Research Article

Daniel Bonet-Solà, Ester Vidaña-Vila*, and Rosa Ma Alsina-Pagès

Prediction of the acoustic comfort of a dwelling based on automatic sound event detection

Abstract: There is an increasing concern about noise pollution around the world. As a first step to tackling the problem of deteriorated urban soundscapes, this article aims to develop a tool that automatically evaluates the soundscape quality of dwellings based on the acoustic events obtained from short videos recorded on-site. A sound event classifier based on a convolutional neural network has been used to detect the sounds present in those videos. Once the events are detected, our distinctive approach proceeds in two steps. First, the detected acoustic events are employed as inputs in a binary assessment system, utilizing logistic regression to predict whether the user's perception of the soundscape (and, therefore, the soundscape quality estimator) is categorized as "comfortable" or "uncomfortable". Additionally, an Acoustic Comfort Index (ACI) on a scale of 1–5 is estimated, facilitated by a linear regression model. The system achieves an accuracy value over 80% in predicting the subjective opinion of citizens based only on the automatic sound event detected on their balconies. The ultimate goal is to be able to predict an ACI on new locations using solely a 30-s video as an input. The potential of the tool might offer data-driven insights to map the annoyance or the pleasantness of the acoustic environment for people, and gives the possibility to support the administration to mitigate noise pollution and enhance urban living conditions, contributing to improved well-being and community engagement.

Keywords: citizen science, acoustic event detection, noise, annoyance evaluation, acoustic comfort, soundscape, convolutional neural networks

* Corresponding author: Ester Vidaña-Vila, Human Environment Research (HER), La Salle – URL, Sant Joan de La Salle, 42, 08022 Barcelona, Spain, e-mail: ester.vidana@salle.url.edu
Daniel Bonet-Solà, **Rosa Ma Alsina-Pagès**: Human Environment Research (HER), La Salle – URL, Sant Joan de La Salle, 42, 08022 Barcelona, Spain

Abbreviations

ACI	Acoustic Comfort Index
ASED	Automatic sound event detection
CNN	Convolutional neural network
DL	Deep learning
EU	European Union
GTCC	GammaTone Cepstrum coefficients
L_{Aeq}	A-weighted equivalent sound level
L_d	Day Noise Level. A-weighted Leq, over the 14-h day period (7 to 21)
L_e	Evening Noise Level. A-weighted Leq, over the 2-h evening period (21 to 23)
L_n	Night Noise Level. A-weighted Leq, over the 8-h night period (23 to 7)
L_{den}	Day-Evening-Night noise level. A-weighted Leq with a penalty for night/evening noise
SNR	Signal-to-noise ratio
SSID	Soundscape indices
WHO	World health organisation

1 Introduction

Noise pollution is a widespread problem affecting millions of people, mostly in urban areas, industrial areas, and the surroundings of transportation hubs such as airports or railway stations. Focusing on Europe alone, a report published in 2020 [1] confirmed that 20% of the European Union (EU) population resides in areas where noise levels surpass the thresholds deemed harmful to health by the World Health Organization (WHO). Exposure to noise levels exceeding unhealthy L_{den} thresholds due to road traffic was estimated to affect over 80 million people within urban areas and more than 30 million people outside urban areas in the countries studied, including the EU and five other European countries. Additionally, a recent study [2] confirmed that more than 3,600 deaths were caused by road traffic exposure could be prevented each year.

There is a long list of unhealthy and pernicious consequences derived from continuous exposure to deteriorated

soundscapes with high noise levels. There are even thousands of reported cases in Europe of premature deaths directly related to noise exposure. Environmental noise can lead to diastolic blood pressure [3] and ischaemic heart diseases [4,5]. It produces high sleep disorders with awakenings that acutely hinder the quality of life of the affected population [6]. It has also been linked to fatigue, headaches and nervousness [7], psychological stress [8,9], decline in working performance [10,11], and learning and cognitive impairment in children and students [12,13]. In addition to these severe effects on the health and the quality of life of citizens, noise exposure is also associated with general annoyance [14], which also has a negative impact on the well-being of those afflicted by it.

