Urban sensor network for characterizing the sound environment in Lorient (France) through an automatic assessment of traffic, voice and bird presence ratios.

Catherine Lavandier[1], Pierre Aumond[2], Arnaud Can[2], Félix Gontier[3], Mathieu Lagrange[3], Gwendall Petit[4]

[1] ETIS, CY Cergy Paris Université, ENSEA, CNRS, Cergy Pontoise, France
catherine.lavandier@cyu.fr
[2] UMRAE, Univ Gustave Eiffel, IFSTTAR, CEREMA, Bouguenais, France
pierre.aumond@univ-eiffel.fr; arnaud.can@univ-eiffel.fr
[3] LS2N CNRS Centrale Nantes, Nantes, France
felix.gontier@ls2n.fr; mathieu.lagrange@ls2n.fr
[4] Lab-STICC, CNRS, UBS, Vannes, France
gwendall.petit@univ-ubs.fr

Abstract
The urban sound quality depends not only on the perceived loudness, but also on the perceived presence ratios of traffic sounds, human sounds and sounds of nature. In the framework of the CENSE project, these perceptual variables (presence ratios of traffic, voices and birds) have been estimated automatically, thanks to a deep learning process, from acoustic data collected on a network of sensors in the city of Lorient. The time evolutions of these sound presence ratios follow the day rhythm of the city (birds singing during the sunrise, regular pattern for the traffic, voices during weekends along the walking areas, etc.). In parallel a questionnaire sent to the population made it possible to propose a sound quality model adapted to the city of Lorient. The crossing of the measurements with the perceptual data is under process, as well as the development of a cartographic representation of the sound quality, with a focus on the presence of these sound sources.

Keywords: urban sound quality, sound source identification, soundscape, sound perception.

1 Introduction

In collaboration with the City of Lorient in the framework of the CENSE project, a sensor network has been deployed on public lighting networks. The spatial distribution of the sensors is quite dense, with about 120 low cost sensors for an area of 1 km². This is denser than similar networks in other cities such as New York City [1], Rome [2], Antwerp [3], or Jena [4]. Third octave bands are calculated every 125 ms on Lorient sensors, in order to acoustically characterize the sound environment, as a monitoring system. But, going further than just the sound levels, these data should allow to perceptually characterize the soundscape as it is defined in the ISO [5] standard “acoustic environment as perceived or experienced and/or understood by people, in context”, as soon as acoustic measurements could correspond to perceptual data.

For outdoor urban contexts, previous studies have already proposed links between sound quality (which corresponds to a pedestrian point of view) and acoustic measurements. In [6] the sound quality defined as the pleasantness of the soundscape depends on percentile sound levels:

$$SQ = 18.67 - 0.20\,L50 - 0.02\,(L10 - L90) \quad (1)$$

In [7] the sound quality (SQ) model (eq.3) introduces a new acoustic indicator TFSD which corresponds to the time and frequency normalized deviations of a sound spectrogram. It aims at describing the time variations within specific octave bands, which are characteristic of the sound source.

$$TFSD(f,t) = \frac{\left|\frac{d^2L(f,t)}{df \cdot dt}\right|}{\sum_{f=31.5\,Hz}^{f=16kHz}\left|\frac{d^2L(f,t)}{df \cdot dt}\right|} \quad (2)$$

where L(f, t) is the octave spectrum of the signal, f varying from 31.5 Hz to 16 kHz. This indicator can be used with third octave bands, and is quite robust to the frequency limits.

The sound quality is then calculated as:

$$SQ = 30.18 - 0.16\, L_{50,1kHz} + 8.92\, TFSD_{500Hz,1s} + 2.99\, TFSD_{4kHz,1/8s} \quad (3)$$

But the variances of the sound quality data explained by these acoustic models are always less important than the explained variance based on perceived source identification models. For example, in [1] for Paris the best perceptual model is:

$$SQ = 8{,}11 - 0{,}38*OL - 0{,}15*T + 0{,}20*V + 0{,}15*B \quad (4)$$

where OL corresponds to the Overall Loudness, T is the percentage of time when the traffic is heard, V and B correspond respectively to this percentage of time for Voices and Birds.

