP: accuracy, reliability and sustainability of dynamic noise maps". *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Vol. 259. 6. Institute of Noise Control Engineering. 2019, pàg. 3653 - 3664.

[14] Zambon, G., Angelini, F., Benocci, R., Bisceglie, A., Radaelli, S., Coppi, P., Bellucci, P., Giovannetti, A. i Grecco, R. "DYNAMAP: a new approach to real-time noise mapping". *EuroNoise 2015 Congress*. 2015, pàg. 1 - 6.

[15] Alías, F., Socoró, J. C., Orga, F. i Alsina-Pagès, R. M. "Characterization of a WASN-based Urban Acoustic Dataset for the Dynamic Mapping of Road Traffic Noise". *Proceedings of the 6th International Electronic Conference on Sensors and Applications (ECSA-6)*. doi:10.3390/ecsa-6-06637. MDPI, 15-30 Novembre de 2019.

[16] Socoró, J. C., Alsina-Pagès, R. M., Alías, F. i Orga, F. "Adapting an Anomalous Noise Events Detector for Real-Life Operation in the Rome Suburban Pilot Area of the DYNAMAP's Project". *Proceedings of EuroNoise2018*. Heraklion, Crete – Greece: EAA – HELINA, 27 - 31 Maig de 2018, pàg. 693 - 698.

[17] Alsina-Pagès, R. M., Alías, F., Socoró, J. C. i Orga, F. "Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN". *Sensors*, vol. 18, núm. 4 (2018), pàg. 1272.

[18] Alsina-Pagès, R. M., Alías, F., Socoró, J. C., Orga, F., Benocci, R. i Zambon, G. "Anomalous events removal for automated traffic noise maps generation". *Applied Acoustics*, vol. 151 (2019), pàg. 183 - 192.

[19] Alías, F., Socoró, J. C., Sevillano, X. i Nencini, L. "Training an Anomalous Noise Event Detection Algorithm for Dynamic Road Traffic Noise Mapping: Environmental Noise Recording Campaign". *TecniAcústica 2015*. Valencia, Spain, oct. de 2015, pàg. 345 - 352.

[20] Mermelstein, P. "Distance measures for speech recognition, psychological and instrumental". *Pattern recognition and artificial intelligence*, vol. 116 (1976), pàg. 374 - 388.

I

# Anomalous events removal for automated traffic noise maps generation

**Rosa Ma. Alsina-Pagès, Francesc Alías, Joan Claudi Socoró, Ferran Orga, Roberto Benocci, Giovanni Zambon**

Publicat: 14 de març del 2019.

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

L'interval de pàgines 33-42 s'han extret de la tesi per motius de confidencialitat

# Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN

II

**Rosa Ma. Alsina-Pagès, Francesc Alías, Joan Claudi Socoró, Ferran Orga**

Publicat: 20 d'abril del 2018.

sensors

MDPI

*Article*

# Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN

**Rosa Ma Alsina-Pagès *, Francesc Alías, Joan Claudi Socoró and Ferran Orga**

GTM—Grup de recerca en Tecnologies Mèdia, La Salle—Universitat Ramon Llull, Quatre Camins, 30, 08022 Barcelona, Spain; falias@salleurl.edu (F.A.); jclaudi@salleurl.edu (J.C.S.); forga@salleurl.edu (F.O.)
* Correspondence: ralsina@salleurl.edu; Tel.: +34-932-902-455

Received: 13 March 2018; Accepted: 18 April 2018; Published: 20 April 2018


**Abstract:** One of the main aspects affecting the quality of life of people living in urban and suburban areas is the continuous exposure to high road traffic noise (RTN) levels. Nowadays, thanks to Wireless Acoustic Sensor Networks (WASN) noise in Smart Cities has started to be automatically mapped. To obtain a reliable picture of the RTN, those anomalous noise events (ANE) unrelated to road traffic (sirens, horns, people, etc.) should be removed from the noise map computation by means of an Anomalous Noise Event Detector (ANED). In Hybrid WASNs, with master-slave architecture, ANED should be implemented in both high-capacity (Hi-Cap) and low-capacity (Lo-Cap) sensors, following the same principle to obtain consistent results. This work presents an ANED version to run in real-time on μController-based Lo-Cap sensors of a hybrid WASN, discriminating RTN from ANE through their Mel-based spectral energy differences. The experiments, considering 9 h and 8 min of real-life acoustic data from both urban and suburban environments, show the feasibility of the proposal both in terms of computational load and in classification accuracy. Specifically, the ANED Lo-Cap requires around $^1/_6$ of the computational load of the ANED Hi-Cap, while classification accuracies are slightly lower (around 10%). However, preliminary analyses show that these results could be improved in around 4% in the future by means of considering optimal frequency selection.

**Keywords:** low capacity; low cost; hybrid wireless acoustic sensor network; real-time signal processing; anomalous noise event; noise; road traffic noise; dynamic noise mapping; μController; μProcessor

## 1. Introduction

Living with continuous exposure to high levels of traffic noise has been found to be harmful to human health, as it affects the quality of life of people living in urban and suburban areas [1]. Several actions have been conducted to address this problem, based on the European Noise Directive 2002/49/EC (END) [2] and the consequent strategic noise mapping assessment CNOSSOS-EU [3], which are the main requirements of current European legislation which request Member States to elaborate specific action plans to mitigate noise pollution, as well as making the public aware of the dangers of noise pollution.

Generally, until recently, noise measurements in cities have been conducted by professionals, who record and analyze the data in specific locations and time periods by using certified sound level meters. Subsequently, noise maps are generated from these noise level measurements by means of the application of complex acoustic models after data post-processing. These maps should be updated and published every five years to fulfill the END requirements for agglomerations with more than 100,000 inhabitants, major roads, major railways and airports [2]. However, this approach becomes difficult to scale when more measurements and/or locations are needed, besides the questionable

representativity of the data and the subsequent lack of accuracy of this kind of predictive models. Internet of Things and Smart City frameworks have led to a change of paradigm for the city noise monitoring and management by means of Wireless Acoustic Sensor Networks (WASNs) [4]. In the last decade, several works focused on the design and implementation of WASNs for environmental noise monitoring have been proposed [5]. The main goal of these approaches has been to develop affordable solutions, maintaining the reliability of the acoustic measures, while improving the scalability of the system through optimum network design. Some WASN-based systems have been developed and tested across Europe, such as the IDEA project in Belgium [6] and the Cense project [7] in France, or the 'Barcelona noise monitoring network' in Spain [8], which follow quite similar approaches. Moreover, it is worth mentioning the SONYC project [9] in the USA, aimed at monitoring noise pollution in New York City, besides providing an accurate description of the surrounding acoustic environment.

The initial development of WASN in several cities has opened several challenges [10,11], especially those derived from acoustic signal processing. As a first step, the aforementioned approximations are only focused on global noise monitoring, without taking into account the type of traffic or the detection of specific acoustic events in the acoustic environment; issues which are mandatory to satisfy the END requirements [2]. In order to provide public bodies with reliable measurements of the noise caused by Road Traffic Noise (RTN) affecting citizens, events not related to RTN—denoted as Anomalous Noise Noise Events (ANE)—should be removed from the noise map computation. In this context, Acoustic Event Detection (AED) algorithms have been designed for several domains of application in urban environments, most of them developed within surveillance applications, which include noise source identification [12–14], together with first works focused on the separation between target and interfering signals for noise monitoring in cities [15,16].

In this context, hybrid WASNs [5] may play a significant role in the large-scale deployment of this kind of noise monitoring networks both in terms of cost and extent of coverage. These networks combine high-capacity (Hi-Cap) nodes with cheaper low-capacity (Lo-Cap) nodes, which operate as masters and slaves in the network, respectively (see Figure 1). This architecture allows sensing places where the power supply cannot easily be provided by means of solar panels or other alternative energy sources that supply the Lo-Cap sensors. Both nodes typically compute the A-weighted equivalent noise level ($L_{Aeq}$) of the monitored acoustic environment [17], being the Hi-Cap nodes also responsible for data communications and any other type of complex processes, e.g., acoustic signal processing, recordings, etc. In [11], the authors present the design of an acoustic sensor network based on this approach, with basic nodes using a low power μController (μC). RUMEUR network [18] is also an hybrid WASN, including both high-accuracy equipment for critical places, combined with less-precise measuring equipment in other locations whose purpose is only updating the noise map in terms of $L_{Aeq}$.

With the same focus, the DYNAMAP LIFE project is aimed at developing a dynamic road traffic noise mapping system to represent the acoustic impact of road infrastructures in real-time in two Italian pilot areas: a suburban environment in Rome and an urban environment in Milan [19]. To do so, the project envisions a hybrid low-cost WASN including both Hi-Cap and Lo-Cap salve sensors, which will be located in places with limited power supply. The DYNAMAP project takes into account a noise monitoring challenge not faced by the aforementioned projects as it is only focused on one specific noise source: road traffic noise. Hence, it considers the inclusion of an Anomalous Noise Event Detector (ANED) [15] to provide a reliable picture of the actual RTN by minimizing the influence of other anomalous noise events [20]. Up to now, the ANED algorithm has been designed as a two-class classifier (ANE vs. RTN) using Mel Frequency Cepstral Coefficients (MFCC) [21] as acoustic parametrization and supervised machine learning classification to run in real-time on the Hi-Cap sensors of the hybrid WASN, showing promising results on real-life data. However, the Hi-Cap ANED algorithm [15] cannot be implemented as originally designed in the slave Lo-Cap sensors due to the computational resources it demands. Nevertheless, it would be desirable to include an adapted version of the algorithm to run on the Lo-Cap sensors in order to provide an homogeneous picture of

the RTN, thus, allowing the hybrid WASN to discard ANE from the $L_{Aeq}$ computation also in those locations where the slave sensors will be placed.

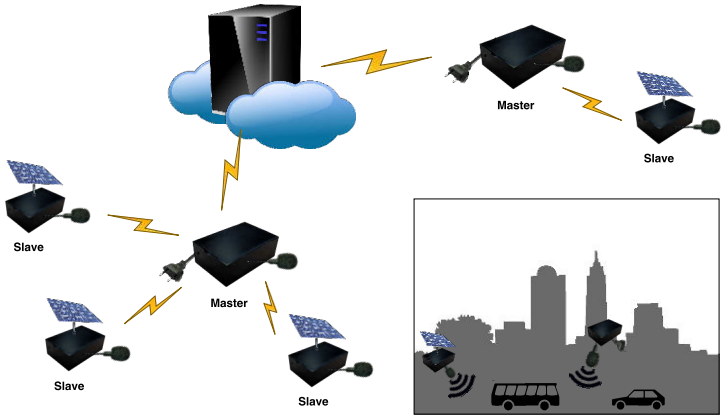



**Figure 1.** Scheme of an hybrid WASN architecture using Hi-Cap master sensor nodes and Lo-Cap slave sensor nodes, with a distributed intelligence.

This paper describes the adaptation of the Hi-Cap ANED algorithm to run in real-time on Lo-Cap sensors (ANED Lo-Cap), analyzing its viability in terms of computational load and classification accuracy. The ANED Lo-Cap algorithm bases on the same principle of the original ANED version to discriminate between ANE and RTN to obtain consistent results across the nodes of the hybrid WASN, but adapted to fit the computational capacity of a Lo-Cap sensor. For this purpose, the approach follows a threshold-based binary classification scheme that considers the spectral energy differences between ANE and RTN [22] for the most discriminatory Mel-based Frequency Subbands (MFS). The experiments are conducted considering 9h and 8 min of real-life acoustic data from the two pilot areas (urban and suburban) of the DYNAMAP project.

This paper is structured as follows. Section 2 reviews the main works about wireless acoustic sensor networks and acoustic event detection in urban areas. In Section 3, we describe the theoretical foundations to design the ANED Lo-Cap proposal. Section 4 evaluates the feasibility of the frequency range selection of the ANED Lo-Cap using a real-life database of urban and suburban acoustic data in the framework of the DYNAMAP project. Section 5 evaluates the computational load necessary to implement the ANED Lo-Cap, considering several commercial hardware platforms. Section 6 discusses the viability of the proposal, considering several open questions foreseen from the obtained results, which should be tackled in future works, and Section 7 details the final conclusions of this paper.

## 2. Related Work

In this section, we describe representative approaches developed to automatically measure the noise levels of the cities in order to tailor noise maps [5]. The first part reviews the environmental sound classification approaches that can be found in the literature to face acoustic event detection in urban environments, with special focus on those works considering real-life operating scenarios. The second part reviews several noise monitoring projects and their platform design and hardware.

### *2.1. Acoustic Event Detection in Urban Environments*

In this section, we review the most significant works focused on Acoustic Event Detection in urban environments based on the one-class novelty detection approach due to their similarity with the

problem our proposal faces. Besides that, some works focused on traffic noise and urban soundscapes based on multiclass classification are also included.

Ntalampiras et al. describe a probabilistic novelty detection approach for acoustic surveillance under *pseudo*-real-life conditions [12]. The study includes normal and abnormal (or anomalous) audio events such as screams, shouting or pleading for help, which are collected in real-life outdoor public security scenarios. The acoustic data is parametrized every 30 ms using a multidomain feature vector including different audio descriptors, such as Mel Frequency Cepstral Coefficients (MFCC), MPEG-7 low-level descriptors (LLD), Intonation and Teager Energy Operator, and Perceptual Wavelet Packets (PWP). The parametrized audio frames are fed into different probabilistic classifiers based on Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM), following a one-class classification (OCC) approach based only on the majority class (i.e., no information from the anomalous events is provided to the classifier): GMM clustering, universal GMM (UGMM), and universal HMM. It is worth mentioning that the hazardous situations are simulated with professional actors, thus, the gathered data cannot be strictly considered as collected in actual real-life conditions.

Later, in [23], Ntalampiras introduced an approach for acoustic surveillance in urban traffic environments following a multi-class AED approach. In that research, the AED is based on a two-stage HMM-based classification system and a window length of 30 ms, which analyzes a multidomain feature set, including again MFCC, LLD and PWP to consider the time, frequency and wavelet domains. The work includes a database composed of nine audio classes: car, motorcycle, aircraft, crowd, thunder, wind, train, horn and crash. The samples are obtained from several professional sound effect collections assuring high quality without being affected by background noise. Moreover, the detection of crash incidents is also studied by merging those sound events with the rest of classes at specific SNRs (i.e., 0, 5 and 15 dB). Thus, experimental configuration of the proposal is far from that obtained in real-life recording conditions since the events of interest are artificially mixed with background noise.

Aurino et al. apply OCC based on Support Vector Machines (SVM) to detect anomalous audio events within an automatic surveillance framework [24]. Specifically, the work is focused on the recognition of three types of burst-like acoustic anomalies: gun-shots, broken glasses or screams, which are defined and used to train one OCC-SVM per class a priori. The system follows a two-stage classification scheme by classifying the short audio segments (of a window length of 200 ms in the experiments) at the first level through an ensemble of OCCs. Then it aggregates these classification outputs into intervals of 1 s, which are subsequently reclassified using a majority voting strategy. The audio events are parametrized using a set of typical audio features such as MFFCs, Fast Fourier Transform (FFT), or Zero Crossing Rate (ZCR), to name a few. Again, it is to note that the events of interest are artificially mixed with background noise acquired in indoor and outdoor environments.

In [25], a real-life urban sound dataset named UrbanSound and its compact UrbanSound8k version are introduced. The dataset is composed of 10 low-level classes organized according to the following taxonomy: construction (drilling and jack hammer), mechanical (air conditioner and engine idling), traffic (car horn), community (dog bark, children play and street music) and emergency (gun shot and siren), whose events are artificially mixed with the background noise considering different SNR values and using a classification window length of 23.2 ms. The audio data are differentiated between foreground and background, depending on their acoustic salience. In order to study the characteristics of both datasets, an AED is also implemented and tested following a 10-fold cross-validation scheme. The AED is based on MFCCs together with other statistically derived features, and some classifiers provided by the Weka data mining platform. Recently, in [14] the authors have presented an AED following a local and global features aggregated based on a mixture of experts model, which is also tested using the UrbanSound8k dataset.

Foggia et al. have recently presented an adapted approach of their AED, focused on the detection of specific anomalous noise events, such as screams, glass breaking or gunshots [26] to detect abnormal sounds within urban traffic noise, such as tire skidding and car crashes using a low-cost hardware platform for urban surveillance [27]. A bag-of-words of sounds representation is used to perform

AED after training a pool of SVM-based classifiers that consider different feature extraction techniques with a window time frame of 32 ms. The authors compare the results over both real-life and synthetic acoustic dataset, and the system shows that the MFCC plus the proposed classification approach presents the best performing configuration according to the ROC curves.

In [16] the authors design a classifier which goal is also to separate between target and interfering noise. The activity of the noise source is detected by means of a binary classifier discriminating between the target, which can be plant or aircraft noise, and the background, which can be traffic, wind, rain, thunder, etc. The algorithm is based on MFCC [21] 100 ms window length feature extraction with the classification using a supervised classifier (GMM and Artificial Neural Networks (ANN)), trained with an annotated real-life dataset.

Finally, and in the framework of the DYNAMAP project [19], a preliminary study on the Anomalous Noise Event Detector (ANED) for high capacity sensors was developed to differentiate between road traffic noise and anomalous noise events in both urban and suburban environments [15]. It was based on a two-class audio event classification approach to be implemented in a low-cost acoustic sensor of a WASN. The algorithm and the experiments were conducted following the project operating specifications, using raw real-life data collected from both urban and suburban environments by means of a recording campaign [28]. The results prove the viability of the two-class classification scheme to detect ANE, using a window of 30 ms, and MFCC to parameterize the audio, and GMM as probabilistic classifier, outperforming the OCC counterpart in both sampled urban and suburban environments.

### *2.2. Networks for Noise Monitoring*

In this section, we review several representative approaches developed to automatically measure the noise levels in cities to tailor noise maps [5]. The analysis considers both the hardware approach and the platform design, with a special focus on hybrid network hardware proposals, where both Hi-Cap and Lo-Cap sensors are included in the network. The main goal of most of the networks is to measure and integrate the calculated $L_{Aeq}$ for a certain interval of time in a map. This analysis has divided the sensors into three categories: (i) high accuracy and low noise floor acoustic sensors, with high price of the sensor nodes in a WASN, (ii) low-cost nodes to design a WASN balancing the accuracy and the price and (iii) hybrid WASN, using both Hi-Cap and Lo-Cap nodes, balancing the cost and the efficiency of the data processing in the network.

The first category of WASN that can be found in the literature is built to achieve high accuracy and reliability, together with low noise floor. To that effect, most of their acoustic sensors are monitoring stations from Bruel and Kjaer [29] or Larson and Davis [30], which are equipped with IEC class 1 microphones. Those projects working with this kind of sensors are mainly focused on performing a detailed study of the acoustic environment of the city of interest. In [13,31], the FI-Sonic project based on the FIWARE platform is described; it consists of an acoustic sensor network based on ambisonics microphones, a multichannel acquisition card (from 2 to 128 GB), a network interface (with a Wi-Fi/3G modem) and a media server, its main processing unit, which runs the audio analyses. The collected information is used to create quasi-real-time dynamic noise and event maps, as well as to identify specific pre-trained sound sources for surveillance purposes. The FI-Sonic project is an example of the application of high accuracy WASNs to noise monitoring, but its pervasive deployment will require a very high investment. The problem associated with this first category of WASN is the price of the deployment of an entire network with several nodes, which may become prohibitive.

A second category of acoustic sensor networks is designed to balance the accuracy and the cost of the network. These WASNs are usually designed to be deployed in large networks, and the priorities in their design are not only price and accuracy, but also allow the possibility of processing real-time the acoustic signal in each node of the network. Some of the networks in this category are based on commercial sound level meters, such as the one used to monitor the traffic noise in Xiamen City (China) [32]. The designed WASN also considers ZigBee technology and GPRS communication, and all the nodes of the network use the same type of device. In the SENSEable project [33], a WASN was

proposed to collect information about the acoustic environment of the city using low-cost acoustic sensors to study the relationship between public health, mobility and pollution through the analysis of citizens' behaviour. Also within this second category, we can find several WASNs which can be deployed in a pervasive manner, such as the ones of the IDEA [6] and the MESSAGE [34] projects. They are based on a single board computer with low computational capacity, using low-cost sound cards. This hardware choice permits the deployment of large sensor networks due to the low economic cost of each node, besides allowing the collection of relevant environmental data from several critical locations in the city. In the IDEA project [35], a cloud-based platform is also developed by integrating an environmental sensor network with an informative web platform, which aims to measure noise and air quality pollution levels in urban areas in Belgium. Most of the aforementioned monitoring projects were only focused on measuring the $L_{Aeq}$ values, therefore, the nodes are only required to conduct their computation. When the application requires higher complexity in the processing of the acoustic signal, the computational capability of the nodes should be increased accordingly. Nevertheless, some of the acoustic sensor designs of this second category are developed ad-hoc for each project. Some projects even design the nodes of the WASN to conduct some signal processing over the acoustic raw data in the proper node of the WASN. In [36], the urban sound environment of New York City is monitored using a low-cost static acoustic sensing network.

The third category would be hybrid WASNs, which arise as a good trade-off between cost and required features of scalability, reliability and flexibility [5]. This typology of acoustic networks include both high capacity nodes, which are able to perform signal processing algorithms in real-time against the recorded data, and low capacity nodes, whose main goal is to compute the $L_{Aeq}$ value. This architecture is cheaper than the one only composed of high-capacity sensors, because the high capacity nodes are only deployed where needed, and it is scalable to wide zones at a lower cost. This hybrid architecturealso makes it possible to sense remote zones with no access to power supply by means of the use of a low-capacity sensor fed by solar panels. In [11], the authors present the design of an acoustic sensor network based on this approach. The hardware platform for a basic node is a low power μC whose main goal is to compute the $L_{Aeq}$ and transmit the collected $L_{Aeq}$ data periodically. The advanced nodes allow far more processing capabilities in comparison with the basic ones, since they use a small PC with a 2 GHz Intel Atom Processor running a Linux operating system. The advanced nodes can both store and process the acoustic data, which are designed to be flexible for the necessary signal processing analyses. Also in [37], the authors obtain the measurements of the RUMEUR project from sound level meters installed in a sensor network that pursues the understanding the measured signal, assess actions to mitigate noise and communicate the information about the soundscape in Ile-de-France. The RUMEUR hybrid networks [18] includes both high accuracy equipment for critical places, like airports, where the focus is to obtain detailed acoustic information due to the intense noise environment, together with less precise measuring equipment in other locations aimed at only updating of the noise map with the corresponding $L_{Aeq}$ level. Achieving a good trade-off between cost and accuracy is also the core idea of the WASN design in the DYNAMAP project [19,38]. This project is aimed at the deployment of a low-cost WASN in two pilot areas in Italy, located in Rome [39] and Milan [40], so as to evaluate the noise impact of road infrastructures in suburban and urban areas, respectively. The sensors designed for the DYNAMAP project are low-cost and use class 2 MEM microphones. This hybrid WASN will deploy two types of sensors: (i) high capacity ARM-based sensors, allowing signal processing techniques to analyze and process the acoustic signals in each node [38], and (ii) low capacity μController (μC)-based sensors, with less computational capabilities and fed by solar panels, but more flexible in terms of sensor positioning, maximizing the coverage of the network.

## 3. ANED Lo-Cap: An Anomalous Noise Event Detector for Low-Capacity Acoustic Sensors

Section 2 describes several AED algorithms developed to identify noise sources. However, the particular requirements of the DYNAMAP project pose the need of designing and implementing an

algorithm capable of detecting ANE within RTN in real-time. The main goal of the Anomalous Noise Event Detector described in [15] is identifying ANE dynamically to allow a reliable representation of the A-weighted equivalent noise level of RTN (see Figure 1) in outdoor acoustic environments. This section describes our proposal for adapting the ANED Hi-Cap [15] to run real-time in Lo-Cap sensors, following a similar classification principle to discriminate between ANE and RTN in order to obtain consistent results across the hybrid WASN.

In this section, we firstly overview a general description of the method; then, we describe the acoustic signal parametrization used to characterize the RTN and ANE acoustic signals. We finally detail the first stage of the optimization process of the ANED Lo-Cap by means of studying the spectral energy distribution of both ANE and RTN acoustic categories.

### *3.1. General Description*

The ANED Lo-Cap algorithm proposal is mainly focused on the reduction of the computational complexity of the ANED algorithm [15], while maximizing its capability of identifying ANE to discard them from the RTN noise level computation. In order to implement the algorithm in an affordable Lo-Cap platform with a µC core, a change of paradigm on the AED is mandatory, both in terms of feature extraction and classification, to assume the severe decrease of the computational load of the Lo-Cap platform with respect to the Hi-Cap counterpart. As studied in [22], the distribution of the energy per subband show signs of separability depending on the frequency between ANE and RTN, and the proposal presented in this piece of research deepens the analysis in this direction.

In Figure 2, the block diagram of the acoustic signal processing within the Lo-Cap acoustic sensor is depicted. It includes both the ANED Lo-Cap and the computation of the $L_{Aeq}$. The upper part of the Figure, corresponding to the ANED Lo-Cap, details that the algorithm starts windowing the acoustic signal registered in the sensor before conducting the Fast Fourier Transform (FFT) [41] on each input acoustic frame. After that, specific Mel frequency subbands [21] of a pre-studied frequency range outputs are computed as a simple frequency-based scalar product between the squared FFT module and the MFS filter. Next, the spectral energy should be computed from the discriminant range in frequency is compared to a pre-calculated threshold (obtained by means of an optimization process), leading to the classification of each input frame as RTN or ANE accordingly. In parallel, the $L_{Aeq}$ is obtained through an A-weighted filter applied to the input acoustic signal and subsequently computing the equivalent energy level in dB.

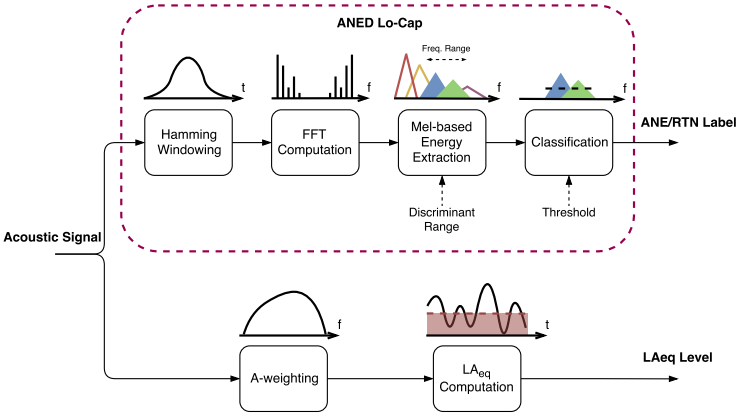



**Figure 2.** Block diagram of the acoustic signal processing within the Lo-Cap acoustic sensor. The upper part details the block diagram of the ANED Lo-Cap, with a binary label as an output, and the lower branch details the evaluation of the $L_{Aeq}$ of the measured acoustic signal.

### *3.2. Acoustic Signal Parameterization*

The ANED designed for the Hi-Cap acoustic sensors [15] used MFCC [21] to parametrize the input acoustic data. With the same basis, the ANED Lo-Cap uses a simplified frequency-based representation based on a Mel filter-bank analysis in order to maintain the homogeneity of the signal analysis in all the nodes of the hybrid WASN while reducing the computational cost of the signal parametrization. The use of the Mel-based frequency scaling simulates the way the human ear works, having a higher resolution in low than in high frequencies [21], which is a good approach for the spectral distribution of the sounds under study [42,43]. The last stage of the MFCC computation which is basically used for dimensional reduction of the obtained feature vector is here omitted, thus, the classifier uses directly the output energies of the pre-selected MFS.

### *3.3. Optimization of the ANED Lo-Cap Configuration*

In this section, we describe the process to obtain the frequency region that maximizes the discrimination between ANE and RTN and also the methodology followed to set the decision threshold. Firstly, the MFS distribution is analyzed to obtain the most suitable frequency range in terms of class separability. This separability is evaluated by means of the results of probability of error for each MFS studied subband, which will be detailed in Section 3.3.1.

The block diagram of the optimization process is shown in Figure 3. The first stage comprises the MFS parameterization of the entire acoustic dataset, which comprises the windowing of the acoustic signal using a Hamming window [44] of length $T_w$, the FFT computation of each acoustic frame and the calculation of the output energies ($E_i, 1 \leq i \leq M$) of a $M$-Mel-based filter bank using a simple scalar product in the frequency domain. Next, two Probability Density Functions (PDFs) of the logarithmic signal level $x$ (in dB) at each $i$ frequency subband are computed, taking into consideration all the signal frames of the dataset that belong to each acoustic class: one for the ANE class, $pdf_{ANE}(x,i)$, and another for the RTN class, $pdf_{RTN}(x,i)$. These two PDFs are the basis to obtain the one-band threshold-based linear discriminators for classification purposes, through which an error probability function can be derived in terms of each of the $M$ frequency subbands, $p_e(i)$. This leads us to obtain the most suitable frequency range to discriminate, and thus, classify an input acoustic frame as ANE or RTN.

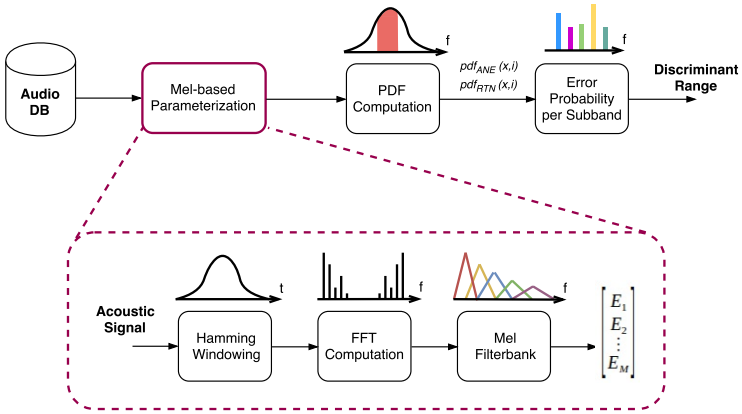



**Figure 3.** Block diagram of the process conducted to select a frequency range with minimum probability of error of classification.

In Section 3.3.1, the procedure used for computing an error probability function per subband ($p_e(i)$) is elaborated, with the goal of obtaining the subbands that improve the classification. Then, Section 3.3.2

explains the criteria used to select the best frequency range to use for the ANED Lo-Cap based on the evaluation of this error probability function.

3.3.1. Computation of the Probability Error Function

Given the pair of PDF distributions, $pdf_{ANE}(x,i)$ and $pdf_{RTN}(x,i)$, which describe the logarithmic energy density distribution of ANE and RTN classes, respectively, in terms of the frequency subband $i$ (see Table 1), two types of total error functions are computed based on two hypotheses:

**Hypothesis 1 (H1).** *class ANE has signal energy levels that are above those of class RTN. For this hypothesis, the total error function is defined as* $p_e^{ANE>RTN}(i)$.

**Hypothesis 2 (H2).** *class RTN has signal energy levels that are above those of class ANE. For this hypothesis, the total error function is defined as* $p_e^{ANE<RTN}(i)$.

**Table 1.** Relation between the $M = 48$ subbands used from the Mel filter-bank parameterization and its central frequency in Hz.

| # of Subband | Freq. (Hz) | # of Subband | Freq. (Hz) | # of Subband | Freq. (Hz) | # of Subband | Freq. (Hz) |
|---|---|---|---|---|---|---|---|
| 1 | 86.7 | 13 | 886.7 | 25 | 2484.6 | 37 | 7231.5 |
| 2 | 153.3 | 14 | 953.3 | 26 | 2715.9 | 38 | 7904.8 |
| 3 | 220 | 15 | 1020 | 27 | 2968.8 | 39 | 8640.9 |
| 4 | 286.7 | 16 | 1115 | 28 | 3242.2 | 40 | 9445.4 |
| 5 | 353.3 | 17 | 1218.8 | 29 | 3547.4 | 41 | 10,324.9 |
| 6 | 420 | 18 | 1332.3 | 30 | 3877.7 | 42 | 11,286.3 |
| 7 | 486.7 | 19 | 1456.3 | 31 | 4238.7 | 43 | 12,337.2 |
| 8 | 553.3 | 20 | 1591.9 | 32 | 4633.4 | 44 | 13,485.9 |
| 9 | 620 | 21 | 1740.2 | 33 | 5064.9 | 45 | 14,741.6 |
| 10 | 686.7 | 22 | 1902.2 | 34 | 5536.5 | 46 | 16,114.3 |
| 11 | 753.3 | 23 | 2079.3 | 35 | 6052 | 47 | 17,614.7 |
| 12 | 820 | 24 | 2272.9 | 36 | 6615.5 | 48 | 19,254.8 |

Both error functions also depend on the energy level threshold $\gamma$ that will be used to discriminate between the two classes, which are defined as follows:

$$\begin{aligned} p_e^{ANE>RTN}(i,\gamma) &= P_{ANE}\int_{-x_{min}}^{\gamma} pdf_{ANE}(x,i)dx + P_{RTN}\int_{\gamma}^{x_{max}} pdf_{RTN}(x,i)dx \\ p_e^{ANE<RTN}(i,\gamma) &= P_{RTN}\int_{-x_{min}}^{\gamma} pdf_{RTN}(x,i)dx + P_{ANE}\int_{\gamma}^{x_{max}} pdf_{ANE}(x,i)dx \end{aligned} \quad (1)$$

where $P_{ANE}$ and $P_{RTN}$ are the *a priori* probabilities of class ANE and class RTN, respectively, and $x_{min}$ and $x_{max}$ are the minimum and maximum observed signal energy level at frequency subband $i$, respectively.

For each $i$th-frequency band, the optimum energy threshold $\gamma$ is defined as the one attaining the minimum value of the corresponding error function for each hypothesis ($\gamma_{opt}^{H1}(i)$ and $\gamma_{opt}^{H2}(i)$). As it can be observed in Equation (1), the total error functions are computed as the sum of two type of errors (*type I*: classifying ANE when class RTN, *type II*: classifying RTN when class ANE), and their evaluation will differ depending on the way the PDF is evaluated. For each $i$th-frequency band, the minimum error is computed for both error functions substituting its corresponding optimum threshold. The final decision is determined by considering the most feasible hypothesis, whether to accept H1 or H2. Thus, the final error probability function $p_e(i)$ is defined as the minimum of both error functions ($p_e^{ANE>RTN}(i,\gamma)$ and $p_e^{ANE<RTN}(i,\gamma)$) for each frequency band (see Equation (2)), and the optimum decision threshold function (see Equation (3)) is computed using the corresponding optimum threshold for each frequency subband index $i$.

$$p_e(i) = \begin{cases} p_e^{ANE<RTN}(i, \gamma_{opt}^{H2}) & \text{if } p_e^{ANE>RTN}(i, \gamma_{opt}^{H1})) > p_e^{ANE<RTN}(i, \gamma_{opt}^{H2}) \\ p_e^{ANE>RTN}(i, \gamma_{opt}^{H1})) & \text{otherwise} \end{cases} \quad 1 \leq i \leq M \qquad (2)$$

$$\gamma_{opt}(i) = \begin{cases} \gamma_{opt}^{H2}(i) & \text{if } p_e^{ANE>RTN}(i, \gamma_{opt}^{H1})) > p_e^{ANE<RTN}(i, \gamma_{opt}^{H2}) \\ \gamma_{opt}^{H1}(i) & \text{otherwise} \end{cases} \quad 1 \leq i \leq M \qquad (3)$$

3.3.2. Selection of Frequency Range

Once the error probability function $p_e(i)$ has been obtained for a given acoustic dataset (see Equation (2)), its evaluation can be used to define a suitable frequency range use in the classification in terms of ANE and RTN discrimination capabilities [22] (see the last module in Figure 3). This selection of frequency range is defined following the next simple rule: including those MFS subbands for which the error probability function is less or equal than a certain probability of error threshold. The computation of the decision threshold of the last stage of the ANED Lo-Cap (see Figure 2) follows a similar procedure of that used for the one-band linear discriminators [45], and should be conducted ad-hoc after the calculus of the probability of error.

## 4. Experimental Section

This section details the experiments conducted to validate the viability of the ANED Lo-Cap algorithm in terms of probability of error and computational requirements within the framework of the DYNAMAP project. The tests are conducted using two real-life acoustic datasets from two different outdoor acoustic environments: urban and suburban. The probability density functions are obtained for both scenarios, and the most discriminative frequency subband ranges are pointed out.

### *4.1. Acoustic Database*

In this section, the acoustic database analyzed to validate the low-capacity version of the ANED is briefly described in [46]. This database includes two different acoustic scenarios regarding road traffic noise monitoring: a suburban dataset including a major road (along the A90 highway surrounding Rome); and an urban dataset (within the district 9 of Milan), which comprises different types of roads, as well as several traffic density conditions. The two acoustic datasets were obtained using low-cost acoustic sensors coupled to a ZOOM H4n digital recorder (ZOOM, Hauppauge, NY, USA), and considering a 48 kHz sampling rate with 24 bits/sample. As a result of the four-day recording campaign between the suburban and urban scenarios, a total of 9 hours and 8 minutes of audio were collected.

Subsequently, a manual labelling process was conducted on the collection of audios gathered from both real-life scenarios, which entailed exhaustive listening by experts and the subjective classification of the acoustic data between RTN (differentiating between road traffic and background city noise when it was not perceived the presence of vehicles) and ANE (considering up to 19 labels to describe different noise sources, such as vehicle horns, airplanes, people talking, sirens, music in the street, noise coming from train or tramways, etc.).

ANEs were also labeled in terms of their acoustic salience with respect to the surrounding noise by firstly obtaining an estimation of the Signal-to-Noise Ratio (SNR in dB), following the computation approach explained in [20]. This provides information which is valuable when it comes to considering or rejecting ANEs with low SNR (e.g., $SNR \leq 0$) when training the ANED as stated in [20], since it does not have a significant impact on the final $L_{Aeq}$ computation.

In this work, the input acoustic signal is segmented into 30 ms frames using a Hamming window with 50% of overlap [44]. The input acoustic frame is transformed to the frequency domain through the FFT [41] after being sampled at 48 kHz. Then a filter-bank of $M = 48$ Mel filters is applied to the spectrum of the signal's square module, covering the entire human audio range from 20 Hz to

20 kHz [47], obtaining the corresponding MFS. The computational load of the signal processing is reduced when applying the filter-bank in the frequency-domain compared to the alternative convolution in the time-domain [48]. The central frequency of each of the 48 MFS used for the ANED Lo-Cap version—which follow a logarithmic distribution—are shown in Table 1.

### *4.2. 2D-PDF Subband Analysis and Selection*

This section details the computation of the PDFs in the framework of the DYNAMAP project, considering the two real-life acoustic datasets from the urban and suburban environments, respectively. These PDFs will permit us to design for each Mel-frequency band a one-band threshold-based linear discriminator for binary classification purposes, and an error probability function will be derived for all subbands in order to observe which frequency range is the most suitable for each environment.

#### 4.2.1. Computation of the PDFs for Both Scenarios

In this section, the resultant energy level PDFs are obtained for both signal classes, RTN and ANE, by considering the real-life data from the two pilot site recordings. The two datasets, already parameterized through the MFS, have been used as the main input data for the computation. Energy level PDFs have been estimated using simple histograms for each of the $M = 48$ frequency bands of the MFS [15], obtaining two-dimensional functions (2D-PDF). For each spectral subband, the PDF is computed using a minimum of 200 samples per class (for accuracy reasons) within the range of observed values of the MFS energy level outputs. For this study, only those ANEs with $SNR \geq 0$ have been considered in order to minimize the potential confusions between RTN and ANE that could minimize the events separation probability. More details about the evaluation of the SNR in all the ANE in the dataset can be found in [20].

In Figure 4, the 2D-PDF matrices corresponding to the MFS distributions of ANE and RTN in the suburban scenario are depicted. For illustrative purposes, all the 2D-PDF visualizations have been scaled at each frequency subband dividing each PDF by its maximum value in order to obtain a more comprehensible plot, so the absolute value of probability is normalized. Also, the optimum decision threshold (see Equation (3)) is plot as a solid white line in both axes across the frequency index.

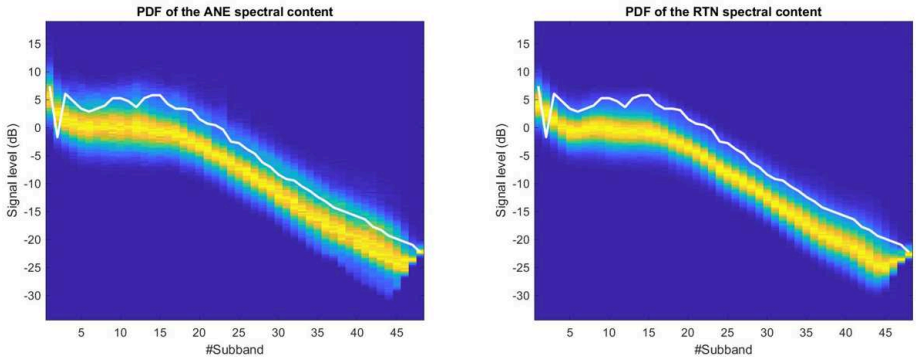



**Figure 4.** 2D-PDF of the ANE (**left**) and the RTN (**right**) of the suburban dataset. The frequency subband index *i* corresponding to the MFS is labelled in the x-axis (the reader can find the corresponding frequency in Table 1), while the corresponding logarithmic signal level at each subband is depicted in the y-axis in dBs. The Figures colormap is blue for lower probabilities while tends to warm colors (with the maximum in red) for higher probabilities. The solid white line represents the optimum decision threshold for the one-band linear discriminant classifier at each frequency bin.

It is worth noting the high similarity of the 2D-PDFs of both ANE and RTN plots for the suburban acoustic environment, showing also an important overlap between them. However, the ANE class presents a slightly wider variance than the RTN class along the signal level axis for most of the MFS (i.e., an increase of about 10 % in the variance values for the ANE class is found with regard the RTN variance). However, a decrease of this signal level dispersion of both distributions can be observed along the frequency subband indexes from **23** to **27**. Then, within those subbands the overlap of both signal level distributions (ANE and RTN) are lower, being the ANE class the one that attains higher signal levels (e.g., its 2D-PDF maximum probabilities, represented in yellow color, are placed closer to the optimum decision threshold). Another subband to analyze is number **2**, which presents a lower value than subbands **1** and **3** in the analysis of both ANE and RTN. Subband **2** corresponds to a central frequency of 153 Hz, which is one of the most common frequency ranges of road traffic noise [42,43], and yet in this location the curve of the threshold adapts to the best possible discrimination between ANE and RTN following Equations (2) and (3).