Annoyance is widely regarded as the primary psychological consequence resulting from noise exposure. It is commonly associated with feelings of nuisance, disturbance, dissatisfaction, and unpleasantness [15]. Moreover, annoyance can significantly interfere with various everyday activities, including mental concentration [16], communication [17], learning [18], work [19], and even recreation [20,21]. Additionally, it is directly linked to the aforementioned sleeping disorders, particularly increased difficulties in falling asleep and frequent awakenings [22]. The starting point for finding a viable solution to this widespread problem should involve an accurate diagnosis of the quality of the acoustic environment. Several studies, both using objective data and psycho-acoustic and non-acoustic parameters, can be found in the literature tackling this issue, as it will be further developed in Section 2.

This present article conducts a different approach to predict the subjective acoustic satisfaction level with a given soundscape neither based on noise level measurements nor psycho-acoustic metrics. The proposed approach can be very useful in making an approximation of the perceived quality of soundscapes (in terms of acoustic comfort) in urban areas without needing expensive dedicated equipment to do so. The goal is to develop a two-stage estimator to predict the level of acoustic comfort in a living environment based on the automatic sound event detection and classification performed on short videos recorded with a smartphone or tablet [23]. This acoustic comfort predictor could be used to rate dwellings for educational or informative purposes in a citizen science context, to map urban areas according to the predicted subjective perception of acoustic satisfaction instead of the normally favoured noise indices or to automatically extract meaningful information from future collecting campaigns among other applications.

Acoustic satisfaction or dissatisfaction is correlated with several perceptual constructs such as pleasantness, calmness, eventfulness, monotony, or annoyance among others. Previous works have successfully explored soundscape modelling and the prediction of perceptual constructs using acoustic and psycho-acoustic indicators as inputs [24,25]. That being said, the difficulty of predicting annoyance or other perceptual constructs using noise levels or psycho-acoustic metrics has also been acknowledged [26,27], primarily due to the complex interplay of variables that influence individual judgments of the polyphonic soundscape. In light of this, the proposed approach emphasizes the significance of considering the type of sound source among the non-sensory variables. Moreover, the proposed solution is easy to implement as it does not require the use of costly sound sensors operated by expert technicians. Instead, anyone can conveniently make a brief recording using a common mobile domestic device. To predict the subjective assessment of the dwelling's soundscapes, two rating scales will be used: (i) a binary assessment and (ii) a 5-point rating scale.

To train and test the design, the authors have used one of the datasets obtained from the citizen science project *Sons al Balcó*. In that project, two campaigns were conducted across Catalonia. The first one took place in 2020 [28], amid lots of mobility and activity restrictions enforced during the lockdown caused by the COVID-19 pandemic. The second one took place in 2021 [29], in a back-to-normal context.

It is important to note that in most previous studies on annoyance modelling or perceptual constructs prediction found in the literature, authors try to predict the perceived annoyance of a particular sound event [30], the subjective perception of a particular audio or video clip that has been assessed by a set of participants [31] or the perceived quality of soundscapes in public spaces by a plural on-site evaluation during a soundwalk [32]. In contrast, the present work aims to predict the global perceived acoustic comfort in a dwelling, reported only by one of its residents. That means that the video used may not be representative and that the opinion reported may not be consensual. This article also assesses the impact of these aspects on the performance with an accurate analysis of the failed predictions and proposes complementary information that could be used to improve it.

In Section 2, a more exhaustive study of the state of the art is developed. In Section 3, relevant information about the *Sons al Balcó* dataset is presented along with the methodology used to predict the subjective perception of the soundscapes and the metrics used to assess the estimator performance. Section 4 gives the results of two studied rating systems (binary assessment and 5-point scale rating). Next, Section 5 offers a more insightful discussion about the results and an analysis of the errors made by the estimator.