It can be noticed that in the equation 3, the 500 Hz octave band supports the voice energy whereas the 4 kHz band supports the energy of bird songs.

For indoor comfort (which corresponds to an inhabitant point of view), most of researches have focused on building insulation, described with R_w [8,9], but this approach could not reveal the load of the semantic behind the sound. It is clear that R_w cannot explain why inhabitants cannot bear the expressive voices in the evening, heard inside their home, but like hearing bird songs. Dedieu et al. [10] showed that sound quality inside a living room depends on the type of sound (traffic, birds, neighbor television, etc.), on personal tastes (preferred style of music, ventilation masking) and on balance between sounds coming from outdoor and sounds coming from neighbors [11].

So, in the CENSE project, we focused on the perceived sound quality and on noise annoyance for pedestrian point of view (soundscape characterization) as well as for resident point of view (at home situation). As these perceptual characteristics depend on perceived sources, we also focused on the possibility to measure automatically the sound source presence ratios on the sensors, in addition to classical level indicators.

2 Questionnaire campaign

The sound quality and the noise annoyance has been assessed for both the pedestrian and inhabitant points of view thanks to a questionnaire that has been sent by postal mail to about 2000 inhabitants in the city center of Lorient between January and March 2019. It was divided into four parts. The first one concerned the sound environment quality. The second part focused on sound sources. The third part concerned the annoyance for both points of view. A last part was dedicated to collect personal data.

2.1 Detailed description of the questionnaire

In the first part, people were asked to assess the quality of the sound environment in their neighborhood and then in their street (when walking or cycling home). This section concerned the Short-Term (ST) assessment and was composed of semantic differential scales (7 levels) inspired on the Swedish protocol [12]. The French semantic anchors (in *italic*) and the proposed English translations are presented below:

- *Désagréable*/Unpleasant …. *Agréable*/Pleasant
- *Inerte*, Amorphe/Inert …. *Animé, mouvementé*/Eventful
- *Bruyant*/Noisy …. *Silencieux*/Silent
- *Ennuyeux, Inintéressant*/Boring …. *Stimulant, Intéressant*/Exciting
- *Agité*, Chaotique /Chaotic …. *Calme, Tranquille*/Calm
- *En inadéquation avec vos attentes*/In inadequacy with your expectations …. *En adéquation*/In adequacy

Then, in a second part, the participants had to fill a table about the perceived time of presence ratio and the perceived loudness of 13 sound sources that they can hear when they come in or out of their home, on foot or by bike, on their street, thinking about a year. So, the presence of the sources has been questioned with a Long-Term (LT) approach with the possibility for respondents to precise the season when the source was specifically heard. The taxonomy had been based on soundwalks on sites made by the authors, and on literature [13,14]. The outdoor sources that had to be assessed were Road traffic, 2-wheel motor vehicles, Sirens or alarms, Rail traffic, Air traffic, Urban maintenance (cleaning, garbage...), Calm voices (conversations), Children's voices (schools, playgrounds), Expressive voices (festive voices, laughter, shouts), Music (from bars, restaurants, shops...), Wind in the vegetation, Small birds, Gulls. This last category was proposed because Lorient is a harbor city with several complaints in the local press about the noise of gulls. Participants could add other sources if needed, and could also add free comments.

A third part focused on the Long-Term annoyance, assessed with questions inspired from ICBEN Guidelines [15]: "Thinking about the last 12 months, when you are (- at home with your windows closed, - at home with your windows open or on your balcony or in your garden, - in the street, when you arrive at home by bike or on foot) how much does (- global noise, - noise from each source) bother, disturb, or annoy you: Extremely, Very, Moderately, Slightly or Not at all?"

The last part was dedicated to personal information: noise sensitivity of the inhabitants based on the 6-item Weinstein's noise sensitivity scale (WNSS) [16], gender, age, socio-professional category, membership (or not) to local groups against noise. People had also to provide their address and a set of questions on their housing.