Figure 5 shows the 2D-PDF matrices corresponding to the urban recordings. This acoustic scenario presents clearer differences between the 2D-PDF plots of the ANE and RTN classes compared to the suburban environment. Moreover, evidences of two clear spectral patterns can be observed in the ANE 2D-PDF: the first one located at high energy levels, and the second one that entails lower energy levels, especially from subband **27** and higher frequencies. In addition, differences between ANE and RTN are more evident in the lower frequency band (e.g., Mel subbands **5** and **6**). Subband **2** in the urban environment follows the same performance as in the suburban, and the threshold is set to a lower value than the neighbour subbands in order to minimize the probability of error.

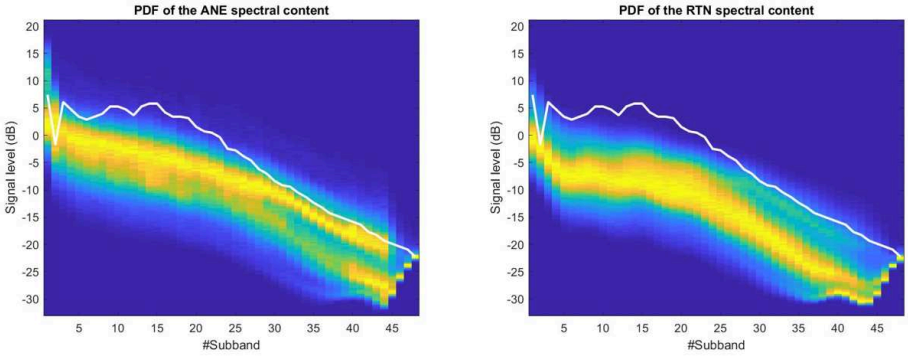



**Figure 5.** 2D-PDF of the ANE (**left**) and the RTN (**right**) of the urban dataset. The frequency subband index $i$ corresponding to the MFS is labelled in the x-axis (the reader can find the corresponding frequency in Table 1), while the corresponding logarithmic signal level at each subband is depicted in the y-axis in dBs. The Figures colormap is blue for lower probabilities while tends to warm colors (with the maximum in red) for higher probabilities. The solid white line draws the optimum decision threshold for the one-band linear discriminant classifier at each frequency bin.

#### 4.2.2. Subband Error Probability Calculation

After the 2D-PDF computation for each use case and type of signal, a threshold optimization is performed for each and every MFS, being $M = 48$ (see Section 3.3). The calculated threshold can be observed as a white solid line in both Figures 4 and 5. This threshold enables us to evaluate the probability of error for each of the MFS, and so conclude in which frequency region the separability between ANE and RTN is optimum to maximize the accuracy of the subsequent classification.

In Figure 4, we can observe a high overlap between the two PDF functions that lead to a hardly stable threshold function, showing a smooth behaviour across the $M = 48$ frequency subbands.

Figure 6 shows the plot of the final error probability function defined in Equation (2) for the suburban dataset. The absolute minimum error probability is achieved in band **23**, being 0.1673. The error probability exhibits also a frequency region with a second local minimum value of 0.17, around subband **26**, and another third local minimum around subband **18**. This leads us to conclude that in the suburban environment, the most suitable frequency region to take into account in terms of separability can be found within the mid-frequency range.

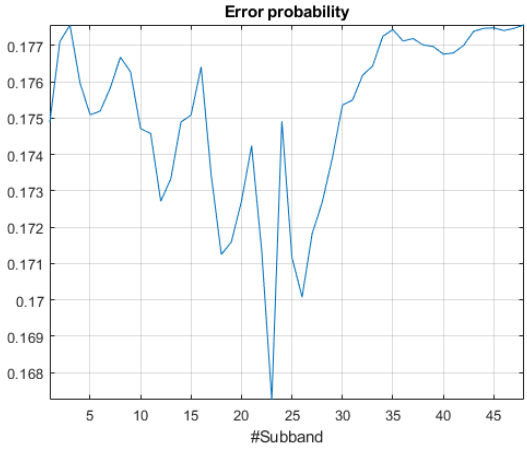



**Figure 6.** Error probability analysis for suburban recordings data.

Figure 5 shows the results of the threshold for the urban acoustic database plotted in solid white line. Figure 7 shows that the minimum error probabilities are obtained for subbands **5** and **6**, presenting the minimum value ($p_e^{ANE>RTN}(x) = 0.31$ in the 6th subband, see Equation (1) for more details). Moreover, subband **1** also yields a third local minimum of the computed error probability. This leads us to conclude that in the urban environment, the low-frequency region is the most suitable to discriminate between ANE and RTN.

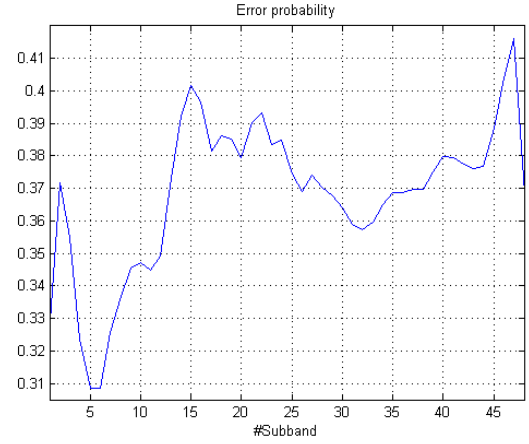



**Figure 7.** Error probability analysis for urban recordings data.

## 5. Operations Cost Analysis of the ANED Lo-Cap

In this section, we evaluate the operations cost of the ANED Lo-Cap to determine the minimum features of the hardware platform required. First, the audio acquisition restrictions are detailed and, after that, the audio processing computational costs are calculated for the three main stages of the ANED Lo-Cap proposal: windowing, Fast Fourier Transform (FFT) [41] computation and subband filtering. After the overall energy that includes the optimized subbands is obtained, a comparison with a decision threshold is conducted to finally classify between RTN and ANE.

### *5.1. Audio Acquisition*

The computational cost of the audio acquisition cannot be simply calculated as the number of floating point operations, since it depends on the architecture of the platform where it has to perform real-time. However, the μC or the processor will require a certain time dedication to the audio data reading process. If the ANED Lo-Cap is implemented on an ad-hoc hardware solution, a buffer can be used to store the audio data, and the data from the buffer would be input to the processing unit. However, if the data should be retrieved directly from the sensor, i.e., using an Analog-to-Digital Converter (ADC), it will require the processor to read periodically the data at the desired sampling rate. Therefore, the time cost of the audio acquisition system will increase.

In the best case, the audio acquisition system should be able to sample the input data at the same rate as the audio sampling frequency, in this case 48 kHz. However, a lower sampling rate could be considered if we assume that the ANED Lo-Cap will only detect low-frequency ANEs and an anti-aliasing filter is implemented before the data acquisition.

### *5.2. Acoustic Signal Processing*

In order to extract the energy for each subband, three steps should be followed. First, a time windowing of $N$ points is applied in order to extract the audio samples of the input frame. After that, the FFT is computed for the desired number of points, in this case, the same as the number of samples, i.e., $N$. Finally, the subband filter of $C$ coefficients is applied to the spectrum of the frame to calculate the energy of the subbands of interest. Table 2 synthesizes the minimum number of operations required to implement the ANED Lo-Cap using the triple stage procedure detailed in this section.

- **Windowing:** In order to analyze short frames of audio, a window function should be applied to the input signal in order to reduce the spectral leakage due to higher frequencies. In this implementation, the hamming window is used [44]. The computational cost associated to any windowing process depends on the number of samples of the analyzed frame if the window function is computed and stored in advance. The ANED Lo-Cap proposal uses time frames 30 ms long, thus, if the sampling frequency is 48 kHz, the window will be 1440 samples long.
- **FFT Computation:** The FFT is one of the most popular algorithms that computes the DFT (Discrete Fourier Transform) of a sequence reducing its complexity by factorizing the DFT matrix. The most used algorithm is the Cooley-Turkey [41], that breaks the down the DFT of $N$ points into smaller ones, typically dividing it in two pieces of $N/2$ at each step. The computational cost of the FFT may vary depending on $N$ (the number of points of the FFT) and on the methodology of implementing the algorithm over a certain hardware platform and its optimization. In our case, the FFT shall be of minimum 1440 points and maybe of 2048 after adding zero-padding if the used algorithm requires a power-of-2 size.
- **Sub-band Filtering:** After the FFT is computed, a triangular-shaped filter is applied to a determined subband. The computational cost of obtaining each filtered subband depends on the number of coefficients ($C$) of the filter, which, in its turn, depends on the sampling frequency and the number of points of the FFT. The filter is used to obtain the energy of the subband, hence, the computational cost should consider the point-to-point multiplication of the vector and the filter and the posterior integration of the resulting vector. In order to reduce the computational

cost, the filter may be designed in advance considering the sampling frequency and the number of points of the FFT. After that, only a product for each bin followed by a sum of all resulting outputs will be needed. The number of operations can be reduced if the filter is only employed in the concerning subbands and all other frequencies are omitted. In our case, two Mel subbands shall be implemented as it is the combination with a lower probability of error.

Depending on the efficiency of the algorithm implementation over every type of hardware platform the total computing time will tend to this evaluation or will be higher, if the implementation is not optimum. It could even be lower if the algorithm is optimized for a determined platform and several operations were omitted or replaced by hardware.

**Table 2.** Computational cost analysis of the audio processing.

| | **Additions** | **Multiplications** | **Floating Point Operations** |
|---|---|---|---|
| Windowing | 0 | $N$ | $N$ |
| FFT | $N \cdot log_2(N)$ | $\frac{N}{2} \cdot log_2(N)$ | $\frac{3}{2}N \cdot log_2(N)$ |
| Subband filtering | $C-1$ | $C$ | $2 \cdot C - 1$ |

### *5.3. Commercial Board Comparison*

In this section, we detail the overview of several commercial platforms that could be used to implement the ANED Lo-Cap. We evaluate their features and their computational capacity to host the performance of the ANED Lo-Cap. Nevertheless, in the framework of the DYNAMAP project an adhoc low-capacity platform is being developed to host the slave nodes of the WASN in both pilots.

The choice of a commercial board as the hardware platform capable to run the ANED Lo-Cap is delimited by its audio acquisition and processing requirements. The typical commercial boards can be classified according to the core processing system, which could be a µC, a µProcessor (µP) or a Field-Programmable Gate Array (FPGA).

The economic µC-based boards present many drawbacks, but they could serve as a first approximation to the ANED Lo-Cap. As an example, an Arduino is used for this viability study. According to the tests conducted in [49], the analog read function in Arduino in a loop takes 39 µs, making it impossible to sample at more than 25 kHz in the ideal scenario and using a full dedication of the controller. In this case, the FFT computation in real-time would be impossible. However, a fine tuning of the board registers could improve the reading speed to a 4 µs and a higher sampling rate could be obtained. Still, in order to work real-time, a cooperative implementation of the FFT and the ADC read should be coded and a deep understanding of the internal registers should be achieved by the developer. The implementation, if possible, would have a double restriction, i.e., the sampling frequency and the number of points of the FFT.

A second approximation could be conducted with ARM-based boards, where ARM stands for Advanced RISC Machine. In this case, the processing capacity allows the FFT computation and some boards, e.g., Raspberry Pi, even allow the acceleration of the FFT using the Graphics Processing Unit (GPU) [50], facilitating the computation and freeing the Central Processing Unit (CPU). However, not all these boards implement an ADC, complicating the task of acquiring the audio. The Raspberry Pi, for example, does not include an internal ADC and an ad-hoc system must be included. Nevertheless, other ARM-based µC boards offered by NXP [51] and STM [52], have ADCs that sample over 1 MSPS, enough for audio acquisition.

A third approach could be using FPGA-based platforms, which entail a completely different architecture and programming paradigm. The processing capacity of the FPGAs is limited basically by the space and it is suitable for parallel executions but it is not optimal when programming complex sequential tasks. In our case, the FFT computation can be easily implemented by using the manufacturer libraries, as the offered by Xilinx (San Jose, CA, USA) [53]. Also, several platforms offer high-speed internal ADCs, as some Xilinx boards or the Arty A7 [54], also built based on a Xilinx FPGA.

From the detailed approaches, the most economic one does not fulfil the basic requirements (e.g., Arduino), a 48 kHz sampling rate with real-time subband energy extraction. However, both microprocessor and FPGA-based boards are able to carry the needed operations in real time. The final decision could be supported by the information in Table 3, which lists the basic price of each commercial hardware platform considered; the first platform that can satisfy restrictions is the Raspberry Pi Model A+, and the price is $20, which is the same price as the Arduino. It includes ad-hoc audio input, which would allow to receive the audio in real-time with no need to constantly stop the processor to read new data.

**Table 3.** Price comparison for the hardware platforms described and algorithm they can assume real-time.

| Hardware Platform | Base Price | Supported ANED Version |
|---|---|---|
| Arduino Uno R3 | from $20 | None |
| Raspberry Pi Model A+ | from $20 | ANED Lo-Cap |
| NXP Semiconductor FRDM-K66F Freedom Board | from $69 | ANED Lo-Cap & Hi-Cap |
| Arty S7: Spartan-7 FPGA | from $109 | ANED Lo-Cap & Hi-Cap |

The ANED Lo-Cap applies only a reduced number of Mel filters and does not compute the Discrete Cosine Transform (DCT) and the GMM of 13 coefficients. Therefore, the computational load is reduced in a factor of around 6 in comparison to the ANED Hi-Cap. In this case, the authors have estimated that the Hi-Cap version of the ANED could no run in this platform as the computational load is six times higher and it requires to have remote access, storage and other monitoring tools to control all the connected Lo-Cap slave nodes. The other platforms, although they could be considered to be low-cost, have higher starting prices, which considering the deployment of a network with a lot of nodes, may alter substantially the global budget.

## 6. Discussion

In this section, we discuss whether the use of an ANED Lo-Cap in the low-capacity nodes of an hybrid WASN is feasible and reach the real-time hardware and accuracy requirements by considering three different aspects. The first one deals with the accuracy results in terms of the Macro-averaged F1 measure; in this case, the evaluation takes into account full-band analysis and random threshold selection as reference and the ANED Hi-Cap counterpart applied to the same acoustic data. The second one deals with the homogeneity of the network performance. Finally, the third one analyses the computational load of the proposed algorithm and its real-time performance in the studied low-cost hardware platforms.

### *6.1. Classification Accuracy of ANED Lo-Cap vs. ANED Hi-Cap*

This article focuses on the design of an ANED Lo-Cap for low-cost platforms and the evaluation of its viability using real-life data, but leaving for future works the development of, exhaustive tests in a real-operation environment. Nevertheless, we have conducted some preliminary analyses to establish a performance baseline for the two studied scenarios, following a 4-fold cross validation scheme.

The results in the suburban scenario considering all the Mel-based subbands ($M = 48$) yield a Macro-averaged F1 measure [55] of 51.6% (with a $\sigma_{F1} = 0.37$). Although this accuracy is quite low, it is to note that it is indeed higher than the one obtained using a non-optimized threshold, i.e., a threshold based on an uniform random selection within the range of the measured signal energies. In this case, the ANED Lo-Cap accuracy decreases up to 50.4%. If we compare the F1 values obtained by the ANED Lo-Cap and the ANED Hi-Cap [15] in this scenario, we can observe that the subband optimized ANED Lo-Cap is around 9% less accurate on average, but allowing a computational load decrease in an order of 6 times (see Section 5). Finally, we want to note that some preliminary tests considering subband selection show promising results, with an averaged increase of around 3% in the F1 measure with respect to the full-band ANED Lo-Cap.

The results in the urban scenario yield a Macro-averaged F1 measure of around 62% (with a $\sigma_{F1} = 0.17$) considering the full Mel-based frequency range. The ANED Lo-Cap presents better results in the urban than in the suburban environment (showing an increase of more than 10% in terms of accuracy); a pattern already observed for the ANED Hi-Cap version [15].

In this scenario, the non-optimized threshold baseline classification system decreases 4.8% the averaged F1 values with respect to the Lo-Cap ANED optimized proposal for the full-band configuration. Therefore, considering this result and the one for the suburban environment, we can conclude that the threshold optimization is a significantly valuable process for the proposed algorithm. Moreover, comparing these results with the ones obtained by the ANED Hi-Cap in the urban environment—reaching 72.7%, the proposed ANED Lo-Cap is also around 10 % less accurate in terms of the F1 measure. Morever, the preliminary tests including some kind of frequency range selection show a relevant improvement with an averaged increase of 5% in the F1 measure. Nevertheless, we shall study in more detail these results for the optimal implementation of the ANED Lo-Cap in a real-life operating framework.

The conducted experiments have considered a 30 ms frame-based classification analysis. In [15], a higher decision level was also included to obtain a binary output every 1 second by means of a majority voting scheme. As a result, some of the noisy decisions of the ANED Hi-Cap were reduced, as the majority vote works as a low-pass filter of the frame-based decisions, thus, improving the F1 results in around 5 %. It is to note that the majority vote has not been yet implemented in the ANED Lo-Cap, being left for a second stage of the development of the proposed algorithm, as it is not critical in terms of computational load in comparison with the entire proposal.

These preliminary results obtained for both environments encourage us to, at least, keep working to improve the algorithm for the urban scenario, where the initial Macro-averaged F1 values are clearly over 60% showing potential improvement after frequency range selection. The viability of the ANED Lo-Cap for the suburban scenario is not so clear due to the poor accuracy results (slightly higher than 50%); deep frequency selection experiments should be conducted to observe whether it is worth implementing it for the suburban scenario or not with the considered classification principle adaptation.

### *6.2. ANED Lo-Cap and Network Homogeneity*

The original idea of the ANED Lo-Cap design was based on the deconstruction of the phases of the ANED Hi-Cap algorithm [15], as the feature extraction, or the classification algorithm, in order to discard those that would involve a high computational cost, as well as maximizing the similarity of the classification approach considered by the original Hi-Cap version.

The study of the spectral differences between the ANE and the RTN [22] allows us to conclude that, for each acoustic scenario, the separability of the two types of signal is possible, especially if the potential algorithm focuses on specific frequency ranges, as observed in Section 4.2. The two evaluated acoustic scenarios, urban and suburban, do not present the same probability of error results in terms of separation of ANE and RTN and its associated MFS.

The analysis of the probability of error in the suburban scenario shows that the most distinctive spectral range correspond to the mid-frequencies (from 2 kHz to 2700 Hz approximately), while the lower discriminating ones comprise higher frequencies (from 3500 Hz to 24 kHz). In the suburban scenario, we mainly found three types of ANE related to natural phenomena (rain, wind and thunder), and four types of ANE [46]: sirens, horns, brakes and birds. All the most typical ANEs in the suburban scenario present spectral activity in the mid-frequencies and some of them even in the high-frequencies; therefore, its reasonable that the most accurate frequency range to distinguish between ANE from RTN in this scenario corresponds to that group of frequencies.

The results of the probability of error in the urban scenario show clear differences in terms of the spectral energy distribution in comparison with the suburban scenario. The most discriminating bands in this scenario correspond to the low-frequency region (from 33 Hz to 1 kHz), which is the range of frequencies where most of the RTN energy is located. The types of ANE found in an urban

environment, apart from the natural phenomena which are the same as in the suburban scenario, are wider than the ones found in the suburban environment [46]: airplanes, doors, people talking, tramways, trains, etc. Some of these ANEs have wider distribution of spectral energy, including low, mid and high frequencies. Airplanes, for example, have their spectral energy distribution centred in the low-frequencies, and so do trams and trains (together with the mid-frequencies). Thus, the low frequency range could be potentially the best discriminating spectral region for the urban environment, which should be confirmed with further experiments.

Nevertheless, the analyses conducted over the MFS encourages us to analyze fully the probability of error depending on the selected frequency region, in order to improve the accuracy results presented in Section 6.1 as well as to maximize the efficiency of the algorithm using the minimum required spectral integration.

### *6.3. Real-Time Implementation in a Low-Cost Platform*

In Section 5, it has been concluded that a commercial platform is able to support the computational load of the ANED Lo-Cap. However, not all boards are capable of processing the audio and making a decision in real time, as the cheapest option µC-based Arduino. Although this development board may be able to sample the audio, compute the FFT and apply the subband filtering, it would not be done in real-time. This condition requires that the FFT and the filtering process are conducted at the same time a new frame is read, implying that the µC should be programmed in a cooperative mode jumping sequentially between the ADC read and the FFT computation. It is possible to add an ad-hoc hardware that reads the audio and puts it in a buffer, easing the task of the µC to read at the sampling frequency. However, the system would still need to compute the FFT and take a decision within the 30 ms-frame rate. As a consequence, we discard this option to implement the ANED Lo-Cap due to the basic computational features that the platform offers.

Other hardware platforms, with higher capacity but not so higher price, offer a viable possibility to run the algorithm real-time. These platforms are mainly built around an ARM processor and usually include ad-hoc peripherals, as timers, ADCs, Ethernet interface and several (even dozens) of General-Purpose Input/Output (GPIOs). One of the most used board within this group is the Raspberry Pi 3 (from $35), which offers the possibility of audio acquisition (through the audio input) and computing the FFT in real time. In the future, several tests should be conducted to study the capability of the board and to find out the maximum number of points of the FFT that it is able to compute as well as the number of subbands.

In Section 5.3, other boards with high-speed internal ADC have been also studied, concluding that they could present a good efficiency in the implementation of the ANED Lo-Cap, e.g., NXP and STM, but at higher cost. The boards in this group have a starting price of $69 and have more specifications than the Raspberry Pi, however, these platforms are also larger, more expensive and with higher power consumption. These boards could implement both the ANED Lo-Cap and Hi-Cap, as they offer more computational capacity. Finally, the FPGA-based hardware platforms have also been analyzed and could be a suitable solution, thanks to the different programming paradigm, closer to hardware real-time signal processing restrictions. However, due to the complexity of the FPGA development boards and the particular specifications they have, they are in a higher price range, with the Arty S7 starting at $109. In this case, the Hi-Cap version of the ANED could be also implemented, as the FFT is a very common operation for the FPGA-based platforms and they do not have problems of parallelization.

In this work, the computational load of the ANED Lo-Cap has been studied in order to mount it using a commercial low-cost platform. However, other factors would influence a final implementation choice, as size, power consumption and supply, and compatibility with other peripherals depending on the needs of the project. Future research will include testing the algorithm over several hardware platforms proposed in this paper and evaluating the power consumption and performance.

## 7. Conclusions

In this work, we have analyzed the viability of an algorithm to implement an ANED Lo-Cap to run real-time in the Lo-Cap acoustic nodes of an hybrid WASN, following the same principle used for implementing the original ANED version designed for the high-capacity nodes. To that effect, the proposal is based on parametrizing the input acoustic signal with its Mel-based spectral energy distribution, and classified as RTN or ANE by means of a one-band threshold-based linear discriminator. The experiments have been conducted considering 9h and 8 min of real-life acoustic data from the suburban and urban pilot areas of the DYNAMAP project.

The main conclusion of the analysis conducted in this research is that the ANED Lo-Cap proposal is viable, both in terms of computational load and classification accuracy (at least, in the urban environment), maintaining a consistent performance with respect to the obtained by the Hi-Cap counterpart (i.e., higher Macro-averaged F1 measures for the urban environment than for the suburban scenario). From the computational analyses, we can conclude that the ANED Lo-Cap – that requires around $^1/_6$ of the computational load of the ANED Hi-Cap – cannot be implemented in the most basic platforms (e.g., Arduino), but the first upgrade of low-cost and Lo-Cap hardware, with small budget differences, can assume the ANED Lo-Cap running real-time (e.g., Raspberry Pi 3). Obviously, any platform with more computational capacity, such as FPGA, can assume the ANED Lo-Cap and even the ANED Hi-Cap. However, the use of high-performance computing platforms would increase the cost of the WASN, when one of the main goals is to deploy a low-cost wireless acoustic sensor network.

Regarding the classification accuracy of the ANED Lo-Cap, the research conducted in this work concludes that the first results in both scenarios decrease the F1 Macro-averaged measure in around 9–10% in comparison with the ANED Hi-Cap, and they need further optimization. Nevertheless, the performance of the ANED Lo-Cap proposal results in both suburban and urban environments is homogeneous with its own performance in ANED Hi-Cap results [15]. In the application in the suburban environment, considering the full Mel-based frequency range, they present values slightly higher than 50% in terms of Macro-averaged F1 measure. If these results cannot be improved, the usefulness of the baseline ANED Lo-Cap in this environment will be questioned. On the contrary, the application of the ANED Lo-Cap working with all the MFS as a baseline in the urban environment shows significantly better results (higher than 60% in terms of Macro-averaged F1 measure), which encourages us to keep working in order to optimize the frequency range selection to improve the obtained results. In this direction, future work will be focused on the study of optimum frequency ranges that allow obtaining improved accuracy values, which now state around 4% lower in the full-band preliminary tests. After that, the subsequent majority voting scheme will be included so as to provide the same output scheme of the ANED Hi-Cap every one second, following the DYNAMAP project requirements for real-life operation.

**Acknowledgments:** This research has been partially funded by the European Commission under project LIFE DYNAMAP LIFE13 ENV/IT/001254 and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (Generalitat de Catalunya) under grant refs. 2017-URL-Proj-013 and 2017-SGR-966. Ferran Orga thanks the support of the European Social Fund and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Catalan Government for the pre-doctoral FI grant No. 2017FI_B00243.

**Author Contributions:** Rosa Ma Alsina-Pagès wrote the paper and contributed to the design of the experiments. Francesc Alías contributed to the writing of the paper and helped in the design of the experiments. Joan Claudi Socoró performed the tests and reviewed the paper and Ferran Orga worked on the real-time capacity estimation of the algorithm.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| ADC | Analog to Digital Converter |
| AED | Acoustic Event Detection |
| ARM | Advanced RISC Machine |
| ANE | Anomalous Noise Event |
| ANED | Anomalous Noise Event Detection |
| CNOSSOS | Common Noise Assessment Methods |
| CPU | Central Processing Unit |
| DCT | Discrete Cosine Transform |
| DFT | Discrete Fourier Transform |
| DYNAMAP | DYNamic Acoustic MAPping |
| EC | European Commission |
| END | Environmental Noise Directive |
| EU | European Union |
| FFT | Fast Fourier Transform |
| FPGA | Field-Programmable Gate Array |
| GPU | Graphics Processing Unit |
| HMM | Hidden Markov Models |
| IIR | Infinite Impulse Response |
| LDD | Low Level Descriptors |
| MFCC | Mel-Frequency Cepstral Coefficients |
| MFS | Mel Frequency Subband |
| MSPS | Mega Samples per Second |
| OCC | One-Class Classifier |
| PWP | Perceptual Wavelet Packets |
| RTN | Road Traffic Noise |
| SNR | Signal to Noise Ratio |
| SVM | Support Vector Machine |
| UGMM | Universal GMM |
| WASN | Wireless Acoustic Sensor Network |
| ZCR | Zero Crossing Rate |

## References

1. Babisch, W. Transportation noise and cardiovascular risk. *Noise Health* **2008**, *10*, 27–33.[CrossRef]
2. E.U. EU Directive: Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise. In *Official Journal of the European Communities*; L 189/12; European Union: Maastricht, The Netherlands, 2002.
3. Kephalopoulos, S.; Paviotti, M.; Ledee, F.A. *Common Noise Assessment Methods in Europe (CNOSSOS-EU)*; European Union: Maastricht, The Netherlands, 2012.
4. Bertrand, A. Applications and trends in wireless acoustic sensor networks: A signal processing perspective. In Proceedings of the 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT), Ghent, Belgium, 22–23 November 2011; pp. 1–6.
5. Basten, T.; Wessels, P. An overview of sensor networks for environmental noise monitoring. In Proceedings of the ICSV21, Beijing China, 13–17 July 2014.
6. Botteldooren, D.; De Coensel, B.; Oldoni, D.; Van Renterghem, T.; Dauwe, S. Sound monitoring networks new style. In Proceedings of the Acoustics 2011: Breaking New Ground: Annual Conference of the Australian Acoustical Society, Gold Coast, Australia, 2–4 November 2011; pp. 1–5.
7. Cense—Characterization of Urban Sound Environments. Available online: http://cense.ifsttar.fr/ (accessed on 14 September 2017).
8. Camps, J. Barcelona noise monitoring network. In Proceedings of the EuroNoise 2015, Maastrich, The Netherlands, 31 May–3 June 2015; EAA-NAG-ABAV: Maastrich, The Netherlands, 2015; pp. 2315–2320.

9. SonYC–Sounds of New York City. Available online: https://wp.nyu.edu/sonyc (accessed on 14 September 2017).
10. Hancke, G.; de Silva, B.; Hancke Jr., G. The role of advanced sensing in smart cities. *Sensors* **2013**, *13*, 393–425, [CrossRef]
11. Wessels, P.W.; Basten, T.G. Design aspects of acoustic sensor networks for environmental noise monitoring. *Appl. Acoust.* **2016**, *110*, 227–234, [CrossRef]
12. Ntalampiras, S.; Potamitis, I.; Fakotakis, N. Probabilistic Novelty Detection for Acoustic Surveillance Under Real-World Conditions. *IEEE Trans. Multimed.* **2011**, *13*, 713–719, [CrossRef]
13. Paulo, J.; Fazenda, P.; Oliveira, T.; Casaleiro, J. Continuos sound analysis in urban environments supported by FIWARE platform. In Proceedings of the EuroRegio2016/TecniAcústica'16, Porto, Portugal, 13–15 June 2016; pp. 1–10.
14. Ye, J.; Kobayashi, T.; Murakawa, M. Urban sound event classification based on local and global features aggregation. *Appl. Acoust.* **2017**, *117 Pt B*, 246–256,[CrossRef]
15. Socoró, J.C.; Alías, F.; Alsina-Pagès, R.M. An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments. *Sensors* **2017**, *17*, 2323.[CrossRef]
16. Maijala, P.; Shuyang, Z.; Heittola, T.; Virtanen, T. Environmental noise monitoring using source classification in sensors. *Appl. Acoust.* **2018**, *129*, 258–267.[CrossRef]
17. Pierre, R.L.S., Jr.; Maguire, D.J.; Automotive, C.S. The impact of A-weighting sound pressure level measurements during the evaluation of noise exposure. In Proceedings of the Conference NOISE-CON, Baltmore, Maryland, 12–14 July 2004; pp. 12–14.
18. Mietlicki, F.; Mietlicki, C.; Sineau, M. An innovative approach for long-term environmental noise measurement: RUMEUR network. In Proceedings of the EuroNoise 2015, Maastrich, The Netherlands, 31 May–3 June 2015; EAA-NAG-ABAV: Maastrich, The Netherlands, 2015; pp. 2309–2314.
19. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; Benocci, R.; Zambon, G. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189.[CrossRef]
20. Orga, F.; Alías, F.; Alsina-Pagès, R.M. On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments. *Int. J. Environ. Res. Public Health* **2018**, *15*, 13.[CrossRef]
21. Mermelstein, P. Distance measures for speech recognition, psychological and instrumental. *Pattern Recogn. Artif. Intell.* **1976**, *116*, 374–388.
22. Alsina-Pagès, R.M.; Socoró, J.C.; Alías, F. Detecting Anomalous Noise Events on Low-Capacity Acoustic Sensor in Dynamic Road Traffic Noise Mapping. *Multidiscip. Digit. Publ. Inst. Proc.* **2018**, *2*, 136.[CrossRef]
23. Ntalampiras, S. Universal background modeling for acoustic surveillance of urban traffic. *Digit. Signal Process.* **2014**, *31*, 69–78.[CrossRef]
24. Aurino, F.; Folla, M.; Gargiulo, F.; Moscato, V.; Picariello, A.; Sansone, C. One-Class SVM Based Approach for Detecting Anomalous Audio Events. In Proceedings of the 2014 International Conference on Intelligent Networking and Collaborative Systems, Salerno, Italy, 10–12 September 2014; pp. 145–151. [CrossRef]
25. Salamon, J.; Jacoby, C.; Bello, J.P. A Dataset and Taxonomy for Urban Sound Research. In Proceedings of the 22Nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 Novemver 2014; ACM: New York, NY, USA, 2014; pp. 1041–1044. [CrossRef]
26. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Reliable detection of audio events in highly noisy environments. *Pattern Recogn. Lett.* **2015**, *65*, 22–28, [CrossRef]
27. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Audio Surveillance of Roads: A System for Detecting Anomalous Sounds. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 279–288.[CrossRef]
28. Alías, F.; Alsina-Pagès, R.M.; Socoró, J.C.; Orga, F.; Nencini, L. Performance analysis of the low-cost acoustic sensors developed for the DYNAMAP project: A case study in the Milan urban area. *J. Acoust. Soc. Am.* **2017**, *141*, 3883–3884.[CrossRef]
29. Kjaer, B. Noise Monitoring Terminal Type 3639. 2015. Available online: https://www.bksv.com/-/media/literature/Product-Data/bp2379.ashx (accessed on 18 April 2018).
30. Davis, L. Model 831-NMS Permanent Noise Monitoring System. 2015. Available online: https://www.johnmorrisgroup.com/NZ/Product/10206/Model-831-NMS-Permenant-Noise-Monitoring-System (accessed on 18 April 2018).

31. Paulo, J.; Fazenda, P.; Oliveira, T.; Carvalho, C.; Félix, M. Framework to monitor sound events in the city supported by the FIWARE platform. In Proceedings of the 46o Congreso Español de Acústica (TecniAcústica), Valencia, Spain, 21–23 October 2015; pp. 21–23.
32. Wang, C.; Chen, G.; Dong, R.; Wang, H. Traffic noise monitoring and simulation research in Xiamen City based on the Environmental Internet of Things. *Int. J. Sustain. Dev. World Ecol.* **2013**, *20*, 248–253.[CrossRef]
33. Nencini, L.; Rosa, P.D.; Ascari, E.; Vinci, B.; Alexeeva, N. SENSEable Pisa: A wireless sensor network for real-time noise mapping. In Proceedings of the EURONOISE, Prague, Czech Republic, 10–13 June 2012; Volume 12, pp. 10–13.
34. Bell, M.C.; Galatioto, F. Novel wireless pervasive sensor network to improve the understanding of noise in street canyons. *Appl. Acoust.* **2013**, *74*, 169–180.[CrossRef]
35. Domínguez, F.; Dauwe, S.; Cariolaro, D.; Touhafi, A.; Dhoedt, B.; Botteldooren, D.; Steenhaut, K.; et al. Towards an environmental measurement cloud: delivering pollution awareness to the public. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, [CrossRef]
36. Mydlarz, C.; Salamon, J.; Bello, J.P. The implementation of low-cost urban acoustic monitoring devices. *Appl. Acoust.* **2017**, *117*, 207–218.[CrossRef]
37. Mietlicki, C.; Mietlicki, F.; Ribeiro, C.; Gaudibert, P.; Vincent, B. The HARMONICA project, new tools to assess environmental noise and better inform the public. In Proceedings of the Forum Acusticum Conference, Krakow, Poland, 7–12 September 2014.
38. Nencini, L. DYNAMAP monitoring network hardware development. In Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy, 12–16 July 2015; pp. 12–16.
39. Bellucci, P.; Peruzzi, L.; Zambon, G. LIFE DYNAMAP project: The case study of Rome. *Appl. Acoust.* **2017**, *117*, 193–206.[CrossRef]
40. Zambon, G.; Benocci, R.; Bisceglie, A.; Roman, H.E.; Bellucci, P. The LIFE DYNAMAP project: Towards a procedure for dynamic noise mapping in urban areas. *Appl. Acoust.* **2016**, *124*, 52–60.[CrossRef]
41. Cooley, J.; Tukey, J. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297–301.[CrossRef]
42. Can, A.; Leclercq, L.; Lelong, J.; Botteldooren, D. Traffic noise spectrum analysis: Dynamic modeling vs. experimental observations. *Appl. Acoust.* **2010**, *71*, 764–770.[CrossRef]
43. Kurze, U. Frequency curves of road traffic noise. *J. Sound Vib.* **1974**, *33*, 171–185.[CrossRef]
44. Blackman, R.B.; Tukey, J.W. Particular pairs of windows. In *The Measurement of Power Spectra, from the Point of View of Communications Engineering*; Nokia Bell Labs: Murray Hill, NJ, USA, 1958; pp. 95–101.
45. Furui, S. Cepstral analysis technique for automatic speaker verification. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 254–272, [CrossRef]
46. Alías, F.; Socoró, J.C. Description of Anomalous Noise Events for Reliable Dynamic Traffic Noise Mapping in Real-Life Urban and Suburban Soundscapes. *Appl. Sci.* **2017**, *7*, 146, [CrossRef]
47. Valero, X.; Alías, F. Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification. *IEEE Trans. Multimed.* **2012**, *14*, 1684–1689.[CrossRef]
48. Nussbaumer, H.J. *Fast Fourier Transform and Convolution Algorithms*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 2.
49. Evaluating Arduino and Due ADCs. Available online: http://www.djerickson.com/arduino/due_adc.html (accessed on 21 February 2018).
50. Accelerating Fourier Transforms Using the GPU. Available online: https://www.raspberrypi.org/blog/accelerating-fourier-transforms-using-the-gpu/ (accessed on 21 February 2018).
51. LPC Microcontrollers—NXP. Available online: https://www.nxp.com/products/processors-and-microcontrollers/arm-based-processors-and-mcus/lpc-cortex-m-mcus:LPC-ARM-CORTEX-M-MCUS (accessed on 21 February 2018).
52. STM32 ARM Cortex Microcontrollers—32-bit MCUs—STMicroelectronics. Available online: http://www.st.com/en/microcontrollers/stm32-32-bit-arm-cortex-mcus.html (accessed on 21 February 2018).
53. LogiCORE IP—Fast Fourier Transform v7.1. Available online: https://www.xilinx.com/support/documentation/ip_documentation/xfft_ds260.pdf (accessed on 21 February 2018).

54. Digilent—Arty Reference Manual. Available online: https://reference.digilentinc.com/reference/programmable-logic/arty/reference-manual (accessed on 21 February 2018).
55. Mesaros, A.; Heittola, T.; Virtanen, T. Metrics for Polyphonic Sound Event Detection. *Appl. Sci.* **2016**, *6*, 162, [CrossRef]

Capítol 4

# Funcionament del detector d'esdeveniments i treball sobre les dades reals

En una primera versió de l'ANED, s'utilitzen dades capturades manualment en els llocs on es desplega la xarxa DYNAMAP. Però una vegada es desplega la xarxa a ambdues àrees (Milà i Roma), els investigadors ja tenen accés a dades obtingudes amb les mateixes condicions en què treballarà la xarxa de sensors. Per això, es desenvolupa una segona versió de l'ANED entrenat i validat amb dades obtingudes a partir dels sensors de la xarxa, utilitzant el mateix maquinari que s'emprarà en la implementació final. Aquesta segona versió disposa de més de 300 h entre Roma i Milà, i amb un mostreig més exhaustiu de les hores i els dies, superant amb escreix les 9 h de la primera base de dades [1, 2].

Això obre les portes a treballar amb una base de dades més realista i extensa que la que contenia les dades capturades manualment. A més a més, ofereix més facilitat de recollida de dades i un mostreig més exacte de les diferents franges horàries, diferenciant dies laborables de no laborables. Així, es poden gravar també determinats moments que interessin per enriquir la base de dades original. Cal dur a terme un segon etiquetatge de les dades per aprofitar les mostres capturades amb la xarxa, emprant les mateixes etiquetes ja definides i completant-ne amb noves quan sigui necessari, però amb una mostra considerablement més gran que l'anterior.

## 4.1 Etiquetatge d'esdeveniments acústics i confecció de la base de dades

Per tal d'entrenar l'ANED, és necessari etiquetar manualment les dades acústiques recopilades. Per tal d'acomplir aquesta tasca, es fa una exploració prèvia de les dades capturades i es defineixen les etiquetes que s'utilitzaran. A la Taula 4.1 hi ha descrits els diferents esdeveniments etiquetats a les dues àrees pilot amb el nom de l'etiqueta utilitzada en la base de dades.

Per tal d'etiquetar, els investigadors utilitzen el programari lliure Audacity[1], que permet assignar una etiqueta de text a un fragment d'àudio i exportar-ho en un fitxer de text pla. El fitxer exportat conté tres columnes: el temps inicial, el temps final i l'etiqueta. Mitjançant un script s'autocompleten els espais no etiquetats com a `rtn`, per facilitar la tasca dels etiquetadors, que

[1] https://www.audacityteam.org/ (visitat: 31/12/2021)

| Nom | Descripció |
|---|---|
| airp | Soroll d'avions i helicòpters |
| alrm | Alarmes i sons de marxa enrere de vehicles |
| bell | Sons de campanes |
| bike | Soroll de bicicletes i cadenes |
| bird | Cantar d'ocells |
| blin | Obertura i tancament de persianes |
| brak | So de frens i discs de fre |
| busd | Obertura i tancament de portes de bus i tramvia |
| cmplx | Sons que no es poden identificar per l'etiquetador o conjunt de sons barrejats |
| dog | Bordar de gossos |
| door | Portes de vehicles i de cases |
| glas | So de vidres trencant-se |
| horn | Botzines de vehicle |
| inte | Soroll d'interferències provinents de la indústria o d'altres màquines |
| musi | Música (de cotxes o del carrer) |
| peop | So de persones parlant, rient, esternudant, tossint... |
| rain | Pluja |
| rtn | Soroll de trànsit rodat |
| rubb | So provinent del servei de recollida d'escombraries |
| sire | Sirenes (ambulància, policia...) |
| sqck | Soroll de grinyol de portes |
| step | Passes de persona |
| thun | Trons |
| trck | Sorolls derivats de la càrrega dels camions o sistemes d'enganxament |
| tram | Parades, arrencades i marxa de tramvies |
| tran | Parades, arrencades i marxa de trens |
| trll | Soroll de rodes de maleta o carretó |
| stru | Sorolls provinents d'estructures (normalment causats per la vibració de l'estructura on es troba el sensor mentre passa un vehicle pesat) |
| wind | Sorolls originats pel vent |
| wrks | Obres i altres sorolls derivats de la construcció |

Taula 4.1: Llistat d'etiquetes utilitzades en l'etiquetatge manual de la base de dades

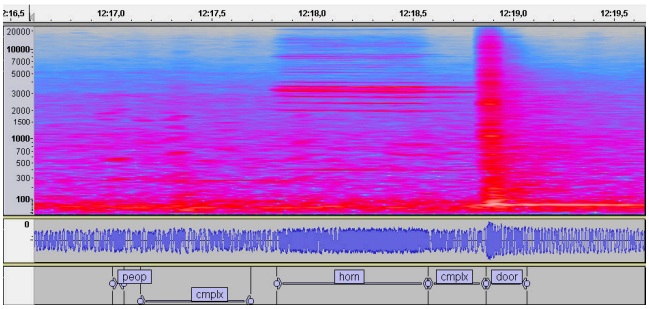



Figura 4.1: Captura de l'etiquetatge utilitzant el programari Audacity.

no cal que incloguin manualment el soroll de trànsit (so més habitual a ambdues àrees pilot). Es pot veure un exemple de l'etiquetatge a la figura 4.1, on s'etiqueten els esdeveniments de `peop` (gent parlant), `horn` (botzina d'un vehicle), `door` (porta) i dos esdeveniments etiquetats com a `cmplx`, que són barreges d'esdeveniments on cap dels sons és més audible que els altres. Es poden trobar més detalls sobre la descripció dels tipus d'esdeveniments a [3].