Finally, Section 6 is dedicated to the final conclusions of this present work.

2 Related work

Studies conducted to assess acoustic comfort in urban locations can be classified into three categories. The first approach (Section 2.1) is restricted to the collection of noise exposure pressure levels using sound sensors (usually deployed in urban areas). The second approach (Section 2.2) focuses on specific types of noise sources and their perception and effects on well-being. Finally, the third approach (Section 2.3) computes psycho-acoustic metrics or uses non-acoustic or even non-sensory variables to try to ascertain the subjective annoyance perceived.

2.1 Urban sound sensors and noise indices

Numerous projects have been undertaken with the objective of mapping different areas within targeted cities based on the measured noise exposure. As the amount of noise mappings published is very large, this subsection will focus only on some of the most recent contributions in the literature.

Most noise mappings done in urban and suburban areas are especially concerned with road traffic noise and use noise indices to assess the noise exposure in each spot. Case studies have been published for many cities such as Piteşti, Romania [33] or Mashhad, Iran [34]. In some instances, a differentiated analysis is conducted depending on the land-use type as in Kigali, Rwanda [35]. In Aburra Valley, Colombia [36] a noise mapping was conducted focused on non-traffic related noise sources, especially leisure noise.

Industrial areas have also been mapped to assess the noise exposure of workers or neighbours. A study conducted on a concrete block-making factory [37] concluded that a Hearing-Loss Prevention Program was advisable due to the elevated sound levels measured. Measurements in the Tarkwa Mining Community of Ghana [38] were correlated with sleep disturbance, hearing problems, and hypertension.

Another study complemented the data obtained through sound sensors with questionnaires about the subjective perception of the noise to evaluate noise pollution and its subjective perception in a university campus in Juiz de Fora, Brasil [39].

Typically, these urban noise mappings rely exclusively on noise indices such as L_{Aeq}, L_{den}, L_d, L_e, or L_n. These indices are reliable for describing the exposure to road traffic noise, which is the primary contributor to the deteriorated sound environment in cities. However, noise indices alone do not provide a complete picture of what constitutes a comfortable or uncomfortable sound environment. For example, there are multiple influencing factors in the noise annoyance perception beyond sound pressure levels, both psycho-acoustic [40] and non-acoustic [41,42], which will be hinted at Sections 2.2 and 2.3.

2.2 Annoyance by type of noise source

Acoustic discomfort is often caused by the presence of annoying sound event. However, not all noise sources are equally annoying. Thus, assessing the subjective perception of the level of annoyance of some of them only taking into account noise indices is usually unreliable. This section will focus on the main types of noise sources that can be found in an urban soundscape beyond the exhaustively researched road traffic noise [43,44].

Neighbourhood noise, for example, can be highly annoying and equally produce harmful effects on health even though the noise indices related are low when they are compared to traffic noise. The subjective experience of this kind of noise stress can lead to inadequate neuroendocrine reactions and regulation diseases [45]. Even though several studies in the literature have focused on assessing neighbour noise, most of the research is still centred on the analysis of traffic noise [46]. A very recent study opted for a qualitative approach to analyse complaints, attitudes, and viewpoints on neighbour noise [47].

Another noise source that has not been thoroughly studied and that has been surprisingly missing in most reports on noise pollution until recently is recreational noise [48]. Many European cities experience increasing noise exposure to daytime and night-time leisure activities which normally involve crowds or outdoor music among other sounds. Again, recreational noise can be very annoying even with moderate noise levels. However, publications centred in leisure noise traditionally have used sound pressure levels to assess the outcomes of its exposure [49].