- Tenant / Owner Courtyard or garden area? (yes/no)
- Has quiet room? (yes /no)
- House/Apartment Living space overlooking the street? (yes/no)
- Double glazing? (yes /no)
- Time of occupancy? (<1 year, 1-3 year, >3 year)
- Living space with a view on natural elements? (no, a little, a lot)
- Insulation of the facade <10 years ago? (yes/no)

More than 300 people completed the questionnaire (exaclty161 sent back the paper form and 157 answered through Internet). A draw has been held for 5 gift certificates of 50 €. Figure 1 shows the study area and a heat map of the responses.

Figure 1: Heat map of the location distribution of the inhabitants who completed the questionnaires. Small points represent the sensors, and the larger light blue point corresponds to the analyzed sensor (see § 3.2).

2.2 Analysis of data: Models for pleasantness and annoyance

2.2.1 Outdoor pleasantness and outdoor annoyance

Analyzing the questionnaire, it has been possible to propose linear regression models [17] to predict Short-Term Outdoor Pleasantness (ST_OP) and Long-Term Outdoor Annoyance (LT_OA) with statistically significant variables. Below are presented the selected variables with the sign of the standardized coefficients (following their rank order) that appear in the Linear Regressions (LR) for Pleasantness and Annoyance, with the corresponding Adjusted R squared.

ST_OP = LR (- Overall Loudness, - Road Traffic, - Air Traffic, - Expressive Voices, + Birds) (5)
Adjusted R squared 44%

ST_OP = LR (- Overall Loudness, - Road Traffic, - Air Traffic, - Expressive Voices, + Birds, - Noise Sensitivity) Adjusted R squared 47% (6)

LT_OA = LR (+ Overall Loudness, + Presence of Two Wheels, + Railway Traffic, - Birds) (7)
Adjusted R squared 29%

LT_OA = LR (+ Overall Loudness, + Presence of Two Wheels, + Railway Traffic, - Birds, + Noise Sensitivity) Adjusted R squared 36% (8)

Overall Loudness (rated on the Noisy/Silent scale) appears in all models at the first place showing that this variable is the most influent one. The louder the environment, the less pleasant and the more annoying. The expressive voices have a negative influence on the sound pleasantness (eq. 5), which is opposed to previous findings (eq. 4). Literature has already revealed that the influence of the human presence depends on the context [18]. Participants that are more sensitive to noise rated the sound pleasantness in a lower part of the pleasantness scale and in an upper side of the annoyance scale, compared to non-sensitive persons. The sound pleasantness is easier to explain with the linear models (R^2adj. $\approx$ 40%) than annoyance (R^2adj. $\approx$ 30%). It shows that annoyance is a complex notion which involves more individual factors than sound pleasantness.

2.2.2 Long-Term Inside Annoyance (LT_IA)

Participants were asked to rate the annoyance when windows are open and closed. For open windows the models consider the difference of calm voices and expressive voices. If participants hear calm voices coming from outside when they are at home, it reduces noise annoyance, but if voices are intrusive like expressive voices, then it increases annoyance.

LT_IA (Open Windows) = LR (+ Loudness, + Presence of Two Wheels, + Expressive Voices, - Calm Voices, - Birds) R^2adj. 35% (9)

LT_IA (Open Windows) = LR (+ Loudness, + Presence of Two Wheels, + Expressive Voices, - Calm Voices, + Noise Sensitivity, + Age) R^2adj. 40 % (10)

LT_IA (Open Windows) = LR (Negative Intercept, + Loudness, + Presence of Two Wheels, + Expressive Voices, - Calm Voices, + Noise Sensitivity, + Age, - Presence of Quiet Room) R^2adj. 44 % (11)

For closed windows, the calm voices are not influent as they are not anymore heard. Expressive voices are then associated with Music from bars and restaurants. Birds keep influence because if they are hear behind closed windows, it means that their presence is quite strong.