A partir d'aquests fitxers d'etiquetes, ja és possible crear el conjunt de dades per l'entrenament de l'algoritme. Per aconseguir una coherència entre els experts que etiqueten, s'unifiquen els criteris prèviament i es defineixen unes directrius comunes. D'això en dependrà la qualitat de l'etiquetatge i el posterior rendiment de l'algoritme de detecció d'esdeveniments.

L'etiqueta `cmplx` indica que el fragment d'àudio en qüestió no pertany únicament a una categoria d'esdeveniment o que el so no és clarament identificable. En el cas que pertanyi a una barreja de dos o més tipus d'esdeveniments, el fragment s'etiqueta amb el següent format: `cmplx:so_1+so_2+...`, per exemple: `cmplx:rtn+bird+peop`.

## 4.2 Bases de dades del projecte DYNAMAP

Dins del projecte DYNAMAP, es creen dues bases de dades que s'utilitzen en les diferents fases de la implementació, i contenen dos conjunts de dades que corresponen a les àrees pilot del projecte: Milà i Roma. Els mapes dels sensors instal·lats es poden veure a les figures 3.5 i 3.6 de la secció 3.

En una primera part del projecte, es creà una primera base de dades utilitzant àudios capturats manualment. Es tracta d'una base de dades reduïda, de 9 h, que va permetre fer una primera aproximació al problema abans que les WASNs estiguessin instal·lades. Aquestes primeres dades es van recopilar entre el 18 i el 21 de maig de 2015, tant a Roma (als portals de l'autovia) com a Milà (a peu de carrer) [1, 4].

Més endavant, es crea la segona base de dades, també amb dades reals, però aquesta vegada capturades mitjançant les dues xarxes de sensors desplegades. Aquesta segona base de dades conté dades de dos dies, un dia entre setmana

(dimarts 28 de novembre de 2017) i un de cap de setmana (dissabte 2 de desembre de 2017) [2]. Es poden veure les durades de cada conjunt de dades a la taula 4.2.

| Lloc | Entorn | Tipus de captura | Durada |
|---|---|---|---|
| Roma (A90) | Suburbà | Manual | 4 h i 24 min |
| Milà (Districte 9) | Urbà | Manual | 4 h i 44 min |
| Roma (A90) | Suburbà | WASN | 153 h i 20 min |
| Milà (Districte 9) | Urbà | WASN | 151 h |

Taula 4.2: Durada i tipus dels conjunts de dades etiquetats durant el projecte DYNAMAP.

El fet de capturar les dades a través de cada una de les xarxes de sensors facilita la tasca de recollida de dades, permetent la creació d'un conjunt de dades més gran i homogeneïtzant les hores i els dies de gravació en diferents punts simultàniament. A més a més, garanteix que es mantinguin les mateixes condicions acústiques en la recollida de dades i en el funcionament normal del sensor. A [1] s'hi poden observar les hores de mostreig en els dies de gravació. Es pot consultar [5] per més detalls sobre la captura de dades.

## 4.3 Publicacions destacades

Tant el procés d'etiquetatge com l'anàlisi exhaustiu de la base de dades de Roma, obtingudes a través de la WASN, es publiquen a l'article *«A WASN-based suburban dataset for anomalous noise event detection on dynamic road-traffic noise mapping»* (en català: Un conjunt de dades suburbà extret d'una WASN per a la detecció d'esdeveniments de soroll anòmals en el mapatge dinàmic del soroll de trànsit rodat). L'article, publicat el maig de 2019, repassa l'estat de l'art en la detecció d'esdeveniments acústics i les bases de dades emprades a la comunitat científica. A continuació, s'hi explica el disseny de la base de dades, des de la descripció fins a la caracterització dels esdeveniments, passant pel seu etiquetatge. També s'hi inclou un estudi exhaustiu de la tipologia dels esdeveniments capturats amb la WASN, que formen part del conjunt de dades. El candidat duu a terme l'anàlisi de les dades i participa en l'etiquetatge de la base de dades; també s'encarrega de la caracterització de les dades, que s'explica en el capítol 5.

## Referències

[1] Socoró, J. C., Alsina-Pagès, R. M., Alías, F. i Orga, F. "Adapting an Anomalous Noise Events Detector for Real-Life Operation in the Rome Suburban Pilot Area of the DYNAMAP's Project". *Proceedings of EuroNoise2018*. Heraklion, Crete – Greece: EAA – HELINA, 27 - 31 Maig de 2018, pàg. 693 - 698.

[2] Alsina-Pagès, R. M., Orga, F., Alías, F. i Socoró, J. C. "A WASN-Based Suburban Dataset for Anomalous Noise Event Detection on Dynamic Road-Traffic Noise Mapping". *Sensors*, vol. 19, núm. 11 (2019), pàg. 2480.

[3] Alías, F. i Socoró, J. C. "Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes". *Applied Sciences*, vol. 7, núm. 2 (2017), pàg. 146.

[4] Sevillano, X., Socoró, J. C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., Coppi, P., Nencini, L., Cerniglia, A., Bisceglie, A., Benocci, R. i Zambon, G. "DYNAMAP – Development of low cost sensors networks for real time noise mapping". *Noise Mapping*, vol. 3 (1 Maig de 2016), pàg. 172 - 189.

[5] Alsina-Pagès, R. M., Alías, F., Socoró, J. C. i Orga, F. "Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN". *Sensors*, vol. 18, núm. 4 (2018), pàg. 1272.

# A WASN-Based Suburban Dataset for Anomalous Noise Event Detection on Dynamic Road-Traffic Noise Mapping

**Rosa Ma. Alsina-Pagès, Ferran Orga, Francesc Alías, Joan Claudi Socoró**

Publicat: 30 de maig del 2019

III

sensors

MDPI

Article

# A WASN-Based Suburban Dataset for Anomalous Noise Event Detection on Dynamic Road-Traffic Noise Mapping

**Rosa Ma Alsina-Pagès *, Ferran Orga, Francesc Alías and Joan Claudi Socoró**

GTM—Grup de recerca en Tecnologies Mèdia, La Salle—Universitat Ramon Llull., C/Quatre Camins, 30, 08022 Barcelona, Spain; ferran.orga@salle.url.edu (F.O.); francesc.alias@salle.url.edu (F.A.); joanclaudi.socoro@salle.url.edu (J.C.S.)

* Correspondence: rosamaria.alsina@salle.url.edu; Tel.: +34-932-902-455

Received: 29 March 2019; Accepted: 27 May 2019; Published: 30 May 2019


**Abstract:** Traffic noise is presently considered one of the main pollutants in urban and suburban areas. Several recent technological advances have allowed a step forward in the dynamic computation of road-traffic noise levels by means of a Wireless Acoustic Sensor Network (WASN) through the collection of measurements in real-operation environments. In the framework of the LIFE DYNAMAP project, two WASNs have been deployed in two pilot areas: one in the city of Milan, as an urban environment, and another around the city of Rome in a suburban location. For a correct evaluation of the noise level generated by road infrastructures, all Anomalous Noise Events (ANE) unrelated to regular road-traffic noise (e.g., sirens, horns, speech, etc.) should be removed before updating corresponding noise maps. This work presents the production and analysis of a real-operation environmental audio database collected through the 19-node WASN of a suburban area. A total of 156 h and 20 min of labeled audio data has been obtained differentiating among road-traffic noise and ANEs (classified in 16 subcategories). After delimiting their boundaries manually, the acoustic salience of the ANE samples is automatically computed as a contextual Signal-to-Noise Ratio (SNR) together with its impact on the A-weighted equivalent level ($\Delta L_{Aeq}$). The analysis of the real-operation WASN-based environmental database is evaluated with these metrics, and we conclude that the 19 locations of the network present substantial differences in the occurrences of the subcategories of ANE, with a clear predominance of the noise of sirens, trains, and thunder.

**Keywords:** road-traffic noise; anomalous noise event; acoustic dataset; noise monitoring; smartcity; WASN; SNR; impact; $L_{Aeq}$; urban sound; noise maps

## 1. Introduction

Presently, cities are growing in both size and population, and the consequent increase in vehicles is making traffic noise problem more present, with a clear effect on the quality of life of their citizens [1]. Noise is one of the main environmental health concerns [2,3], and its impact on social and economic aspects has been proved [4]. To face this issue, European authorities have driven the European Noise Directive (END) [5], focused on the creation of noise-level maps to inform citizens of their exposure to noise, and aided the authorities to take appropriate action to minimize its impact.

Noise maps have been historically generated by means of costly expert measurements using certified sound-level meters, with a basis of short-term periods aimed at being sufficiently representative. This approach is presently overcome by the technological advances of the Internet of Things in the framework of smart cities, which has allowed the emergence of Wireless Acoustic Sensor Networks (WASN) [6,7]. In the literature, several different WASNs have been designed for

urban sound monitoring, some of them focused on security and surveillance and others on city noise management, involving noise mapping, the development of action plans, and public awareness campaigns. For example, the SENSEable project [8] deployed a WASN to collect information from the acoustic environment by means of a set of low-cost acoustic sensors with the goal of analyzing that data together with public health information. Other similar projects are the IDEA project in Belgium [9], or the RUMEUR network in France [10] with special focus on aircraft noise, or even the Barcelona noise-monitoring network [11], whose data is integrated in the Sentilo city management platform [12]. Recently, the SONYC project has deployed 56 low-cost acoustic sensors across New York City to monitor urban noise and perform a multi-label classification of urban sound sources in real time [4], but not on site. Finally, the LIFE DYNAMAP project [13] aims to monitor the noise level generated by road infrastructures by means of two WASNs installed in two pilot areas, one within an urban environment in Milan (District 9), and another in a suburban area surrounding Rome (A90 highway). To monitor Road-Traffic Noise (RTN) levels reliably, all Anomalous Noise Events (ANE) unrelated to regular RTN (e.g., sirens, horns, speech, etc.) should be removed before updating the corresponding noise maps [13].

The deployment of these projects has shown that the WASN paradigm entails several challenges, from technical issues [14,15], to other aspects related to the WASN-based application, such as the automation of data collection and the subsequent signal processing [16–18], especially if the system intends detect acoustic events in real operation and locally in each sensor. Acoustic event detection and classification belongs to the Computational Auditory Scene Analysis (CASA) paradigm [19], and it is usually based on the segmentation of the input acoustic data into slices that represent a single occurrence of the target class, and focus on individual simultaneous events [20]. To do so, Acoustic Event Detection (AED) algorithms are typically trained with databases designed ad hoc in each of the problems to be solved, hence typically considering a finite set of predefined acoustic classes [21,22].

Therefore, the development of AED-based applications entails representative audio databases with all kinds of sounds of interest, as in the one obtained in the SONYC project [4], with data from 56 sensors deployed in different neighborhoods of New York, which considers 10 different kinds of common urban sound sources labeled in an urban soundscape. As a first attempt to create an acoustic dataset to model the acoustic environments of urban and suburban pilot areas in the framework of the DYNAMAP project, an expert-based recording campaign was conducted before the two WASNs were deployed [23]. The analyses showed the highly local, unpredictable, and diverse nature of ANEs in real acoustic environments can be far different to previous models obtained by means of synthetically generated datasets [23]. After labeling the gathered acoustic data, the dataset was used to train the AED-based algorithm designed to detect ANEs, known as Anomalous Noise Event Detector (ANED) [18]. Although that preliminary dataset collected a representative number of acoustic events of interest from both acoustic environments, it missed several key aspects, such as different RTN patterns observed during day–night, weekday–weekend and the effect of diverse weather conditions [24]. This work describes the generation of the acoustic dataset to model the Rome's acoustic environment in real-operation conditions, after deploying the 19-node WASN in its final location. The paper describes the conducted recording campaign and the subsequent labeling of ANEs in 16 different subcategories (without considering combined sounds in a sample), as well as the analysis of their occurrences, duration, Signal-to-Noise Ratio (SNR), and impact on the A-weighted equivalent noise level ($L_{Aeq}$) computation.

The remainder of this paper is the following. Section 2 details the most relevant previous attempts to generate environmental audio databases. Section 3 describes the generation and labeling of the real-operation conditions environmental audio database in the suburban scenario. Section 4 analyses the ANE of the dataset in terms of occurrences, duration, SNR, and impact on the $L_{Aeq}$. Finally, Section 5 discusses in detail the results obtained in the analysis and the future applications of the designed dataset.

## 2. Related Work

In the literature, several audio databases related to machine hearing (or machine listening) have been unveiled for benchmarking purposes under the umbrella of the so-called CASA, and mainly oriented to evaluate the performance of acoustic scene classification and AED. This section reviews the literature about environmental acoustic databases and the datasets designed for challenges (e.g., DCASE), and describes their characteristics and limitations.

### *2.1. Environmental Acoustic Databases*

The environmental acoustic databases employed by the machine-hearing research community have been generally created from live recordings directly and/or synthetically generated by artificially mixing sound events with certain acoustic environments (i.e., background noise). The latter allows control of the SNR of the mixture, and dealing with data scarcity of specific audio events in real-life contexts—which is one of the key problems when trying to gather representative data from live environments [25]—while the former entails a huge effort for data collection and subsequent manual annotation to generate the labeled database or ground truth [23,26,27].

Regarding real-life environmental acoustic databases, in [28] a 1133-min audio database including 10 different acoustic environments, both indoor and outdoor was introduced. On the other hand, the MIVIA audio events dataset was designed for surveillance applications focused on the identification of glass breaking, gun shots, and screams (https://mivia.unisa.it/datasets/audio-analysis/mivia-audio-events/) [21]. The training dataset is about 20 h, while the test set is about 9 h. Moreover, the same research laboratory developed a smaller dataset of about 1 h duration also for surveillance purposes focused on road audio events, which contains sound events from tire skidding and car crashes (https://mivia.unisa.it/datasets/audio-analysis/mivia-road-audio-events-data-set/) [29].

In [23], a real-life acoustic database collected from the urban and suburban environments of the pilot areas of the Life DYNAMAP project [13] is described. The database composed of 9 h and 8 min was obtained through an in situ recording campaign. This acoustic database was developed for discriminating road-traffic noise from ANE through the ANED [18]. The ANEs, which only represented 7.5% of the annotated data, were subsequently classified in 19 different subcategories after manual inspection, showing SNR levels with respect to background noise that ranged from −10 dBm to +15 dBm and showing a high heterogeneity of intermediate SNR levels. When comparing both environments, it can be observed that the number of ANEs in the urban area is approximately four times higher than in the suburban area, also including events with larger acoustic salience. Nevertheless, it is worth mentioning that the recordings in the urban area were conducted at the street level of the preselected locations [30] within District 9 of Milan, while the recordings in the suburban area were conducted on the A90 ring-road portals surrounding Rome (see [31] for further details).

One of the main sources employed to build acoustic databases is the well-known Freesound online repository (https://www.freesound.org). For instance, *freefield1010* is a database composed of 7690 audio clips tagged as "field recording" in the metadata of the original recordings uploaded in this online repository, totaling over 21 h of audio [32]. In [33], 60 h of real field recordings uploaded in Freesound from urban environments were used to build the UrbanSound database. The database is composed of 27 h, which includes 18.5 h of verified and annotated sound event occurrences classified in 10 sound categories (i.e., air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren, and street music). Moreover, the authors also provide an 8.75 h subset—denoted as UrbanSound8k—designed to train sound classification algorithms and obtained after arbitrarily fixing the number of items to 1000 slices per class. In [34], a mixture of sound sources from Freesound mixed with real-life traffic noise was considered to train and evaluate an AED algorithm (considering two SNRs levels: +6 dB and +12 dB). Finally, "ESC: Dataset for Environmental Sound Classification" [35] is composed of three subsets: *(i)* ESC-50, a strictly balanced 50 classes of various environmental sounds obtained through manual annotation, *(ii)* ESC-10, as a reduced 10 classes subset of the former as a proof-of-concept dataset, and *(iii)* EC-US, which contains 250,000 recordings directly extracted

from the "field recording"-tagged category of Freesound. Nevertheless, due to the uncontrolled origin of the sound sources uploaded to Freesound and similar online repositories (see [23] for further examples), involving a wide variation in the recording conditions and quality, the derived environmental acoustic databases may become unsuitable for reliable CASA- and AED-based systems evaluation purposes [27,32].

In [27], the described acoustic dataset covers both indoor and outdoor environments, including real-life recordings of predefined acoustic event sequences and individual acoustic events synthetically mixed with background recordings by considering specific SNRs levels, such as $-6$ dB, 0 dB, and $+6$ dB. In [25], the authors developed a mixed acoustic database composed of acoustic data from real-life recordings, which was subsequently extended with synthetic mixtures of extra events of interest to increase database diversity. In [21], an acoustic database for surveillance purposes that includes sound events such as screams, glass breaking, and gunshots was also artificially generated from indoor and outdoor environments considering different SNR levels (from $+5$ to $+30$ dB). Moreover, a small environmental acoustic database containing 20 scenes mixing background noise with car, bird, and car horn samples synthesized with SimScene software [36] was described in [37]. Following a similar approach, in [38], the TUT Sound Events Synthetic 2016 (TUT-SED-2016 for short) was introduced. The 566 min dataset is composed of synthetic mixtures created by mixing isolated sound events from 16 sound event classes from the original TUT database (The reader is referred to http://www.cs.tut.fi/sgn/arg/taslp2017-crnn-sed/tut-sed-synthetic-2016 for a detailed explanation).

Finally, it is worth mentioning the recent development of an open-source library for the synthesis of soundscapes named Scaper, mainly focused on SED-related applications [39]. This library provides an audio sequencer to generate synthetic soundscapes following a probabilistic approach including isolated sound events. The proposal allows control of the characteristics of the sound mixtures, such as the number of events, and their type, timing, duration, and SNR level with respect to the background noise. The authors validate their proposal through the development of the URBAN-SED database from UrbanSound8k as an example of the result of this kind of data augmentation. The synthetic generation of acoustic databases is a potential solution to address data scarcity when training Deep Neural Networks (DNN) for CASA-related problems (e.g., see [38,40]). Nevertheless, although the artificial generation of sound mixtures allows the creation of controlled training and evaluation environments, it has been also stated that these artificially generated databases could not represent the variability encountered in real-life environments accurately enough [18,26].

### *2.2. Challenge-Oriented Acoustic Datasets*

Over the last decade, the CASA research community has provided publicly available datasets and standard metrics to evaluate the development of their investigations. One of the seminal international efforts that emerged to evaluate systems developed to model the perception of people, their activities, and interactions was the CLassification of Events, Activities and Relationships (CLEAR) competition, which included in its 2006 and 2007 editions specific data for acoustic event detection and classification mainly collected from indoor environments—specifically, from meeting rooms (see e.g., [41–43]). Although other attempts emerged to provide evaluation material for CASA-focused systems, such as DARESounds.org initiative [44] or TRECVID Multimedia Event Detection competition (focused on audiovisual and multi-modal event detection (https://www.nist.gov/itl/iad/mig/multimedia-event-detection)), neither them nor CLEAR led to the establishment of a reference evaluation challenge for the CASA research community [20].

Later, the IEEE AASP supported a new competition named Detection and Classification of Acoustic Scenes and Events (DCASE), which started in 2013 at WASPAA conference (http://www.waspaa.com/waspaa13/d-case-challenge/index.html). The database that was provided for that challenge contained both live and synthetic recordings [45]. The results of that competition can be found in [27]. Since 2016, DCASE has become an annual competition including different challenges covering acoustic scene classification, sound event detection in synthetic and in real-life audio, domestic

audio tagging, or the detection of bird or rare sound events, to name a few (see [22,46,47] for further details) (http://dcase.community/events), mainly thanks to the contribution of several researchers who have made available different datasets for public evaluation.

Among them , it is worth mentioning the creation of the "TUT Database for Acoustic Scene Classification and Sound Event Detection" from real-life recordings [26]. This database allows the evaluation of automatic event or acoustic scene detection systems within 15 different real-life acoustic environments, such as lakeside beach, bus, cafe/restaurant, car, city center, forest path, grocery store, home, library, metro station, office, urban park, residential area, train, and tram. It is worth mentioning that the sound events subset, which covers both indoor and outdoor environments, it is mainly focused on surveillance and human activity monitoring. Finally, it is worth noting that in the recent announcement of the DCASE2019 competition, an Urban Sound Tagging (UST) challenge has been presented. The goal of this task is to predict whether each of 23 sources of noise pollution is present or absent in a 10-s scene. In this challenge, the audio in the dataset has been acquired with the acoustic sensor network of the SONYC project: SOunds of New York City [4], and it provides a simplified taxonomy of the sounds of the city in two levels, 8 coarse categories, and 23 fine labels. The challenge dataset includes 2351 recordings in the training split and 443 in the validation split, making a total of 2794 10-s audios. The full taxonomy and details of the SONYC project dataset can be found in [33].

## 3. Design of an Environmental Database

In this section, the real-operation environmental audio database recorded and built in for the suburban area of the DYNAMAP project is detailed. First, the conducted recording campaign is described. Second, the subsequent generation of the audio database is detailed, which includes the labeling and the computation of the acoustic salience of the anomalous noise events (in terms of SNR) and their impact on the $L_{Aeq}$. These are computed in this study and used jointly with duration, number of occurrences and other database descriptors to analyze in detailed nature of the ANEs of the suburban area.

### *3.1. Description of the WASN-based Suburban Recording Campaign*

The main goal of any recording campaign is to collect representative samples of the acoustic environment under study through the WASN in real operation. Taking advantage of the experience gained from the preliminary recording and analysis of that acoustic environment [23], a second recording campaign was designed. The main reason for it was three-fold: *(i)* all the nodes of the WASN had been already deployed in their definitive operative location, which increased the number of recording points and changed slightly the sensor position in the portals, *(ii)* the sampling completeness, because the previous recording campaign did not include nighttime, weekend data, or different meteorological conditions, and *(iii)* the total amount of time recorded was quite short (4 h and 44 min), including only 12.2% of ANEs, which led us to the conclusion that ANEs were misrepresented in the dataset, after discarding the augmentation of the dataset by means of synthetic samples according to [23].

The WASN deployed in the suburban area of DYNAMAP project is located on the A90 highway surrounding Rome, and comprises 24 acoustic nodes, 5 of which are low-capacity sensors without enough computational resources to run the ANED algorithm. The locations of the 19 high-capacity sensors of the WASN in the Rome's suburban pilot area are shown in Figure 1. The set of basic specifications [48] that are defined to satisfy DYNAMAP requirements for each monitoring station are the following: *(i)* 40–100 dB(A) broadband linearity range, *(ii)* 35–115 dB working range with acceptable Total Harmonic Distortion (THD), and *(iii)* narrowband floor noise level. The project also requires the possibility of audio recording, as well as Virtual Private Network (VPN) connection and GPRS/3G/WiFi connection. The precision of the sensors is a key issue for the system reliability [49]. During the developing stage, all the elements that could increase the uncertainty of the measurement were taken into account following the requirements of IEC 61672 [50]. Several tests have been

conducted with both the hardware and the software, using a climate chamber with different operation temperatures. Electromagnetic Compatibility tests were also conducted as well as atmospheric agent simulations over the designed equipment [51].

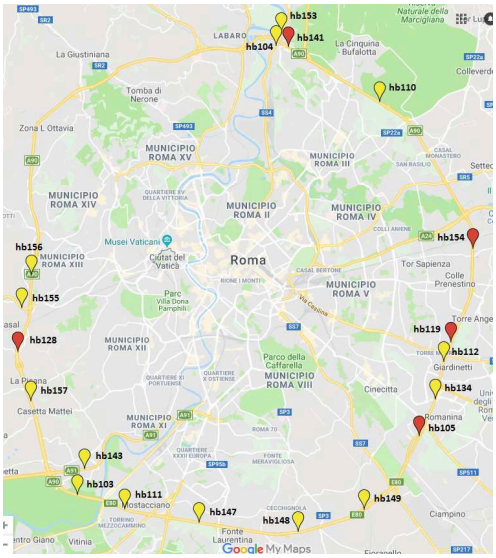



**Figure 1.** Map with sensors' location information within the WASN of DYNAMAP project in the suburban area of Rome, those locations also sensed during the initial recording campaign described in [23] shown in red.

Some of the recording locations are the same as the ones used in the previous recording campaign [23] (see Figure 2). However, although conducted in the same portal both the sensor and the exact location measurement differ. The microphone location is slightly different with respect to the entire structure of the portal (see Figure 2a,b).

(**a**) (**b**)

**Figure 2.** Location of the sensor used in the recording campaign in the portals. (**a**) Recording campaign deployment in [23]; (**b**) WASN deployment in this work (picture property of ANAS S.p.A.).

Achieving a complete and exhaustive dataset is a challenging task since the amount of available resources is limited, e.g., processing and storage capabilities, data collection using 3G modems, availability of all nodes in fully operative conditions, etc. In Figure 3, it can be observed that the $L_{Aeq}$

presents a diurnal variation that suggests sampling the recording differentiating day and night to obtain data from several patterns of traffic noise, and so of ANEs.

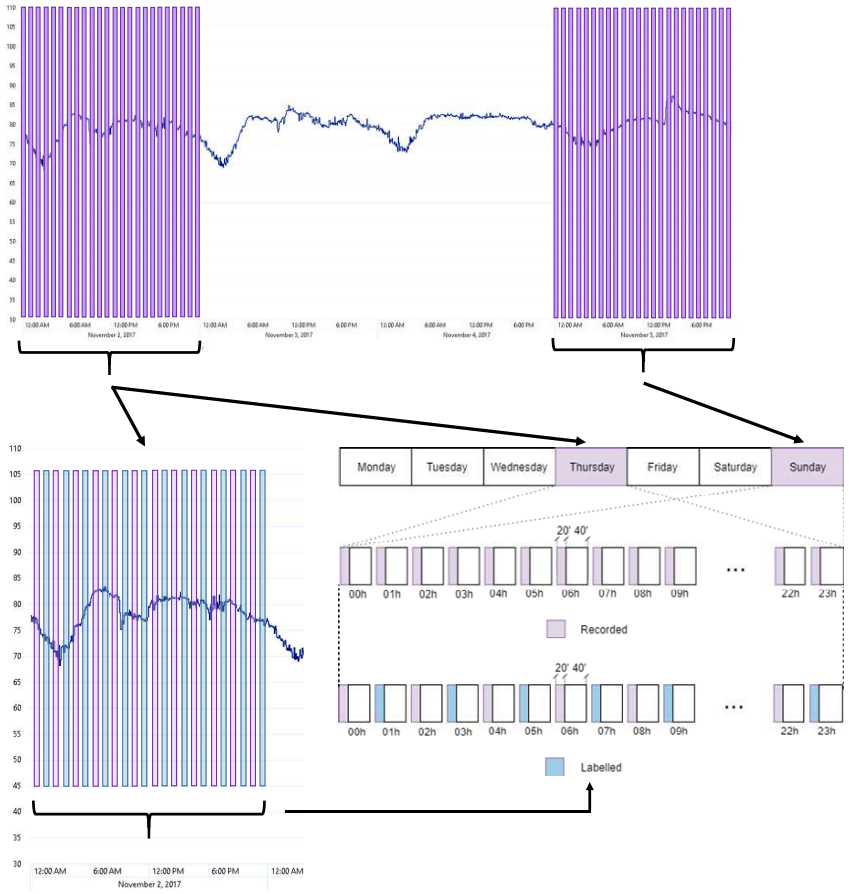



**Figure 3.** Daily curve of $L_{Aeq}$ for sensor hb147 for the 2nd and the 5th of November 2017, and the recordings conducted. Diagram of the recording days and duration for each sensor, and scheme of the labeled files to build the dataset.

To this aim, one-day real-operation recordings were planned through all nodes of the suburban WASN considering different days of variating traffic conditions, and assuming a trade-off between completeness and available computation and data communication resources, which were limited by the resources available in each of the nodes of the WASN (storage capacity, throughput of data, etc.). The following data sampling approach was proposed: Thursday and Sunday were selected as representative weekday and weekend days, respectively, being the 2nd and the 5th of November 2017, specifically. From each sensor, 20 min have been recorded per hour which was limited by the storage capacity and communication resources of each of the nodes. Figure 3 shows the recordings over the values of $L_{Aeq}$, while Figure 3 shows a schematic diagram with the recording process during the selected weekday and weekend day. As a result, 16 h of acoustic data were collected from each sensor to cover the diversity of the acoustic environment in a workday and a weekend day in this suburban environment.

It is worth mentioning that the high-capacity sensors used to conduct the recording campaign using the WASN in real operation were low-cost acoustic sensors designed ad hoc for the DYNAMAP project [51].

### *3.2. Data Labeling*

After recording representative acoustic data for building the suburban environmental database, a labeling process was conducted. The manual annotation of ANEs becomes particularly complex when dealing with real-operation data from raw recordings. Thus, this process must be conducted by experts, since it is very important to precisely determine the occurrence and boundaries of each event, e.g., indicating the start and end points in the mixed audio [23].

The labeling process was not exhaustive because of the excessive burden that such a task would represent if considering all recorded data, meaning that only 50% of gathered audio signals were finally labeled. This represented a total of 156 h and 20 min of labeled audio data. From the labeling process 94.8% was labeled as RTN, 1.8% as ANE and the remaining 3.4% was categorized as *others* when the audio passage was difficult to categorize in one or other class due to the complexity of the audio scene. These last passages were not included in the subsequent analyses presented throughout this paper, but they have been left for future analyses that will focus, e.g., on the impact they can have for ANED assessment.

In Figure 4, an example of the labeling process using Audacity software is shown for illustrative purposes. The example of ANE is a siren, which is a long event, recorded in sensor hb149 the 2nd of November 2017.

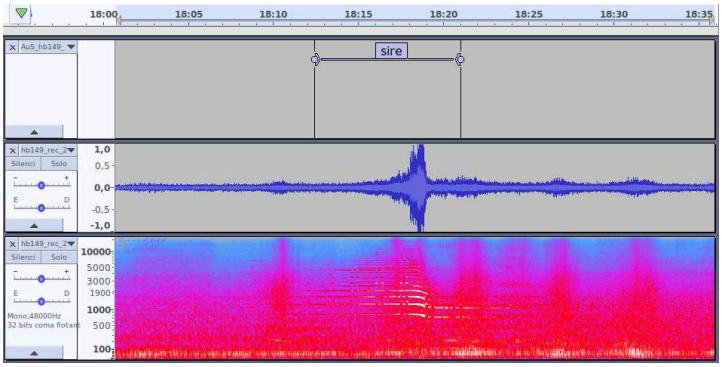



**Figure 4.** Example of the labeling of a siren using Audacity, with the label on top, the signal in the middle and the spectrum on the bottom.

From the labeling of all the collected data through the 19 nodes of the Rome's WASN, the following list of ANEs were observed:

- *airp*: airplanes.
- *alrm*: sounds of cars and houses alarm systems.
- *bike*: noise of bikes.
- *bird*: birdsong.
- *brak*: noise of brake or cars' trimming belt.
- *busd*: opening bus or tramway, door noise, or noise of pressurized air.
- *door*: noise of house or vehicle doors, or other object blows.
- *horn*: horn vehicles noise.
- *inte*: interfering signal from ad industry or human machine.
- *musi*: music in car or in the street.
- *rain*: sound produced by heavy rain.

- *sire*: sirens of ambulances, police, fire trucks, etc.
- *stru*: noise of portals structure derived from its vibration, typically caused by the passing-by of very large trucks.
- *thun*: thunderstorm.
- *tran*: stop, start, and pass-by of trains.
- *trck*: noise when trucks or vehicles with heavy load passed over a bump.

Another label has been used for the annotation process, the *cmplx* label that indicates that the piece of raw acoustic audio was the result of more than one subcategory of ANE or that the sound was not identifiable by the labeler. This is not considered to be a subcategory of ANE because it cannot be assigned to a single noise source.

### 3.3. Characterization of the ANEs

In previous works, two parameters were considered by the authors to figure out the effects of the ANEs on noise-map generation [52]. The first of the parameters is based on the classical SNR calculation, consisting of the ratio of power of the ANE in relation to the power of the surrounding RTN. The second metric determines the impact of the ANE on the equivalent noise level used to build the noise map. The calculation of the two parameters is described below.

#### 3.3.1. SNR Calculation

As aforementioned, the SNR is calculated as the classical signal-to-noise ratio used in signal processing, considering that the ANE corresponds to the signal and the RTN is the noise. The acoustic power of the ANE and the RTN are calculated as follows:

$$P_x = \sum_{n=1}^{N} \left( \frac{x[n]^2}{N} \right) \tag{1}$$

where $x[n]$ is the recorded audio with $N$ samples that belongs to either the ANE of the RTN.

After the power calculations of the ANE and the surrounding RTN, the SNR is calculated as:

$$SNR = 10\, log_{10} \left( \frac{P_{ANE}}{P_{RTN}} \right) \tag{2}$$

where $P_{ANE}$ belongs to the anomalous event in question and $P_{RTN}$ is the power of the surrounding RTN.

All the casuistry of the calculation is detailed in [52]. Finally, it is worth mentioning that the SNR of a particular ANE could be negative if the power of the surrounding RTN is higher than the power of the ANE itself. This may happen in cases where RTN masks other low-energy sounds, e.g., birds, because of the fluctuation of the road pass-bys.

#### 3.3.2. Impact Calculation

The computation of the ANE impact consists of determining the contribution of a particular event to the equivalent noise level of the recorded audio after applying the A-weighting filter [53], i.e., $L_{Aeq}$. This metric has been defined to evaluate the effect of each ANE on the noise map $L_{Aeq}$. It is calculated as the difference between the $L_{Aeq}$ computed with all the raw data and the $L_{Aeq}$ after removing the ANE by means of a linear interpolation (see Equation (3)). The event should be replaced by a lower-period linear interpolation to maintain the weight of the surrounding RTN level. That way, the road-traffic noise measurement is as accurate as possible in the whole integration time.

$$\Delta L_{Aeq} = L_{Aeq,ANE} - L_{Aeq,\overline{ANE}} \tag{3}$$

where $\Delta L_{Aeq}$ is the contribution of this ANE to the $L_{Aeq}$, $L_{Aeq,ANE}$ is the A-weighted equivalent noise level of the raw audio given an evaluation period and $L_{Aeq,\overline{ANE}}$ is the equivalent A-weighted equivalent noise level of the same audio after removing the ANE.

In the DYNAMAP project, noise-map values are updated every 5 min, thus the contribution of ANEs on the $L_{Aeq}$ will be evaluated in this integration time, i.e., $\Delta L_{A300s}$. The low-level interpolations are carried out in a 1-s integration window, i.e., $L_{A1s}$, as it heuristically proved to be a good trade-off between representing all audible short events and not adding imperceptible changes to the equivalent sound measurement. As the goal is to obtain the equivalent noise level completely unaffected by the ANE, a span of 500 ms is left before and after the exact ANE label. More details on the casuistry of this calculation may be found in [52].

Finally, it is worth mentioning that the $L_{A300s}$ measurement is conducted in a 5-min sliding window where the ANE is centered. This is a better approximation than using a 5-min fixed window as the mean distance between RTN samples and ANE samples is reduced. The fact that future samples are needed (plus the fact that the exact labeling can only be provided after listening and labeling the recordings) implies that this measurement can only be applied in an off-line analysis and not in the real-time operation mode of the WASN.

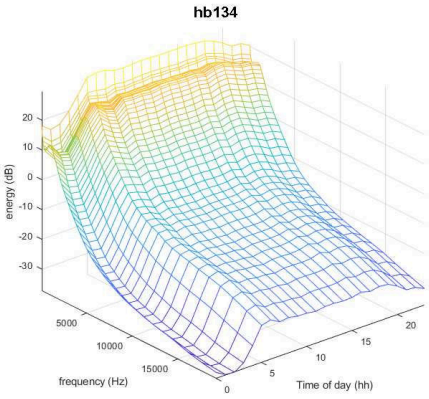
## 4. Dataset Analysis

In this section, a detailed analysis of the ANEs present in the labeled WASN-based audio database is described. Specifically, the distribution of the occurrences and duration of the ANE subcategories together with their contextual SNR distribution and the impact of each of them to the $L_{Aeq}$ value are analyzed. First, an analysis was conducted regarding the general characteristics of the new audio dataset considering the spectral and time behavior of nodes recordings along one day. Secondly, two more detailed analyses were performed: *(i)* aggregating values of occurrence, duration, and SNR, as well as $L_{Aeq}$ impact considering all network sensors at a whole and, *(ii)* highlighting the particularities of sensor's locations of the WASN in terms of the same ANE statistical measures.

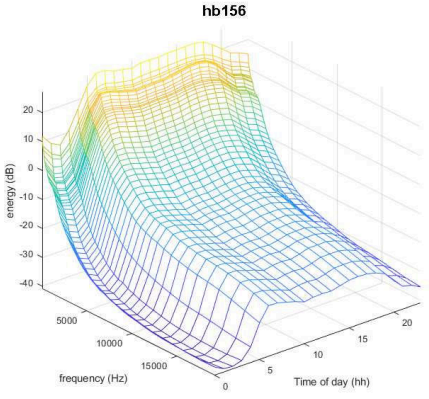
### *4.1. General Characteristics*

The global trend of the audio database obtained from the sensor network has been firstly analyzed through focusing mainly in RTN, disregarding specific characteristics of the ANE observed during recordings. In this regard, spectrum-time profiles [54] defined as the hourly time evolution of the mean spectrum of the audio has been computed for each sensor and day, following the same approach as in [24]. Their computation has been performed using the 48 frequency sub-bands of the Mel-Frequency Cepstral Coefficients (MFCC) features of the incoming audio signal, following the setup explained in [55], and computing the mean spectrum along the 20 min of audio gathered every hour. This mean spectrum obtained for each of the 24 h in a day conforms the spectrum-time profile of a complete measured day, which mainly includes RTN raw signal.

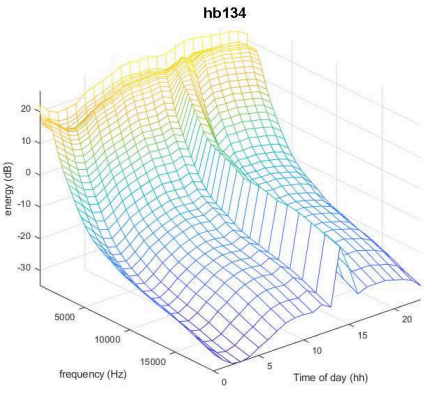
Figure 5 shows four examples of the measured spectrum-time profiles, for two sensors hb134 and hb156 and for the two recorded days of the campaign. As can be observed, the general trend of the spectrum-time profile during weekday (Figure 5a,b) is quite similar for these two sensors, while sensor hb156 presents a slightly higher energy at high frequencies and during a narrower daily period. During the weekend (Figure 5c,d) a very persistent ANE of hard rain was present in almost all sensors, which can be seen as a high spectrum-profile value around 14 h. Additionally, another aspect that can be highlighted comparing week spectrum-time profiles with those of the weekend is that during weekend the raising and decreasing of acoustic energy during daily period is smoother than during the week day, which can be explained by the fact that on work days there are traffic rush hours at the beginning and ending of working hours.



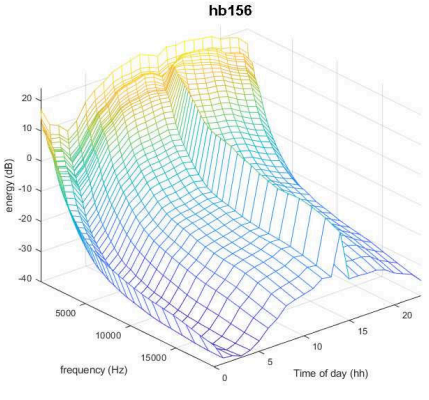

(**a**) Spectrum-time profile of sensor hb134 at week recording day.




(**b**) Spectrum-time profile of sensor hb156 at week recording day.




(**c**) Spectrum-time profile of sensor hb134 at weekend recording day.




(**d**) Spectrum-time profile of sensor hb156 at weekend recording day.

**Figure 5.** Spectrum-time profiles of sensors hb134 and hb156 at weekday (the 2nd of November 2017) and at weekend day (the 5th of November 2017).

### *4.2. Overall Analysis*

Table 1 lists the distribution of the ANEs in terms of their number of occurrences observed within the recorded database, showing both the aggregated and segregated distributions for all sensors, respectively. The total amount of ANEs labeled in the dataset generation represents a 1.8% of all the recorded time, together with around 3.4% of raw data labeled as *cmplx*, which is the data result of the mixture of noises or with unidentifiable sound.

**Table 1.** Aggregated ANE occurrences and their total duration distribution per category.

| Category | Number of Occurrences | Total Duration (s) |
|---|---|---|
| ANE | 3170 | 10,752.9 |
| rain | 754 | 6413.5 |
| brak | 737 | 1387.2 |
| thun | 236 | 753.4 |
| tran | 76 | 655.7 |
| trck | 380 | 382.4 |
| bird | 482 | 338.9 |
| sire | 55 | 327.8 |
| horn | 210 | 165.7 |
| alrm | 5 | 84.3 |
| inte | 8 | 69.4 |
| busd | 88 | 56.6 |
| stru | 46 | 45.8 |
| musi | 3 | 31.6 |
| door | 83 | 16.3 |
| airp | 4 | 14.5 |
| bike | 3 | 9.6 |

Table 1 shows the distribution of total ANE duration segregated by ANE subcategory. Most frequently observed ANEs were rain (6413.5 s) and thunder (753.4 s), due do the thunderstorm episode observed during the weekend day in the city of Rome at 14:00. Non-meteorological ANEs with higher total duration values were sounds of vehicles brakes (1387.2 s), trains (655.7 s), birds (338.9 s), sirens (327.8 s), horns (165.7 s), and sounds of trucks (382.4 s). ANEs with mid-total duration values were alarm sounds (84.3 s), interfering sounds (69.4 s), structure movement sounds (45.8 s), noise of pressurized air or *busd* (56.6 s), and music (31.6 s), while ANEs with lowest presence during the recordings were door sounds (16.3 s), airplanes (14.5 s), and bikes (9.6 s).

In Figure 6, the boxplots of the ANE durations are shown. The longer ANE correspond to *musi* and *alrm*, while the shorter are the *door*, *bird* and *busd*.