It is also tricky to base the assessment of the annoyance of construction sites only on L_{Aeq} measurements [50]. There are several individual noise sources in construction sites, including several machines from pile drivers or earth augers to bulldozers and excavators. Combined noise produces a higher annoyance than individual noise sources for L_{Aeq} above 65 dBA. However, little is known about the

real factors that could predict the annoyance level associated with construction sites according to a recent report by van Kamp *et al.* [51]. In China, an initiative was launched consisting of mapping and analysing the construction noise annoyance using data mining on social media platforms [52].

Other noise sources that have a negative effect on the psychological state and well-being of citizens are dogs barking or babies crying [53]. Not all people are equally annoyed by these kinds of sounds. A recent study [54] proved that young adults found the high-pitched barks more annoying than other age groups. As with other noise sources, duration is also a very relevant factor related to the annoyance produced. Koffi [55] proved that after sound intensity in dBA, duration was the second most important determinant of annoyance.

Further studies have also established relationships between other individual noise sources that are not necessarily well represented by noise-level measurements, such as air traffic, floor impact, or drainage and overall dissatisfaction with indoor soundscapes in residences [56]. Therefore, the evaluation of the annoyance perceived in a soundscape and, by extension, its acoustic comfort can greatly be improved when the sound event present are known [32,57].

2.3 Psychoacoustic and non-acoustic factors

Some psychoacoustic factors have also been correlated with subjective annoyance judgments. Kim *et al.* [58] studied the psycho-acoustic effect of the level variation, the duration and the number of impacts on the floor and determined that the duration and total energy level are more suitable predictors than maximum sound pressure level when assessing the annoyance produced by children's impact sounds.

Psychoacoustic metrics (including loudness, sharpness, or roughness) have been used to predict the perceived noise annoyance. One publication by Orga *et al.* [31] used a multi-level psychoacoustic model that combined sharpness, roughness, impulsiveness, and tonality. However, approaches only based on psycho-acoustic metrics do not take into account the nature of the sound which can add emotional and cognitive variables that have an impact on the subjective assessment of the noise.

There are non-acoustic and non-sensory variables that clearly influence the subjective perception of the noise environment, including familiarity, preferences, or even expectations. Annoyance judgements by people revolve around an internal representation of the noise situation.

In many cases, the same noise level causes different degrees of annoyance depending on their occurrence during day or night-time. However, this does not happen for all kinds of noise sources. For example, no differences between day and night-time annoyance were found regarding traffic noise. On the contrary, reactions to rail or air traffic noise differ depending on the time of day [59].

Another study revealed that even in the case of railway and road traffic noise there are non-acoustical variables that explain part of the variance in noise annoyance beyond the noise indices (L_{den}). Some of these variables are the individual noise sensitivity, the coping capacity or the concern about the harmful effects [60].

Neighbourhood characteristics also modify the subjective perception of the soundscape. Surrounding greenery, especially garden and wetland parks usually reduce noise annoyance perception in the living environment [61,62]. Facade and building orientation are other influential factors in the perceived noise annoyance [63]. Even socioeconomic status is related to noise pollution perception. A study conducted in Germany [64] concluded that younger people and those with lower socioeconomic status have higher probabilities of being affected by noise pollution because they live in areas with more deteriorated soundscapes. However, it has also been stated that people with high socioeconomic status appear to be more noise-sensitive, maybe because they have higher expectations of quiet in the living environment [65].

There have been imaginative approaches to noise annoyance assessment that have combined acoustic and non-sensory variables. De Muer *et al.* [66] included the type of activity conducted along indoor background level and signal-to-noise ratio (SNR) measurements. Bravo-Moncayo *et al.* [30] used noise exposure levels but added other variables such as noise perception and demographics but focused only on road traffic noise annoyance. Finally, González *et al.* [67] combined meteorological and noise measurements, objective urban variables and *in situ* surveys to evaluate the effects of road traffic noise on pedestrians.

3 Materials and methods

As stated in Section 1, this project will be structured in two parts. The first set of experiments will give a binary assessment of the quality of the dwelling's soundscape. Next, the second set of experiments will offer an Acoustic Comfort Index (ACI), which uses a 5-point rating scale, for the same soundscape.