LT_IA (Closed Windows) = LR (+ Loudness, + Presence of Air Traffic, + Expressive Voices, + Music from bars, - Birds) R^2adj. 34% (12)

LT_IA (Closed Windows) = LR (+ Loudness, + Presence of Air Traffic, + Expressive Voices, + Music from bars, - Birds, + Noise Sensitivity, + Age) R^2adj. 37% (13)

LT_IA (Closed Windows) = LR (Negative Intercept, + Loudness, + Presence of Expressive Voices, + Noise Sensitivity, + Age, - Presence of Quiet Room in the housing) R^2adj. 40% (14)

Noise Sensitivity and Age increase the adjusted explained variance of the models. In the Lorient corpus, these variables are not correlated ($r = -0.08$). What is interesting is the presence of a housing variable in the models. It has already been shown that the presence of a quiet side reduces the annoyance, increasing the possibility to cope with the noise for the residents [19]. Moreover, in equation 11 and 14, the reduction of the annoyance with the quiet room variable is expressed by the presence of a significant negative intercept in the models.

It is clear that the presence of sound sources is very important to characterize the outdoor environment as well as the indoor sound ambiance. So, in the frame of the CENSE project, the calculation of the source presence ratios has been automatically implemented on the low-cost sensors, for relevant sources such as Road traffic, 2-wheel motor vehicles, Voices and Birds for examples.

3 Sound identification

The source identification is realized every second for each source of interest on each sensor. As the source should be identified like a person would do, the masking effect due to source superposition should be considered. Moreover, we saw that this is the time ratio of sources (that are heard) which has an impact on pleasantness or on annoyance. So, what would be the good duration for the calculation of these ratios? Time ratios of source presence change over the time. For the duration of the acoustic measurements. Brocolini et al [20] showed that a 10-minute measurement is a good compromise: not too long in order to reveal the rhythm of the days, but not too short to characterize the soundscape of the city location

3.1 Process

Source identification is conducted using a deep convolutional architecture that first extracts time-frequency patterns relevant to the identification of sound sources from each 1s third-octave segment. The convolutional layers are followed by batch normalization [21] and rectified linear unit activations. At inference, presence or absence labels predicted for each sound source are aggregated over time to obtain the time of presence in third-octave measurements of arbitrary duration. In our case, a 10-minute duration has been chosen to be in line with the optimum measurement duration.

The model is trained on a fully synthetic set of 400 sound scenes of 45s each as described in [22]. Sound scenes are simulated with the simScene Matlab[1] library by combining background noise recordings and extracts representing sound events from sources of interest recorded in Lorient. The source categories, signal-to-noise ratios, and inter-onsets characteristics of sound events are drawn semi-randomly from normal distributions. The corresponding parameters, as well as the overall sound level of each scene, are conditioned on a desired type of sound environment: quiet street, noisy street, very noisy street, park, or square. The ground truth composition, i.e. separate channels for each active sound source enables the training of the deep neural architecture in a supervised approach. Synthetic scene generation and automatic annotation processes are further detailed in [23]. Figure 2 shows an example of source identification by the trained model. Only the third-octave spectrogram of the mixed scene (top) is visible to the network.

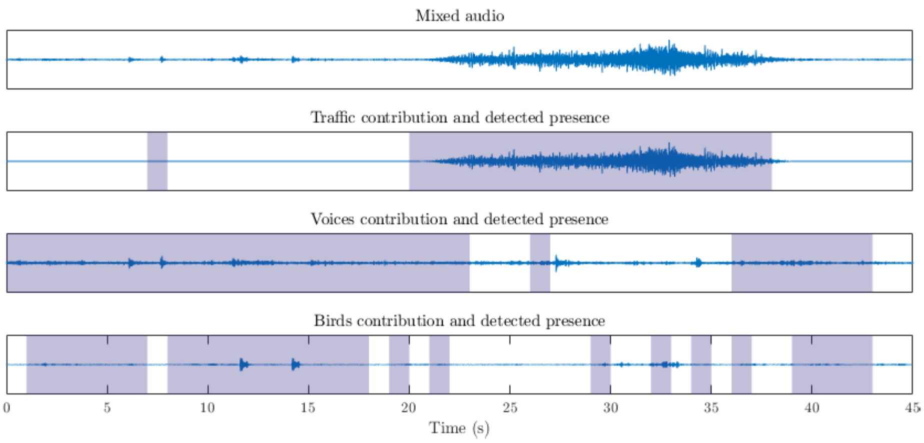

Figure 2: Example of the source (Traffic, Voices, and Birds) identification for a 45s excerpt. The network is just fed by the third octave band of the top wav form. The grey parts correspond to the identified sources.