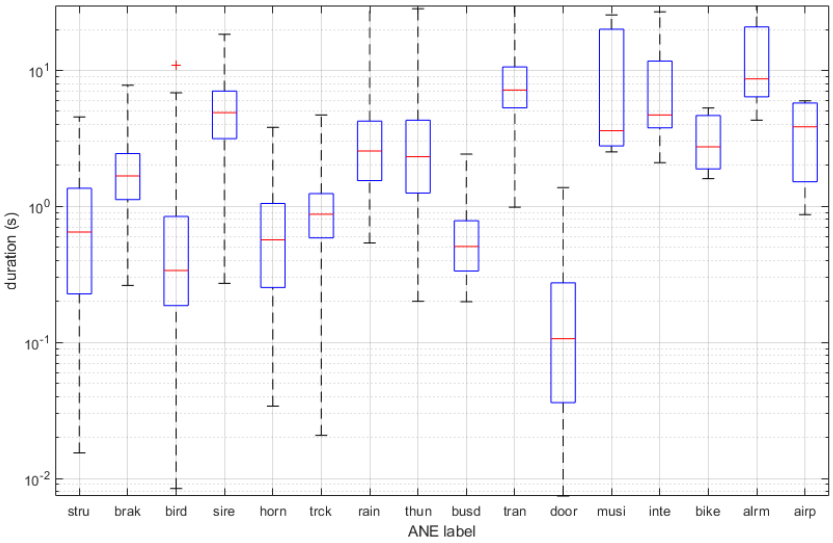



**Figure 6.** Boxplots of duration for each ANE category. Logarithmic axis of duration is used for a better observation of the duration values.

To determine the salience of the ANEs, the SNR measure is calculated by following the steps in Section 3.3.1. Figure 7 shows the SNR distributions for each ANE category. It can be observed that ANEs with higher SNR are sounds of trains, bikes, sirens, trucks, and horns.

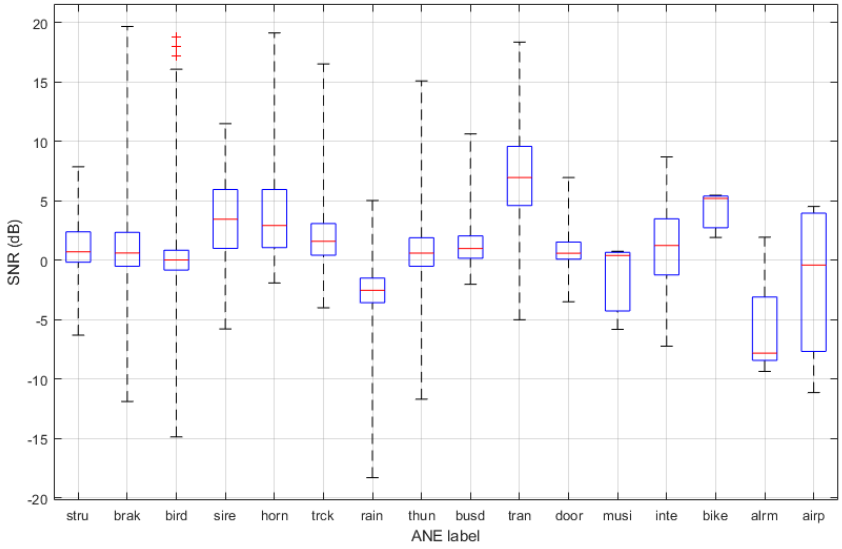



**Figure 7.** Boxplots of SNR for each ANE subcategory.

### *4.3. ANEs' Impact on the $L_{Aeq}$*

The ANE contribution in the noise map can be analyzed in several ways. In this study, each ANE has been characterized with the two variables analyzed in the previous section: duration and SNR. Nevertheless, the impact of each ANE to the noise map should be quantified, and preliminary studies with a smaller dataset show that has a strong dependence on the SNR and the duration values [52]. The impact consists of the $L_{A300s}$ measurement of the raw audio minus the same measurement after removing the noise event, which allows discovery of the final contribution of the event to the noise map (more details are given in Section 3.3.2).

To find a first approximation to the ANE subcategories and impact analysis, all the recorded ANEs are depicted according to their characteristics in Figure 8. The SNR is plotted on the vertical axis and the duration on the horizontal axis in a logarithmic scale, for illustration purposes. Also, the size of the marker represents the ANE impact, in a scale indicated in the legend. Besides, the class of the event is depicted in a color scale detailed in the legend, designed to distinguish events with similar parameters easily (the reader is referred to Section 3.2 for a list of all ANE subcategories).

In Figure 8 the reader may appreciate that the class distribution of the events is not uniform, as seen in Table 1. Events shorter than 1 s does not appear to have a significant impact on the $L_{A300s}$ when evaluated individually, as they represent an impact near 0 dB. However, events that last more than 4 s and have a positive SNR may contribute to the $L_{A300s}$ level with impacts of more than 3 dB. Among these significant ANEs, train pass-bys and sirens are the events presenting a higher overall impact score and presence, as also depicted in Figure 7 (where bike noise also appears to have a high median impact because it has only three occurrences). The events presenting a higher impact on the $L_{A300s}$ are trains and sirens, mainly, being only the trains the ANE that surpass the 3 dB level.

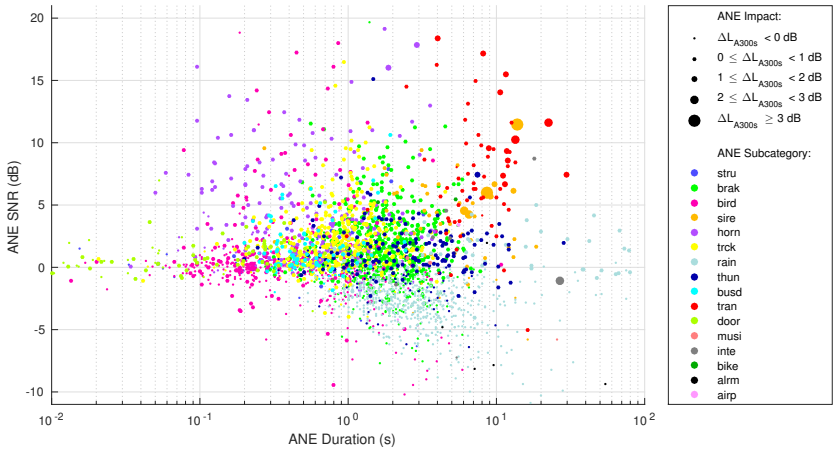


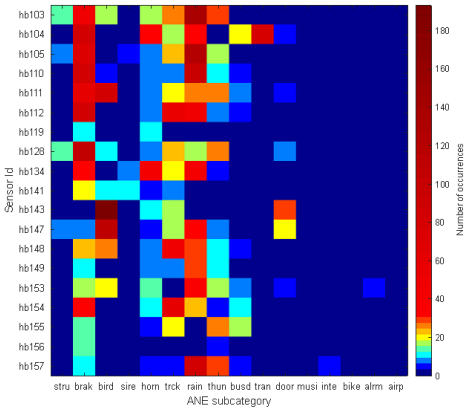

**Figure 8.** Scatterplot of all ANE parameters separated in subcategories (recorded and labeled in the 2nd and the 5th of November 2017).

It is worth mentioning that in some long events, the SNR and the impact on the $L_{A300s}$ have computation problems. Hence, in Figure 8, only ANEs with both a quantifiable SNR and impact are depicted. Of the total of 2014 events, only three have a duration of more than 300 s, hence, making the impact on the $L_{A300s}$ impossible to calculate. In addition, 61 events give SNR calculation problems, mainly due to high ANE density segments where the interpolation between RTN labels is not possible. All those events have been discarded from these representations.

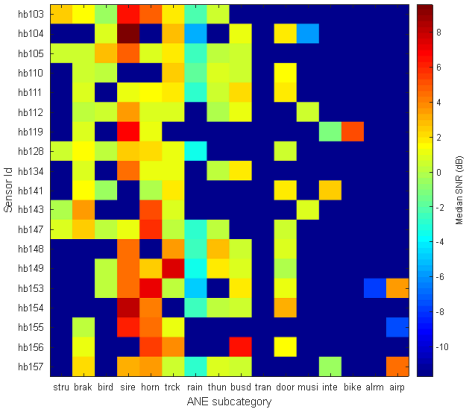
### *4.4. Node-Based Analysis*

The main upgrade of the dataset detailed in this work, apart from the time distribution of the recordings and the total amount of time labeled, is the fact that the data gathered corresponds to 19 different locations and nodes in a WASN. This leads us to detail a spatial study of the collected data, assuming that not all nodes will observe the same subcategories of ANEs and, of course, the same number of occurrences. This is a key study for the final usage of this dataset, which is the training of the ANED, to detect the ANE in all the sensors of the WASN. A first approach to the homogeneity of this network will be given by the results of the cross analysis between ANE subcategory and sensor Id.

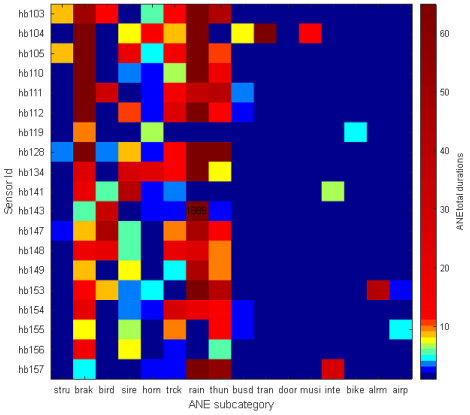
Figure 9a shows the ANE occurrences distribution segregated by sensors Id and ANE subcategory as an image, where it can be appreciated that birds in sensor hb143 are the ones more frequently observed, as birds produce short noise bursts and they can be very repetitive in certain locations and hours. Otherwise, the rain episode during the weekend day of the recording campaign is the second mostly observed event. From the same figure, it can be observed that noise events with quite uniform distribution across all the sensors network are some which are more related to traffic (horns, brakes, and sounds of trucks), and meteorological sounds during the weekend (rain and thunder). The rest of ANEs show a more irregular distribution across the sensor network during the recording campaign.



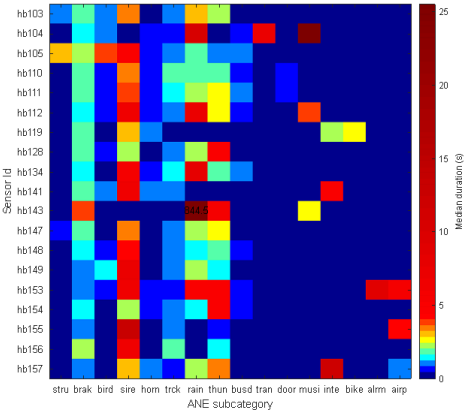

(**a**) ANE occurrences distribution for each ANE subcategory and sensor Id.




(**b**) ANE SNR distribution. Median SNR value is depicted for each ANE subcategory and sensor Id.




(**c**) ANE total duration distribution. The median duration value is depicted for each ANE subcategory and sensor Id.




(**d**) ANE duration distribution. The median duration value is depicted for each ANE subcategory and sensor Id.

**Figure 9.** ANE parameters distribution per subcategory for each WASN node.

Figure 9b shows the median SNR values segregated by sensors and ANE subcategory. There are several nodes and ANE subcategories that exhibit positive SNR values, and their maximum values are attained for sirens, horns, trucks, and *buses* while medium SNRs are observed for brakes, birds, thunder, doors, and airplanes. The negative median SNR values of *rain* can be basically explained by the fact that the used computation methodology of SNR (see [23] for further details) is imprecise when the ANE duration is too high because of the underlying stationary assumption of the RTN assumed within this method.

Figure 9c shows the total duration of ANE recorded for each ANE type and sensor Id. The colormap scale leaves the maximum value as an outlier, depicted as a numeric value 1689 for rain in sensor hb143. Regarding the other values, it can be appreciated that *brak*, *sire*, *horn*, *trck*, *rain* and *thun* are the ANEs with more regular presence across the entire sensor network, while other ANEs like *stru*, *tran*, *musi*, *inte*, *bike*, and *alrm* are more irregularly observed. Figure 9d shows the ANE mean duration

distribution per category for each of the WASN sensors, i.e., the median ANE duration statistic has been computed for each cell of the depicted matrix. The color bar legend also presents an outlier, which is reached by sensor hb143 and rain ANE category. The following maximum mean duration is related to sensor hb104 and *musi*.

As a conclusion of this WASN-based analysis, *siren* is the subcategory of ANEs with longest duration and with presence in most of the sensors, and so are *horns*, but the latter have shorter duration; both ANEs present high values of SNR. Another two subcategories are present in most of the sensors in this WASN-based recording campaign; *rain* and *thunder* have a wide presence in the recording, due to the fact that on the 5th of November nearly all the WASN suffered heavy rain in the afternoon. The main difference between them is that while *rain* presents mainly low SNR values, *thunder* presents mid SNR values. *birds* present values of high occurrence and duration for several sensors, but the SNR associated with the birdsong is moderate or low. Nevertheless, the main output of this analysis is that there are few events with uniform appearance in all the sensors, and that most of the subcategories labeled in this work correspond to recordings of fewer groups of sensors. This leads us to the conclusion that the WASN-based recording considering all the sensors in the network was a requirement to observe the variety of the distribution of the events occurring in the entire network.

## 5. Discussion and Conclusions

In this section, several key aspects of this work are discussed and concluded after the recording, labeling, and analysis of the WASN-based suburban scenario audio samples.

### *5.1. WASN-Based Dataset Analysis vs. Expert-Based Dataset*

As previously stated, a preliminary suburban environment dataset in Rome was published in [23], which consisted of a set of expert-based recordings of limited duration and scope. The work presented in this article is based on the knowledge acquired from that first dataset. It responds to the need for increasing the coverage of the RTN and the ANEs at all hours of the day and night, the weekend, and even when elements external to the noise appear in the measurements, such as adverse weather conditions, i.e., in real-operation conditions.

The WASN-based dataset presented in this work has been enriched with seven classes in comparison to the previous one: *rain*, *thun*, *tran*, *bird*, *alrm*, *inte* and *airp*. On the contrary, it has not been possible to record the noise of people talking (*peop*) as in the previous dataset caused by the presence of workers in the portals. All the rest of the ANEs were already part of the expert-based dataset, but with fewer occurrences because the dataset was much smaller. The preliminary dataset contained 3.2% of ANE of the total recorded time, and this new dataset contains 1.8%. A possible explanation to these differences is that the expert-based dataset recording was centered in daytime and this WASN-based dataset has recorded day and night, where night shows low presence of ANE with respect to the day.

The longest ANEs have been found within the *sire* subcategory, while the shortest ones are found in the *door* subcategory, as also happened in the expert-based recording. Moreover, the ANEs labeled as *horn* and *sire* present the highest SNR in both datasets. However, it is worth noting that in this new dataset there are samples of *tran* and *inte* subcategories that also entail high SNRs in many occurrences, a characteristic that was not found in [23].

From this comparative analysis, it can be deduced that the data captured in the expert-based dataset was suitable enough for the first characterization of the suburban soundscape. Nevertheless, the WASN-based recording campaign has shown that there were several noise subcategories that in the preliminary recording campaign had not been recorded and labeled, which present critical characteristics in terms of SNR and duration.

### *5.2. WASN-Based Dataset and Node Homogeneity*

The final use of the presented dataset is the training of the ANED algorithm for a precise detection of the ANEs in all the nodes of the WASN already deployed in the Rome pilot area. Although the developed dataset contains a significantly larger sample of both RTN and ANEs—around 19 times the data gathered in the preliminary expert-based dataset [23]—the analysis of the acoustic data confirms the heterogeneous distribution of the ANEs subcategories in real-life environments already observed in [23]. This heterogeneity has also been observed across the nodes as detailed in Section 4.4. Although some of the subcategories occurred in most of the nodes (e.g., *brak*, *rain*, *truck* and *horn*), there are others found particular in some sensors of the network (e.g., *airp*, *bike* and *train*).

The design of a WASN-based dataset raises the hypothesis of homogeneity of the raw data captured in each of the nodes of the network. This hypothesis was analyzed by means of the distribution of the ANEs in the previous recording campaign [56], with an analysis of the five recording locations. To ensure that the ANED will operate properly in all the nodes, their acoustic environments should present a certain homogeneity in terms of frequency distribution. To collect the data in similar conditions, all nodes have been installed maintaining the same distances and orientations to the road. Nevertheless, a study to evaluate the homogeneity should be carried out taking into account both the locations of the nodes and the occurrences of the ANEs with the final goal of a generalist training of ANED for all the network.

### *5.3. Impact on the $L_{A300s}$ of the ANE Subcategories*

In this paper, the impact on the computation of $L_{A300s}$ has been evaluated for every individual ANE every 5 min. The analysis presents interesting results to be discussed considering the SNR and duration of individual ANEs. The results in Section 4 show the existence of several ANEs with high impact, which present high SNR and long duration. After the individual analysis of the impact of ANEs on the $L_{A300s}$ computation, future work will also take into account the fact that dynamic acoustic mapping in real-life conditions face a more complex operating scenario. In a real-operation scenario, several ANEs can occur in a predefined integration time, so the ANE impact must be evaluated in an aggregated way for each period.

Another relevant result of the individual ANE analysis is the presence of ANEs with negative SNR. As detailed in Section 3.3.1, the SNR is evaluated taking into account each ANE in relation to its surrounding RTN signal level. In certain cases, the RTN decreases as the ANE occurs, so a negative SNR is obtained. Therefore, only those events with positive SNR should be removed from the $L_{Aeq}$ computation, as also concluded in [52].

From this work, it can be concluded that working with data recorded in a real operating scenario is crucial to obtain a reliable modelling of the nodes' acoustic environment, according to the differences observed between the expert- and WASN-based datasets. Moreover, the analysis of the WASN-based collected data again shows the important role played by SNR and duration of individual ANEs in their impact on the $L_{Aeq}$ computation, obtaining ANEs that should be considered for their high impact on the equivalent level.

**Author Contributions:** R.M.A.-P. has written part of the paper and has participated in the recording and labeling of the dataset. F.O. implemented the SNR and impact calculations, conducted the part of the dataset analysis and participated in the audio labeling and writing of the paper. F.A. has written the related work and reviewed the entire paper, with a special support on the contributions. J.C.S. has worked in the technical part, the recording, the labeling, and the analysis of the dataset.

**Funding:** The research presented in this work has been partially supported by the LIFE DYNAMAP project (LIFE13 ENV/IT/001254). Francesc Alías thanks the Obra Social La Caixa for grant ref. 2018-URL-IR2nQ-029. Rosa Ma Alsina-Pagès thanks the Obra Social La Caixa for grant ref. 2018-URL-IR2nQ-038. Ferran Orga thanks the support of the European Social Fund and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Catalan Government for the pre-doctoral FI grant No. 2019_FI_B2_00168.

**Acknowledgments:** The authors would like to thank Marc Hermosilla, Ester Vidaña, Alejandro González and Sergi Barqué for helping in the audio labeling. The authors would like to thank ANAS S.p.A. for the picture of the sensor installed in the portal and Bluewave for the recording of the audio files in the sensors of the WASN.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AED | Audio Event Detector |
| ANE | Anomalous Noise Event |
| ANED | Anomalous Noise Event Detection |
| BCK | Background Noise |
| CASA | Computational Scene and Event Analysis |
| CLEAR | Classification of Events, Activities and Relationships |
| DNN | Deep Neural Networks |
| DYNAMAP | Dynamic Noise Mapping |
| END | European Noise Directive |
| GTCC | Gammatone Cepstral Coefficients |
| MFCC | Mel-Frequency Cepstral Coefficients |
| RTN | Road-Traffic Noise |
| SED | Sound Event Detector |
| SONYC | Sounds of New York City |
| SNR | Signal-to-Noise Ratio |
| VPN | Virtual Private Network |
| WASN | Wireless Acoustic Sensor Network |

## References

1. Goines, L.; Hagler, L. Noise Pollution: A Modern Plague. *South. Med. J. Birm. Ala.* **2007**, *100*, 287–294. [CrossRef]
2. Guite, H.; Clark, C.; Ackrill, G. The impact of the physical and urban environment on mental well-being. *Public Health* **2006**, *120*, 1117–1126. [CrossRef]
3. Hänninen, O.; Knol, A.B.; Jantunen, M.; Lim, T.A.; Conrad, A.; Rappolder, M.; Carrer, P.; Fanetti, A.C.; Kim, R.; Buekers, J.; et al. Environmental burden of disease in Europe: Assessing nine risk factors in six countries. *Environ. Health Perspect.* **2014**, *122*, 439. [CrossRef] [PubMed]
4. Bello, J.P.; Silva, C. SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution. *Commun. ACM* **2019**, *62*, 68–77. [CrossRef]
5. 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. *Off. J. Eur. Commun.* **2002**, *L189*, 12–25.
6. Manvell, D. Utilising the Strengths of Different Sound Sensor Networks in Smart City Noise Management. In Proceedings of the EuroNoise 2015, Maastrich, The Netherlands, 1–3 June 2015; EAA-NAG-ABAV: Maastrich, The Netherlands, 2015; pp. 2305–2308.
7. Alías, F.; Alsina-Pagès, R.M. Review of Wireless Acoustic Sensor Networks for Environmental Noise Monitoring in Smart Cities. *J. Sens.* **2019**. [CrossRef]
8. Nencini, L.; De Rosa, P.; Ascari, E.; Vinci, B.; Alexeeva, N. SENSEable Pisa: A wireless sensor network for real-time noise mapping. In Proceedings of the EuroNoise 2012, Prague, Czech Republic, 10–13 June 2012; pp. 10–13.
9. Botteldooren, D.; De Coensel, B.; Oldoni, D.; Van Renterghem, T.; Dauwe, S. Sound monitoring networks new style. In Proceedings of the Acoustics 2011, Gold Coast, Australia, 2–4 November 2011; Mee, D., Hillock, I.D., Eds.; Australian Acoustical Society: Queensland, Australia, 2011; pp. 93:1–93:5.
10. Mietlicki, F.; Mietlicki, C.; Sineau, M. An innovative approach for long-term environmental noise measurement: RUMEUR network. In Proceedings of the EuroNoise 2015, Maastrich, The Netherlands, 1–3 June 2015; EAA-NAG-ABAV: Maastrich, The Netherlands, 2015; pp. 2309–2314.

11. Camps-Farrés, J.; Casado-Novas, J. Issues and challenges to improve the Barcelona Noise Monitoring Network. In Proceedings of the EuroNoise 2018, Crete, Greece, 27–31 May 2018; EAA—HELINA: Heraklion, Crete, Greece, 2018; pp. 693–698.
12. Bain, M. SENTILO—Sensor and Actuator Platform for Smart Cities. 2014. Available online: https://joinup.ec.europa.eu/document/sentilo-sensor-and-actuator-platform-smart-cities (accessed on 25 June 2018).
13. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189. [CrossRef]
14. De la Piedra, A.; Benitez-Capistros, F.; Dominguez, F.; Touhafi, A. Wireless sensor networks for environmental research: A survey on limitations and challenges. In Proceedings of the 2013 IEEE EUROCON, Zagreb, Croatia, 1–4 July 2013; pp. 267–274.
15. Rawat, P.; Singh, K.D.; Chaouchi, H.; Bonnin, J.M. Wireless Sensor Networks: A Survey on Recent Developments and Potential Synergies. *J. Supercomput.* **2014**, *68*, 1–48. [CrossRef]
16. Bertrand, A. Applications and trends in wireless acoustic sensor networks: A signal processing perspective. In Proceedings of the 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT), Ghent, Belgium, 22–23 November 2011; pp. 1–6.
17. Griffin, A.; Alexandridis, A.; Pavlidiand, D.; Mastorakis, Y.; Mouchtaris, A. Localizing multiple audio sources in a wireless acoustic sensor network. *Signal Process.* **2015**, *107*, 54–67. [CrossRef]
18. Socoró, J.C.; Alías, F.; Alsina-Pagès, R.M. An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments. *Sensors* **2017**, *17*, 2323. [CrossRef] [PubMed]
19. Wang, D.; Brown, G.J. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*; Wiley-IEEE Press: New York, NY, USA, 2006.
20. Giannoulis, D.; Benetos, E.; Stowell, D.; Rossignol, M.; Lagrange, M.; Plumbley, M.D. Detection and classification of acoustic scenes and events: An IEEE AASP challenge. In Proceedings of the 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 20–23 October 2013; pp. 1–4.
21. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Reliable detection of audio events in highly noisy environments. *Pattern Recognit. Lett.* **2015**, *65*, 22–28. [CrossRef]
22. Mesaros, A.; Diment, A.; Elizalde, B.; Heittola, T.; Vincent, E.; Raj, B.; Virtanen, T. Sound event detection in the DCASE 2017 Challenge. *IEEE/ACM Trans. Audio Speech Language Process.* **2019**, *27*, 992–1006. [CrossRef]
23. Alías, F.; Socoró, J.C. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Appl. Sci.* **2017**, *7*, 146. [CrossRef]
24. Socoró, J.C.; Alsina-Pagès, R.M.; Alías, F.; Orga, F. Adapting an Anomalous Noise Events Detector for Real-Life Operation in the Rome Suburban Pilot Area of the DYNAMAP's Project. In Proceedings of the EuroNoise, Crete, Greece, 27–31 May 2018.
25. Nakajima, Y.; Sunohara, M.; Naito, T.; Sunago, N.; Ohshima, T.; Ono, N. DNN-based Environmental Sound Recognition with Real-recorded and Artificially-mixed Training Data. In Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (InterNoise 2016), Hamburg, Germany, 21–24 August 2016; German Acoustical Society (DEGA): Hamburg, Germany, 2016; pp. 3164–3173.
26. Mesaros, A.; Heittola, T.; Virtanen, T. TUT database for acoustic scene classification and sound event detection. In Proceedings of the 24th European Signal Processing Conference (EUSIPCO 2016), Budapest, Hungary, 28 August–2 September 2016; Volume 2016, pp. 1128–1132.
27. Stowell, D.; Giannoulis, D.; Benetos, E.; Lagrange, M.; Plumbley, M.D. Detection and Classification of Acoustic Scenes and Events. *IEEE Trans. Multimed.* **2015**, *17*, 1733–1746. [CrossRef]
28. Heittola, T.; Mesaros, A.; Eronen, A.; Virtanen, T. Context-dependent sound event detection. *EURASIP J. Audio Speech Music Process.* **2013**, *2013*, 1–13. [CrossRef]
29. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Audio Surveillance of Roads: A System for Detecting Anomalous Sounds. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 279–288. [CrossRef]
30. Zambon, G.; Benocci, R.; Bisceglie, A.; Roman, H.E.; Bellucci, P. The LIFE DYNAMAP project: Towards a procedure for dynamic noise mapping in urban areas. *Appl. Acoust.* **2017**, *124*, 52–60. [CrossRef]
31. Bellucci, P.; Peruzzi, L.; Zambon, G. LIFE DYNAMAP project: The case study of Rome. *Appl. Acoust.* **2017**, *117*, 193–206. [CrossRef]

32. Stowell, D.; Plumbley, M.D. An open dataset for research on audio field recording archives: Freefield1010. *arXiv* **2013**, arXiv:1309.5275.
33. Salamon, J.; Jacoby, C.; Bello, J.P. A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 1041–1044.
34. Socoró, J.C.; Ribera, G.; Sevillano, X.; Alías, F. Development of an Anomalous Noise Event Detection Algorithm for dynamic road traffic noise mapping. In Proceedings of the 22nd International Congress on Sound and Vibration (ICSV22), Florence, Italy, 12–16 July; The International Institute of Acoustics and Vibration: Florence, Italy, 2015; pp. 1–8.
35. Piczak, K.J. ESC: Dataset for Environmental Sound Classification. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 1015–1018.
36. Rossignol, M.; Lafay, G.; Lagrange, M.; Misdariis, N. SimScene: A web-based acoustic scenes simulator. In Proceedings of the 1st Web Audio Conference (WAC), Paris, France, 26–28 January 2015; pp. 1–6.
37. Gloaguen, J.R.; Can, A.; Lagrange, M.; Petiot, J.F. Estimating Traffic Noise Levels using Acoustic Monitoring: A Preliminary Study. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 (DCASE'2016), Budapest, Hungary, 3 September 2016; pp. 40–44.
38. Çakır, E.; Parascandolo, G.; Heittola, T.; Huttunen, H.; Virtanen, T. Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 1291–1303. [CrossRef]
39. Salamon, J.; MacConnell, D.; Cartwright, M.; Li, P.; Bello, J.P. Scaper: A library for soundscape synthesis and augmentation. In Proceedings of the 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 15–18 October 2017; pp. 344–348.
40. Salamon, J.; Bello, J.P. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [CrossRef]
41. Temko, A.; Malkin, R.; Zieger, C.; Macho, D.; Nadeu, C.; Omologo, M. CLEAR Evaluation of Acoustic Event Detection and Classification Systems. In *Multimodal Technologies for Perception of Humans: First International Evaluation Workshop on Classification of Events, Activities and Relationships, CLEAR 2006*; Stiefelhagen, R., Garofolo, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 311–322.
42. Temko, A. Acoustic Event Detection and Classification. Ph.D. Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, 2007.
43. Heittola, T.; Klapuri, A. TUT Acoustic Event Detection System 2007. In *Multimodal Technologies for Perception of Humans. International Evaluation Workshops CLEAR 2007 and RT 2007*; Stiefelhagen, R., Bowers, R., Fiscus, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 364–370.
44. Van Grootel, M.W.W.; Andringa, T.C.; Krijnders, J.D. DARES-G1: Database of Annotated Real-world Everyday Sounds. In Proceedings of the NAG/DAGA Meeting 2009, Rotterdam, The Netherlands, 23–26 March 2009.
45. Giannoulis, D.; Stowell, D.; Benetos, E.; Rossignol, M.; Lagrange, M.; Plumbley, M.D. A database and challenge for acoustic scene classification and event detection. In Proceedings of the 21st European Signal Processing Conference (EUSIPCO 2013), Marrakesh, Marocco, 9–13 September 2013; pp. 1–5.
46. Mesaros, A.; Heittola, T.K.; Benetos, E.; Foster, P.; Lagrange, M.; Virtanen, T.; Plumbley, M.D. Detection and Classification of Acoustic Scenes and Events: Outcome of the DCASE 2016 Challenge. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 379 – 393. [CrossRef]
47. Mesaros, A.; Heittola, T.; Diment, A.; Elizalde, B.; Shah, A.; Vincent, E.; Raj, B.; Virtanen, T. DCASE 2017 challenge setup: Tasks, datasets and baseline system. In Proceedings of the DCASE 2017-Workshop on Detection and Classification of Acoustic Scenes and Events, Munich, Germany, 16–17 November 2017
48. Nencini, L. DYNAMAP monitoring network hardware development. In Proceedings of the 22nd International Congress on Sound and Vibration (ICSV22), Florence, Italy, 12–16 July 2015; The International Institute of Acoustics and Vibration (IIAV): Florence, Italy, 2015; pp. 1–4.
49. Nencini, L. Progetto e realizzazione del sistema di monitoraggio nell'ambito del progetto Dynamap. In Proceedings of the 44° Congress of the Italian Acoustic Association, Pavia, Italy, 7–9 June 2017.
50. International Electroacoustics Commission. *Electroacoustics-Sound Level Meters-Part 1: Specifications (IEC 61672-1)*; International Electroacoustics Commission: Geneva, Switzerland, 2013.

51. Nencini, L.; Bisceglie, A.; Bellucci, P.; Peruzzi, L. Identification of failure markers in noise measurement low cost devices. In Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (InterNoise 2016), Hamburg, Germany, 21–24 August 2016; pp. 6362–6369.
52. Orga, F.; Alías, F.; Alsina-Pagès, R.M. On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments. *Int. J. Environ. Res. Public Health* **2017**, *15*, 13. [CrossRef] [PubMed]
53. Pierre, R.L.S.; Maguire, D.J. The impact of A-weighting sound pressure level measurements during the evaluation of noise exposure. In Proceedings of the NOISE-CON 2004, Baltimore, MD, USA,12–14 July 2004; pp. 1–8.
54. Socoró, J.C.; Alsina-Pagès, R.M.; Alías, F.; Orga, F. Acoustic Conditions Analysis of a Multi-Sensor Network for the Adaptation of the Anomalous Noise Event Detector. *Proceedings* **2019**, *4*, 51. [CrossRef]
55. Valero, X.; Alías, F. Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification. *IEEE Trans. Multimed.* **2012**, *14*, 1684–1689. [CrossRef]
56. Orga, F.; Alsina-Pagès, R.M.; Alías, F.; Socoró, J.C.; Bellucci, P.; Peruzzi, L. Anomalous Noise Events Considerations for the Computation of Road Traffic Noise Levels in Suburban Areas: The DYNAMAP's Rome Case Study. In Proceedings of the 44° Congress of the Italian Acoustic Association, Pavia, Italy, 7–9 June 2017.


Capítol 5

# Anàlisi de l'impacte en el nivell equivalent

Els esdeveniments acústics, fragments de so d'un origen determinat, es poden caracteritzar de diverses maneres. En aquest capítol es presenten primer dos paràmetres que pretenen mesurar la rellevància del so en la mesura del nivell equivalent. En concret, es consideren dues mètriques, la durada de l'esdeveniment i la relació senyal-soroll (SNR, de l'anglès *Signal-to-Noise Ratio*); la primera parametritza la rellevància d'un esdeveniment en el temps, i la segona en caracteritza la prominència.

També es proposa un càlcul, l'impacte en el nivell equivalent, que mesura la contribució dels esdeveniments sobre el nivell equivalent sonor dins d'un període temporal predefinit. Aquesta mètrica s'utilitzarà, en última instància, per analitzar la tipologia dels esdeveniments que poden esbiaixar el càlcul del nivell equivalent d'un mapa de soroll de trànsit.

Finalment, s'analitza l'impacte dels esdeveniments a les dues àrees pilot del projecte DYNAMAP, Milà i Roma. Això es fa de dues maneres, calculant l'impacte individual, analitzant esdeveniment per esdeveniment; i calculant l'impacte agregat, que mesura l'impacte de tots els esdeveniments en un període determinat.

## 5.1 Caracterització dels esdeveniments acústics

Per una banda, la durada de l'esdeveniment dona una indicació de la seva presència en el temps; i, per l'altra la SNR que quantifica la rellevància acústica de l'esdeveniment respecte al soroll de fons. A més a més, es proposa una mesura per avaluar la implicació de les dues mètriques calculades, l'impacte en el nivell equivalent.

### 5.1.1 La durada

La durada (s) es calcula com la diferència entre l'instant inicial i el final de l'esdeveniment. En els casos on l'esdeveniment acústic té un inici i un final molt marcats, podrem obtenir una mesura precisa de la durada d'un so; però en casos on l'inici i/o el final siguin progressius, la durada de l'etiqueta pot variar entre els etiquetadors. L'etiquetatge, per tant, condiciona el càlcul de la durada i caldrà establir unes pautes entre tots els tècnics abans d'iniciar el procés d'etiquetatge.

En el cas de DYNAMAP, el criteri que s'ha seguit ha estat el d'etiquetar l'objecte mentre estigui en primer pla. D'aquesta manera, els esdeveniments

etiquetats com a tal sempre tenen el so de la classe que correspon amb una presència majoritària respecte als altres sons. Tots els fragments on hi havia barreges de diferents classes, han estat etiquetats com a `cmplx:so_1+so_2+...` tal com s'ha indicat en el capítol 4.

### 5.1.2 L'SNR

L'SNR (dB) es computa com la ràtio entre la potència de l'esdeveniment i la potència del soroll que hi ha al voltant. Seguint el càlcul clàssic de la SNR que s'utilitza en processament del senyal [1], es pretén mesurar la prominència d'un esdeveniment respecte al nivell de soroll del fons. En aquest cas, pel context del problema, el senyal que es mesura equival al nivell sonor de l'esdeveniment i el soroll equival al nivell sonor de trànsit que hi ha en aquell moment. Pel fet que el soroll de fons no és estacionari, és complicat de mesurar la potència de l'esdeveniment en qüestió, per tant, s'ha de cercar la manera d'aproximar-lo al màxim, mesurant-lo en intervals de temps pròxims al senyal acústic. Concretament, en aquest càlcul el soroll de fons és el mesurat en l'interval de temps abans de l'esdeveniment i el de després, a parts iguals. D'aquesta manera, es quantifica la prominència d'un esdeveniment respecte al soroll de trànsit de fons de forma aproximada.

En el cas DYNAMAP, els edeveniments acústics són els ANEs, tal com s'han definit en apartats anteriors. I el soroll de fons és l'RTN, el so més present i constant tant en l'àmbit urbà com en el suburbà. D'aquesta manera, només podrem mesurar l'SNR, en aquest cas, dels ANEs i sempre que al voltant hi hagi RTN.

A l'equació (5.1) s'hi mostra el càlcul de l'SNR, adaptat de la clàssica SNR, substituint la potència del senyal per la potència de l'ANE a mesurar i la potència del soroll per la del RTN al voltant. El detall del càlcul de la potència és a l'equació (5.2), que serveix per a les dues potències.

$$SNR = 10 \log_{10} \left( \frac{P_{ANE}}{P_{RTN}} \right), \tag{5.1}$$

on

$$P_x = \sum_{n=1}^{N_x} \left( \frac{x^2[n]}{N_x} \right), \tag{5.2}$$

la $x[n]$ és el vector del senyal acústic de $N_x$ mostres de longitud, sigui l'ANE o el soroll de trànsit.

En el cas de l'RTN, es consideren les $N_x/2$ mostres anteriors a l'ANE i les $N_x/2$ mostres següents a aquest, sempre que sigui possible. Quan no és possible, es consideren només les anteriors o les posteriors i si el nombre de mostres d'RTN disponibles és menor a l'interval de durada de l'ANE, es calcula la potència amb menys mostres. Per a més detall en les diverses casuístiques, el lector pot consultar [2].

## 5.2 Càlcul de l'impacte individual dels esdeveniments acústics

Recuperem la segona de les hipòtesis plantejades en aquesta tesi:

> P2: *Provocarien, els esdeveniments amb durada i prominència significatives, un impacte considerable en el càlcul del nivell equivalent? Serà factible classificar-los en regions d'impacte en funció de la prominència i la durada?*

La durada ens dona la rellevància de l'esdeveniment en el temps, però no la prominència i, en canvi, l'SNR, en mesura la prominència respecte a l'entorn, però no té consciència de la rellevància de l'esdeveniment a l'escala temporal. Per això, es creu necessària la implementació d'una eina nova: l'impacte determina la contribució que té un esdeveniment sonor particular en una mesura del nivell de so equivalent d'un interval donat $T$, o sigui l'$L_{eq,T}$.

Com que no és possible de separar les fonts sonores amb precisió, cal fer una aproximació per esbrinar com seria el nivell equivalent sonor si l'esdeveniment no hagués succeït. Es calcula com la diferència entre l'$L_{eq}$ calculat amb les mostres de so crues i l'$L_{eq}$ després d'eliminar l'esdeveniment avaluat. Per eliminar l'esdeveniment es proposa la interpolació del període etiquetat, des d'abans de l'inici i fins passat el so. Però en cap cas es pot fer amb les mostres d'àudio, ja que en tractar-se d'ones de pressió, la interpolació generaria un so sense cap sentit aparent i amb un nivell equivalent sonor incoherent. Com que el rellevant no és substituir l'esdeveniment, sinó el seu nivell acústic, es procedeix a interpolar l'$L_{eq}$ en lloc de les mostres acústiques obtenint així el nivell equivalent del període sense l'esdeveniment, *i.e.* $L_{eq,\overline{ANE}}$. Així s'obté l'estimació de la contribució de l'esdeveniment concret en el nivell equivalent sonor, expressat com a $\Delta L_{eq}$. És a dir: $\Delta L_{eq} = L_{eq,ANE} - L_{eq,\overline{ANE}}$.

Recordem que l'impacte només té sentit si s'avalua en un interval de temps donat, i que aquest interval ha de ser prou gran per contenir l'esdeveniment en qüestió. Per tant, el càlcul de l'$\Delta L_{eq}$ s'ha de fer amb una finestra temporal que pugui contenir els esdeveniments etiquetats, i la interpolació s'haurà d'efectuar amb mesures d'$L_{eq}$ calculades en períodes tan curts com sigui possible, per tal de tenir prou granulació en el punt inicial i el punt final de l'esdeveniment. Aquest càlcul de l'$L_{eq}$ a nivell inferior al de l'interval de mesura, es fa amb una finestra lliscant per totes les mostres de l'àudio. I per tal de generar el nivell de so d'una mostra contreta en un període de temps $T$, la finestra d'integració queda centrada en la mostra on s'efectua el càlcul, considerant la meitat de mostres anteriors i la meitat següents, sumant un total d'$N$ mostres. Per tant, si volem que els dos punts interpolats no continguin mostres de l'esdeveniment a avaluar, caldrà prendre la mostra $\frac{N}{2}$ anterior a l'inici de l'esdeveniment i la mostra $\frac{N}{2}$ següent al final d'aquest. Això suposa perdre una finestra d'integració sencera quan avaluem l'impacte d'un esdeveniment sonor. El càlcul de l'impacte individual està detallat a [2].

### 5.2.1 L'anàlisi de l'impacte individual dels esdeveniments acústics en el cas DYNAMAP

En el cas del projecte DYNAMAP, l'interval d'actualització dels mapes acústics canvia en funció de l'àrea pilot i del període de dia [3]. Amb la finalitat de mesurar si l'impacte esbiaixa el nivell presentat en el mapa sonor, caldrà avaluar-lo en l'interval de mesura. En treballs anteriors [2], s'estudia l'impacte en tres intervals diferents: 1, 5 i 15 minuts, dels esdeveniments de la primera base de dades (9 h). L'interval d'un minut no recull tots els esdeveniments, ja que alguns duren més d'aquest període, i a més, no s'utilitza en l'actualització dinàmica dels mapes de DYNAMAP. I l'interval de 15 minuts dil·lueix els impactes individuals, ja que no representa bé els esdeveniments de curta durada. Aprofitant que el moment de refresc del mapa de DYNAMAP més curt és de 5 minuts, s'ha agafat aquest interval com al més crític. Així, en aquesta tesi es detallen els impactes dels esdeveniments de DYNAMAP en aquest període.

A més a més, abans de calcular el nivell equivalent, s'aplica el filtre A sobre l'àudio, que filtra algunes freqüències del so [4]. Aquest filtre, que permet emular el nivell de sonoritat que percebem les persones, atenua sobretot les freqüències més greus, on l'oïda humana té menys sensibilitat, i s'utilitza freqüentment en els mesuraments de soroll ambiental. Per tant, ja que el mapa acústic de DYNAMAP presenta la mesura del nivell equivalent de so després del filtre A, en intervals de 5 minuts ($L_{A300s}$), ens interessarà mesurar els impactes individuals dels esdeveniments en aquestes mateixes condicions, *i.e.* $\Delta L_{A300s}$.

En el cas de DYNAMAP, la finestra d'integració a nivell inferior on s'efectua la interpolació, és d'un segon, *i.e.* $L_{A1s}$. Aquest interval està fixat pel mateix funcionament del sistema, que calcula el nivell equivalent en finestres d'un segon. De manera heurística, s'ha demostrat que els esdeveniments més curts de la base de dades queden representats correctament i que és prou gran per no afegir mesures de soroll imperceptibles. Per aquest motiu, la interpolació s'efectua entre la mostra 500 ms abans de l'esdeveniment i la de 500 ms després [2].