The following subsections detail the data gathering process framed in the *Sons al Balcó* project, the processing

of the data collected and the experimental pipeline designed including the setup for both estimators, i.e. the binary estimator and the ACI estimator.

3.1 Data gathering and *Sons al Balcó*

In this research article, the authors used data collected from the *Sons al Balcó* citizen science project. As a part of this project, two Catalonia-wide campaigns were conducted in 2020 and 2021. The project asked participants to make a double contribution. First, they had to record a short video of a minimum 30 s from their balconies using their smartphones or tablets and upload them to a server. Additionally, they had to answer a questionnaire about the perception of their soundscapes. The questions in the survey included a global subjective assessment of the soundscapes from their balconies (acoustic satisfaction), a description of the sound event present, and their respective level of annoyance or pleasantness according to their opinion, the frequency of appearance of the mentioned sound classes, and other useful information.

Even though the original goal of the project was to study the changes in the perceived soundscape during the lockdown, the data obtained are valuable and are currently being used with a broader scope in mind. For this current work, the dataset offers a combination of real-life video clips from living environments with a subjective assessment of the acoustic comfort by the dwellers themselves.

Both campaigns were advertised on social media and it was open to all people living in Catalonia. All the videos collected were manually reviewed to guarantee that three requirements were satisfied: (i) they should be recorded from a balcony or window, (ii) they should not contain human faces, and (iii) they should be recorded in Catalonia.

Data collected are very relevant as an important percentage of the *Sons al Balcó* contributions come from the biggest city in the region, Barcelona, which is particularly affected by noise pollution with over 210,000 people suffering serious psychological, emotional, or social effects caused by noise exposure and more than 60,000 with sleep disorders. Noise mapping by the Barcelona Public Health Agency reported that 57% of the population live in areas with traffic noise levels considered detrimental to health [68].

All the videos collected came from Catalonia minimizing the possible cultural differences in the subjective appreciation of the annoyance of noise sources. As these differences do exist [69], if the system were to be used in another cultural framework, it would be recommended to train the algorithm with locally acquired contributions.

3.2 Data processing

The first two campaigns of *Sons al Balcó* conducted in 2020 and 2021 received 365 and 237 contributions, respectively. Two complex polyphonic datasets were obtained from them, one for 2020 and one for 2021. Both datasets were manually labelled using a hierarchical taxonomy and analysed by the authors in the previous work [70]. They were annotated considering polyphony because there were frequent overlapping sound event.

For this present study, only the dataset of the 2021 campaign could be used. The 2020 campaign was handled during the lockdown caused by the COVID-19 pandemic in order to study the effects of the restrictions on the soundscape in Catalonia. This severe activity and mobility restrictions shaped a quieter sound environment in the cities across Catalonia [71,72]. In this context, the subjective perception of citizens during the lockdown also drastically changed to the point where there were virtually no examples of negative soundscapes reported in the 365 videos collected, making them unsuitable for the purpose of the current study.

In contrast, the 237 surveys answered in the 2021 campaign offer a realistic variety of positive, neutral, and negative scenarios. The videos from this campaign came from different spots representing a wide extension of the Catalan geography, as seen in Figure 1. About half of them were collected in big cities, especially in the metropolitan area of Barcelona, and the other half were collected in smaller cities or towns. Figure 1 shows that most of the scenarios reported as negative are found in large urban areas, a significant number of them in the metropolitan area of Barcelona, which suffers especially from noise pollution, as it was already stated in Section 3.1.

Of the 237 videos collected in 2021, two of them were discarded because they were too short. The mean duration of the actual videos collected was close to the specifications: 32.44 s. However, there were some outliers ranging from 3.2 to 85.2 s. Almost all of them could be used to characterize the soundscape, even if they were shorter, but 8 s were chosen as the minimum duration necessary to have enough relevant information on the scenario. Therefore, a total of 235 videos have been used in the present work. Almost all videos were recorded during the daytime. In fact, no video was recorded between 22:30 and 05:00. Consequently, video contents are more representative of daytime noise sources.