3.2 Results

An extraction of the ratios has been done for different sources of interest, for several sensors, and for different periods. Only results for the sensor presented in Figure 1, for the four sources (Road traffic, 2-wheel motor vehicles, Voices and Birds) and for the period between the 11th of January 2020 and the 11th of February 2020 are presented in this paper, as an illustration of the results.

The ratios are calculated over the first 10 minutes of each hour. During the four weeks of the month, 744 ratios have been calculated for each type of source (one ratio for each hour). Figure 3 presents the evolution of the median values for the 7 days of a week (about 4 values for one specific hour of a specific day).

It shows that the percentage of time when birds are heard is maximum in the morning. The automatic identification and ratio calculation correspond quite well to the fact that birds sing specifically at the sunrise.

[1] https://bitbucket.org/mlagrange/simscene/downloads

For voices, it can be seen that its percentage of presence is more important on the Wednesdays due to the fact that children do not have school on that specific day in France, and days of the week-end when people do not work.

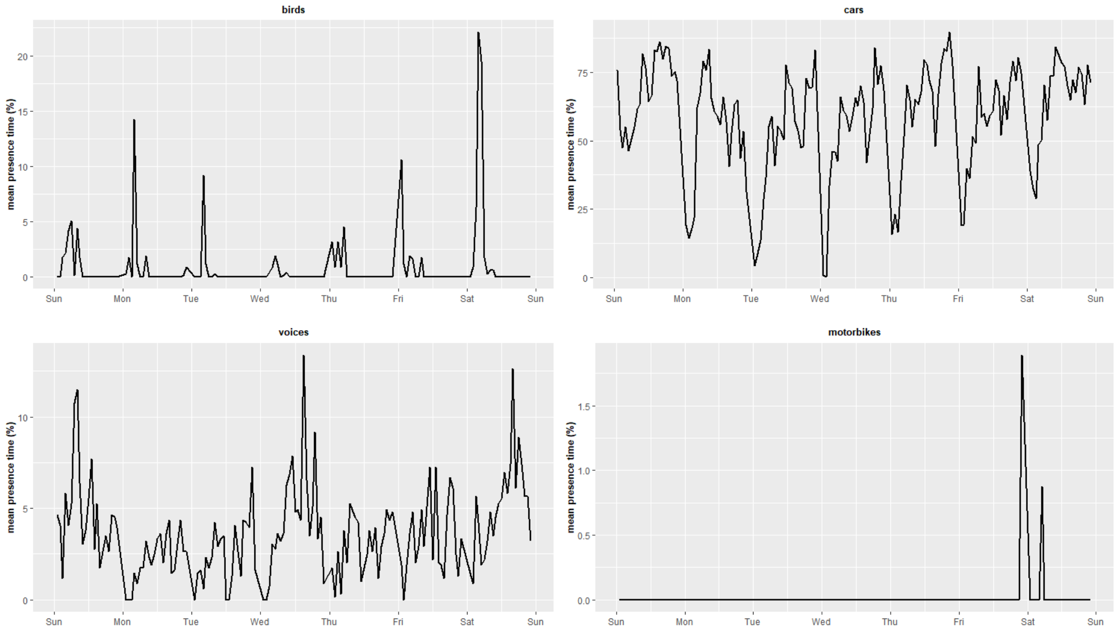

Figure 3: Median values of the source presence ratios for birds, cars, voices and motorbikes, calculated over a month period.

For cars, the time evolution shows the rhythm of the days, with a drop of passbys during the night. The nights between Fridays and Saturdays, as well as the nights between Saturdays and Sundays, present a larger activity than during the other nights. The detection of motorbikes (1% and 2% of the time) on the Friday nights has to be validated with resident comments in the questionnaires.

These results show that the automatic identification of sources and the extraction of the presence of source ratios are very well in line with the activity of the Lorient city. Including these ratios into the equation (4) or into the perceptual equations presented in §2.2 for Lorient makes it possible to measure not only the sound level as a pollution metrics, but also the sound pleasantness as a quality metrics, in the streets as well as inside the housings of this city.