Finalment, cal destacar que l'impacte dels esdeveniments en l'interval de 5 minuts, s'efectua amb finestres lliscants d'aquesta mida. D'aquesta manera, l'esdeveniment analitzat queda centrat en l'interval de 300 s i s'aconsegueix aproximar millor el nivell acústic del soroll de trànsit en els casos on l'esdeveniment cauria a l'inici de la mesura d'$L_{A300s}$ o al final.

La Comissió Europea té un grup de treball encarregat de l'avaluació de l'exposició al soroll (ECWG-AEN de les sigles en anglès *European Commission Working Group - Assessment of Exposure to Noise*), que considera que 2 dB és l'error màxim tolerat en la mesura del nivell equivalent sonor d'un determinat segment [5]. Per aquest motiu, en aquesta tesi i en el projecte DYNAMAP, es considera que un esdeveniment que implica un biaix de 2 dB o més a la corba de l'$L_{eq}$ té un impacte alt, ja que esbiaixa directament la mesura del segment per damunt del llindar d'error tolerat.

A les figures 5.1 i 5.2 s'hi pot observar l'impacte de tots els esdeveniments

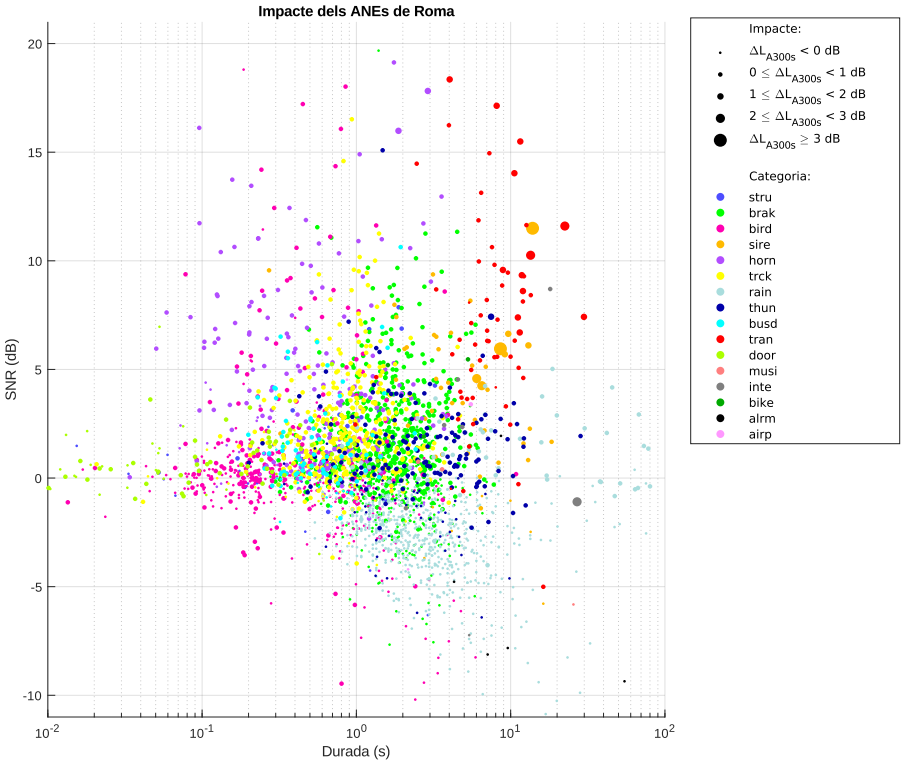



Figura 5.1: Impacte dels esdeveniments de Roma, per tipus, SNR i durada.

anòmals capturats a Milà i Roma, respectivament. Aquestes figures s'han construït utilitzat els conjunts de dades registrats amb la WASN, de 151 h i 153 h i 20 min, respectivament. S'hi poden observar tots els ANEs classificats en l'eix $x$ per la durada i en l'eix $y$ per l'SNR; la tipologia de l'esdeveniment s'hi troba representada per una llegenda de colors i l'impacte és representat en diferents mides en el diagrama de punts. Per motius il·lustratius, s'ha hagut de representar l'escala de la durada en l'eix logarítmic.

A la figura 5.1, pertanyent a Roma, podem observar només 6 esdeveniments que superen el llindar dels 2 dB d'impacte, pertanyents a tres sirenes, dos trens i un so d'interferència. Excepte en la interferència, tots mesuren una SNR per damunt de 4 dB i com a mínim 6 segons de durada; la interferència, en canvi, té una durada de 28 segons i una SNR negativa, que significa que els sons d'abans i després són més potents que l'esdeveniment en si.

En canvi, a Milà, a la figura 5.2 observem una gran quantitat d'esdeveniments que superen el llindar de 2 dB i fins i tot el de 3 dB. El so més habitual que supera el llindar dels 3 dB són les sirenes, una vintena ho fan amb una SNR positiva i amb durades superiors als 10 s. D'altres sorolls també superen aquest llindar: dues botzines de 2 s de durada però amb SNR al voltant

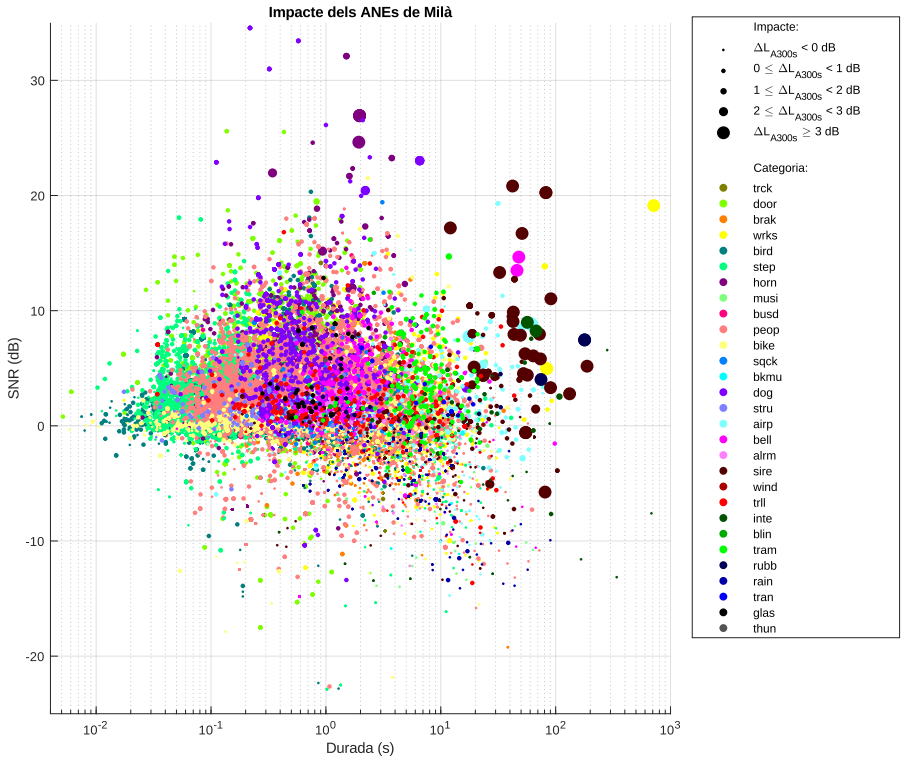



Figura 5.2: Impacte dels esdeveniments de Milà, per tipus, SNR i durada.

de 25 dB, dos sorolls provinents de campanars de durada prop d'un minut i SNR de 14 dB, dues interferències d'1 minut amb una SNR de 9 dB, dos serveis d'escombraries d'1 i 3 minuts amb SNR de 4 i 8 dB, dos sorolls d'obres de més d'un minut i SNRs de 5 i vora els 20 dB i un avió de 8 dB i 20 segons de durada. Entre els 2 i 3 dB hi tenim dos sorolls de gossos bordant de més de 20 dB d'SNR, alguna sirena de prop els 20 s amb SNR d'entre 5 i 10 dB, una botzina similar a les ja descrites i algun soroll d'alarma sense durada ni SNR significatives.

En ambdós casos, es pot concloure que efectivament hi ha un patró clar que estableix que perquè un esdeveniment tingui un impacte rellevant, cal que tingui una durada i SNR significatives. I també es donen molts casos en què l'esdeveniment només destaca en un dels dos paràmetres, per exemple les botzines en l'SNR i no en la durada, i la majoria de sirenes en la durada, però no per la seva SNR.

També s'identifiquen conjunts d'esdeveniments clars, visibles gràcies als colors, que caracteritzen els tipus d'ANE que s'han registrat a Milà i a Roma, i que permeten reconèixer l'ambient sonor de cada àrea. Si ens fixem en els esdeveniments més sorollosos, a l'entorn urbà de Milà n'hi ha més varietat, mentre que Roma només destaquen algunes sirenes i algun soroll de tren que

passa prop de la zona monitorada.

## 5.3 Càlcul de l'impacte agregat dels esdeveniments acústics

L'Impacte Agregat (IA) de diversos esdeveniments acústics es defineix com la contribució acumulada de tots els impactes individuals de cada esdeveniment comprès en un interval de temps donat. Aquest impacte, equivaldria a la suma de tots els impactes que ocorren en un període determinat, tal com s'indica a l'equació (5.3).

$$IA = \sum_{n=1}^{N_E} \Delta L_{eq}(n), \tag{5.3}$$

on $n = \{1, 2, ..., N_E\}$ és l'índex que identifica tots els esdeveniments que ocorren dins el període i $\Delta L_{eq}(n)$ és l'impacte individual de l'esdeveniment $n$.

### 5.3.1 L'anàlisi de l'impacte agregat dels esdeveniments acústics en el cas DYNAMAP

Com en l'anàlisi de l'impacte individual, es calculen els impactes després d'aplicar el filtre A [4] i en intervals de 5 minuts, mantenint el criteri d'anàlisi dels impactes individuals.

Per poder analitzar ordenadament els impactes agregats de cada període de 5 minuts i de cada sensor de la xarxa de sensors, es defineix una matriu d'impactes, que segueix l'equació (5.4). A l'equació, $IA^i_T(t)$ representa el resultat del càlcul de l'impacte agregat d'un sensor $i$ de la xarxa per un període $t$ d'integració on els períodes tenen longitud $T$, que en el cas DYNAMAP són 300 s (5 min) de durada.

$$IA^i_T(t) = \sum_{n=1}^{N^i_E(t)} \Delta L^i_{Aeq,T}(n, t), \tag{5.4}$$

on $i = \{1, 2, ..., N_S\}$ ($N_S$ és el nombre de sensors que s'estan analitzant); $t = \{1, 2, ..., N^i_T\}$ ($N^i_T$ és el nombre total de períodes d'integració de longitud $T$ del sensor $i$; i $n = \{1, 2, ..., N^i_E(t)\}$ ($N^i_E(t)$ és el nombre d'esdeveniments present en un període $t$ d'un sensor $i$). Llavors, $\Delta L^i_{Aeq,T}(n, t)$ és l'impacte de l'esdeveniment $n$, del període d'integració $t$ del sensor $i$, mesurat com la contribució de l'esdeveniment en l'$L_{Aeq}$ de temps d'integració $T$.

A la figura 5.3, s'hi poden observar dos exemples de Milà recopilats durant la primera campanya de gravació (9 h), on es mostra l'impacte que poden acabar provocant diversos esdeveniments que es trobin en el mateix interval de mesura de 5 minuts, i.e., $T = 300s$. En el primer cas, dues botzines, un tramvia, i algun camió provoca que l'impacte agregat dels esdeveniments sigui d'1,5 dB, que és equivalent a la suma dels impactes individuals mesurats en aquell interval de temps. En el segon cas, una sirena de 3,9 dB és suficient

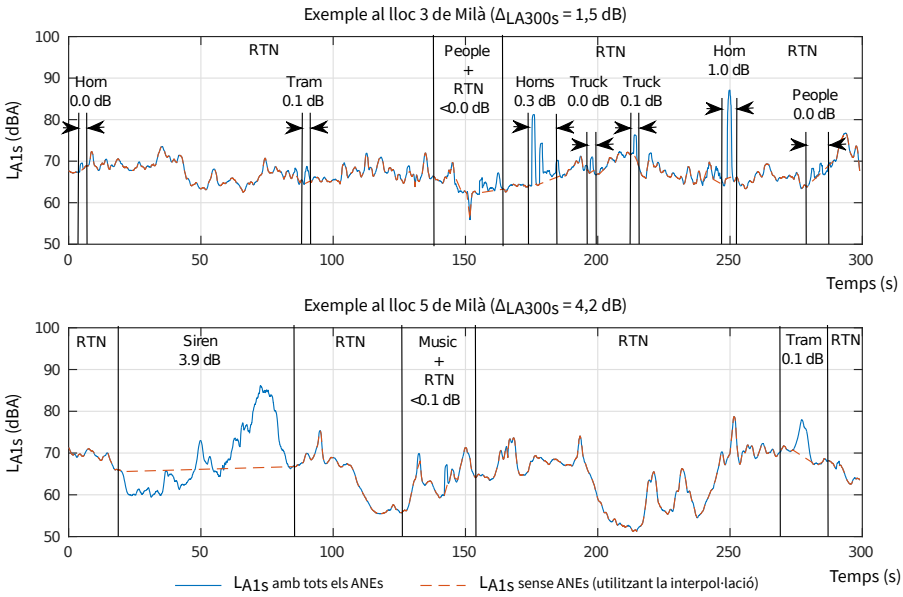



Figura 5.3: Dos intervals d'exemple del càlcul de l'impacte agregat que corresponen als llocs 3 i 5 de Milà.

per a superar el llindar de 2 dB, però s'hi afegeixen dos esdeveniments petits que provoquen que l'impacte final sigui de 4,2 dB. A la figura també s'hi pot observar el detall de la mesura de l'impacte per cada esdeveniment, on es veu la línia blava que marca el nivell equivalent de l'esdeveniment a intervals d'1 s, i la vermella discontínua que interpola l'esdeveniment per eliminar-lo de la mesura i així calcular el nivell del tram seguint el descrit a l'apartat 5.2.

L'impacte individual calcula la rellevància d'un esdeveniment puntual en la corba de l'$L_{eq}$, permetent que s'estudiïn els impactes en funció de les tipologies i els escenaris. Però els escenaris urbans són complexes i els sons no són sempre aïllats. Per exemple, les sirenes (siguin d'ambulància, de policia o de serveis d'emergències) poden anar lligades a cops de clàxon o frenades en una situació determinada, en un espai de temps molt pròxim. També hi ha casos de sons repetitius o discontinus, que no tenen rellevància individualment, però en tenen si es té en compte el període de temps en què ocorren, per exemple: els sons d'ocells piulant o el so de maquinària provinent de les obres. Per representar correctament l'impacte dels sons en el nivell equivalent, cal tenir en compte aquestes aglomeracions de sons provinents de la mateixa causa i les repeticions d'un mateix so, com a part del mateix impacte, creant el concepte d'impacte agregat.

## 5.4 La percepció de l'impacte dels esdeveniments acústics

La contribució dels esdeveniments en el nivell sonor equivalent no avalua la percepció dels ciutadans en un entorn sonor concret. Per aquest motiu, en aquesta tesi se segueix indagant en la possibilitat de caracteritzar els esdeveniments sonors d'una manera més fidel a la percepció de les persones. Per aconseguir-ho, és necessari de recollir dades relacionades amb la molèstia d'esdeveniments determinats.

### 5.4.1 L'avaluació perceptiva dels esdeveniments de DYNAMAP

Utilitzant dades recollides en l'entorn urbà de DYNAMAP durant la primera campanya de gravació, s'han confeccionat un conjunt de tests que permeten esbrinar la percepció que tenen els ciutadans d'una sèrie d'esdeveniments acústics [6]. L'objectiu és determinar la molèstia dels sons més presents en el paisatge urbà i relacionar aquesta molèstia amb dos indicadors psicoacústics de Zwicker: la potència acústica (volum, o *loudness*) i la nitidesa [7] (o *sharpness*). Els tres tests s'han efectuat íntegrament per una mostra de 100 participants, vinculats entre si sense comprometre dades personals, utilitzant una plataforma pròpia en un servidor de La Salle-URL [8]. Els tres tests, que s'han efectuat per recollir dades són els següents:

**Comparació per parells.** Els participants han d'escoltar dos àudios pertanyents al mateix tipus d'esdeveniment acústic i escollir quin dels dos li genera més molèstia. Els dos sons s'han recopilat utilitzant el mateix node de la xarxa de sensors i tenen una pressió acústica similar, però presenten nivells de nitidesa diferents [9].

**Test de sensacions.** Els participants han d'escoltar una sèrie de sons típics d'un entorn urbà i avaluar com de bé cinc adjectius el descriuen en una escala de Likert [10] de 5 punts entre gens d'acord i molt d'acord. Els adjectius que s'han d'avaluar són: fort, estrident, pertorbador, nítid i agradable.

**Test MUSHRA.** De l'anglès: *MUltiple Stimuli with Hidden Reference and Anchor)*, aquest test consisteix a avaluar els sons en conjunts. Per cada conjunt, el participant haurà d'escoltar el so i presentar-lo en una escala de molèstia entre «absolutament gens molest» i «extremadament molest», tal com s'indica en la figura 5.4. Per més detalls sobre el test de MUSHRA, el lector pot consultar [11]. Cada conjunt dels 5 avaluats consta de diversos àudios (entre 4 i 7) pertanyents a esdeveniments acústics amb una pressió acústica i nitidesa similars, recollits en el mateix sensor de la xarxa, fent un total de 27 àudios.

A [9] s'hi recullen els resultats preliminars dels tests efectuats per una mostra de 100 participants. La comparació per parells mostra una gran dependència entre la molèstia percebuda i el tipus d'esdeveniment al qual pertany el fragment d'àudio. El test de sensacions ajuda a caracteritzar la percepció dels tipus d'esdeveniments més habituals, i se'n poden treure alguns resultats clars: el so de persones parlant (`peop`) és el més agradable, el més pertorbador és el so generat per camions (`trck`) i el més estrident són les sirenes (`sire`). Finalment, el test de MUSHRA demostra que la sirena (`sire`),

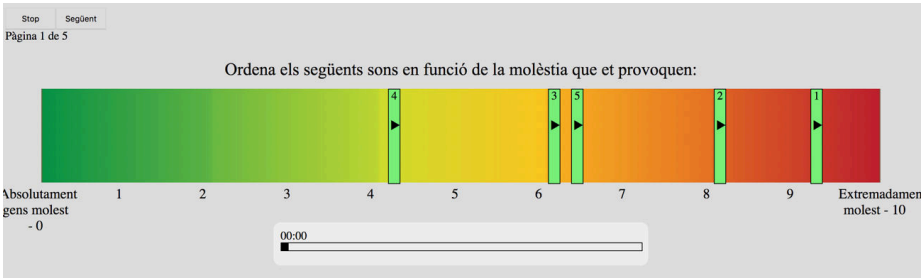



Figura 5.4: Exemple d'un dels conjunts avaluats en un test de MUSHRA, on cada barra verda conté un so que es pot reproduir clicant a la mateixa barra. El sistema col·loca els sons arbitràriament en la barra i no permet d'avançar al següent test fins que no s'han reproduït tots els sons i no s'han mogut totes les barres.

si és present al conjunt de comparació, sempre és l'esdeveniment més molest, seguit dels sons de construcció (`wrks`) i botzines (`horn`). Els sons menys molests corresponen habitualment a ocells (`bird`) o persones (`peop`). Però més enllà d'aquests resultats preliminars, el test de MUSHRA conté molts tipus de sons classificats en funció de la molèstia, en una escala de 0 a 100, que permetria de crear un model de molèstia en funció de la tipologia del so.

### 5.4.2 Construcció d'un model de molèstia

A partir de les dades provinents del test de MUSHRA, s'ha generat una primera versió de model multinivell mitjançant regressions lineals. Aquest model s'utilitza típicament per estudis de mesures repetides [12, 13] i per models de predicció aplicats [14, 15] i permet la incorporació de dades de diferents nivells. Inicialment, les dades s'han agrupat en dos grups no dependents per formar un model de dos nivells, que són els mateixos resultats agrupats per etiqueta i agrupats per usuari. També s'ha afegit una variable d'usuari en el segon nivell, per corregir el nivell de molèstia mitjà de cada participant, ja que el test de MUSHRA permet molta llibertat per indicar la molèstia, com hem vist a la figura 5.4. Els predictors que s'incorporen en el model són les mesures psicoacústiques definides per Zwicker [7], calculades a cada so: *loudness* (pressió acústica), *sharpness* (nitidesa), *roughness* (aspror), *tonality* (tonalitat) i *impulsiveness* (impulsivitat).

Els primers resultats analitzats indiquen que ni l'àrea de residència del participant ni el gènere són estadísticament significatius en la predicció del model. I en aplicar el model, es revela que la nitidesa (*sharpness*) és el predictor principal de la molèstia, amb una significança estadística molt elevada: com més nítid el so, més molest acaba essent. Aquest model, que té un coeficient de determinació de $R^2 = 0,64$, és una primera prova de concepte de com es pot predir la molèstia a partir de les mesures psicoacústiques [16].

## 5.5 Publicacions destacades

El primer càlcul d'impacte individual es presenta a l'*International Journal of Environmental Research and Public Health* (IJERPH) el desembre de 2017. L'article duu per títol *«On the impact of anomalous noise events on road traffic noise mapping in urban and suburban environments»* (en català: Sobre l'impacte dels esdeveniments de soroll anòmals en mapes de soroll de trànsit d'entorns urbans i suburbans). A l'article, s'hi exposa l'estat de l'art en la influència del soroll a la salut pública i l'estat actual dels projectes de monitoratge de soroll a diferents països. A continuació, presenta la metodologia emprada, que consisteix a comparar l'impacte d'un esdeveniment amb la seva durada i el seu SNR. Així, es presenta una hipòtesi, que coincideix amb la segona de les plantejades en aquesta tesi:

> P2: *Provocarien, els esdeveniments amb durada i prominència significatives, un impacte considerable en el càlcul del nivell equivalent? Serà factible classificar-los en regions d'impacte en funció de la prominència i la durada?*

Per comprovar la hipòtesi, es calculen els paràmetres de tots els esdeveniments recopilats a Roma i a Milà en la primera campanya de gravació. Per fer l'estudi, s'utilitzen tres períodes d'integració diferents, trobant l'impacte en 1, 5 i 15 minuts. Es demostra que tant l'SNR com la durada són clau per determinar l'impacte d'un esdeveniment i s'observa com els esdeveniments que presenten més impacte tenen o bé un SNR o bé una durada significatives.

En un segon treball, publicat a la revista *Sensors* el gener del 2020, es presenta la mesura de l'impacte agregat i s'avalua també en els dos entorns de DYNAMAP: l'urbà i el suburbà. L'article duu per títol *«Aggregate impact of anomalous noise events on the WASN-based computation of road traffic noise levels in urban and suburban environments»* (en català: L'impacte agregat dels esdeveniments de soroll anòmals en els nivells de soroll de trànsit capturats mitjançant una WASN en els entorns urbà i suburbà). En aquest treball, el candidat duu a terme l'anàlisi de les dades i participa en l'escriptura dels resultats i la discussió.

En l'article es presenta una mesura de l'impacte agregat que permet sistematitzar l'anàlisi dels impactes per tots els sensors d'una xarxa de sensors. Mitjançant l'estudi i gràcies al fet de disposar de més de 300 h de dades, es permet observar com esdeveniments amb un impacte baix o mitjà són capaços d'esbiaixar una mesura de nivell equivalent de manera agregada tant com un esdeveniment d'impacte alt, cosa que no s'havia pogut observar emprant la base de dades inicial de 9 h recopilada manualment. Això ressalta la importància de detectar els esdeveniments no relacionats amb el soroll de trànsit, encara que individualment presentin impactes reduïts, pel seu correcte monitoratge.

## Referències


[1] Johnson, D. H. "Signal-to-noise ratio". *Scholarpedia*, vol. 1, núm. 12 (2006), pàg. 2088.

[2] Orga, F., Alías, F. i Alsina-Pagès, R. M. "On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments". *International Journal of Environmental Research and Public Health*, vol. 15, núm. 1 (2017), pàg. 13.

[3] Sevillano, X., Socoró, J. C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., Coppi, P., Nencini, L., Cerniglia, A., Bisceglie, A., Benocci, R. i Zambon, G. "DYNAMAP – Development of low cost sensors networks for real time noise mapping". *Noise Mapping*, vol. 3 (1 Maig de 2016), pàg. 172 - 189.

[4] Pierre, R. L. S. i Maguire, D. J. "The impact of A-weighting sound pressure level measurements during the evaluation of noise exposure". *NOISE-CON 2004*. Baltimore, Maryland, des. de 2004, pàg. 1 - 8.

[5] WG-AEN (European Commission Working Group Assessment of Exposure to Noise). *Good Practice Guide for Strategic Noise Mapping and the Production of Associated Data on Noise Exposure Ver.2, 13 Agost*. Inf. tèc. 2007.

[6] Alsina-Pagès, R. M., Freixes, M., Orga, F., Foraster, M. i Labairu-Trenchs, A. "Perceptual evaluation of the citizen's acoustic environment from classic noise monitoring". *Cities & Health*, vol. 5, núm. 1-2 (2021), pàg. 145 - 149.

[7] Zwicker, E. i Terhardt, E. *Facts and Models in Hearing: Proceedings of the Symposium on Psychophysical Models and Physiological Facts in Hearing, Held at Tutzing, Oberbayern, Federal Republic of Germany, Abril 22–26, 1974*. Vol. 8. Springer Science & Business Media, 2013.

[8] Alsina-Pagès, R. M., Orga, F., Freixes, M. i Foraster, M. "Citizens' perceptual evaluation of noise events in an urban environment". *13$^{th}$ ICBEN Congress on Noise as a Public Health Problem*. Stockholm – Sweden, 14 - 17 Juny de 2021.

[9] Labairu-Trenchs, A., Alsina-Pagès, R. M., Orga, F. i Foraster, M. "Noise Annoyance in Urban Life: The Citizen as a Key Point of the Directives". *Multidisciplinary Digital Publishing Institute Proceedings*. Vol. 6. 1. 2018, pàg. 1.

[10] Likert, R. "A technique for the measurement of attitudes." *Archives of psychology* (1932).

[11] ITU-R. "BS.1534-3: Method for the subjective assessment of intermediate quality level of coding systems". *International Telecommunication Union* (2015).

[12] Quené, H. i van den Bergh, H. "On multi-level modeling of data from repeated measures designs: a tutorial". *Speech Communication*, vol. 43, núm. 1 (juny de 2004), pàg. 103 - 121.

[13] Volpert-Esmond, H. I., Page-Gould, E. i Bartholow, B. D. "Using multilevel models for the analysis of event-related potentials". *International Journal of Psychophysiology*, vol. 162 (abr. de 2021), pàg. 145 - 156.

[14] Gelman, A. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do". *Technometrics*, vol. 48, núm. 3 (2006), pàg. 432 - 435.

[15] Frees, E. W. i Kim, J.-S. "Multilevel Model Prediction". *Psychometrika*, vol. 71, núm. 1 (2006), pàg. 79 - 104.

[16] Orga, F., Mitchell, A., Freixes, M., Aletta, F., Alsina-Pagès, R. M. i Foraster, M. "Multilevel Annoyance Modelling of Short Environmental Sound Recordings". *Sustainability*, vol. 13, núm. 11 (2021), pàg. 5779.

# On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments

**Ferran Orga, Francesc Alías, Rosa Ma. Alsina-Pagès**

Publicat: 23 de desembre del 2017.

IV

International Journal of
*Environmental Research and Public Health*

MDPI

*Article*

# On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments

**Ferran Orga *, Francesc Alías and Rosa Ma Alsina-Pagès**

GTM-Grup de recerca en Tecnologies Mèdia, La Salle-Universitat Ramon Llull, C/Quatre Camins, 30, 08022 Barcelona, Spain; falias@salleurl.edu (F.A.); ralsina@salleurl.edu (R.M.A.-P.)
* Correspondence: forga@salleurl.edu; Tel.: +34-93-290-2427

Received: 3 November 2017; Accepted: 21 December 2017; Published: 23 December 2017

**Abstract:** Noise pollution is a critical factor affecting public health, the relationship between road traffic noise (RTN) and several diseases in urban areas being especially disturbing. The Environmental Noise Directive 2002/49/EC and the CNOSSOS-EU framework are the main instruments of the European Union to identify and combat noise pollution, requiring Member States to compose and publish noise maps and noise management action plans every five years. Nowadays, the noise maps are starting to be tailored by means of Wireless Acoustic Sensor Networks (WASN). In order to exclusively monitor the impact of RTN on the well-being of citizens through WASN-based approaches, those noise sources unrelated to RTN denoted as Anomalous Noise Events (ANEs) should be removed from the noise map generation. This paper introduces an analysis methodology considering both Signal-to-Noise Ratio (SNR) and duration of ANEs to evaluate their impact on the A-weighted equivalent RTN level calculation for different integration times. The experiments conducted on 9 h of real-life data from the WASN-based DYNAMAP project show that both individual high-impact events and aggregated medium-impact events bias significantly the equivalent noise levels of the RTN map, making any derived study about public health impact inaccurate.

**Keywords:** anomalous noise event; wireless acoustic sensor network; noise pollution; $L_{Aeq}$; acoustic impact; road traffic noise; noise map; health effects

## 1. Introduction

Environmental noise pollution is increasing year after year because of population growth and the consequent expansion of transportation systems, including highways, railways and airways [1]. It is not merely an annoyance, since several studies warn about its adverse effects on people pointing to health-related problems [2]. Most of the conducted studies address the effects of long-term exposure to environmental noise, mainly focused on concentration issues, sleep disturbance and stress [3], emphasizing the negative effects on children [4].

Research also specifically analyzes the association between road traffic noise and several diseases in suburban and urban areas. Öhrström states that the influence of road-traffic noise implies an increase in tiredness and disturbs the sleep [5]. In addition, Botteldooren et al. analyze the influence of road traffic on noise annoyance in neighborhoods [6], while Jakovljevic et al. conclude that the most significant noise source in urban areas is road traffic noise, according to the interviewed residents [7].

In order to ensure that these studies only evaluate the impact of road traffic noise (RTN) on the well-being of citizens, those noise sources unrelated to RTN should be removed from the study as they could alter the conclusions significantly. This is usually assured by the experts conducting the acoustic measurements.

The European Union has reacted to this alarming increase of environmental noise pollution, especially in large agglomerations, approving the Environmental Noise Directive 2002/49/EC (END) [8]. In accordance with the END, the CNOSSOS-EU methodological framework pretends to improve the consistency and comparability of noise assessment results across the EU Member States [9]. The main pillars of the END are the following: (i) determining the noise exposure, (ii) making the updated information related to noise available to citizens, and (iii) preventing and reducing the environmental noise where necessary. Moreover, the END requires the European Member States to publish noise maps and action plans for agglomerations with more than 100,000 inhabitants and major roads, railways and airports every five years, and introduces the need to discern between the different sound sources [8].

Recent technological advances have posed a significant change of paradigm to address the END regulatory requirements, mainly based on the design and development of Wireless Acoustic Sensor Networks (WASNs). Authors have suggested their WASN-based solutions involving cities as Barcelona (Spain) [10] and Pisa (Italy) [11]. However, most of these projects do not identify the noise typology of the area of interest, thus addressing noise monitoring in a holistic way. As an exception, in [12], sound recognition is applied together with a subjective survey in order to cross both acoustic and subjective perception of noise sources.

Nevertheless, since the identification of sound sources is conducted after computing the noise levels, this information cannot be used to modify the noise map calculation accordingly. To this aim, the DYNAMAP project [13] pretends to deploy a WASN to tailor dynamic noise maps of RTN in suburban and urban areas [13]. In an attempt to monitor only the RTN, those events unrelated to RTN denoted as Anomalous Noise Events (ANEs) (e.g., birds, people, sirens, etc.) have to be detected and removed automatically before computing the A-weighted equivalent noise levels ($L_{Aeq}$) of RTN so as to obtain a reliable picture of citizens' exposure to this pollutant [14]. In addition, the contribution of these ANEs to the $L_{Aeq}$ calculation should be studied in order to obtain reliable RTN maps, from which accurate studies of the impact of the RTN on health can be derived from WASN-based approaches.

The goal of this paper is to evaluate the impact of the ANEs on the computation of RTN $L_{Aeq}$ in real-life suburban and urban scenarios. The datasets contain 17 recording locations in total, defined in [15] with the purpose of capturing real-life ANE diversity and sampling correctly both pilot areas. To that effect, we introduce an analysis methodology that takes into account the duration and Signal-to-Noise Ratio (SNR) of the ANEs with respect to the surrounding RTN levels. A priori, it seems reasonable that the higher the duration and the SNR of an individual ANE, the higher the impact on the final $L_{Aeq}$, but the importance of the impact of an ANE with short duration and high SNR, or ANE with long duration and low SNR is not so foreseeable. For this reason, a joint study of duration and SNR is also conducted considering the impact of individual ANEs on the $L_{Aeq}$ computation for different integration times, besides evaluating their aggregated impact for a given integration span. integration times, and validating the impact for several aggregated ANEs in a one-time integration span.

## 2. Related Work

In this section, first, we review several representative investigations focused on the study of the impact noise to health, together with those projects focused on measuring the quality of life of citizens due to road traffic noise. Second, we describe different approaches to address automatic noise monitoring by means of low-cost WASN and the use of their measurements to tailor noise maps.

### *2.1. Influence of Noise on Public Health*

Some studies warn about the noise exposure of citizens in certain cities, with some critical examples as the case of Tainan (Taiwan). According to [16], over 90% of Tainan City inhabitants are exposed to unacceptable noise levels—62 dB(A), as defined by US Department of Housing and Urban Development. In addition, 93.3% of the inhabitants of Curitiba (Brazil) are exposed to a sound level over 65 dB(A), as stated by [17]. According to [18], 90% of the population of Cáceres (Spain) is also

affected by a sound level of 65 dB(A) during working time. The reader may find other examples related to sound exposure in the aforementioned countries seeing the cited references. Other cases include India [19] and China [20].

The World Health Organization (WHO) quantifies the healthy life years lost in Europe in terms of "Disability-Adjusted Life-Years" (DALYs) [21], concluding that the diseases related to noise exposure produce a loss of a million healthy life years in western Europe every year. Furthermore, research has found that noise exposition does not only affect health, but also social and economic aspects [1]. In addition, a permanent hearing loss and possible tinnitus is associated with noise exposure, as stated in [22], especially caused by a prolonged listening of loud music through PLDs (Personal Listening Devices). Another relation between environmental noise and adverse birth outcomes can be found in [23], concluding that a low quality association exists between aircraft noise and preterm birth. Furthermore, other studies connect the noise pollution to mental illnesses [24], diabetes [25] and other heart diseases [26].

Moreover, some authors remark on the importance of road traffic noise to the living quality of the neighborhoods, conducting several subjective studies: Botteldooren et al. compare a set of indicators related to sound exposure in [6], Van Renterghem et al. conclude the importance of having a quiet facade in the dwellings [27]. In addition, RTN increased the tiredness and disturbs the sleep, says Öhrström in [5], who even states that access to quiet parts of the residence contributes to physiological and psychological well-being. Finally, a study by Jakovljevic et al. determines the principal factors for high noise annoyance in adults using several indicators, as stress scores, age and other location variables (e.g., orientation of the rooms towards the street). After conducting perception surveys, most of the interviewed residents found road traffic the most significant noise source. A review of the transport noise interventions and their impacts on health is conducted in [28], studying the European region in particular.

### 2.2. Noise Monitoring

In order to satisfy the increasing demand of automatically monitoring the noise levels in urban areas, several WASN-based projects are being developed in different countries. The DREAMSys project monitors several UK areas with a distributed sensor network [29]. Other projects pretend to monitor also the urban noise in real-time, as the UrbanSense project in Canada [30], which also pretends to monitor other pollutants as carbon dioxide ($CO_2$) and carbon monoxide (CO). The Senseable project in Pisa is based on the concept of real-time city and smart city to measure the sound level in several points [11]. In addition, a noise monitoring network is also being deployed in Barcelona in order to manage the resources efficiently and to reduce the impact of urban infrastructure on the environment [10] and, recently, in Monza, by a LIFE project that implements also a low-cost system [31]. Finally, some projects focus in other areas, such as, for example, highways. In [32], five points along the National Highway of Burdwan have been monitored with an audiometer in order to register the acoustic equivalent level ($L_{eq}$) and carrying a statistical analysis. In addition, the location strategy to evaluate multiple noise sources has been studied in [33,34].

Other noise monitoring projects take into account other data further than the noise equivalent level. In the Smart Sound Monitoring project, De Coensel et al. conducted a study in [12] that crosses acoustic information with subjective perception surveys, allowing for considering the typology of the acoustic information. Furthermore, a sound recognition system is applied in order to give information about the detected sounds and establish a relation between the identified events and the perception surveys. However, the system identifies events only to give information but not to remove these sounds from the noise map.

The aforementioned projects pretend to monitor the noise in determined areas using a low-cost sensors network; however, as far as we know, none of them intend to remove the anomalous events biasing the traffic noise map measurement. With this purpose, the DYNAMAP project pretends to

monitor road traffic noise in suburban and urban areas reliably, after removing the anomalous noise events from the road traffic noise map computation [13].

## 3. Analysis Methodology

This section describes the methodology applied to analyze the impact of the ANEs on the equivalent noise level computation of road traffic noise, which details the ANE parametrization and the calculation of their acoustic impact for a certain integration time.

Figure 1 depicts the proposed analysis methodology used to determine the impact of a particular ANE in the equivalent noise level computation. The ANE is parametrized by means of their SNR and duration. The analysis takes the raw acoustic data corresponding to an integration time, and computes both the $L_{Aeq}$ considering the ANE and after removing the event. For the latter, the ANE is replaced with a linear interpolation connecting the previous RTN sample with the following. It is worth mentioning that the A-weighted filter is applied before the $L_{eq}$ calculation in an effort to account for the relative loudness perceived by the human ear. After that, the particular impact of an ANE is calculated by deducting the $L_{Aeq}$ of the interpolated ANE from the $L_{Aeq}$ calculation of the entire piece of audio, for a given integration time. The last stage of the analysis methodology associates the impact of the ANE under study with the corresponding SNR and duration. The high-impact ANEs are defined as the ones surpassing the predefined limit.

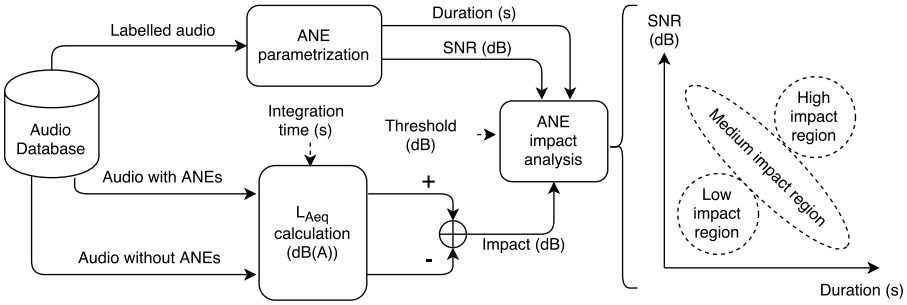



**Figure 1.** Block diagram of the methodology used to analyze the impact contribution of the Anomalous Noise Events (ANEs) to the $L_{Aeq}$ considering the Signal-to-Noise Ratio (SNR) and the duration of the anomalous event. The analysis includes three regions: (**a**) high-impact, due to ANEs with high SNR and long duration; (**b**) low-impact, due to ANEs with low SNR and short duration; and (**c**) medium impact, which integrates those ANEs falling in between the previous two regions.

The impact analysis consists in validating the hypothesis, which states that long ANEs with high-SNR would entail high impact on the $L_{Aeq}$ computation, while short and low-SNR ANEs would affect less in the impact for a given integration time. However, a wide variety of events may be captured from real-life data, including ANEs presenting different possible combinations of duration and SNR and making it possible that long events with low SNR could have a similar impact than those short events with high SNR. To this aim, the study has to be completed by evaluating the aggregated impact of the ANEs for a specific integration time, considering their SNR and their duration. As a consequence, the aggregated impact of several ANEs is expected to be higher than any of their individual impacts by itself, even coming from the low or medium impact region of Figure 1, which could entail the removal from the $L_{Aeq}$ computation of individual ANEs with individual moderate SNR and/or duration.

In the following sections, we describe the main elements of the introduced analysis methodology.

### 3.1. ANE Parametrization

A set of parameters have been defined in order to analyze the impact of the ANEs in detail. Both duration and SNR are considered to analyze the impact of ANEs in the $L_{Aeq}$ computation. The SNR considers the energy of the event in relation to the surroundings; however, it does not take into consideration the duration of the event, which could also impact the $L_{Aeq}$. By considering both parameters, we derive the hypothesis depicted in Figure 1.

- **SNR Calculation.**

  First, the SNR of an ANE is defined as a classical Signal-to-Noise Ratio, but considering that the RTN noise is not stationary. It is evaluated in order to obtain the impact of a particular ANE in relation to the surrounding RTN signal level. The acoustic power of the ANE with respect to the surrounding RTN is calculated as expressed by Equation (1):

$$P_x = \sum_{t=1}^{N} \left( \frac{x(t)^2}{N} \right), \tag{1}$$

  where $N$ is the number of samples and $x(t)$ is the recorded audio during a certain integration period.

  After the powers of the ANE and the RTN are evaluated, the SNR is calculated as follows:

$$SNR = 10\, log_{10} \left( \frac{P_{ANE}}{P_{RTN}} \right), \tag{2}$$

  where $P_{ANE}$ belongs to the anomalous event and $P_{RTN}$ belongs to the power of $RTN_1$ and $RTN_2$, the former is the previous RTN to the ANE and the latter is the next portion of RTN sound.

  Due to the relative nature of the measure, it is worth mentioning that the SNR of the event could result in a negative value if the energy of the ANE is less than the previous and following sound. This is normally caused by events that can only be heard in moments when the road traffic noise decreases. In addition, to compute the SNR, previous and following samples to the event are considered; thus, this calculation is only meant for analysis and it is not applicable for real-time decision purposes due to the need of future samples of the signal.

- **Duration of the ANE.**

  The duration of the events is an important factor, since not only the salience of the event is important in what concerns its impact on the $L_{Aeq}$ computation, but also the time the anomalous situation lasts. The duration of the ANE is obtained by means of the difference between the time stamp of the first and the last sample of the labelled dataset. As published in the previous study of the dataset, the duration of the ANEs may vary depending on the typology of the sound and the circumstances of the location [15]. One of the shorter samples observed corresponds to 40 ms length in Rome, and the longer 53 s in Milan, with 0.6 s being the mean of the duration of the ANE.

Figure 2 shows a piece of audio where the SNR and the duration of an ANE has been measured, with the tagged RTN and ANE used to evaluate the SNR, for illustrative purposes.