As shown in Figure 2, 50 participants (21.28%) deemed that the soundscape from their balconies has poor quality. On the other hand, 144 participants (61.28%) considered that they are surrounded by positive soundscapes.
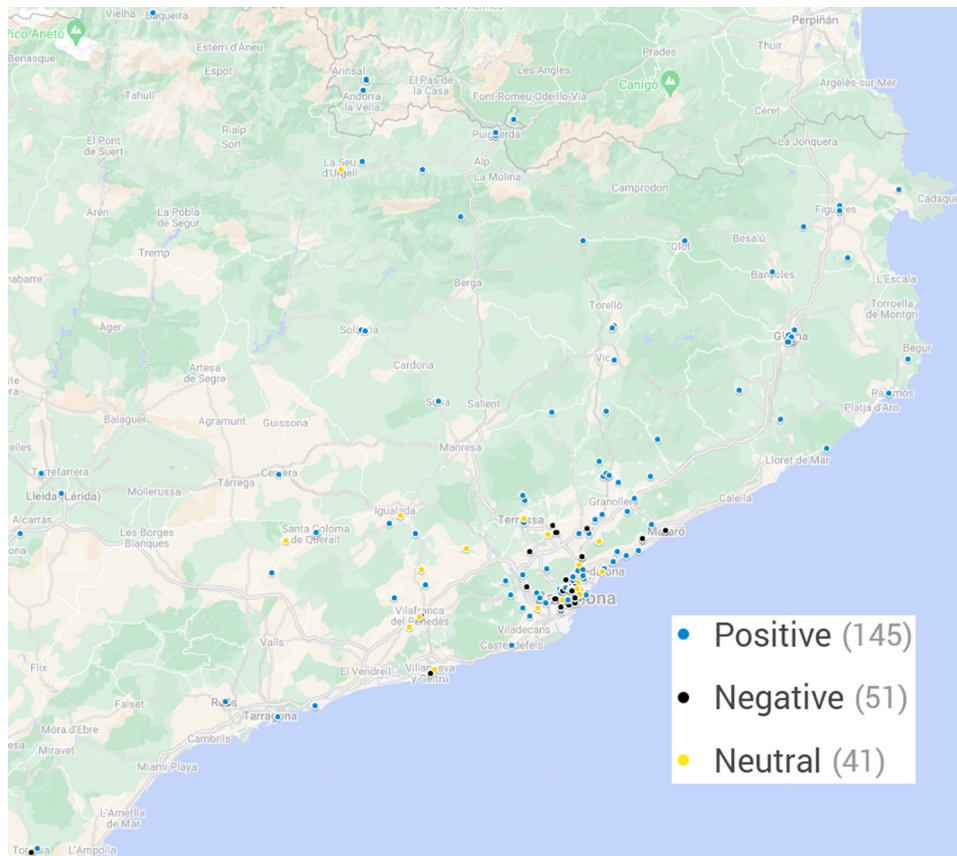
Figure 1: Distribution of the contributions for the 2021 *Sons al Balcó* campaign classified by global level of acoustic satisfaction (negative, neutral or positive).

3.3 Experimental pipeline

A two-stage soundscape quality predictor has been designed (Figure 3). The first stage consists of an automatic sound event classifier and the second stage includes two soundscape quality estimators. A process of aggregation and normalization followed by a predictor selection is applied between both stages.