4 The cartographic web portal

To highlight the results on a map, a cartographic portal has been set up. Publicly available at http://cense.noise-planet.org/, this website allows users to navigate the map (with the classic zoom and address search functions) and to consult the calculated indicators in detail.

To facilitate consultation of the data, the results have been organized under headings:

- Infrastructure: the layers presenting the sensor network
- Territory (*Territoire* in French): the buildings in the study area
- Noise map of L_{DEN} indicator
- Soundscape (*Paysage sonore*): the layers resulting from the survey, divided into two sub-sections:
    - Sound Quality (*Agrément*): outdoor annoyance, pleasant or not, inert or animated, agitated or calm layers (Figure 4)
    - Perceived sources (*Sources perçues*): birds, road traffic, quiet voices and expressive voices layers (Figure 5)

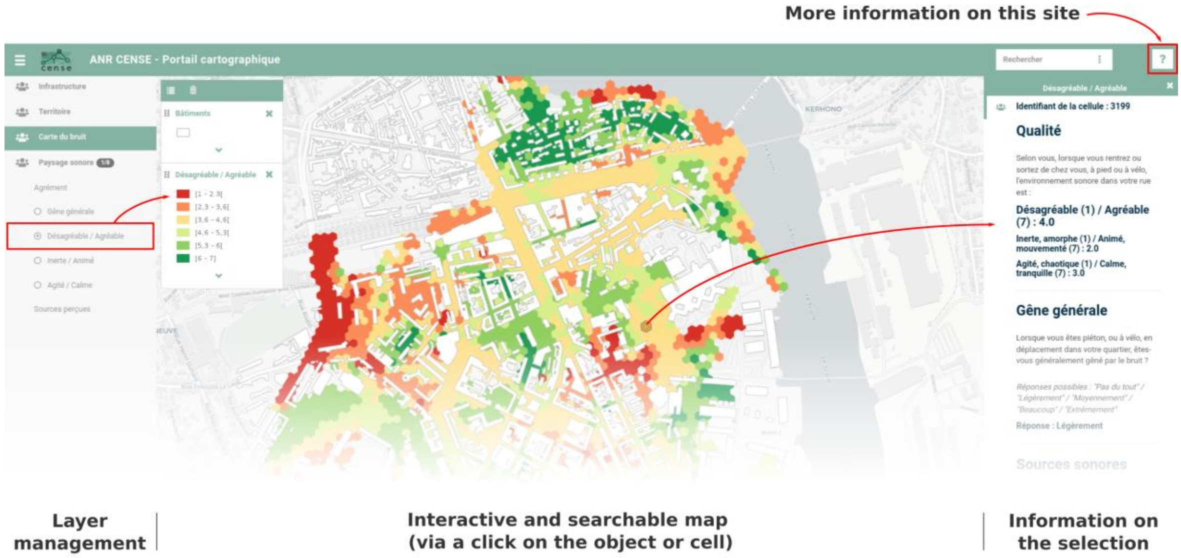

Figure 4: Welcome page of the portal, explaining how to visit the different layers.

For each of the layers from the “Soundscape” part, the user can click on a cell to bring up a whole series of indicators for the layer in question but also for the others, thus allowing comparisons to be made in a very simple manner. The layers where derived from an Inverse Distance Weighting interpolation between data collected through the questionnaires.