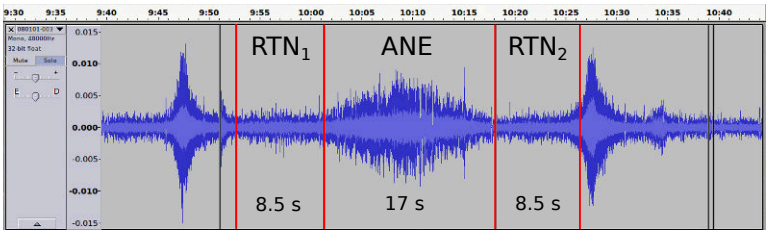



**Figure 2.** An example of the SNR and duration measurement of a raw acoustic signal corresponding to a 17 s long train ANE. The ANE presents a SNR of 6.6 dB, resulting in an impact of 1.1 dB when the $L_{A5min}$ is calculated, and uses a 5 min time interval evaluation. The left and right RTN regions used to calculate the SNR are marked as $RTN_1$ and $RTN_2$, respectively.

### 3.2. Impact Calculation

The impact calculation takes into account the integration period of the recording in order to obtain the repercussions of the analyzed event or set of events on the $L_{Aeq}$ computation. It is computed as the difference between the $L_{Aeq}$ considering the individual ANE and the $L_{Aeq}$ replacing the event with the linear interpolation from the last and the following RTN sample of the original raw data. For this reason, in this work, it is referred to as $\Delta L_{Aeq}$.

As a guideline for the evaluation of the impact over the $L_{Aeq}$, and, according to the European Commission Working Group, the maximum tolerated change is 2 dB for a given segment [35]. Therefore, a 2 dB threshold is considered in this work to discern high-impact from low-impact ANEs. The analysis can be conducted for several integration times in order to observe the homogeneity of the ANE impact results; the most critical ANE appear as so in all integration times, despite having different impact values. Thus, the ANE evaluation sorted depending on the impact should maintain the same order for all integration periods, despite presenting different impact values for each of the different periods of integration.

## 4. Experiments

In this section, we present the conducted experiments to evaluate the ANE impact on the $L_{Aeq}$ computation. First, the main characteristics of the considered real-life audio dataset are described. Next, the individual analysis of all ANE impact is detailed for three different integration times: 1, 5 and 15 min. The minimum integration time has been set to 1 min since it is the shortest period necessary to include the longest ANE. The 5 min interval is chosen according to the DYNAMAP project specifications [13], and the 15 min span is also considered as the maximum time interval defined by the shortest period of a recording location. Finally, the aggregated impact of he ANEs on the $L_{Aeq}$ is also computed for the 5 min and the 15 min integration time, in order to evaluate a real-life situation, where several ANEs usually occur in a time integration period. Literature suggests that estimation of daily indicators can be extrapolated from short time spans when road traffic is the studied source. In [36], the study of the minimum measurement time interval is conducted for several noise sources including road traffic. However, in our project, continuous measurements are provided; thus, it is not necessary to use short time periods to obtain the hourly and daily indicators.

### 4.1. Real-Life Dataset

The dataset used for this experiments was recorded by means of two measuring campaigns conducted in the two pilot areas of the DYNAMAP project. Specifically, the recordings were conducted on the A90 motorway surrounding Rome as a suburban area, and the district 9 of Milan as an urban area. The used real-life dataset contains 4 h and 44 min of suburban audio and 4 h and 24 min of urban recordings. The former consists of five locations captured in the ring-road of Rome, where the shorter

recording is 50 min and the longest is 1 h and 35 min. In addition, the latter consists of environmental sounds from 12 different locations within the Milan urban area, where the minimum audio duration is 15 min and the maximum 47 min. For more details about the real-life dataset and the urban and suburban recording campaign, the reader is referred to [15].

In the suburban scenario, 261 anomalous events were recorded with a total duration of 543 s, while, in the urban scenario, 711 ANEs were labelled with a total duration of 1932 s. In the suburban area, the ANEs occupy the 3% of the total recorded time, while, in the urban scenario, the presence of ANEs increase to 12%. To classify the ANE diversity in subcategories, these labels were used as defined in [15]:

*airp*: airplanes, *musi*: music coming from cars, *bike*: bikes, *peop*: people talking, *bird*: birdsong, *sire*: sirens, *brak*: vehicle brakes, *stru*: structure sounds and vibrations, *busd*: bus or tram doors, *thun*: thunders, *chai*: chains (e.g., from bikes), *tram*: tramway pass-by, *dog*: dog barks, *tran*: train pass-by, *door*: house or vehicle doors, *trck*: hitch and towing system sounds of heavy-load vehicles, *horn*: vehicle horns, *wind*: noise of wind or leaf movements and *mega*: public address system.

Regarding the SNR and duration measures of the ANEs, there is a variety of options recorded. On the one hand, the duration of the ANEs differ from 40 ms (a door recorded in Rome) to 53 s (music sound heard in Milan), with 0.6 s being the average duration value of the corpus. On the other hand, the SNR of the ANEs differ from −9.5 dB (a 3-s conversation recorded in Milan) to 27.3 dB (an extremely short door sound of Milan), becoming the average SNR value of all ANEs, 1.2 dB. However, since the study aims to evaluate the additive impact of the ANEs, we remove from the analysis those ANEs presenting subtracting impact, which, in turn, allows the use of the logarithmic representation, more illustrative to observe the impact of each individual ANE. However, as different integration times may affect the impact calculation, the number of positive-impact ANEs vary from 258 to 347 in Milan and from 129 to 141 in Rome.

### 4.2. Individual ANE Impact on the $L_{Aeq}$ Computation

The first part of the analysis studies and evaluates the impact of each ANE of the dataset on the $L_{Aeq}$ for a particular integration time. To that effect, the methodology explained in Section 3 is applied for each ANE belonging to the urban and suburban corpus, Milan and Rome, respectively. Both SNR and duration are calculated and the impact is measured as the individual contribution of each ANE to the given integration period. Figures 3 and 4 show the impact of each ANE in a scatter plot in both urban and suburban scenarios, respectively. The $x$-axis represents the SNR of the ANE according to the previously-defined calculation (see Section 3.1). The color bar depicts the duration of the event in seconds, and the colormap has been modified to distinguish easily the shorter ANEs. Finally, the $y$-axis displays the contribution that this particular ANE has in the 1, 5 and 15 min computation of $L_{Aeq}$ in a logarithmic scale, which depicts the distribution of the impact by decades.

In both Figures 3 and 4, the ANEs present a similar pattern, where high-impact ANEs have at least a long duration or a high SNR. In addition, the reader may observe that the higher the integration time, the lower is the impact of the ANEs. This is because of the nature of the impact calculation, the individual contribution decreasing in longer integration times as the duration of the ANE remains the same, and the background noise has more presence. Thus, more impact ANEs are found in the 1-min integration time.

In order to quantify the contribution, these ANEs could have in the whole $L_{Aeq}$ computation, three impact regions have been defined, discerning between those ANEs affecting the total noise level and those which not. The low-impact region is defined between 0 and 0.5 dB, the medium-impact region from 0.5 to 2 dB and, the high-impact region comprises the ANEs that surpass the 2 dB threshold. Following the defined regions, both suburban and urban scenarios are described in detail below.

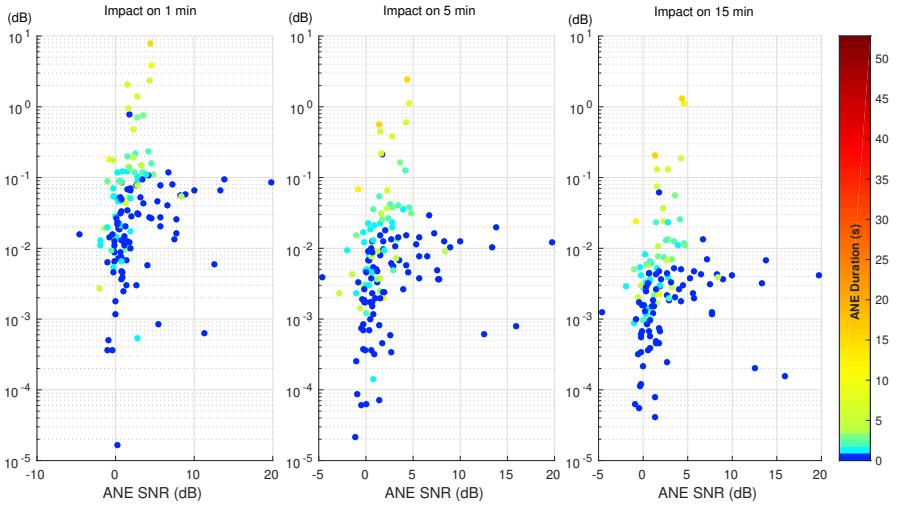



**Figure 3.** Impact of individual ANEs on the $L_{Aeq}$ value for the 1, 5 and 15 min integration times in the suburban scenario.

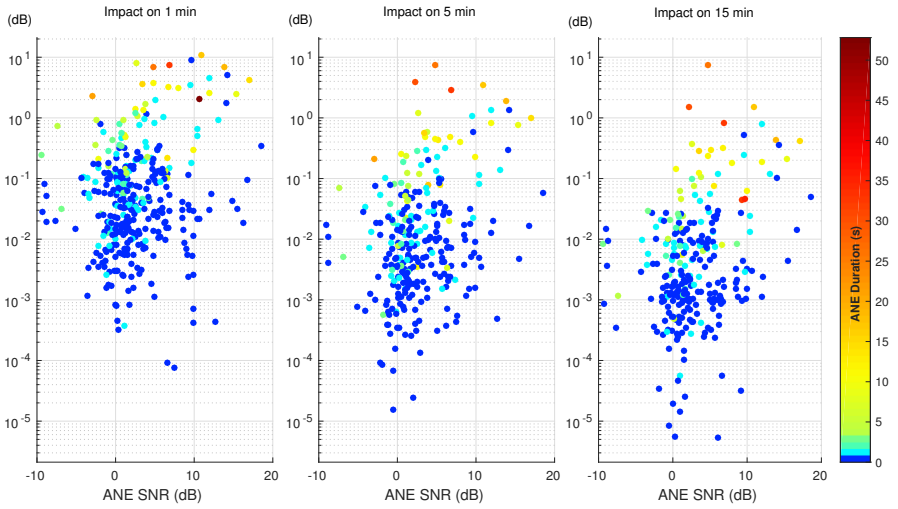



**Figure 4.** Impact of individual ANEs on the $L_{Aeq}$ value for the 1, 5 and 15 min integration times in the urban scenario.

In the suburban scenario, Figure 3, four ANEs surpass the 2 dB limit when analyzing the impact in 1 min, and only one surpasses the same limit in the 5-min integration time. In the 15-min integration time, no ANEs are above the 2 dB threshold. Another range can be defined from 0.5 dB and the 2 dB limit, embracing the medium-impact ANEs. In the 1-min integration time, five ANEs are comprised

in this region, while three belong to the same region in a 5-min span. Only two ANEs belong to the medium-range impact region in the 15-min integration time, both being above 1 dB.

Regarding the urban scenario, Figure 4, which has a 12% presence of ANEs in comparison to the 3% of the previous scenario, more ANEs are recorded to have a higher impact. In the 1-min integration time, 18 ANEs are above the 2 dB limit, the value decreases to four in the 5-min case and only one ANE is higher than the 2 dB threshold in the 15-min span. When analyzing the medium-impact range, 22 ANEs are comprised between 0.5 dB and 2 dB in the 1-min integration time, 11 ANEs in the 5-min case and only five in the 15-min integration period.

Furthermore, a representation of the ANEs belonging to each region is presented in Figure 5. The reader may observe the classification of the ANEs in three defined regions of impact, represented as a percentage in respect to the total of the correspondent scenario and integration time.

As seen in Figures 3 and 4, and summarized briefly in Figure 5, few ANEs may potentially affect the $L_{Aeq}$ computation of a certain integration time. A difference between the two scenarios can be observed, as the number of high and medium impact ANEs recorded in the urban area of Milan is larger (proportionally to the total captured ANEs). Furthermore, the integration time affects deeply on the impact of the ANE; most ANEs have a great impact if measured in a shorter integration time, while the the impact decreases as the integration time increases, so more presence of background noise is also included in the $L_{Aeq}$ evaluation.

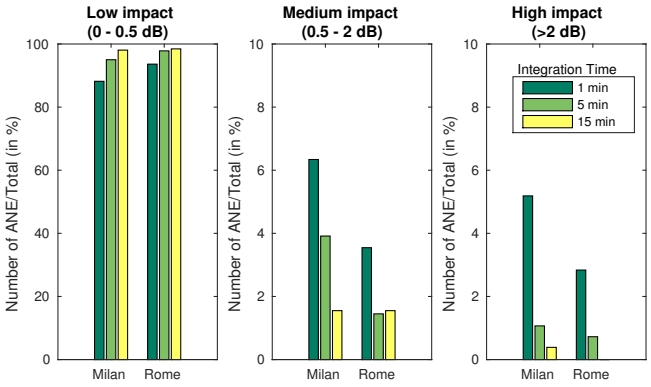



**Figure 5.** Distribution of the individual ANEs according to their impact on the $L_{Aeq}$ computation for the urban (Milan) and suburban (Rome) scenarios.

However, the defined integration time is 5 min [13] for the requirements of the DYNAMAP project, yielding to a detailed analysis of this particular integration time. In Table 1, the ANEs belonging to high and medium impact regions, given a 5-min integration time, are described.

The reader may observe that the urban scenario presents more significant ANEs with higher impact on the $L_{Aeq}$ computation than the suburban one, trains, sirens and truck pass-bys being the ones attaining the highest impact. The duration of the ANEs varies from 16 to 32 s and their SNR varies from 2.2 to 11 dB. The ANE with the highest impact is a train recorded in the urban scenario, but it is not the longest event, since the maximum duration is 53 s. It has a 4.8 dB SNR, which is below the 27.3 dB of the highest ANE's SNR. The ANEs comprised in the medium impact region belong to sirens, in the suburban scenario and to horns, trains and tramway pass-bys. The SNR of the sirens captured in Rome, i.e., suburban scenario, differs from 1.5 to 4.5 dB and the ANEs last from 6 to 16 s. However, the typology of the ANEs changes in the urban scenario of Milan, as the horns are shorter than 2 s and have a SNR above 9 dB. The train and tramway sounds last from 6 to 18 s and comprise a great variety of SNRs, from 2.6 to 15.4 dB.

In addition, the duration and the SNR have an impact on the $L_{Aeq}$ calculation. In both the suburban and the urban scenario, the pattern of the ANEs is similar and the basic difference is that the contribution of the ANE is higher in shorter integration times and higher in longer periods. The suburban scenario in Table 1 presents only one siren with 4.4 dB of SNR and duration of 16 s, generating an impact of 2.4 dB after 5 min $L_{Aeq}$ evaluation. The urban scenario presents more variate types of ANE and longer duration, as the siren with only 2.2 dB of SNR but lasting 30 s, with an impact of 3.9 dB, a high value in the range of the results of our analysis. The urban scenario propitiates the occurrences of ANE with higher SNR (the measuring point is usually closer to the noise source than in a suburban environment), and urban environments present more actors living in the street (e.g., people, animals, sounds coming from houses).

**Table 1.** Anomalous Noise Events (ANEs) of the high- and medium-impact regions from the suburban (Rome) and urban (Milan) scenarios.

| Region | Scenario | Type | Signal-to-Noise Ratio (dB) | Duration (s) | $\Delta L_{A5min}$ (dB) |
|---|---|---|---|---|---|
| High Impact (>2dB) | Suburban | Siren | 4.4 | 16 | 2.4 |
| | Urban | Train | 4.8 | 26 | 7.4 |
| | | Siren | 2.2 | 50 | 3.9 |
| | | Truck | 11.0 | 17 | 3.5 |
| | | Train | 6.9 | 32 | 2.8 |
| Medium Impact (0.5–2 dB) | Suburban | Siren | 4.5 | 9 | 1.1 |
| | | Siren | 4.3 | 6 | 0.6 |
| | | Siren | 1.5 | 16 | 0.6 |
| | Urban | Train | 13.8 | 18 | 1.9 |
| | | Horn | 12.0 | 2 | 1.4 |
| | | Horn | 14.3 | 1 | 1.4 |
| | | Train | 2.6 | 7 | 1.2 |
| | | Horn | 9.6 | 1 | 1.1 |
| | | Tramway | 17.1 | 15 | 1.0 |
| | | Tramway | 12.0 | 10 | 0.8 |
| | | Tramway | 15.4 | 9 | 0.8 |
| | | Tramway | 3.8 | 6 | 0.6 |
| | | Tramway | 3.5 | 15 | 0.6 |
| | | Horn | 9.6 | 1 | 0.6 |

### 4.3. Aggregated Impact of All ANEs

From the previous section, it can be observed that several individual ANEs should be removed from the road traffic noise level calculation due to their high impact—more than 2 dB—over the $L_{Aeq}$ value. However, this individual analysis should be verified to address what has been observed in real-life acoustic data, where several ANEs can occur within a predefined integration time, and depending on the density of ANEs, a series of low-impact and medium-impact events could also exceed the 2 dB threshold. In this consideration, ANEs with high SNR and small duration, or ANEs with moderate SNR and longer duration could become a key component of this aggregated impact value (see Figure 1 for more detail). Next, several examples showing the aggregated impact of ANEs on $L_{A5min}$ are included as a previous step to the 15-min impact study.

Four particular ANEs belonging to urban and suburban scenarios are depicted in Figure 6, where a representative combination of duration and SNR parameters have been chosen to illustrate the low-, medium- and high-impacts in the $L_{A5min}$ calculation. For illustrative purposes and to improve the comparative, the $x$-axis has been established to 80 s. The bottom left ANE belongs to a door sound and has nearly no impact on the $L_{A5min}$ measurement. Bottom right example corresponds to a train pass-by with long duration and moderate SNR; it represents an individual impact of 1.2 dB, which is under the 2 dB threshold, but together with other occurring ANEs in the same integration time can lead the global impact to a value over 2 dB. Something similar happens with the top left example, which contains a horn with a short duration but a high SNR, whose individual impact on the $L_{A5min}$

sums up to 1.1 dB. Finally, the top right example is a siren with an impact of 3.9 dB on the $L_{A5min}$ value, which could be even higher if other ANEs were found in the surrounding 5-min piece of sound, and it could also have a relevant impact for larger integration times.

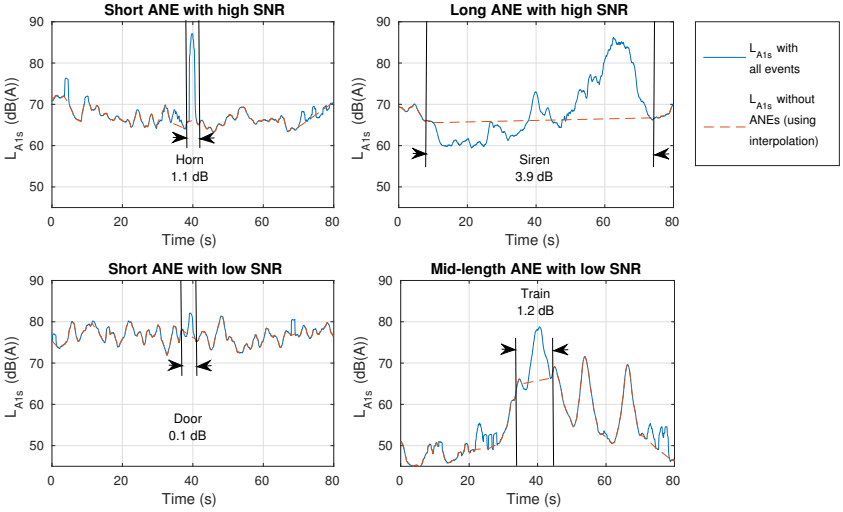



**Figure 6.** Four examples of ANEs showing different individual impact on the $L_{A5min}$ computation, considering a reference integration time of 1 s for resolution purposes.

In Figure 7, the reader may observe a $L_{A5min}$ of two different recordings in the urban area. The upper example contains several horns, two truck hitching system noises, two conversations near the microphone and a tramway pass-by. From these ANEs, the ones biasing the $L_{A5min}$ more significantly are two horns of 0.3 and 1.0 dB each. Once the impact is evaluated, considering all of the involved ANEs, it rises to 1.5 dB, becoming an aggregated set of ANEs with a medium-impact on the $L_{A5min}$ computation. This is an example of why the medium-impact ANEs should also be removed from the data to compute the $L_{Aeq}$.

Moreover, in Figure 7, the lower example belongs to a 5 min fragment that contains a siren yielding a high individual $L_{Aeq}$ impact (around 3.9 dB). The joint impact of the three ANEs existing in this example sums up to 4.1 dB of impact over the total $L_{A5min}$, which is clearly over the threshold of 2 dB.

Below, the aggregated impact is also analyzed in the 15-min time span as the maximum integration time including all recordings. In Table 2, the highest impact in terms of $\Delta L_{A15min}$ within the Milan area are shown, including a brief description of the contained ANEs. A high presence of tramway and train pass-bys are recorded in Site 4 and Site 5. In addition, several high-SNR horns and hitching systems have been captured in Site 3, whose ANEs represent an impact of 1.9 dB on the $L_{A15min}$ calculation. In Rome, the maximum impact calculation in the 15-min interval is 0.2 dB, a non-relevant contribution to the $L_{Aeq}$ calculation.

It is worth mentioning that several ANEs have proven to have a significant impact on the $L_{Aeq}$ value for all the considered integration times. For instance, the siren depicted in the top-right example of Figure 6 belongs to the urban scenario and results as a high-impact individual ANE in Figure 4 and in Table 1 (with an $\Delta L_{A15min} = 3.9$ dB). Moreover, the same siren is the main one responsible for the $\Delta L_{A15min} = 1.9$ dB impact of the second row of Table 2, from this table, the reader may also notice that the second and the third examples show more heterogeneity in the variety of occasionally aggregated

ANEs, not as in the first and the last rows, where the impact on the $L_{A15min}$ value is caused by the same typology of ANEs, i.e., trains and tramways, respectively.

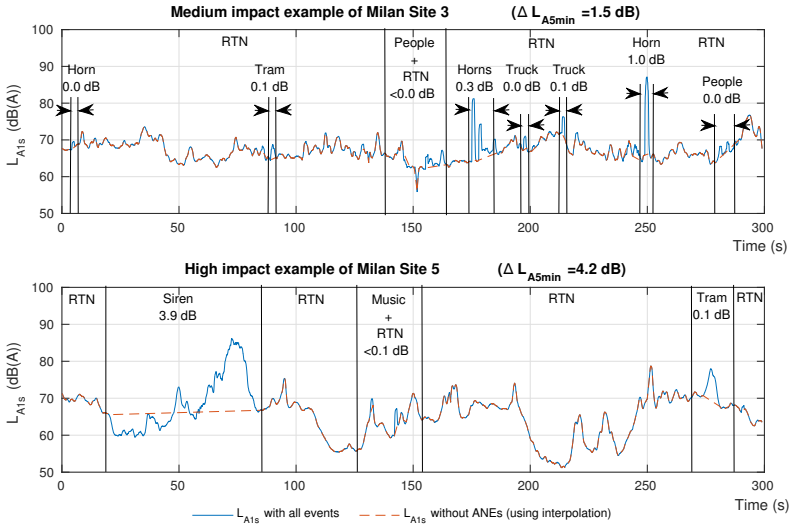



**Figure 7.** Joint impact of all the existing ANEs in the $L_{A5min}$measurement for two different recording locations in the urban environment.

**Table 2.** Detail of the highest impact of aggregated ANEs on $L_{A15min}$ within the Milan urban area for different recording sites.

| Scenario | Location | $\Delta L_{A15min}$ (dB) | Description |
|---|---|---|---|
| Milan | Site 4 | 2.7 | Train pass-bys of 140 s in total. |
| | Site 5 | 1.9 | Tramway pass-bys of 40 s in total and a 50-s high-SNR siren. |
| | Site 3 | 1.9 | Many high-SNR horns and three high-impact hitching system noises. |
| | Site 5 | 1.3 | Tramway pass-bys of 80 s in total. |

As a general conclusion, the aforementioned examples show how the presence of low- and/or medium-impact ANEs can modify the aggregated impact calculation regardless of the homogeneity or heterogeneity found in the time interval, thus, all ANEs that could bias the equivalent noise level of the map must be taken into account. The high-impact ANEs affect greatly the $L_{Aeq}$ and must be considered, but also intervals with several medium-impact ANEs can be biased. Hence, both medium- and high-impact ANEs must be detected and identified if public health studies have to be derived from the noise map.

## 5. Discussion

The performed analyses on real-life urban and suburban acoustic data have led us to conclude that the impact of the ANEs on the calculation of the $L_{Aeq}$ value can become very significant (higher than 2 dB), making the ANE removal mandatory for deriving a reliable noise map from RTN. In the following sections, we discuss several questions that can be derived from these results.

### 5.1. Evaluation of the Impact of ANEs on $L_{Aeq}$ Considering Real-Life Data

In this paper, the impact on the computation of $L_{Aeq}$ has been evaluated for every individual ANE over three different integration times using both urban and suburban acoustic data. The results show that the pattern of the results of impact for most ANEs remain the same—the impact of each ANE remains sorted with the same order—despite the final value of impact decreasing as the integration time increases (see the comparisons in Figures 3 and 4). The individual ANE analysis presents interesting results to be discussed using both SNR and duration of the ANE; the results in Section 4 present several ANEs with critical impact, usually related to high SNR and long duration. Several details are also presented about the medium impact ANEs, which, despite presenting impacts from 0.5 dB to 2 dB, can not be neglected in terms of this study. Nevertheless, the other scenarios obtained from the values of SNR and duration have to be studied deeply.

After the individual analysis of the impact of ANEs on the equivalent noise level computation, the work has also taken into account that dynamic acoustic mapping in *real-life* conditions face a more complex operating scenario. Within a predefined integration time, several ANEs can occur in real-life data, so the ANE impact has to be evaluated in an aggregated way. Figure 7 illustrated a couple of examples where the final aggregated impact is significantly higher than the individual impact of each ANE by itself in the $L_{A5min}$ , following the DYNAMAP project specifications. In this scenario, the aforementioned medium individual impact of a particular ANE can lead to a significant impact on $L_{Aeq}$ if surrounded by other ANEs entailing low and/or moderated individual impacts. In this sense, the medium-impact ANE group has been defined in the range [0.5, 2] dB, but all positive impact ANEs should be taken into account in terms of aggregated results in an integration time period.

Another relevant result of the individual ANE analysis is the presence of ANEs with negative SNR. As detailed in Section 3.1, the SNR is evaluated taking into account a particular ANE in relation to the surrounding RTN signal level; in certain cases, the RTN noise decreases as the ANE occurs, so a negative SNR is obtained. After the individual study, it can be concluded that only events with positive SNR should be removed from the $L_{Aeq}$ computation, otherwise, we could face an unrealistic situation of a negative impact on the $L_{Aeq}$: the equivalent level could be lower taking into account the ANEs. In this case, the closer location to the real $L_{Aeq}$ value is to consider the ANEs to perform the integration. The reason is due to the RTN or the background noise equivalent level during an ANE, which is lower, and this value should be considered as the most suitable in the total $L_{Aeq}$ value for the period.

Finally, the 15-min joint study shows that many cases exist in the urban area where the labeled ANEs bias the $\Delta L_{A15min}$ calculation in more than 1 dB and, in one case, in almost 3 dB. Thus, a presence of medium- and high-impact ANEs could distort the noise map also in larger time periods if needed in any evaluation.

### 5.2. Impact of the ANEs on the Population

The results presented in this paper detail a clear impact of the ANEs over the evaluation of the $L_{Aeq}$ value. However, a substantial impact in terms of dB may not correspond to a proportional annoyance to the people hearing these anomalous noise events. Many studies have been conducted detailing the adverse effects of traffic noise on health, but, to complete the characterization of the ANEs in the streets of an urban and suburban environment, we should test whether the impact on the quantitative measures used in this paper are correlated with the annoyance of the affected population in terms of their subjective perception. In addition, analyzing the acoustic measure every 5 min allows for informing the authorities and population in a maximum delay of 5 min. Several key questions still remain open. Is a siren more annoying than a train, having both the same acoustic impact? Or maybe the duration of the ANEs is a key factor in the disturbance? Or the salience in comparison with the background noise? The next steps of this study pretend to make way towards subjective analysis of the noise impact and, eventually, to contribute to a method of determining the impact of certain sound events to the citizens taking into account the effects of ANEs in these kinds of studies.

## 6. Conclusions

This work has introduced an analysis methodology to evaluate the impact of the ANEs on the $L_{Aeq}$ computation for different integration times in both real-life urban and suburban areas. The conducted experiments allow for concluding that any automatic approach to noise monitoring and control should take into account the impact of anomalous noise events on the road traffic noise $L_{Aeq}$ computation. To that effect, both SNR and duration have been proven to be key variables to determine the impact of either individual or aggregated ANEs on the equivalent noise level for a predefined integration time. In this sense, the results show that the impact of individual ANEs can be substantial, especially in an urban environment, but when the study is widened by taking into account several ANEs occurring within the same period simultaneously, the analysis shows that even those ANEs attaining individual low- or medium-impact may contribute to surpass the 2 dB defined threshold that determines a substantial change in the $L_{Aeq}$ evaluation. For this purpose, a deeper analysis on the evaluation of the $L_{Aeq}$ value when considering all positive impact ANEs for a certain integration period will be conducted in the future. The goal is to determine the minimum values of SNR and duration for the ANEs to suggest its removal and to find the minimum number of ANEs needed to have an impact of more than the threshold of 2 dB.

Another conclusion obtained from this study is that negative SNR can be occasionally obtained when parametrizing the ANE database. Due to the relative nature of the SNR parameter definition, these results could be obtained when the ANE is quieter than the surrounding background noise. The detailed analysis of this kind of situations is left for future works.

The duration and the SNR have been proven to be key values to model the nuisance that an ANE causes to the neighbourhood in real-time (e.g., every 5 min as agreed within the DYNAMAP project). On the one hand, when working in short time spans, all ANEs non-related to the source under study, traffic noise in this case, should be removed from the noise map computation to avoid biasing the results. However, on the other hand, it allows for making fast decisions to inform the exposed citizens or the local authorities if unusual noise levels are recorded.

We would like to close this work with a question. Is the impact of the ANEs on $L_{Aeq}$ a good measure for annoyance in the neighbourhood? We know from literature that the equivalent noise level is a good indicator to find a relationship between high noise measurements and several illnesses. The study of the degree of annoyance in the citizens considering anomalous noise events would allow us to study if the impact, the SNR and/or the duration, are key values to modelling the nuisance that an ANE causes to the neighbourhood. This research will entail the next step of our investigations.

**Acknowledgments:** 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. Ferran Orga is thankful for the support of the European Social Fund and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Catalan Government for the pre-doctoral FI Grant No. 2017FI_B00243.

**Author Contributions:** Ferran Orga conducted the calculations of SNR and impact on the $L_{Aeq}$, as well as most of the writing. Francesc Alías and Rosa Ma Alsina-Pagès reviewed the paper and had an important role in the design of the analysis methodology and discussion of the results.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| ANE | Anomalous Noise Event |
| CNOSSOS-EU | Common Noise Assessment Methods in Europe |
| DALYs | Disability-Adjusted Life-Years |
| END | Environmental Noise Directive |
| RTN | Road Traffic Noise |

SNR Signal-to-Noise Ratio
WASN Wireless Acoustic Sensor Networks
WHO World Health Organization

## References


1. Goines, L.; Hagler, L. Noise pollution: A modern plague. *South. Med. J.* **2007**, *100*, 287–294.
2. Stansfeld, S.A.; Matheson, M.P. Noise pollution: Non-auditory effects on health. *Br. Med. Bull.* **2003**, *68*, 243–257.
3. Babisch, W. Transportation noise and cardiovascular risk. *Noise Health* **2008**, *10*, 27–33.
4. Regecová, V.; Kellerová, E. Effects of urban noise pollution on blood pressure and heart rate in preschool children. *J. Hypertens.* **1995**, *13*, 405–412.
5. Öhrström, E.; Skånberg, A.; Svensson, H.; Gidlöf-Gunnarsson, A. Effects of road traffic noise and the benefit of access to quietness. *J. Sound Vib.* **2006**, *295*, 40–59.
6. Botteldooren, D.; Dekoninck, L.; Gillis, D. The influence of traffic noise on appreciation of the living quality of a neighborhood. *Int. J. Environ. Res. Public Health* **2011**, *8*, 777–798.
7. Jakovljevic, B.; Paunovic, K.; Belojevic, G. Road-traffic noise and factors influencing noise annoyance in an urban population. *Environ. Int.* **2009**, *35*, 552–556.
8. EUR-Lex. Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise. *Off. J. Eur. Communities* **2002**, *L 189/12*, 12–26.
9. Kephalopoulos, S.; Paviotti, M.; Ledee, F.A. *Common noise assessment methods in Europe (CNOSSOS-EU)*; Publications Office of the European Union: Brussels, Belgium, 2012.
10. Camps, J. Barcelona noise monitoring network. In Proceedings of the Euronoise, Maastricht, Netherlands, 31 May – 3 June 2015; pp. 218–220.
11. Nencini, L.; De Rosa, P.; Ascari, E.; Vinci, B.; Alexeeva, N. SENSEable Pisa: A wireless sensor network for real-time noise mapping. In Proceedings of the EURONOISE, Prague, Czech Republic, 10–13 June 2012; pp. 10–13.
12. De Coensel, B.; Botteldooren, D. Smart sound monitoring for sound event detection and characterization. In Proceedings of the 43rd International Congress on Noise Control Engineering (Inter-Noise 2014), Ghent University, Gent, Belgium, 16–19 November 2014.
13. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189.
14. Socoró, J.C.; Alías, F.; Alsina-Pagès, R.M. An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments. *Sensors* **2017**, *17*, 2323, doi:10.3390/s17102323.
15. Alías, F.; Socoró, J.C. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Appl. Sci.* **2017**, *7*, 146.
16. Tsai, K.T.; Lin, M.D.; Chen, Y.H. Noise mapping in urban environments: A Taiwan study. *Appl. Acoust.* **2009**, *70*, 964–972.
17. Zannin, P.H.T.; Diniz, F.B.; Barbosa, W.A. Environmental noise pollution in the city of Curitiba, Brazil. *Appl. Acoust.* **2002**, *63*, 351–358.
18. Morillas, J.B.; Escobar, V.G.; Sierra, J.M.; Gómez, R.V.; Carmona, J.T. An environmental noise study in the city of Cáceres, Spain. *Appl. Acoust.* **2002**, *63*, 1061–1070.
19. Chakrabarty, D.; Chandra Santra, S.; Mukherjee, A.; Roy, B.; Das, P. Status of road traffic noise in Calcutta metropolis, India. *J. Acoust. Soc. Am.* **1997**, *101*, 943–949.
20. Brown, A.; Lam, K. Levels of ambient noise in Hong Kong. *Appl. Acoust.* **1987**, *20*, 85–100.
21. World Health Organization. *Burden of Disease from Environmental Noise: Quantification of Healthy Life Years Lost in Europe*; World Health Organization: Geneva, Switzerland, 2011; p. 126.
22. Śliwińska-Kowalska, M.; Zaborowski, K. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Permanent Hearing Loss and Tinnitus. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1139.

23. Nieuwenhuijsen, M.J.; Ristovska, G.; Dadvand, P. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Adverse Birth Outcomes. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1252.
24. Guite, H.; Clark, C.; Ackrill, G. The impact of the physical and urban environment on mental well-being. *Public Health* **2006**, *120*, 1117–1126.
25. Sørensen, M.; Andersen, Z.J.; Nordsborg, R.B.; Becker, T.; Tjønneland, A.; Overvad, K.; Raaschou-Nielsen, O. Long-term exposure to road traffic noise and incident diabetes: A cohort study. *Environ. Health Perspect.* **2013**, *121*, 217–222.
26. Passchier-Vermeer, W.; Passchier, W.F. Noise exposure and public health. *Environ. Health Perspect.* **2000**, *108*, 123–131.
27. Van Renterghem, T.; Botteldooren, D. Focused study on the quiet side effect in dwellings highly exposed to road traffic noise. *Int. J. Environ. Res. Public Health* **2012**, *9*, 4292–4310.
28. Brown, A.L.; van Kamp, I. WHO environmental noise guidelines for the European region: A systematic review of transport noise interventions and their impacts on health. *Int. J. Environ. Res. Public Health* **2017**, *14*, 873.
29. Barham, R.; Chan, M.; Cand, M. Practical experience in noise mapping with a MEMS microphone based distributed noise measurement system. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*; Institute of Noise Control Engineering: Reston, VA, USA, 2010; Volume 6, pp. 4725–4733.
30. Rainham, D. A wireless sensor network for urban environmental health monitoring: UrbanSense. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2016; Volume 34.
31. Bartalucci, C.; Borchi, F.; Carfagni, M.; Furferi, R.; Governi, L. Design of a prototype of a smart noise monitoring system. In Proceedings of the 24th International Congress on Sound and Vibration (ICSV24), London, UK, 23–27 July 2017.
32. Gupta, S.; Ghatak, C. Environmental noise assessment and its effect on human health in an urban area. *Int. J. Environ. Sci.* **2011**, *1*, 1954–1964.
33. Licitra, G.; Gallo, P.; Rossi, E.; Brambilla, G. A novel method to determine multiexposure priority indices tested for Pisa action plan. *Appl. Acoust.* **2011**, *72*, 505–510.
34. Miedema, H.M. Relationship between exposure to multiple noise sources and noise annoyance. *J. Acoust. Soc. Am.* **2004**, *116*, 949–957.
35. European Commission Working Group (Assessment of Exposure to Noise) Good Practice Guide for Strategic Noise Mapping and the Production of Associated Data on Noise. 2006. Available online: http://sicaweb.cedex.es/docs/documentacion/Good-Practice-Guide-for-Strategic-Noise-Mapping.pdf (Accessed on 22 December 2017).
36. Liguori, C.; Ruggiero, A.; Russo, D.; Sommella, P. Estimation of the minimum measurement time interval in acoustic noise. *Appl. Acoust.* **2017**, *127*, 126–132.


# Aggregate Impact of Anomalous Noise Events on the WASN-Based Computation of Road Traffic Noise Levels in Urban and Suburban Environments

**Francesc Alías, Ferran Orga, Rosa Ma. Alsina-Pagès, Joan Claudi Socoró**

Publicat: 22 de gener del 2020.

sensors

MDPI

*Article*

# Aggregate Impact of Anomalous Noise Events on the WASN-Based Computation of Road Traffic Noise Levels in Urban and Suburban Environments

**Francesc Alías *, Ferran Orga, Rosa Ma Alsina-Pagès and Joan Claudi Socoró**

GTM—Grup de recerca en Tecnologies Mèdia, La Salle—Universitat Ramon Llull. c/Quatre Camins, 30, 08022 Barcelona, Spain; ferran.orga@salle.url.edu (F.O.); rosamaria.alsina@salle.url.edu (R.M.A.-P.); joanclaudi.socoro@salle.url.edu (J.C.S.)

* Correspondence: francesc.alias@salle.url.edu; Tel.: +34-932902440

Received: 10 December 2019; Accepted: 20 January 2020; Published: 22 January 2020

**Abstract:** Environmental noise can be defined as the accumulation of noise pollution caused by sounds generated by outdoor human activities, Road Traffic Noise (RTN) being the main source in urban and suburban areas. To address the negative effects of environmental noise on public health, the European Environmental Noise Directive requires EU member states to tailor noise maps and define the corresponding action plans every five years for major agglomerations and key infrastructures. Noise maps have been hitherto created from expert-based measurements, after cleaning the recorded acoustic data of undesired acoustic events, or Anomalous Noise Events (ANEs). In recent years, Wireless Acoustic Sensor Networks (WASNs) have become an alternative. However, most of the proposals focus on measuring global noise levels without taking into account the presence of ANEs. The LIFE DYNAMAP project has developed a WASN-based dynamic noise mapping system to analyze the acoustic impact of road infrastructures in real time based solely on RTN levels. After studying the bias caused by individual ANEs on the computation of the A-weighted equivalent noise levels through an expert-based dataset obtained before installing the sensor networks, this work evaluates the aggregate impact of the ANEs on the RTN measurements in a real-operation environment. To that effect, 304 h and 20 min of labeled acoustic data collected through the two WASNs deployed in both pilot areas have been analyzed, computing the individual and aggregate impacts of ANEs for each sensor location and impact range (low, medium and high) for a 5 min integration time. The study shows the regular occurrence of ANEs when monitoring RTN levels in both acoustic environments, which are especially common in the urban area. Moreover, the results reveal that the aggregate contribution of low- and medium-impact ANEs can become as critical as the presence of high-impact individual ANEs, thus highlighting the importance of their automatic removal to obtain reliable WASN-based RTN maps in real-operation environments.

**Keywords:** road traffic noise; noise monitoring; dynamic noise maps; anomalous noise events; individual impact; aggregate impact; WASN; sensor nodes; urban and suburban environments.

## 1. Introduction

Environmental noise can be defined as the accumulation of noise pollution caused by sounds generated by human activity outdoors, mainly produced by transport, road traffic, rail traffic, air traffic and industrial activities [1]. According to the World Health Organization, noise exposure produces a loss of around one million healthy life years in Western Europe every year due to different types of derived diseases [2,3]. Focusing on this public health problem, the European (EU) authorities published the Environmental Noise Directive (END) [1] in 2002, which requires the EU member states to tailor noise maps and to develop the subsequent action plans to mitigate noise every five years for

major agglomerations and key infrastructures [4]. To address this issue in a harmonized manner, the Common Noise Assessment Methods in Europe (CNOSSOS-EU) was also developed, defining the measurement guidelines to allow comparable noise assessments across the EU [5]. However, as one of the first set of results obtained after the implementation of the END regulation showed [6], noise pollution continues to be one of the principal causes of health problems in Europe. This premise was further endorsed by [7,8], which led to the development of an updated version of the CNOSSOS-EU [9].

The aforementioned dramatic effects of noise pollution on citizens are mainly caused by traffic noise, as it is the main noise source in urban and suburban areas [10,11]. Road Traffic Noise (RTN) maps have been historically created from expert-based measurements using certified devices during specific time periods and locations, considering vehicle flows averaged over long periods of time [12]. During the recordings, the presence of acoustic events non-related to road traffic (e.g., sirens, horns, works, dogs' barks, airplanes flyovers, etc.) may occur [13]. As a consequence, the collected acoustic data should be cleaned of these undesired events before feeding the noise map creation software [13] to avoid biasing the computation of the A-weighted equivalent sound levels ($L_{Aeq}$) beyond 2 dB, as recommended by the European Commission Working Group Assessment of Exposure to Noise (WG-AEN) [14]. In this context, the Signal-to-Noise Ratio (SNR) of these acoustic events becomes a crucial parameter to evaluate and model [15,16]. Although some researchers have opted to control the SNR of the events by creating artificially mixed datasets (see e.g., [17–20]), their accurate characterization remains as an open research question as it is almost unfeasible to represent the wide diversity of acoustic data for real world [21].