The automatic sound event classifier used in Stage 1 is the same one previously published by the same authors [70] that had already been tested with the *Sons al Balcó* datasets. It is fed with a 30-s video that is subsequently framed using 30 ms Hamming windows. Then, 100 GammaTone Cepstral Coefficients (GTCC) are extracted [73]. GTCC were chosen as they outperformed other feature extraction methods in a survey comparison conducted by authors [74]. They are formatted into a 10 × 10 matrix which is what is expected by the deep learning (DL) algorithm chosen to detect and classify the sound classes. Specifically, a convolutional neural network (CNN) was chosen [75]. Finally, an array of 34 probabilities corresponding to the 34 classes in the taxonomy is binarized using a threshold

of 0.5 to obtain an array of 34 Booleans indicating which sound categories are detected in each frame. The exact setting of the classifier can be found in the aforementioned work [70].

Afterwards, the classified sound event detected in all the frames of a given video are aggregated and normalized
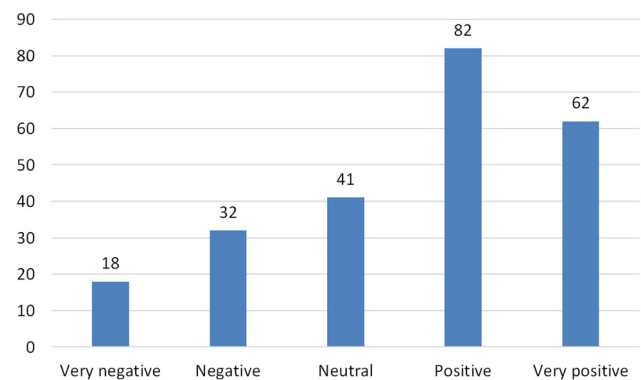

Figure 2: Global subjective assessment of the soundscape according to its dwellers (number of videos).
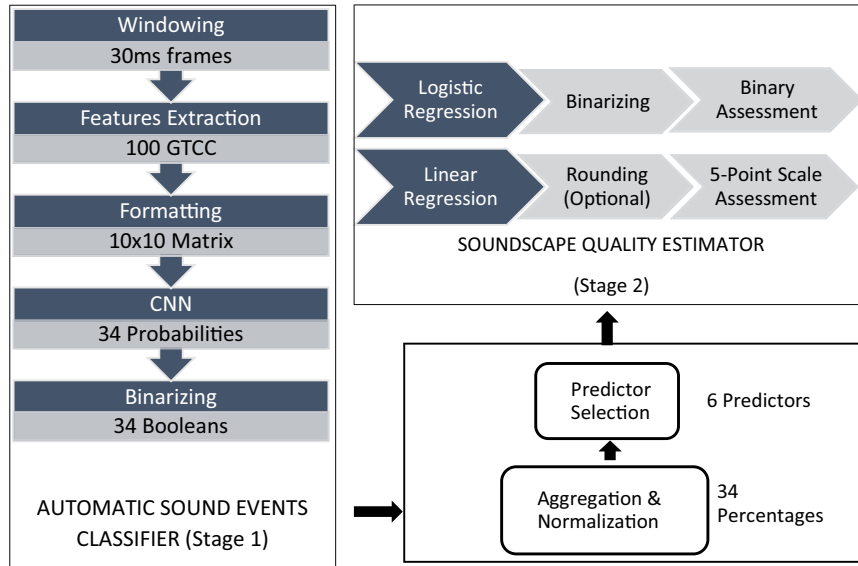
Figure 3: Two-stage soundscape quality predictor.

to obtain an array of percentages of presence for each sound class in the studied audio file.

Even though there are up to 30 different sound classes spotted in the 2021 dataset (the taxonomy is described in a previous work by the authors [70]), not all of them are suitable to be used as predictors. In fact, only those that have an impact on the prediction metrics of the assessment of the soundscape quality are considered. A hypothesis has been made that the less prevalent sounds would be irrelevant and that some sound classes that are not homogeneously considered pleasant nor unpleasant by the general population can be counterproductive as predictors. These hypotheses have been tested by comparing the performance of the estimator with different sets of predictors, starting with all the sound categories and subsequently removing the ones suspected to have a negative effect on the prediction. Particularly, the first sound categories removed were those with less than 1% of prevalence in the dataset according to Bonet-Solà *et al.* [70]. Next, the remaining sound categories were removed one by one, starting from the ones

less correlated (either positively or negatively) with the reported acoustic comfort as stated in Figure 4, until the performance started to improve.