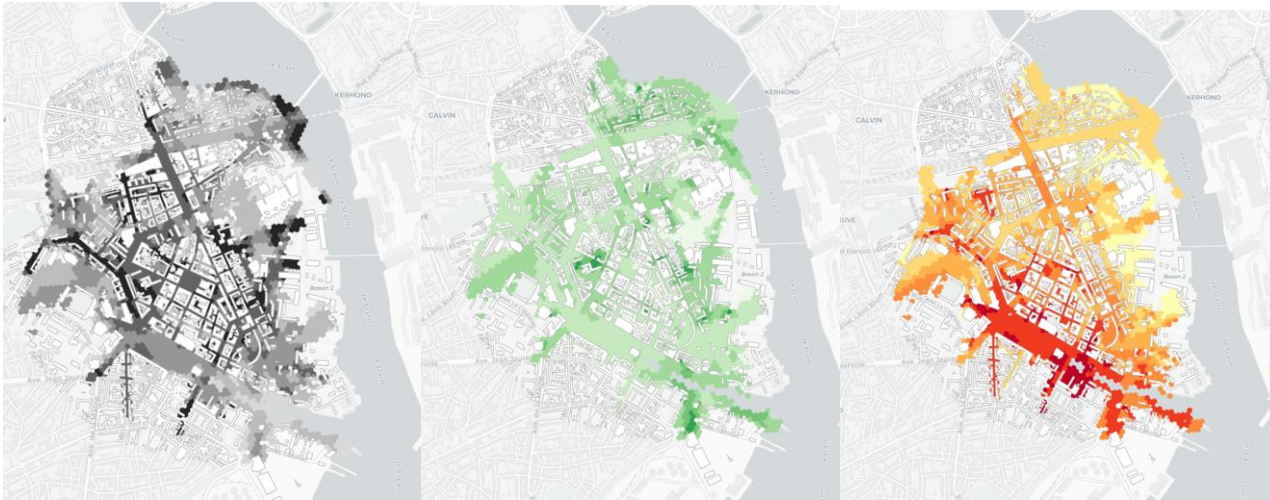
Figure 5: Traffic (left), birds (middle), and expressive voices (right) layers which represent the presence ratios of different sources.

From a technical point of view, this platform is based on the following free and open source components:
- PostgreSQL[2] and PostGIS[3]: for data storage and manipulation,
- Geoserver[4]: for data distribution notably in the form of WMS5 streams,
- MViewer[6]: for the generation of the mapping platform and all associated functionalities.

[2] https://www.postgresql.org/
[3] https://postgis.net/
[4] http://geoserver.org/
[5] https://www.ogc.org/standards/wms
[6] https://github.com/geobretagne/mviewer

5 Conclusion

In addition to the classical acoustic indicator measurements (A weighted equivalent sound levels, percentile levels), an automatic identification of the sound sources is now possible to measure their presence ratios, directly on the low-cost sensors. Thanks to simple regressions, it is possible to calculate the sound quality of the environment, not only with a negative point of view (sound as pollution), but also with a positive attitude (sound as quality).

During the CENSE project, the storage of the acoustic data extracted from all the sensors has been available after the perceptual campaign, so it was not possible to cross all the acoustic measurements with all the questionnaires, by lack of synchronicity. The evolutions of sound source presence ratios have been studied for a limited number of sensors, making cartographic representation impossible with interpolation. Moreover, the link between sensor data (acoustic measurements and sound source ratios) and the cartographic portal was not possible in real time, rendering the web portal static instead of dynamic.

Most of these issues are not scientific problems but need much more time of engineering development than initially planned. Nevertheless, the scientific issue about the crossing of sensor data with perceptual data is still not solved. Additional perceptual experiments are still to be organized.

Acknowledgements

The authors want to thank the city of Lorient and in particular J.C. Baron and P. Crépeaux from the “Pôle cadre de vie et development durable” (Life and sustainable development center). This work was funded under the ANR project CENSE No. ANR-16-CE22-0012.

References

[1] C. Mydlarz, M. Sharma, Y. Lockerman, B. Steers, C. Silva, J. P. Bello, “The life of a New York City noise sensor network,” Sensors, 19(6): 1415 (2019).

[2] P. Bellucci, L. Peruzzi, G. Zambon, “LIFE DYNAMAP project: The case study of Rome,” Applied Acoustics, 117: 193–206 (2017).

[3] D. Botteldooren, L. Dekoninck, C. Meeussen, T. Van Renterghem, “Early stage sound planning in urban re-development: The Antwerp case study,” presented at the Proceedings of the International Congress and Exposition on Noise Control Engineering (Inter-Noise) (2018).

[4] J. Abeßer, M. Gotze, S. Kuhnlenz, R. Grafe, C. Kuhn, T. ClauB, H. Lukashevich, “A distributed sensor network for monitoring noise level and noise sources in urban environments,” presented at the Proceedings of the IEEE International Conference on Future Internet of Things and Cloud (FiCloud), pp. 318–324 (2018).