The so-called Wireless Acoustic Sensor Networks (WASNs) have become an alternative to the creation of noise maps using real-life data, since they allow the ubiquitous monitoring of environmental noise [22–24]. During the last decade, several WASNs have been deployed in different smart cities such as Barcelona [25], Algemesí [26], Pisa [27], Monza [28], Halifax [29] and Milan and Rome [30] in Europe, or New York city [31], to name a few. In this WASN-based approach, the traditional manual cleaning of the Anomalous Noise Events (ANEs) on the noise pattern [32] becomes unfeasible due to the huge volume of data that have to be processed in real time [13]. As a consequence, the first generation of these WASN-based environmental noise monitoring systems have mainly been focused on measuring the global sound levels of the sensed locations, without considering the impact of the presence of specific acoustic events on the $L_{Aeq}$ computation. To address this issue, some projects have started incorporating acoustic event detection techniques within the WASN-based noise monitoring pipeline. The Sounds of New York City (SONYC) project includes the real-time identification of 10 common classes of urban sound sources [31] through a machine listening system trained after artificially mixing the events with background noise in the UrbanSound dataset [16]. Moreover, the DYNAMAP project aims at developing a WASN-based dynamic noise mapping system to monitor the acoustic impact of road infrastructures through the creation of noise maps in real time [30]. The project includes two pilot areas: one in the District 9 of Milan as urban area [33], and another in the A90 highway surrounding Rome as a suburban area [34,35]. As the system focuses on measuring RTN levels solely, the ANEs present in the acoustic environments should be automatically removed. To that effect, a machine listening algorithm denoted as Anomalous Noise Events Detector [36] was designed and initially trained using a 9-h expert-based dataset collected from the two pilot areas before installing both sensor networks [21]. The analysis of that preliminary dataset highlighted the importance of the removal of individual ANEs based on their duration and SNR [37]. However, no evidence of a critical impact was yet observed in that dataset due to the presence of several ANEs within the same period of time, probably because the expert-based dataset missed several key aspects from real operation, such as different RTN patterns between day-night and weekday-weekends, or variable weather conditions, among others [38].

After the deployment of the two WASNs in the urban and suburban pilot areas, this paper evaluates the aggregate impact of ANEs on the $L_{Aeq}$ computation of RTN in both environments in real operation. Besides analyzing the individual impact of ANEs on the measurements, the analysis

methodology focuses on evaluating the bias caused by the presence of several ANEs within a given period of time, taking into account their impact range (low, medium or high) and sensor location. The study is conducted on 304 h and 20 min of WASN-based labeled acoustic data collected through both sensor networks, before proceeding to update the ANED algorithm with both WASN-based datasets (see [39,40] for a detailed description of the general characteristics of the urban and suburban datasets, respectively).

The paper is structured as follows. Section 2 reviews practices in acoustic environments where the salience and the impact of the events is a key issue. Section 3 presents the impact analysis methodology and impact-related measurements. Section 4 presents the conducted experiments and the results obtained from the analysis of the WASN-based urban and suburban acoustic datasets. Finally, after discussing several key aspects of this work in Section 5, the main conclusions and future work are described in Section 6.

## 2. Related Work

In this section, we review several works from the literature dealing with the identification of salient acoustic events regardless of the noise source; this issue together with the duration of the event sets the basis for the evaluation of the actual impact of these events on the $L_{Aeq}$ computation.

One of the most challenging issues when working with environmental acoustic data recorded in real-life is their accurate characterization, which is supervised by experts. More precisely, this process deals with the parameterization of the data by means of several representative features, among which are the temporal limits of each sound event—i.e., its *actual* duration—by setting up its start and end boundaries [41,42], and its acoustic salience with respect to the background noise [15,16], i.e., the SNR of the event, which is a key parameter to consider. To properly address this issue, it should be taken into account that the events that need to be detected are usually independent one from each other, and typically present a variable duration and SNR. Furthermore, no temporal correlation can be found among them, which makes the challenge of parameterizing audio events particularly more complex compared to speech or music signal [43]. Consequently, the accurate characterization of environmental sound remains as an open research question in real world environments [21].

To work with a controlled environment, artificially-mixed datasets are usually built taking into account a predefined range of SNRs when mixing the events with the background noise during the dataset process generation. Some examples can be found in Foggia et al. [17], Stowell et al. [18] and Socoró et al. [19] (see [21] for further examples). The measurements of SNRs in audio fragments makes it possible to sort events by their degree of acoustic salience with respect to their environment. Moreover, datasets containing synthetic or artificially modified samples also respond to the need to generate more samples of a particular type of noise that is scarce , which is yet today one of the main limitations of acoustic event detection [44]. The explicit SNR measure can be evaluated by means of a closed set of saliency levels, such as $-6$ dB, 0 dB or $+6$ dB, as suggested by Stowell et al. in [18]; the authors also propose to record live scripted monophonic event sequences in acoustic environments under control. Foggia et al. [17] mixes several sounds related to surveillance (e.g., scream, glass breaking and gunshots) with both indoor and outdoor environments with six different levels of SNR (from 5 dB to 30 dB, with a step of 5 dB), after the observation of the occurrences of these events in a real-life environment. Socoró et al. [19] presents a dataset composed of a mixture of sound sources considering road traffic noise plus other type of sound events generated using two different SNRs ($+6$ dB and $+12$ dB) in order to assess the performance of an anomalous noise event detector. The original non-traffic-noise related audio fragments were extracted from Freesound (https://freesound.org/) while road traffic noise was recorded in a city ring road in real-life conditions. Nakajima et al. [45] works with a dataset recorded in real operation with several examples of noise sources of interest (e.g., cicadas, outside air conditioner, road traffic noise, and neighborhood noise). The work complements the dataset with artificial mixtures to increase the sound source diversity by means of varying the salience of the events using the SNR of three sound sources in the dataset,

adapting the margins from −6 dB to +6 dB depending on the characteristics of the noise source. Finally, in Koizumi et al. [46], the authors conduct an objective evaluation on a synthetic dataset, using an open toy-car-running sound dataset; the dataset includes four types of factory noises, and it was generated by mixing synthetically those audio samples at a SNR = 0 dB, together with the audio files of less than 5-s duration from the Task-2 dataset of DCASE 2018 Challenge [47].

Following a different approach, several research works consider auditory attention when evaluating the impact of sound events on acoustic measurements through the evaluation of their SNR levels, whose focus may vary depending on the domain of application (e.g., noise monitoring or surveillance) or the signal of interest (see [48] and references therein for further details). These works analyze the perceptual relevance of audio events according to human response, as in [15], where De Coensel and Bootteldooren design a salience-based map to simulate the capability of humans to switch the attention among several auditory stimuli along time, considering noise examples of means of transportation. This research approach is focused on the identification of the salient event. However, it ignores both its origin and its relative energy with respect to background noise. Following this approach, Salamon et al. [16] included a perceptually based binary descriptor in their dataset to discriminate whether the event was perceived as the main noise source or in the background of the recording. Afterwards, the dataset was used to evaluate the performance of a sound event classification algorithm, getting better accuracy results on foreground events rather than those perceived in the background. Annotating and evaluating a recorded set of audio files is a very time-consuming task. To address these limitations, Salamon et al. published Scaper [20], whose goal is to conduct soundscape synthesis together with data augmentation given a soundbank, controlling characteristics such as the number and type of events, their timing, duration and SNR with respect to a background sound. The final goal is to ease the dataset generation process but also to ensure that the sets of data evaluated present suitable statistical characteristics for training and test of acoustic event detection algorithms.

Finally, it is worth mentioning that a couple of WASN-based projects have recently incorporated the detection of acoustic events in urban and suburban environments in the environmental noise monitoring pipeline. To that effect, the SONYC project [31] has developed a representative dataset with diverse sounds of interest, using the data gathered from the 56 sensors deployed in different neighborhoods of New York, considering up to 10 different common urban sound sources from the urban soundscape (highly frequent in urban noise complaints). The UrbanSound dataset was created after artificially mixing the events coming from Freesound with the background noise collected in the project [16]. Our team, in the framework of the DYNAMAP project [30] made its first attempt to create an acoustic dataset of the urban and suburban pilot areas (District 9 in Milan and A90 highway surrounding Rome) before the sensors of the two WASN were deployed in those scenarios, by means of an expert-based recording campaign [21]. The analysis of those datasets showed the highly local and unpredictable nature of anomalous noise events, which were manually labeled and used to train the preliminary version of the ANED algorithm [36]. Recently, the deployment of the two WASNs in both pilot areas has led to the generation of a suburban acoustic dataset through the 19-nodes WASN in Rome [40], together with the completion of the first steps of the creation of an urban dataset through the 24-node WASN installed in Milan in real operation [39]. From these two experiences, it can be concluded that the evaluation of the acoustic salience of any environmental acoustic event is relevant in order to improve the accuracy of the derived machine listening approaches [43], an issue that was justified in [37] after evaluating the individual impact of the detected events on the overall equivalent noise level computation considering 9 h of real-life acoustic data collected through an expert-based recording campaign. However, as far as we know, no specific analysis has been conducted to assess to what extent the concentration of ANEs with low SNRs within a period of time may bias the WASN-based computation of the $L_{Aeq}$ measurements.

## 3. Impact Analysis Methodology

This section describes the methodology followed to analyze the bias caused by ANEs on the $L_{Aeq}$ computation for a given integration time $T$ (hereafter denoted as $L_{Aeq,T}$), building on the analysis methodology presented in [37]. The impact analysis methodology permits the study of both individual and aggregate contributions of the anomalous noise events present within a specific period of time. To that effect, individual and aggregate impact histograms are obtained from the labeled data for each sensor of the network according to the considered impact ranges. As depicted in Figure 1, the analysis starts with the labeled acoustic data collected from a WASN of $N_S$ sensors in real operation. After windowing the audio streams into frames of $T$ seconds, the individual and aggregate impacts of the ANEs present in each period of time $t$ are computed and stacked. Finally, both individual and aggregate impact histogram matrices are derived to account for the occurrences belonging to each impact range defined by a set of impact thresholds. The following paragraphs explain the key elements of the proposed analysis methodology in detail.

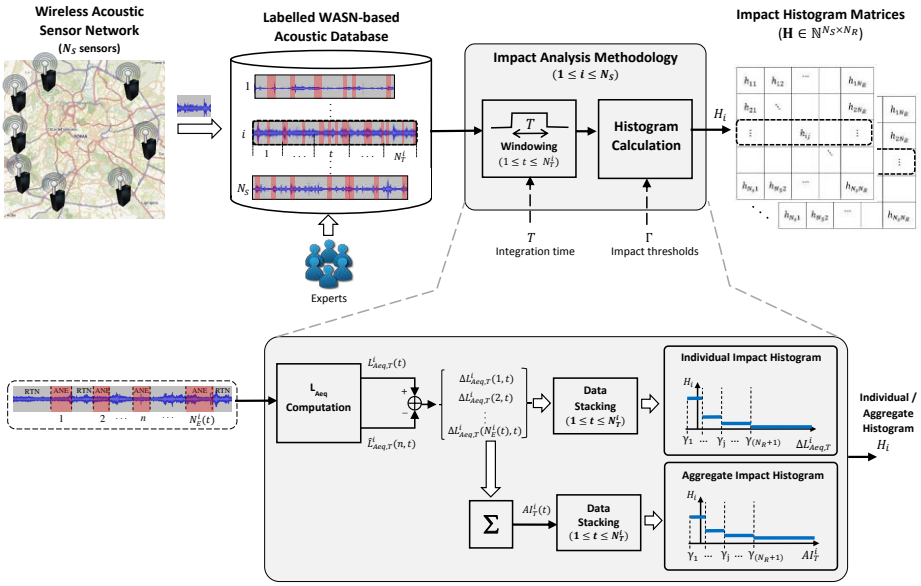



**Figure 1.** Block diagram of the impact analysis methodology on a labeled WASN-based acoustic dataset obtained from a $N_S$ sensors network, where $T$ is the integration time considered to compute $L^i{}_{Aeq,T}(t)$ and $\hat{L}^i_{Aeq,T}(n,t)$ for each sensor $i$ and event $n$. Moreover, $\Delta L^i_{Aeq,T}(n,t)$ and $AI^i_T(t)$ denote the individual and aggregate impacts of the ANEs, respectively. Finally, $h_{ij}$ represents the components of the histogram matrices $\mathbf{H}$ derived from the individual and aggregate impact histograms $H_i$, which account for the impact values according to $N_R$ impact ranges defined by a set of impact thresholds $\Gamma = \{\gamma_1, \gamma_2, ..., \gamma_{(N_R+1)}\}$.

- **Aggregate impact computation per sensor**

  The *Aggregate Impact* (AI) of several acoustic events can be defined as the accumulated contribution of the individual impacts of all the ANEs present within a period of time and sensor node.

  It is denoted as $AI^i_T(t)$, where indexes $i$ and $t$ respectively represent the sensor number, for $i = \{1, 2, ..., N_S\}$, and the integration time period, for $t = \{1, 2, ..., N^i_T\}$, $N^i_T$ being the total number of integration time periods of length $T$ considered for its computation given a sensor $i$, and it is defined as

$$AI_T^i(t) = \sum_{n=1}^{N_E^i(t)} \Delta L_{Aeq,T}^i(n,t), \tag{1}$$

where $\Delta L_{Aeq,T}^i(n,t)$ is the individual impact of the $n$-th ANE on the $L_{Aeq,T}$ computation within the integration time period $t$, $N_E^i(t)$ being the total number of ANEs present in that time period for sensor $i$, and it is computed as

$$\Delta L_{Aeq,T}^i(n,t) = L_{Aeq,T}^i(t) - \hat{L}_{Aeq,T}^i(n,t), \tag{2}$$

$L_{Aeq,T}^i(t)$ being the total A-weighted equivalent sound level in the integration period of interest $t$ for the $i$-th sensor (i.e., considering RTN and all ANEs found in that $t$), and $\hat{L}_{Aeq,T}^i(n,t)$ the corresponding noise level after removing the $n$-th ANE from the measurement through the linear interpolation of the $L_{Aeq,1s}$ values of the previous and subsequent RTN samples (the reader is referred to [37] for further details).

To that effect, first, the audio data collected from sensor $i$ is divided into $N_T^i$ windows of $T$ seconds length (see Figure 1). Next, the A-weighted equivalent noise levels with and without ANEs are computed, whose difference gives the $n$-th individual ANE impact $\Delta L_{Aeq,T}^i(n,t)$. Then, the aggregate impact of window $t$ is obtained by accumulating the individual impacts of all the ANEs it contains.

- **Range-based impact analysis per sensor**

The analysis methodology also aims at categorizing the relevance of both individual and aggregate impacts according to $N_R$ impact ranges $\Theta = \{\theta_1, \theta_2, ..., \theta_{N_R}\}$ (in dB) delimited by a predefined set of impact thresholds $\Gamma = \{\gamma_1, \gamma_2, ..., \gamma_{(N_R+1)}\}$, and it is computed as

$$\Theta = \bigcup_{j=1}^{N_R} \theta_j = \bigcup_{j=1}^{N_R} [\gamma_j, \gamma_{j+1}), \tag{3}$$

where $\theta_j$ is defined as the impact range where $\gamma_j \leq \Delta L_{Aeq,T}^i(t) < \gamma_{j+1}$, for $j = \{1, 2, ..., N_R\}$.

This information is statistically analyzed through the histograms obtained for each sensor (see Figure 1) in the *impact histogram matrix* $\mathbf{H} = (h_{ij}) \in \mathbb{N}^{(N_S \times N_R)}$, $h_{ij}$ being the number of occurrences of ANEs that account for an impact within $\theta_j$ observed in the $i$-th sensor as follows

$$\mathbf{H} = \begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_i \\ \vdots \\ H_{N_S} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & \cdots & \cdots & h_{1N_R} \\ h_{21} & h_{22} & \cdots & \cdots & \cdots & h_{2N_R} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & h_{ij} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ h_{N_S1} & h_{N_S2} & \cdots & \cdots & \cdots & h_{N_SN_R} \end{pmatrix}, \tag{4}$$

where

$$h_{ij} = \begin{cases} \sum_{t=1}^{N_T^i} \sum_{n=1}^{N_E(t)} \mathbf{1}_{\theta_j}\left(\Delta L_{Aeq,T}^i(n,t)\right) & \text{for individual impact,} \\ \sum_{t=1}^{N_T^i} \mathbf{1}_{\theta_j}\left(AI_T^i(t)\right) & \text{for aggregate impact,} \end{cases} \tag{5}$$

with $\mathbf{1}_{\theta_j}(\cdot)$ being the indicator function defined for the interval range $\theta_j$ as

$$\mathbf{1}_{\theta_j}(x) = \begin{cases} 1 & \text{if } x \in \theta_j, \\ 0 & \text{if } x \notin \theta_j. \end{cases} \tag{6}$$

Notice that rows of $\mathbf{H}$ (denoted as $H_i$ in Equation (4)) correspond to the impact histograms obtained from each $i$ sensor.

- **Analysis of the critical aggregate impacts per impact range and sensor**

  To complement the previous analyses, it is also interesting to identify the origin of critical AIs for those cases that surpass the critical threshold $\gamma_c$. To that effect, the aggregate impact of ANEs for a given integration time period and sensor is computed considering only those individual ANEs which $\Delta L^i_{Aeq}(n,t)$ belongs to a particular impact range (i.e., $\Delta L^i_{Aeq}(n,t) \in \theta_j$) as follows

$$AI^i_T(\theta_j, t) = \sum_{n \in \Psi(\theta_j, t)} \Delta L^i_{Aeq,T}(n,t), \tag{7}$$

  where $\Psi(\theta_j, t)$ represents the subset of ANE indices within $t$ which individual impact belongs to impact range $\theta_j$.

  Finally, the *critical AI histogram matrix* $\mathbf{H_c} = (h^c_{ij}) \in \mathbb{N}^{N_S \times N_R}$ is defined as a particular case of $\mathbf{H}$ (see Equation (4)) considering the matrix components as

$$h^c_{ij} = \sum_{t=1}^{N^i_T} \mathbf{1}_{\theta_c}\left(AI^i_T(\theta_j, t)\right), \tag{8}$$

  the $\mathbf{1}_{\theta_c}(x)$ being a particular case of the indicator function defined by $\theta_j = \theta_c$ (see Equation (6)), where $\theta_c = [\gamma_c, +\infty)$ defines the range of critical impacts, as $\gamma_c$ represents the threshold of a non-tolerable deviation of the A-weighted equivalent road traffic noise levels.

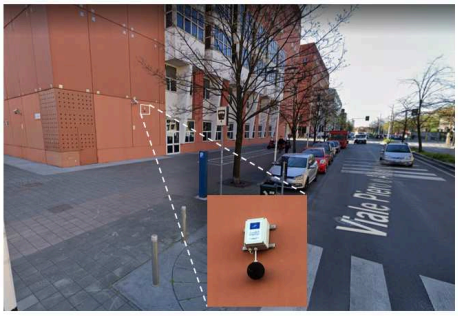
## 4. Experiments and Results

This section describes the results of the experiments from the impact analysis conducted on the two environmental WASN-based audio databases from the DYNAMAP's Milan and Rome pilot areas [39,40]. According to the project specifications, the considered integration time to update the $L_{Aeq,T}$ values of the RTN maps is 5 min [30], i.e., $T = 300$ s. To analyze to what extent the collected ANEs from each sensor location bias the $L_{Aeq,300s}$ measurement, the impacts are categorized within three impact ranges (i.e., $N_R = 3$) [37], accounting for those occurrences (from either individual or aggregate ANEs) causing a low-impact in $\theta_1 = (-\infty, 0.5)$ dB, a medium-impact in $\theta_2 = [0.5, 2)$ dB, and, finally, a high-impact in $\theta_3 = [2, +\infty)$ dB, $\theta_3 = \theta_c$ being as this last interval collects those cases that surpass the critical threshold $\gamma_c = 2$ dB according to the WG-AEN [14]. Regarding the two WASNs, the number of sensors $N_S$ considered for the subsequent analyses is 19 for the suburban network, and 23 for the urban one, whereas the total number of evaluated segments of 5 min is 1812 in Milan and 1840 in Rome, respectively.

### *4.1. WASN-Based Environmental Databases*

After the deployment of the sensor networks in the urban and suburban pilot areas of the DYNAMAP project, two WASN-based databases were obtained from environmental acoustic data in real-operation conditions. On the one hand, the nodes distribution across the urban area of Milan is based on the clustering of traffic noise profiles in order to place the best sensor locations for different road categories [33]. On the other hand, in the Rome suburban area, the sensor nodes have been

spread along the A90 highway, considering several scenarios of different complexity (single road, crossings, nearby railways and multiple connections) [34,35]. Figure 2 depicts two examples of the sensor placements in both urban and suburban areas, and Appendix A details the sensors' Ids as well as the description of their locations within Tables A1 and A2 for the urban and suburban environments, respectively.



(**a**) Example of the sensor in its location in the urban area of Milan.

(**b**) Example of the sensor in its location in the suburban area of Rome (picture property of ANAS S.p.A.).

**Figure 2.** Examples of the location of the low-cost acoustic sensors in the DYNAMAP's urban and suburban pilot areas.

In both cases, the recorded databases include data from two days with different traffic conditions: one from a weekday (on Tuesday, the 28th of November 2017 for the urban area, and on Tuesday, the 2nd of November 2017 for the suburban environment), and another during the weekend (on Sunday, the 3rd of December 2017 on the urban area, and on Sunday, the 5th of November 2017 in the suburban environment). The audio recordings were collected in continuous raw audio clips from the first 20 min of each hour (considering a sampling frequency of 48 kHz) , as a trade-off between the storage capacity and communications resources of the nodes, and obtaining a representative sub-sampling of the $L_{Aeq}$ measurements along the day [40]. The gathered acoustic data were manually labeled by experts in audio signal processing (see [39,40] for further details). As a result, up to 28 ANE subcategories were identified. Table 1 lists the 16 types of ANEs observed during the manual labeling process in the suburban environment (subcategories being *stru* and *trck* only specifically detected in this scenario), together with the 26 subcategories identified during the annotation of the urban dataset (being *bell*, *blin*, *dog*, *glas*, *peop*, *rubb*, *sqck*, *step*, *tram* and *wrks* those ANE subcategories typically found within this environment). Meteorological-related ANEs like *thun*, *rain* and *wind* cannot be attributed to any specific acoustic environment since they are highly dependent on the weather during the days of the WASN-based data collection. Finally, audio excerpts that contained a mixture of different sound sources (e.g., diverse ANEs together with RTN as background) were labeled as complex sound mixtures or CMPLX. Both CMPLX and ANEs are considered for the subsequent impact-related analyses as both contain undesired acoustic events, after windowing the audio streams into $N_T^i$ frames of length $T$ (see Figure 1).

**Table 1.** Description and % of occurrences of the 28 sound subcategories attributed to anomalous noise events found throughout the manual labeling process of the WASN-based urban and suburban acoustic databases.

| Label | Suburban Counts (%) | Urban Counts (%) | Description |
|---|---|---|---|
| *airp* | 0.1 | 1 | Noise of airplanes and helicopters |
| *alrm* | 0.2 | 0.3 | Sound of an alarm or a vehicle beep moving backwards |
| *bell* | 0 | 1.2 | Church bells |
| *bike* | <0.1 | 3.6 | Sound of bikes and bike chains |
| *bird* | 15.1 | 14.7 | Birdsong |
| *blin* | 0 | <0.1 | Opening and closing of a blind |
| *brak* | 23.1 | 12.7 | Brakes and conveyor belts |
| *busd* | 2.8 | 1.1 | Opening bus door (or tramway), depressurized air |
| *dog* | 0 | 2.5 | Barking of dogs |
| *door* | 2.6 | 14.7 | Closing doors (vehicle or house) |
| *glas* | 0 | 0.1 | Sound of glass crashing |
| *horn* | 6.7 | 3.7 | Horns of vehicles (cars, motorbikes, trucks, etc.) |
| *inte* | 0.3 | 0.2 | Interfering signal from an industry or human machine |
| *musi* | <0.1 | 0.6 | Music in car or in the street |
| *peop* | 0 | 22.2 | Sounds of people chatting, laughing, coughing, sneezing, etc. |
| *rain* | 23.7 | 0.4 | Sound of heavy rain |
| *rubb* | 0 | 0.1 | Rubbish service (engines and grabbing system) |
| *sire* | 1.8 | 0.7 | Sirens (ambulances, police, etc.) |
| *sqck* | 0 | 0.8 | Squeak sound of door hinges |
| *step* | 0 | 13.7 | Sounds of steps |
| *thun* | 7.4 | <0.1 | Thunderstorm |
| *trck* | 11.9 | 0 | Noise when trucks or vehicles with heavy load passed over a bump. |
| *tram* | 0 | 0.7 | Stop, start and passby sounds of tramways |
| *tran* | 2.7 | <0.1 | Sound of trains |
| *trll* | 0 | 1 | Sound of wheels of suitcases (trolley) |
| *stru* | 1.4 | 0 | Noise of highway portals structure caused by vibration of trucks passbys |
| *wind* | 0 | <0.1 | Noise of wind (movement of the leaves of trees,...) |
| *wrks* | 0 | 4.1 | Works in the street (e.g., saws, hammer drills, etc.) |

As a result, the subsequent analyses evaluate 153 h and 20 min of audio data obtained from the 19 sensors placed on the A90 highway portals along the Rome suburban environment, and 151 h obtained from 23 different sensors placed in the building façades of several public buildings across the District 9 of Milan, after discarding node hb114 due to technical problems during the data recording process, but keeping sensor hb119 despite missing some data from the Sunday recordings to 75%$N_T^i$.

Table 2 summarizes the general characteristics of both analyzed datasets. As can be observed, RTN is the majority class in both cases, as identified 83.7% of the time in the urban environments, while this value raised to 96.5% in the suburban scenario. Accordingly, ANEs were more frequently observed in the urban than in the suburban dataset, being more than four times detected in this environment compared to the suburban one (8.7% of ANE in urban while 1.9% of ANE in suburban). It should be also noticed that the increase of ANE occurrences in the urban environment also fostered the presence of highly complex audio passages.

**Table 2.** General characteristics of the WASN-based urban and suburban acoustic databases evaluated considering the impact analysis methodology.

| Acoustic Environment | Total Duration | RTN (%) | ANE (%) | CMPLX (%) |
|---|---|---|---|---|
| Milan (Urban) | 151 h | 83.7% | 8.7% | 7.6% |
| Rome (Suburban) | 153 h 20 min | 96.5% | 1.9% | 1.6% |

### *4.2. Individual Impact of ANEs*

To understand the relevance of the events, first, a study of the individual ANE impact is conducted following the aforementioned impact analysis methodology. As an overall analysis, Table 3 details the number of occurrences and sensor activation ratios for each environment and recording day.

**Table 3.** Number of occurrences and sensor activation ratios per sensor for low, medium and high individual impact ranges.

| **Individual Impacts** | | **Low Impact** $(-\infty, 0.5)$ **dB** | | **Medium Impact** $[0.5, 2)$ **dB** | | **High Impact** $[2, +\infty)$ **dB** | |
|---|---|---|---|---|---|---|---|
| | | **Occurrences Count (%)** | **Activation Count/$N_S$** | **Occurrences Count (%)** | **Activation Count/$N_S$** | **Occurrences Count (%)** | **Activation Count/$N_S$** |
| **Milan** | Tuesday | 21,264 (99.5%) | 23/23 | 76 (0.4%) | 21/23 | 28 (0.1%) | 16/23 |
| | Sunday | 15,215 (99.4%) | 23/23 | 58 (0.4%) | 20/23 | 29 (0.2%) | 16/23 |
| **Rome** | Tuesday | 2105 (98.1%) | 19/19 | 33 (1.6%) | 13/19 | 7 (0.3%) | 5/19 |
| | Sunday | 3415 (99.0%) | 19/19 | 31 (0.9%) | 11/19 | 5 (0.1%) | 3/19 |

As can be observed, the presence of anomalous noise events is common in both environments, particularly in Milan which records 10 times more ANEs on Tuesday and 4 times more on Sunday than Rome. Specifically, all recording days have yielded a high percentage of low-impact ANEs, but in Milan, particularly, the presence of low-impact events in relation to the other impact ranges, is higher than in Rome, rising from 98.1 to 99.5% on Tuesday, and from 99.0 to 99.4% on Sunday. In Rome, however, the percentage of medium-impact events is higher than in Milan on both days, with a total of 134 ANEs in Milan and 64 in Rome, respectively. This implies that the sensors in Milan can detect this kind of event in almost all sensors, while only 60% of the sensors in Rome can detect these ANEs. Finally, concerning high-impact events, the percentage of occurrences is similar in both locations, despite Milan has 57 high-impact events detected in 16 sensors and Rome only 12, which activate few sensors.

In Figure 3, the corresponding impact histogram matrices for individual ANEs are detailed for each sensor location according to the three impact range intervals (low, medium and high). Notice that the number of occurrences in the low-impact intervals is depicted separately from the medium and high-impact intervals for illustration purposes, as it is more than two orders of magnitude larger.

It can be observed that the maximum number of low-impact ANEs has been found in sensor hb123 of Milan on Tuesday, with 2374 occurrences. In contrast, the maximum number of low-impact events in Rome is 379 for sensor hb143 on Sunday. Concerning the medium-impact events in Milan, the first day accounts for the highest number of events, coming from hb139, which obtains the maximum number of medium-impact ANEs, with 9 occurrences, also presents a significant number in Sunday, with 6 events. In the rest of the cases in Milan, no clear pattern is observed relating both recording days. In Rome, however, sensor hb104 attributes for the maximum number of medium-impact events, with 18 occurrences on Tuesday and 17 on Sunday. This is a particularly relevant case in the suburban area as the second closest sensor is hb134 with only 3 medium-impact events on Sunday. When looking at the column depicting high impact ANEs, it can be observed that a maximum of 5 events were captured on Sunday in sensor hb133 of Milan, while also a significant presence on Tuesday with 4 occurrences. In Rome, sensor hb104 accounts for the highest number of high-impact ANEs on Tuesday, with 3 events, which also recorded one of the highest number of occurrences on Sunday, with 2 events.

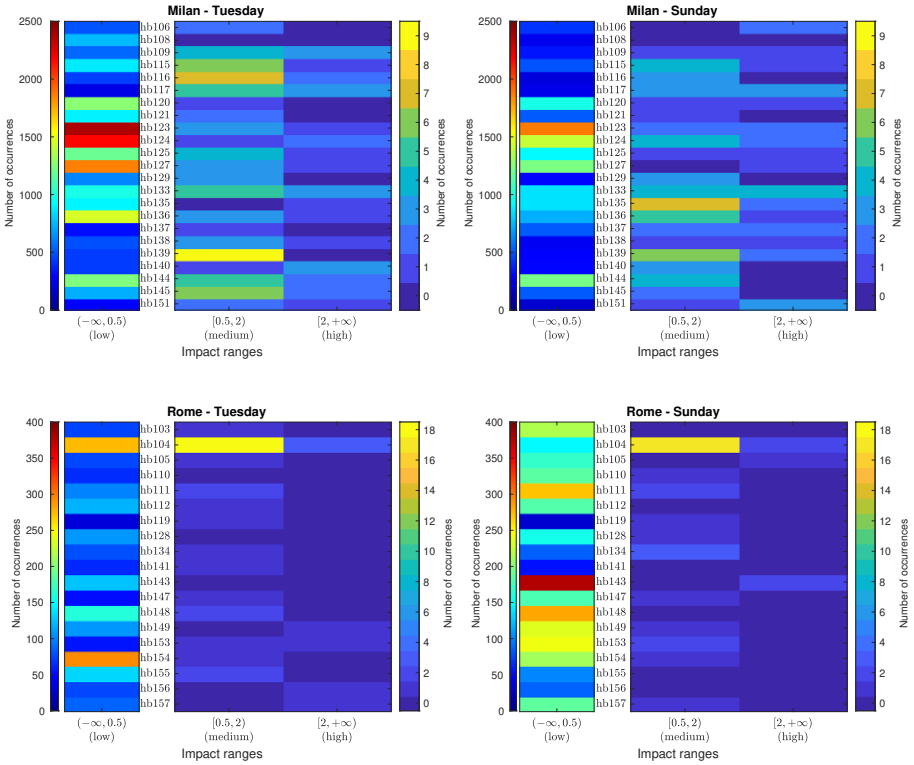



**Figure 3.** Individual impact histogram matrices (obtained using integration time $T = 300$ s) categorized in three impact ranges (low, medium and high) for the urban (Milan) and suburban (Rome) environments obtained from a weekday (Tuesday) and weekend day (Sunday).

### *4.3. Aggregate Impact of ANEs*

This section details the results obtained from the analysis of the labeled data in order to find to what extent the presence of several ANEs with low and medium individual impacts within the same integration period can bias the $L_{Aeq,300s}$ computation.

First, Table 4 shows the number of occurrences and sensors activation ratios of the AI for environment and recording day. As it can be observed from the table, the overall presence of occurrences and activation ratios are similar for both days within each location. However, when comparing Milan with Rome, the distribution of the impact ranges differs. In the case of Milan, near 85% of the AIs entail a low impact on the $L_{Aeq300s}$. This percentage increases to almost 96% in Rome. For this reason, the presence, as well the sensor activation, of medium and high-level AIs in Rome is lower than in Milan. In Milan, only one sensor on Tuesday and two on Sunday fail to detect a medium-impact AI. However, in Rome, on Tuesday 7 sensors were not capable of detecting any event and on Sunday the number was 6. In the particular case of high-impact aggregates, their presence is reduced from near 4% in Milan to less than 1% in Rome. Most Milan sensors activate (18 on Tuesday and 17 on Sunday), but only 5 and 3 sensors detect ANEs of this category in Rome in the weekday and during the weekend, respectively.

Following the same analysis scheme described in the previous section, Figure 4 depicts the AI histogram matrices showing the number of occurrences of aggregate ANEs for each impact range and sensor location for both pilot areas. Again, the number of occurrences in the low-impact range is

separated from the rest of occurrences for illustration purposes, due to the same reason indicated in the previous analysis. As can be observed, in Milan, low-impact AIs range from 28 to 43 on Tuesday, and from 24 to 36 on Sunday. A total of 107 intervals on the first day and 88 in the second day contain a medium-impact AI, highlighting sensor hb115 in Milan, with 11 occurrences on Tuesday and hb124 with 10 occurrences on Sunday. However, high-impact AIs record a lower presence of occurrences, with a highest value of 4 in sensors hb109 and hb140 on Tuesday, and in sensor hb133 on Sunday.

**Table 4.** Number of occurrences and sensor activation ratios per sensor for low, medium and high aggregate impact ranges.

| **Aggregate Impacts** | | **Low Impact** $(-\infty, 0.5)$ **dB** | | **Medium Impact** $[0.5, 2)$ **dB** | | **High Impact** $[2, +\infty)$ **dB** | |
|---|---|---|---|---|---|---|---|
| | | **Occurrences Count (%)** | **Activation Count/$N_S$** | **Occurrences Count (%)** | **Activation Count/$N_S$** | **Occurrences Count (%)** | **Activation Count/$N_S$** |
| **Milan** | Tuesday | 855 (85.5%) | 23/23 | 107 (10.7%) | 22/23 | 38 (3.8%) | 18/23 |
| | Sunday | 693 (85.4%) | 23/23 | 88 (10.8%) | 21/23 | 31 (3.8%) | 17/23 |
| **Rome** | Tuesday | 874 (95.8%) | 19/19 | 29 (3.2%) | 12/19 | 9 (1.0%) | 5/19 |
| | Sunday | 887 (95.6%) | 19/19 | 35 (3.8%) | 13/19 | 6 (0.6%) | 3/19 |

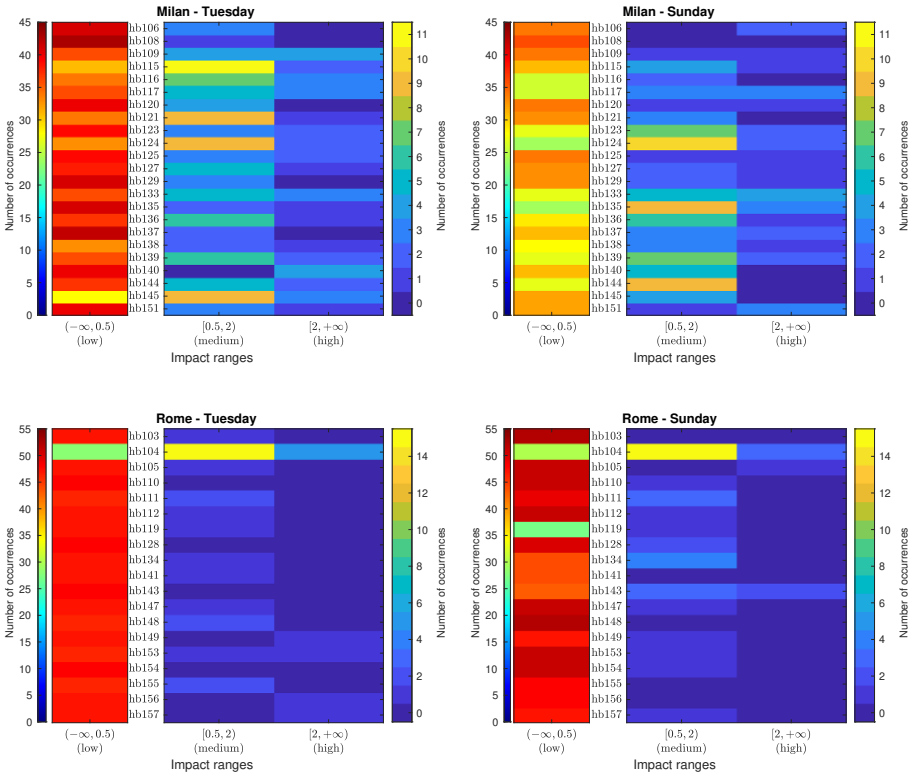



**Figure 4.** Aggregate impact histogram matrices (obtained using integration time $T = 300$ s) categorized in three impact ranges (low, medium and high) for the urban (Milan) and suburban (Rome) environments obtained from a weekday (Tuesday) and weekend day (Sunday).

Regarding the pilot area in Rome, the presence of low-impact AIs is clearly dominant. However, it is worth mentioning that sensor hb104 presents a completely different pattern, with 15 medium-impact AIs on Tuesday and Sunday. This reduces significantly the low-impact occurrences in that sensor in comparison to other nodes. Finally, as aforementioned, it is to note that sensor hb119 failed in recording several hours of Sunday.

### *4.4. Critical Aggregate Impacts Per Level*

In this section, the occurrences that surpass the critical threshold $\gamma_c = 2$ dB, are analyzed in detail. First, the individual ANEs that bias the $L_{Aeq,300s}$ beyond threshold $\gamma_c$ by themselves belong to the high-impact range. To analyze their distribution in detail, the critical individual ANEs observed in Section 4.2 (see Figure 3) are divided in 2-dB spans for each sensor in Figure 5. When analyzing this kind of anomalous noise events, Milan credits for most of the high-impact individual ANEs, most of them within the range of 2 to 4 dB, without belittling their presence in the other ranges for both days. Concerning Rome, sensor hb104 is the one that recorded the largest number of high-impact events, most of them belonging to the $[2, 4)$ dB range. Finally, it is to note that 10 events surpass the 10-dB impact range are *sirens*, being the event with the highest impact a 3-min siren with 29.4 dB of impact, recorded in sensor hb137 on Sunday. In contrast, no events surpassing the 10-dB threshold are present in Rome.

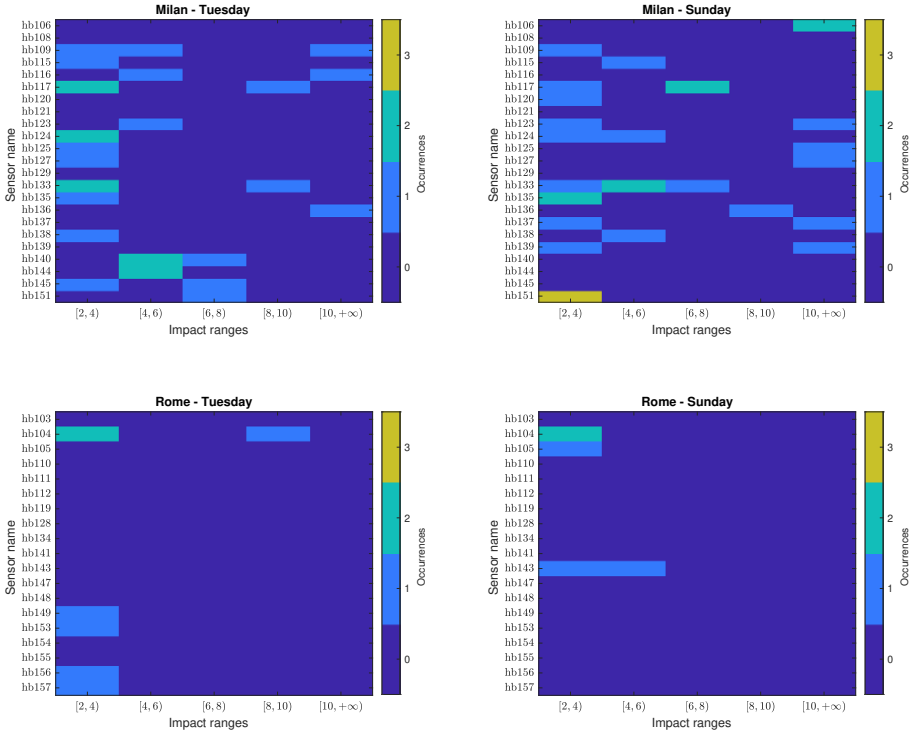



**Figure 5.** Critical AI histogram matrices ($\mathbf{H_c}$) of individual ANEs for the urban (Milan) and suburban (Rome) environments obtained from a weekday (Tuesday) and weekend day (Sunday).

On the other hand, in order to evaluate if the presence of several ANEs may contribute to the surpassing of the $\gamma_c$ threshold, Figure 6 shows the critical AI histogram matrices $\mathbf{H_c}$ obtained for each network for different impact intervals. That is to say, it depicts the number of times the AI of

ANEs contribute to bias the $L_{Aeq,300s}$ of RTN critically for both pilot areas and recording day according to the type of impact range. To that effect, besides considering $\theta_1$ (low), $\theta_2$ (medium) and $\theta_3$ (high) impact intervals to analyze the critical aggregate impacts, two more intervals are considered: $\theta_1 \bigcup \theta_2$ to account for co-occurring low and medium individual impact ANEs, and $\theta_1 \bigcup \theta_2 \bigcup \theta_3$ to quantify all the critical cases, disregarding the type of the ANE's individual impact.