[5] ISO 12913-1:2014: Acoustics - soundscape - part 1: definition and conceptual framework. Standard, International Organization for Standardization, Geneva, CH, (2014).

[6] P. Ricciardi, P. Delaitre, C. Lavandier, F. Torchia, P. Aumond, “Sound quality indicators for urban places in Paris cross-validated by Milan data”. J. Acoust. Soc. Am. 138: 2337–2348. (2015).

[7] P. Aumond, A. Can, B. De Coensel, D. Botteldooren, C. Ribeiro, C. Lavandier, “Modeling soundscape pleasantness using perceptive assessments and acoustic measurements along paths in urban context”. Acta Acust. unit. Acust. 103: 430–443 (2017).

[8] Hongisto V., Oliva D., Keränen J., “Subjective and Objective Rating of Airborne Sound Insulation–Living Sounds”. Acta Acust. unit. Acust. 100: 848–863 (2014).

[9] Hongisto V., Oliva D., Rekola L., “Subjective and objective rating of the sound insulation of residential building façades against road traffic noise”, The Journal of the Acoustical Society of America, 144: 1100-1112 (2018).

[10] Dedieu R., Lavandier C., Camier C., Berger S., “Pleasantness of Typical Acoustic Environments inside a Living Room in a European Residential Context”, Proceedings of ICA 2019, pp. 4170-4177, Aachen, (2019).

[11] Dedieu R., Lavandier C., Camier C., Berger S., “Sound design strategies in composing pleasant and unpleasant acoustic environments inside a living room in a European residential context”, Proceedings of Forum Acusticum 2020, (2020).

[12] Axelsson O., Nilsson M. E., Berglund B., “The Swedish soundscape-quality protocol”, The Journal of the Acoustical Society of America, 131(4): 3476 3476, (2012).

[13] Brown, A. L., Kang, J., Gjestland, T., “Towards standardization in sound-scape reference assessment”. Applied Acoustics, 72(6): 387–392, (2011).

[14] Berglund B, Nilsson M., “On a tool for measuring soundscape quality in urban residential areas”. Acta Acust United Acust, 92: 938–44, (2006).

[15] Fields J. M. et al., “Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation”, Journal of Sound and Vibration, 242 (4): 641-679, (2001).

[16] Kishikawa H., Matsui T., Uchiyama I., Miyakawa M., Hiramatsu K., Stansfeld S. A., “The development of Weinstein’s noise sensitivity scale”, Noise and Health, 8(33): 154-160, (2006).

[17] Aumond P., LavandierC., “Links between representation of outdoor soundscape, noise annoyance at home and outdoor acoustic measurement”, Proceedings of Internoise 2019, 10 pages, (2019).

[18] Lafon J., Lavandier C., “Soundscape quality of urban parks and gardens”, proceedings of Euro Noise, pp. 2383-2388, (2015).

[19] Öhrström E., Skånberg A., Svensson H., Gidlöf-Gunnarsson A., “Effects of road traffic noise and the benefit of access to quietness”, Journal of Sound and Vibration, 295 (1–2): 40-59, (2006).

[20] Brocolini L., Lavandier C., Quoy M., Ribeiro C., “Measurements of acoustic environments for urban soundscapes: choice of homogeneous periods, optimization of durations and selection of indicators”, Journal of the Acoustical Society of America, 134(1-Pt.2): 813-821, (2013).

[21] Ioffe S., Szegedy C., “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, Proceedings of the 32nd International Conference on Machine Learning, PMLR 37: 448-456, (2015).

[22] Gontier F., Lostanlen V., Lagrange M., Fortin N., Lavandier C., Petiot J.F., Polyphonic training set synthesis improves self-supervised urban sound classification, The Journal of the Acoustical Society of America, 149, 4309 (2021).

[23] Gontier F., Lavandier C., Aumond P., Lagrange M., Petiot J.F., "Estimation of the perceived time of presence of sources in urban acoustic environments using deep learning techniques", Acta Acustica united with Acustica, Vol. 105 (6), pp.1053–1066, (2019).