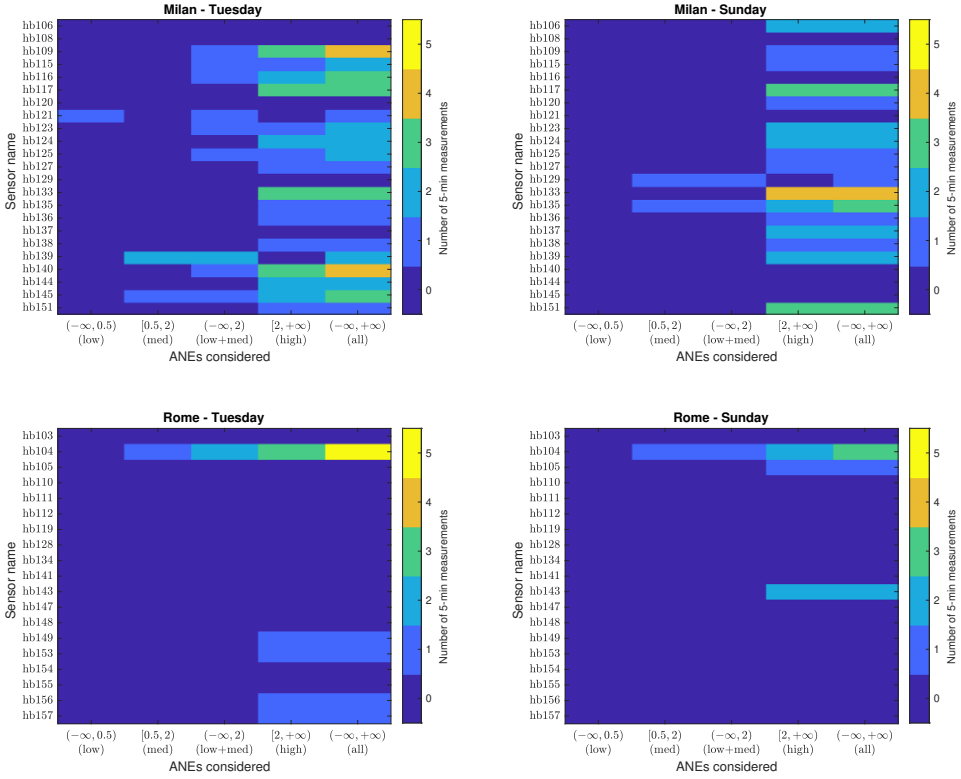



**Figure 6.** Critical AI histogram matrices ($\mathbf{H_c}$) categorized in the defined impact ranges for the urban (Milan) and suburban (Rome) environments obtained from a weekday (Tuesday) and weekend day (Sunday).

The first column of each $\mathbf{H_c}$ matrices depicted in Figure 6 shows those low-impact AIs causing a critical impact. It can be observed that there is one case accounting for a deviation of the AI higher than 2 dB for a particular period of time $t$ of 5 minutes in sensor hb121 installed in Milan. It is due to 13 *wrks* sounds recorded on Tuesday ranging from 0.01 dB to 0.4 dB, i.e., all of them belong to the individual low-impact range $\theta_1$, but due to their co-occurrence within the same period of time their AI becomes critical.

Likewise, the second column plots critical medium-impact AIs. In Milan, the threshold $\gamma_c$ is surpassed three times on Tuesday and twice on Sunday, whereas in Rome, purely medium-impact occurrences cause a critical AI once each day. Specifically, sensor hb139 collected two of these pieces of evidence on Tuesday. In the first case, the two most significant ANEs are *horns*, with individual impacts of 0.8 and 1.2 dB, respectively (the third one is a *dog* bark with an impact of 0.03 dB). The second is composed of a *horn* of 1.3 dB and two CMPLX sounds, consisting on a mix of RTN and an undetermined beep noise of 0.8 and 0.5 dB. Moreover, sensor hb145 also recorded a period in which individual ANEs

bias the $L_{Aeq,300s}$ critically on Tuesday, where the most important event is a *tram* passby of 1.5 dB and the second one is a 1.6-dB CMPLX event consisting of a mix of a *tram* passby and *birds* tweeting near the sensor. On Sunday, two of the periods recorded in the urban environment contain a combination of medium-impact ANEs that surpass the threshold: one in sensor hb129, composed of two distant *sirens* mixed with other sounds, and another due two CMPLX sounds in sensor hb135, containing unidentified mechanical sounds. In what concerns Rome, sensor hb104 presents critical impact evidence due the co-occurrence of purely medium-impact events for both week and weekend periods. On Tuesday two *train* passbys of 1.3 and 1.1 dB bias the $L_{Aeq,300s}$ more than 2 dB. On Sunday, the critical bias is caused by the presence of two *horns* of 1.2 and 1.9 dB, respectively.

The third column of the four AI critical matrices of Figure 6 show the number of times $\gamma_c$ is surpassed for ANEs when considering low and medium-impact ANEs, i.e., it collects the occurrences of aggregate low-impact ANEs from $\theta_1$ and the aggregate medium-impact ANEs from $\theta_2$, as well as the the number of times that the critical threshold is surpassed as a result of the combination of the medium- and low-impact events. This last case is only observed during the weekday 6 times in Milan and once at sensor hb104 in Rome. The latter happens on Tuesday and it consists of the sum of several *train* passbys, with the most salient event an impact of 1.9 dB and the other ones of about 0.1 dB. The six cases in Milan have all been found on Tuesday in different sensors: in hb109, three CMPLX sounds have been found that consist of *train* passbys mixed with RTN of 1.8, 0.2 and 0.2 dB; in hb115, a sum of 13 *wrks* sounds with impacts from 0.01 dB to 0.9 dB; in hb116, a 1.9-dB siren co-occurring with a 0.4-dB CMPLX sound of *birds* mixed with RTN; in hb123, an *airp* of 1.9 dB and other *peop* and *brak*-related sound with impacts smaller than 0.02 dB; in hb125, all significant events are *dog* barks, with impacts of 0.9, 0.6, 0.4, 0.3 dB and decreasing; and in hb140, a *siren* of 1.9 dB has been found, jointly with *people*-related sounds of 0.2 dB.

The next column of critical AI matrices presents high-impact ANEs. For the data at hand, the aggregate high-impact occurrences coincide with the number of individual high-impact events depicted in Figure 3 (see also Table 3, where the number of occurrences in this level is quantified).

Finally, the last column of matrices $\mathbf{H_c}$ shows the critical AI histogram caused by the co-occurrence of ANEs of any individual impact range altogether. If we focus on the last three columns of Figure 6, it can be appreciated that in all cases, the sum of the low and medium-impact ANEs with the high-impact ANEs results in the total number of times the 2 dB threshold is surpassed. This result could have differed in the case that aggregate low and medium ANEs co-occurred with high-impact ANEs. Therefore, Figure 6 clarifies the fact that high-impact events have not co-occurred at the same 5-min interval for the datasets at hand, besides showing there is no situation in our datasets where low and medium impact aggregated surpass $\gamma_c$ at the same 5-min slot $t$ in which a high-impact ANE occurs.

To summarize, in Milan, the threshold has been surpassed due to low and medium aggregate impacts in 12 of the 69 critical cases, which correspond to 17% of cases. Likewise, in Rome, the ratio is 3 to 15, corresponding to 20% of the critical cases. Therefore, according to these results, it can be stated that the removal of low and medium-impact ANEs becomes as relevant as high-impact events in order to preserve the accuracy of the RTN level measurements in both urban and suburban environments.

## 5. Discussion

This section discusses several relevant aspects related to the results obtained after applying the impact analysis methodology to the two WASN-based datasets collected from the urban and suburban areas. First of all, it is to note that the individual analysis of the impact of each ANE of those co-occurring within the same integration period has been conducted as a baseline study, since the individual view of the impact of acoustic events unrelated to traffic noise is a straightforward but unrealistic approach to the problem at hand. However, this study has been useful to set the basis for the subsequent aggregate analyses. In this sense, it is worth noting that although the datasets have been collected during specific time periods, the analyzed data show the regular presence of anomalous events across all the days and locations in a real-operation context. Specifically, the number of ANEs

found in the urban area is seven times greater than in the suburban environment on average (this ratio being ten times on the weekday). In the suburban environment, the weekday pattern is very similar to what is observed in during the weekend, although a larger number of events have been recorded during the weekend, which should be studied in the future with more detail.

In terms of the acoustic categories, it is worth mentioning that 7.7% of the urban WASN-based dataset and 1.6% of the suburban one has been annotated as CMPLX. As aforementioned, the CMPLX acoustic category can be either caused by a mix of RTN and ANEs or by unidentified ANEs by the experts. The conducted analyses have shown that these kinds of acoustic events can also have a significant impact on the $L_{Aeq,300s}$ computation, showing a similar presence in both datasets as the corresponding ANE acoustic category. Therefore, as well as ANEs, CMPLX audio passages should also be removed from the computation of road traffic noise levels to tailor reliable RTN maps.

When comparing the individual and aggregate impact occurrences for low, medium and high-impact ranges, the analyzed environments present a different distribution. In the case of the urban area, a larger number of low-impact events have been recorded than in the suburban environment. However, as far as AIs are concerned, the percentage of low-impact pieces of evidence are lower in the former than in the latter. In addition, medium and high-impact aggregate ANEs have a significant presence in the urban environment, being near the 15% of occurrences; however, in the suburban area, this value decreases to 5%, probably because also the high-impact ANEs present a lower number of instances. From these results it can be concluded that the detection and removal of ANEs will be more usual in a urban than in a suburban environment, since a significantly higher number of $L_{Aeq,300s}$ values can be biased critically. Furthermore, it is worth mentioning that the number of individual high-impact ANEs may not always coincide with the number of times these events bias the 2 dB threshold. This is because it could happen that two or more high-impact events co-occurred in the same evaluated period of time. However, as shown in the results of this work, this is not the case for the data at hand, thus, all high-impact ANEs occur in different integration times.

The impact patterns observed on both environments present different trends. From the analysis conducted in the suburban area, it was observed that sensor hb104 presents a clearly different pattern of the impact of ANEs compared to the rest of the nodes of that WASN for both week and weekend days. This sensor was installed on a major road with two lanes in each direction with a crossing highway under the bridge (see Table A2), which makes this location substantially different from the other sensors locations in Rome (as they do not correspond to major crossroads). For this sensor, the aggregate ANEs are more likely to bias the $L_{Aeq,300s}$, as a 40% of the analyzed measurements contain a medium or high aggregate impact considering both days. This result leads to the preliminary conclusion that in a suburban area, a crossroad is more susceptible to collect anomalous noise events that may distort the RTN level measurements critically. On the contrary, the data analyzed from the other sensor locations in Rome show that the AIs of the ANEs do not usually have a significant impact on the A-weighted equivalent RTN level measurements. In Milan, however, it becomes difficult to identify specific impact patterns according to the sensor locations due to the great variability of occurrences observed from the recordings of both week and weekend days. Nevertheless, note that all sensors have recorded ANEs with a significant impact—both evaluated individually and in an aggregate manner—being relevant enough to bias the RTN map representation in certain periods of time. Given the fact that the recordings were taken over two days, a relevant number of $L_{Aeq,300s}$ measurements could have been computed with an inaccuracy of more than 2 dB, we can conclude that is necessary to remove all kind of anomalous noise events from the final computation of the noise map.

Briefly, the results drawn from this work present a non-negligible number of anomalous noise events that occur randomly both in the DYNAMAP's pilot urban and suburban acoustic environments. This is a relevant issue, as we have to mention that the analyzed data correspond only to a recording campaign of two different days, which provide a relevant but limited scope of *all* the possible issues that may occur in all streets and ring road portals during any day of the year at any time. Nevertheless, although the amount of evidence observed in the gathered data may result statistically poor (i.e., only

84 critical pieces of evidence have been observed), their mere presence demonstrates the importance of their automatic removal to obtain reliable dynamic RTN maps through WASN-based approaches. That is, if the sub-sampling done in two days for several 20-min long audio files has led us to this conclusion, what will be the real impact on the measurements in a 24-h $\times$ 7-day WASN-based monitoring system? How many works around the city and the highway can occur throughout the year together some horns and sirens? How many sensors can be located close to a school (with the children in the playground) or next to a church with its bells?... This opens a much wider research goal, focused on the detailed analysis of the sensors location and the consequences it entails in terms of anomalous noise events detection and removal, as the election of the sensor's installation place is usually based on spatial coverage to draw the acoustic map, being also limited by the actual location of the portals and public buildings where the sensors are finally installed.

## 6. Conclusions

In this work, we have analyzed more than 300 h of labeled acoustic data collected through two WASNs after being deployed in the pilot urban and suburban areas of the DYNAMAP project. The study shows that ANEs can be widely found in acoustic environments when monitoring RTN levels in real-operation conditions, being particularly common in the data gathered from the urban area. Moreover, through the impact analysis methodology, it has been also concluded that the aggregate contribution of low and medium-impact ANEs can deviate the $L_{Aeq,300s}$ as critically as high-impact individual ANEs. Therefore, the obtained results highlight the importance of the automatic removal of low, medium and high-impact events to obtain reliable WASN-based RTN maps in real-operation environments.

Future work will be focused on the detailed analysis of the particularities of each acoustic environment and ANEs subcategories together with complex passages, not only to consider their global impact patterns in the urban and suburban, but also to study the spatio-temporal particularities of all the locations and periods of time. Finally, we plan to adapt the preliminary version of the ANED algorithm by using the two WASN-based datasets to improve its performance in both urban and suburban environments in real operations.

**Author Contributions:** F.A. has written most parts of the paper, besides participating in the design and formulation of the analysis methodology and in the study of the results. F.O. has worked and led the analysis of the data and participated in the writing of the results and discussion sections. R.M.A.-P. has written the related work, besides contributing to the discussion and reviewing the entire paper. J.C.S. has worked in the design and formulation of the analysis methodology, and its writing, besides summarizing the main characteristics of the WASN-based acoustic datasets. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** The authors would like to thank ANAS S.p.A. for the picture of the sensor installed in the portal. The research presented in this work has been partially supported by the LIFE DYNAMAP project (LIFE13 ENV/IT/001254). Joan Claudi Socoró thanks the Obra Social La Caixa for grant ref. 2019-URL-IR1rQ-053. Ferran Orga thanks the support of the European Social Fund and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Catalan Government for the pre-doctoral FI grant No. 2019_FI_B2_00168.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Aggregate Impact |
| ANE | Anomalous Noise Event |
| ANED | Anomalous Noise Event Detection |
| CNOSSOS-EU | Common Noise Assessment Methods in Europe |
| DYNAMAP | Dynamic Noise Mapping |
| END | European Noise Directive |
| EU | European Union |

| | |
|---|---|
| RTN | Road Traffic Noise |
| SNR | Signal-to-Noise Ratio |
| SONYC | Sounds of New York City |
| WASN | Wireless Acoustic Sensor Network |
| WG-AEN | European Commission Working Group Assessment of Exposure to Noise |

## Appendix A

This section includes the description of the sensor locations for both urban and suburban environments by means of Tables A1 and A2, respectively.

**Table A1.** Sensor locations description for the urban environment. X-lane/Y-lane road stands for a two-way road that has X lanes in one sense and Y lanes in the opposite sense. X-lane road stands for a street with X lanes in the same sense.

| Sensor Id | Sensor Location Description |
|---|---|
| **hb106** | 1-lane/1-lane road with connection with 1 line road, area with parks nearby, no shops |
| **hb108** | 1-lane/1-lane road, in front University exit, no shops |
| **hb109** | 3-lane/3-lane road, near crossing with tramway and 1 line+2 line/2 line+1 line road, shopping and coffe/restaurant area |
| **hb115** | 1-lane road with shopping in front |
| **hb116** | 1-lane/1-lane road with connection with 1-lane road, residential area |
| **hb117** | 3-lane/3-lane road, near school, area with parks nearby, no shops |
| **hb120** | 1-lane/1-lane road, residential area, no shops |
| **hb121** | 2-lane/2-lane road, connection with 1-lane road, University area, no shops |
| **hb123** | 2-lane/2-lane road with hotel and traffic light nearby |
| **hb124** | 1-lane road, no shops |
| **hb125** | 1-lane road with connection with 1-lane/1-lane road, mix of residential with some shops |
| **hb127** | 1-lane road near bifurcation with 1 line road, some shop nearby |
| **hb129** | 1-lane/1-lane road, bike line, connection with 1-lane road, some shop |
| **hb133** | 1-lane road, residential area, no shops, little park area in front |
| **hb135** | 1-lane road with connection with 1-lane road (low speed), near University campus (students), no shops, in front of park area |
| **hb136** | 1-lane/1-lane road with connection with 1-lane road, area with parks nearby, no shops |
| **hb137** | 1-lane road with connection with 1 line road, in front of park, residential area, no shops |
| **hb138** | 1-lane road near connection with other 1-lane road, no shops |
| **hb139** | 1-lane road, residential area, some shop/enterprise |
| **hb140** | 2-lane/2-lane road with parking area and traffic light with crossing nearby, no shops near and high traffic |
| **hb144** | 1-lane road in residential area, one shop far away |
| **hb145** | 1-lane road, in front of park |
| **hb151** | 1-lane/1-lane road, bike line, some shop and restaurant |

**Table A2.** Sensor locations description for suburban environment. X-lane/Y-lane road stands for a two-way road that has X lanes in one sense and Y lanes in the opposite sense. X-lane road stands for a street with X lanes in the same sense.

| Sensor Id | Sensor Location Description |
|---|---|
| **hb103** | Highway with 3-lane/3-lane |
| **hb104** | Major road with 2-lane each direction crossing a highway under bridge (out of major ring) |
| **hb105** | Highway with 4-lane (only 1 direction, and near exits/crossings) |
| **hb110** | Highway with 3-lane/3-lane |

**Table A2.** *Cont.*

| Sensor Id | Sensor Location Description |
|---|---|
| **hb111** | Highway with 3-lane/3-lane |
| **hb112** | Highway with 3-lane/3-lane (near exit and near crossings) |
| **hb119** | Highway with 3-lane/3-lane |
| **hb128** | Highway with 3-lane/3-lane |
| **hb134** | Highway with 4-lane/4-lane (near bridge and crossings) |
| **hb141** | Highway with 5-lane/5-lane (near crossings) |
| **hb143** | Highway with 2-lane/2-lane (out of major ring) |
| **hb147** | Highway with 3-lane/3-lane |
| **hb148** | Highway with 3-lane/3-lane |
| **hb149** | Highway with 3-lane (near tunnel) |
| **hb153** | Major road with 2-lane each direction crossing a highway under bridge (out of major ring) |
| **hb154** | Highway with 4-lane/4-lane |
| **hb155** | Highway with 2-lane (near connection but out major ring) plus 1 road same sense next to |
| **hb156** | Highway with 3-lane/3-lane |
| **hb157** | Highway with 5-lane/5-lane |

## References


1. Directive, E.U. Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise. *Off. J. Eur. Communities* **2002**, *L 189/12*.
2. World Health Organization Regional Office for Europe. Burden of Disease from Environmental Noise: Quantification of Healthy Life Years Lost in Europe 2011. Available online: https://www.who.int/quantifying_ehimpacts/publications/e94888/en (accessed on 21 January 2020).
3. World Health Organization Regional Office for Europe. Environmental Noise Guidelines for the European Region 2018. Available online: https://apps.who.int/iris/handle/10665/279952 (accessed on 21 January 2020).
4. Licitra, G.; Ascari, E.; Fredianelli, L. Prioritizing Process in Action Plans: a Review of Approaches. *Curr. Pollut. Rep.* **2017**, *3*, 151–161. [CrossRef]
5. Kephalopoulos, S.; Paviotti, M.; Anfosso-Lédée, F. *Common Noise Assessment Methods in Europe (CNOSSOS-EU)*; Report EUR 25379 EN; Publications Office of the European Union: Luxembourg, 2002; pp. 1–180.
6. King, E.A.; Murphy, E.; Rice, H.J. Implementation of the EU environmental noise directive: Lessons from the first phase of strategic noise mapping and action planning in Ireland. *J. Environ. Manag.* **2011**, *92*, 756–764. [CrossRef] [PubMed]
7. Carrier, M.; Apparicio, P.; Séguin, A.M.; Crouse, D. School locations and road transportation nuisances in Montreal: An environmental equity diagnosis. *Transp. Policy* **2019**, *81*, 302–310. [CrossRef]
8. Tsai, K.T.; Lin, M.D.; Lin, Y.H. Noise exposure assessment and prevention around high-speed rail. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 4833–4842. [CrossRef]
9. Morley, D.; de Hoogh, K.; Fecht, D.; Fabbri, F.; Bell, M.; Goodman, P.; Elliott, P.; Hodgson, S.; Hansell, A.; Gulliver, J. International scale implementation of the CNOSSOS-EU road traffic noise prediction model for epidemiological studies. *Environ. Pollut.* **2015**, *206*, 332–341. [CrossRef] [PubMed]
10. Botteldooren, D.; Dekoninck, L.; Gillis, D. The influence of traffic noise on appreciation of the living quality of a neighborhood. *Int. J. Environ. Res. Public Health* **2011**, *8*, 777–798. [CrossRef] [PubMed]
11. Alberts, W.; Roebben, M. Road Traffic Noise Exposure in Europe in 2012 based on END data. In Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (INTER-NOISE 2016), Hamburg, Germany, 21–24 August 2016; pp. 1236–1247.
12. Garcia, A.; Faus, L. Statistical analysis of noise levels in urban areas. *Appl. Acoust.* **1991**, *34*, 227–247. [CrossRef]
13. Alsina-Pagès, R.M.; Alías, F.; Socoró, J.C.; Orga, F.; Benocci, R.; Zambon, G. Anomalous events removal for automated traffic noise maps generation. *Appl. Acoust.* **2019**, *151*, 183–192. [CrossRef]
14. European Commission Working Group Assessment of Exposure to Noise (WG-AEN). *Good Practice Guide for Strategic Noise Mapping and the Production of Associated Data on Noise Exposure Ver.2, 13 August*; Technical Report; 2007. Available online: https://www.lfu.bayern.de/laerm/eg_umgebungslaermrichtlinie/doc/good_practice_guide_2007.pdf (accessed on 21 January 2020).

15. De Coensel, B.; Botteldooren, D. A model of saliency-based auditory attention to environmental sound. In Proceedings of the 20th International Congress on Acoustics (ICA-2010), Sydney, Australia, 23–27 August 2010; pp. 1–8.
16. Salamon, J.; Jacoby, C.; Bello, J.P. A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 1041–1044.
17. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Reliable detection of audio events in highly noisy environments. *Pattern Recognit. Lett.* **2015**, *65*, 22–28. [CrossRef]
18. Stowell, D.; Giannoulis, D.; Benetos, E.; Lagrange, M.; Plumbley, M.D. Detection and Classification of Acoustic Scenes and Events. *IEEE Trans. Multimedia* **2015**, *17*, 1733–1746. [CrossRef]
19. Socoró, J.C.; Ribera, G.; Sevillano, X.; Alías, F. Development of an Anomalous Noise Event Detection Algorithm for dynamic road traffic noise mapping. In Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy, 12–16 July 2015; pp. 1–8.
20. Salamon, J.; MacConnell, D.; Cartwright, M.; Li, P.; Bello, J.P. Scaper: A library for soundscape synthesis and augmentation. In Proceedings of the 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 15–18 October 2017; pp. 344–348.
21. Alías, F.; Socoró, J.C. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Appl. Sci.* **2017**, *7*, 146. [CrossRef]
22. Botteldooren, D.; De Coensel, B.; Oldoni, D.; Van Renterghem, T.; Dauwe, S. Sound monitoring networks new style. In Proceedings of Acoustics 2011: Breaking New Ground – Proceedings of the Annual Conference of the Australian Acoustical Society, Gold Coast, Australia, 2–4 November 2011.
23. Manvell, D. Utilising the Strengths of Different Sound Sensor Networks in Smart City Noise Management. In Proceedings of the EuroNoise 2015, EAA-NAG-ABAV, Maastrich, Netherlands, 31 May–3 June 2015; pp. 2305–2308.
24. Alías, F.; Alsina-Pagès, R.M. Review of Wireless Acoustic Sensor Networks for Environmental Noise Monitoring in Smart Cities. *J. Sens.* **2019**, *2019*, 13. [CrossRef]
25. Camps, J. Barcelona noise monitoring network. In Proceedings of the EuroNoise 2015, EAA-NAG-ABAV, Maastrich, The Netherlands, 1 May–3 June 2015; pp. 2315–2320.
26. Segura-Garcia, J.; Pérez Solano, J.J.; Cobos Serrano, M.; Navarro Camba, E.A.; Felici-Castell, S.; Soriano Asensi, A.; Montes Suay, F. Spatial Statistical Analysis of Urban Noise Data from a WASN Gathered by an IoT System: Application to a Small City. *Appl. Sci.* **2016**, *6*, 380. [CrossRef]
27. Nencini, L.; De Rosa, P.; Ascari, E.; Vinci, B.; Alexeeva, N. SENSEable Pisa: A wireless sensor network for real-time noise mapping. In Proceedings of the EuroNoise 2012, Czech Acoustical Society, Prague, Czech Republic, 10–13 June 2012.
28. Bartalucci, C.; Borchi, F.; Carfagni, M.; Furferi, R.; Governi, L. Design of a prototype of a smart noise monitoring system. In Proceedings of the 24th International Congress on Sound and Vibration (ICSV24); The International Institute of Acoustics and Vibration, London, UK, 23–27 July 2017.
29. Rainham, D. A wireless sensor network for urban environmental health monitoring: UrbanSense. *IOP Conf. Ser. Earth Environ. Sci.* **2016**, *34*, 012028. [CrossRef]
30. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189. [CrossRef]
31. Bello, J.P.; Silva, C.; Nov, O.; Dubois, R.L.; Arora, A.; Salamon, J.; Mydlarz, C.; Doraiswamy, H. SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution. *Commun. ACM* **2019**, *62*, 68–77. [CrossRef]
32. Zambon, G.; Benocci, R.; Bisceglie, A.; Roman, H.E.; Bellucci, P. The LIFE DYNAMAP project: Towards a procedure for dynamic noise mapping in urban areas. *Appl. Acoust.* **2017**, *124*, 52–60. [CrossRef]
33. Zambon, G.; Benocci, R.; Bisceglie, A. Development of optimized algorithms for the classification of networks of road stretches into homogeneous clusters in urban areas. In Proceedings of the 22nd International Congress on Sound and Vibration (ICSV22). International Institute of Acoustics and Vibrations, Florence, Italy, 12–16 July 2015; pp. 1–8.
34. Bellucci, P.; Peruzzi, L.; Zambon, G. LIFE DYNAMAP project: The case study of Rome. *Appl. Acoust.* **2017**, *117*, 193–206. [CrossRef]

35. Benocci, R.; Bellucci, P.; Peruzzi, L.; Bisceglie, A.; Angelini, F.; Confalonieri, C.; Zambon, G. Dynamic Noise Mapping in the Suburban Area of Rome (Italy). *Environments* **2019**, *6*, 79. [CrossRef]
36. Socoró, J.C.; Alías, F.; Alsina-Pagès, R.M. An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments. *Sensors* **2017**, *17*, 2323. [CrossRef] [PubMed]
37. Orga, F.; Alías, F.; Alsina-Pagès, R.M. On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments. *Int. J. Environ. Res. Public Health* **2017**, *15*, 13. [CrossRef] [PubMed]
38. Socoró, J.C.; Alsina-Pagès, R.M.; Alías, F.; Orga, F. Adapting an Anomalous Noise Events Detector for Real-Life Operation in the Rome Suburban Pilot Area of the DYNAMAP's Project. In Proceedings of the EuroNoise 2018, EAA—HELINA, Heraklion, Crete, Greece, 27–31 May 2018; pp. 693–698.
39. Alías, F.; Socoró, J.C.; Orga, F.; Alsina-Pagès, R.M. Characterization of a WASN-Based Urban Acoustic Dataset for the Dynamic Mapping of Road Traffic Noise. Available online: https://www.researchgate.net/publication/338103848_Characterization_of_A_WASN-Based_Urban_Acoustic_Dataset_for_the_Dynamic_Mapping_of_Road_Traffic_Noise (accessed on 22 January 2020). [CrossRef]
40. Alsina-Pagès, R.M.; Orga, F.; Alías, F.; Socoró, J.C. A WASN-Based Suburban Dataset for Anomalous Noise Event Detection on Dynamic Road-Traffic Noise Mapping. *Sensors* **2019**, *19*, 2480. [CrossRef] [PubMed]
41. Giannoulis, D.; Stowell, D.; Benetos, E.; Rossignol, M.; Lagrange, M.; Plumbley, M.D. A database and challenge for acoustic scene classification and event detection. In Proceedings of the 21st European Signal Processing Conference (EUSIPCO 2013), Marrakech, Morocco, 9–13 September 2013.
42. Mesaros, A.; Heittola, T.; Virtanen, T. Metrics for Polyphonic Sound Event Detection. *Appl. Sci.* **2016**, *6*, 162. [CrossRef]
43. Alías, F.; Socoró, J.C.; Sevillano, X. A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Appl. Sci.* **2016**, *6*, 143. [CrossRef]
44. Zhang, J.; Ding, W.; He, L. Data augmentation and prior knowledge-based regularization for sound event localization and detection. Tech. Report of Detection and Classification of Acoustic Scenes and Events (DCASE2019 Challenge), Online, 4 March–30 June 2019. Available online: http://dcase.community/documents/challenge2019/technical_reports/DCASE2019_He_97.pdf (accessed on 21 January 2012).
45. Nakajima, Y.; Sunohara, M.; Naito, T.; Sunago, N.; Ohshima, T.; Ono, N. DNN-based Environmental Sound Recognition with Real-recorded and Artificially-mixed Training Data. In Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (INTER-NOISE 2016), Hamburg, Germany, 21–24 August 2016; pp. 3164–3173.
46. Koizumi, Y.; Saito, S.; Yamaguchi, M.; Murata, S.; Harada, N. Batch Uniformization for Minimizing Maximum Anomaly Score of DNN-Based Anomaly Detection in Sounds. In Proceedings of the 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 20–23 October 2019.
47. Fonseca, E.; Plakal, M.; Font, F.; Ellis, D.P.; Favory, X.; Pons, J.; Serra, X. General-purpose Tagging of Freesound Audio with AudioSet Labels: Task Description, Dataset, and Baseline. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), Surrey, UK, 19–20 November 2018; pp. 69–73.
48. Schauerte, B.; Stiefelhagen, R. Wow! Bayesian surprise for salient acoustic event detection. In Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2013), Vancouver, BC, Canada, 26–31 May 2013; pp. 6402–6406.


Capítol 6

# Conclusions i línies de futur

L'increment de les vies de transport i l'ús intensiu d'aquestes durant les últimes dècades ha representat un increment en la pol·lució acústica. El soroll de trànsit és la font majoritària de soroll ambiental en els entorns urbans, afectant greument la salut de les persones. Això va fer saltar les alarmes d'organismes governamentals com la Unió Europea, que exigeix un monitoratge acústic regular de les vies de transport més concorregudes i dels principals nuclis de població. Aquest fet, juntament amb els avenços recents en xarxes de sensors, ha donat fruit a un seguit de projectes de monitoratge acústic en diferents nuclis urbans i vies arreu del món. Aquests sistemes han permès validar la viabilitat d'oferir a les autoritats un monitoratge dels nivells acústics a temps real, i també faciliten l'accés a la informació pels ciutadans; cosa impensable emprant mapes estratègics de soroll estàtics que s'actualitzen cada 5 anys, els quals requereixen un processament manual dels mesuraments realitzats. A més a més, aquestes xarxes obren la possibilitat de fer un processament de les dades acústiques a temps real, permetent fins i tot la detecció d'esdeveniments acústics de forma automàtica. Això és una tasca necessària per mesurar la contaminació ambiental real, ja que certes casuístiques com la pluja o altres esdeveniments no relacionats amb el trànsit esbiaixen la mesura del soroll de trànsit.

Per aquest motiu, les xarxes de sensors acústics poden esdevenir una eina eficaç per oferir informació dels nivells acústics en temps real de manera desatesa. Aquest seria el cas del projecte DYNAMAP, que a més a més de calcular el nivell de so equivalent, també identifica si el senyal acústic prové del soroll de trànsit rodat o bé d'algun dels esdeveniments anòmals etiquetats.

Però per tal de poder detectar i identificar la presència d'esdeveniments que ocorren en un entorn, cal tenir una base de dades molt completa. Idealment, cal que tingui totes les tipologies d'esdeveniments i incloure tota la diversitat acústica d'aquests, a més a més de cobrir totes les franges horàries del paisatge acústic. Per això, al projecte DYNAMAP, s'ha utilitzat la base de dades provinent dels sensors de les xarxes desplegades, que són els mateixos que després detectaran i identificaran els esdeveniments acústics. Aquesta és una tasca que requereix moltes hores d'etiquetatge i un esforç de curació posterior. Tot i que es comparteix el criteri d'etiquetatge, en participar diversos etiquetadors, pot existir cert biaix en el procés, provinent del factor humà. La precisió d'aquest procés marcarà la precisió del detector d'esdeveniments.

Tot i això, els esdeveniments de les diferents categories poden arribar a tenir unes característiques acústiques molt diferents. Per exemple: una sirena és un so típicament prominent i amb una certa durada, mentre que una botzina, que té un espectre similar és forta, però curta; per altra banda, el so de persones és inconstant i més imprevisible, diverses durades i volums, com el so de frens, que també té molta variabilitat en les durades i volums, però en una franja de freqüències molt diferent (més aguda). Per estudiar la tipologia dels sons de l'ambient acústic i avaluar-ne l'impacte, en aquesta tesi s'ha exposat un mètode de càlcul d'impacte, que es pot aplicar tant en els esdeveniments individuals com en el conjunt dels esdeveniments de forma agregada dins un període de temps determinat. Mitjançant una anàlisi en profunditat dels dos entorns –el de Milà i el de Roma–, es poden determinar millor les necessitats del sistema i entrenar millor el detector d'esdeveniments en funció d'aquestes necessitats.

A continuació, es respon cada pregunta d'investigació presentada a la introducció. La primera:

> P1: *Poden, els esdeveniments acústics no relacionats amb el trànsit, esbiaixar la mesura del nivell equivalent de so en entorns urbans i suburbans? I, doncs, cal tenir-los en compte per confeccionar un mapa acústic que representi el trànsit de forma fiable?*

Gràcies a una anàlisi en profunditat dels sons presents en l'entorn d'operació real, s'ha pogut demostrar que realment certs esdeveniments acústics esbiaixen el càlcul del nivell equivalent sonor emprat per actualitzar dinàmicament el mapa de soroll. No només si el que es pretén és monitorar el soroll de trànsit, com en el cas DYNAMAP; el reconeixement d'esdeveniments també pot ajudar a fer millors mapes acústics encara que només es vulgui mesurar el nivell de soroll general. Això és degut a la proximitat de certes fonts als sensors acústics, que poden interferir amb els mesuraments, com per exemple un niu d'ocells al costat del sensor, una gota d'aigua que impacta a prop o una interferència d'algun aparell proper. A DYNAMAP, aquest últim cas també es pot observar, per exemple el soroll de la pluja quan impacta prop del sensor, el vent que interfereix o en un altre cas que només afecta la xarxa de Roma, la vibració de les estructures dels portals que s'acaba propagant en un so només perceptible des de la mateixa estructura. Els sons agregats en un mateix període també poden esbiaixar substancialment el nivell equivalent sonor, com per exemple els treballs de construcció, que produeixen sons inconstants, fenòmens meteorològics repetitius o altres esdeveniments intermitents.

Per aquest motiu, es corrobora la necessitat d'utilitzar un detector d'esdeveniments quan es mesuren nivells acústics de forma automàtica mitjançant xarxes de sensors; però també la necessitat d'utilitzar dades reals capturades amb els mateixos sensors de la xarxa que s'utilitzarà.

Aquests esdeveniments que esbiaixen les mesures de nivell equivalent tenen diferents durades i prominències i, fent referència a la segona pregunta:

P2: *Provocarien, els esdeveniments amb durada i prominència significatives, un impacte considerable en el càlcul del nivell equivalent? Serà factible classificar-los en regions d'impacte en funció de la prominència i la durada?*

El mètode de càlcul de l'impacte permet de mesurar la contribució que tindran aquests esdeveniments en els nivells acústics. I gràcies a l'anàlisi en profunditat que s'ha presentat al capítol 5, podem afirmar que existeix una relació entre la prominència i la durada de l'esdeveniment i la seva contribució a l'$L_{eq}$. Tot i que depèn de la tipologia dels sons, s'ha pogut corroborar una tendència general que indica que els esdeveniments que tenen una prominència i/o una durada significatives acaben tenint un impacte més gran. A més a més, l'estudi també ha permès de veure quins tipus poden causar un biaix més important a les mesures de soroll i que, per tant, establir una preferència de categories a tenir en compte en l'entrenament del detector d'esdeveniments.

Finalment, pel que fa a la relació entre l'impacte i la molèstia percebuda per les persones en aquell entorn sonor, l'última pregunta és la següent:

P3: *És suficient mesurar l'impacte dels esdeveniments en el nivell equivalent de so per tal d'analitzar la molèstia percebuda per les persones?*

L'índex HARMONICA, contempla el soroll de fons i els pics dels esdeveniments puntuals, i ho tradueix a una escala del 0 a 10, però es calcula a partir de nivells equivalents sonors, i no té en compte la tipologia de cada tipus d'esdeveniment. Gràcies als tests perceptius duts a terme, podem assegurar que l'impacte en el nivell equivalent no té una relació directa amb la percepció que en té la persona que l'escolta i ens demostren que certes tipologies d'esdeveniments acústics són més molests per la naturalesa del so en si, per exemple el soroll de camions, les sirenes o les obres. Per això, incorporar la informació sobre la percepció dels sons pot ajudar a generar un mapa de soroll que representi la molèstia percebuda en un entorn acústic i que podria ser utilitzat com un indicador de salut de l'àrea monitorada.

Per comprovar la viabilitat de modelitzar-la, a partir dels resultats dels tests perceptius s'ha creat un primer model multinivell que integra un conjunt de mesures psicoacústiques dels sons analitzats i els resultats de molèstia percebuts per la mostra que ha participat en els tests. El model, que presenta un coeficient de determinació de $R^2 = 0,64$, demostra que és possible de predir la molèstia de certs sons mitjançant una sèrie de caracteritzacions prèvies, esdevenint una primera prova de concepte sobre la modelització de la molèstia dels esdeveniments acústics. Es tracta d'una primera prova de concepte, ja que per poder incorporar informació sobre el grau de molèstia dels esdeveniments presents en l'ambient sonor estudiat, caldria estendre l'estudi perceptiu dels sons a tots els tipus d'esdeveniments que hi ocorren.

Amb aquest últim estudi recollit a aquesta tesi, s'obren les portes a una nova generació de mapes acústics. Aquests podrien oferir, a més a més dels nivells de soroll, l'estimació sobre la molèstia percebuda al carrer. Primerament, caldria ampliar les mostres d'esdeveniments acústics amb tots els sons del

paisatge sonor, integrant una gran variabilitat de condicions dins de cada tipologia. A continuació, s'hauria de mesurar tots els paràmetres psicoacústics que s'han demostrat representatius i tornar a fer una sèrie de tests perceptius per recollir les caracteritzacions dels sons i la molèstia. Aquestes noves dades s'incorporarien al model per tal que fos més acurat. Finalment, caldria incorporar el model en els nodes de la xarxa, que a partir podrien oferir una predicció de la molèstia percebuda, conjuntament amb els nivells equivalents objectius de forma dinàmica. Aquest sistema suposaria una millora pel que fa a les prestacions dels mapes acústics automàtics, ja que oferiria un valor de molèstia que no tindria per què anar relacionat amb el nivell de soroll en decibels. A més, el model de molèstia seria independent del detector d'esdeveniments, perquè prediu el valor de molèstia a partir d'una sèrie de mesures psicoacústiques.

## 6.1 Dedicació del candidat

El treball d'aquesta tesi es duu a terme en paral·lel al projecte LIFE DYNAMAP, en el qual el candidat col·labora.

Des de la fase inicial de la seva tesi, el candidat participa en algunes de les tasques de desenvolupament de l'ANED, mitjançant l'etiquetatge de dades reals que s'utilitzen en l'entrenament i la validació de l'algoritme. Aquest treball li aporta un aprenentatge en profunditat en el camp dels mapes de soroll i els algoritmes detectors d'esdeveniments. També col·labora en la migració del codi desenvolupat originalment en MATLAB al que utilitza el sensor, en C++, i ajuda en la validació del codi migrat. Finalment, en un estudi de l'equip del projecte que comprova la viabilitat d'un detector d'esdeveniments de baixa capacitat, en realitza l'estimació de la càrrega computacional consumida, l'implementa en un ordinador de butxaca i el testeja. Quan l'equip de La Salle obté els primers resultats del detector d'esdeveniments funcionant en l'entorn real d'operació, el candidat col·labora en el testatge de les dades obtingudes i elabora un estudi exhaustiu de la tipologia dels esdeveniments detectats.

Amb la vocació de millorar el detector d'esdeveniments per poder mesurar amb més precisió els nivells de soroll de trànsit, el candidat comença a treballar la caracterització dels esdeveniments i el seu impacte en els nivells sonors. El candidat implementa, amb la col·laboració de l'equip, el càlcul d'SNR, adaptat del càlcul utilitzat en comunicacions, que permet de trobar la prominència de l'esdeveniment respecte al soroll que l'envolta. També implementa la mesura de l'impacte en l'$L_{eq}$, que permet de calcular el biaix produït per un esdeveniment a la mesura del nivell sonor. Finalment, elabora un estudi de com afecten la SNR i la durada de l'esdeveniment als nivells sonors, que permetrà de discernir entre les diferents tipologies d'esdeveniments. En una etapa posterior, s'estén l'estudi dels impactes individuals per mesurar també impactes agregats. El candidat duu a terme l'anàlisi de les dades i ajuda a sintetitzar l'equació d'impacte agregat.

A més a més, el candidat també expandeix el seu coneixement en detecció

d'esdeveniments acústics en una estada a Telefònica Research implementant un detector d'esdeveniments mitjançant una xarxa neuronal sota de supervisió dels Drs. Joan Serrà i Carlos Segura. Participa en un *Challenge* de detecció, el DCASE i posteriorment el prova amb dades del projecte DYNAMAP.

En l'última etapa de la seva tesi, el candidat participa en un estudi per avaluar la percepció dels esdeveniments urbans de DYNAMAP, que mitjançant mesures psicoacústiques serveix per caracteritzar una gran part dels esdeveniments acústics tal com els perceben els ciutadans. D'aquest estudi en deriva posteriorment, i més recentment, el modelatge de la molèstia de les tipologies d'esdeveniments més freqüents en entorns urbans. Això suposa un pas cap a la implementació d'una escala de mesura dissenyada per mapes de soroll i complementària als nivells sonors per representar de forma més fidel la molèstia percebuda pels ciutadans.