Effectively, most of the sound classes with less than 5% of prevalence in the dataset are detrimental to the performance of the algorithm except for **rail** and, to some extent, **construction**. Furthermore, some of the sound classes with higher presence such as **wind** or **voice** are not correlated with the subjective perception of the quality of the soundscape. They appear in both positively and negatively perceived scenarios. In some instances, they contribute to a more negative assessment while in other instances they do the opposite. This fact can be observed in the correlation map shown in Figure 4. Thus, they are not reliable for prediction purposes. Consequently, the only five sound classes originally annotated in the dataset that are relevant for the present study and will be used as predictors are as follows: **bird**, **road traffic**, **rail**, **water**, and **construction**.

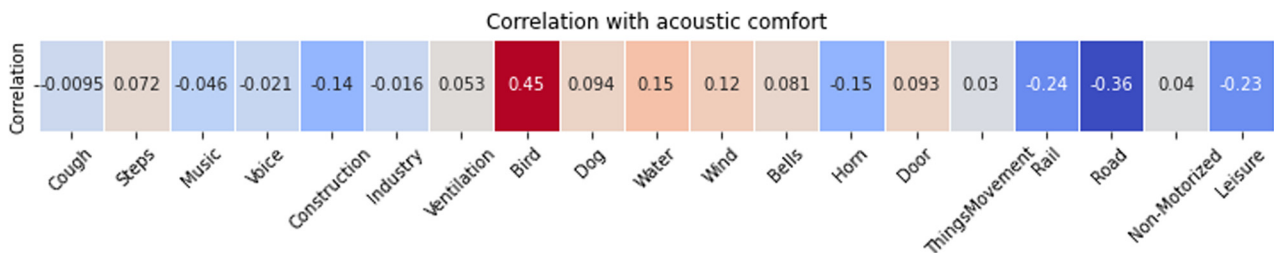Initial experiments and the analysis of the survey results revealed a new noise source with a significant

Figure 4: Correlation between the different predictors of the dataset (sound classes appearing at a minimum of four videos) and the acoustic comfort marked by the participants.

impact on the assessment of several soundscapes that was not previously annotated: leisure activities (especially, nightlife and restaurants). This noise source is a composite class made up of voices, music, and other basic sound event that, when they are integrated, are especially annoying compared to the individual sound classes mentioned. Therefore, the 2021 campaign dataset's labels were manually updated to add the **leisure** category as the sixth relevant sound class in the present study that will also be used as a predictor.

These sound event are consistent with the main noise sources detected in Barcelona in a previous study [76], which shows that road traffic is the main contributor to noise exposure in the city with more than 85% of exposure, followed by night-time leisure with less than 10%. The other noise sources detected are rail and industrial/construction noise, with a residual exposure below 2%.

Once all the videos were annotated and the predictors were chosen, an estimator of the quality of a given urban soundscape was designed (corresponding to stage 2 in Figure 3). The goal was to try to predict the subjective quality perceived by the participants using objective data, i.e. the specific noise sources present in the short video clips they sent. This estimator was initially tested using the real sound event manually annotated in the videos for the 2021 *Sons al Balcó* campaign to assess the performance of the estimator independently, without the possible error added by a classification algorithm.

Afterwards, the designed estimator was added to an automatic sound event classifier (Stage 1) to implement the two-stage system capable of automatically assessing the quality of the soundscape, as depicted in Figure 3.

Two rating scales were chosen to assess the level of acoustic comfort of the dwelling:
(1) A binary assessment
(2) A continuous 5-point rating scale (ACI).

For the first approach, the global subjective assessment of each contribution has been binarized. The dwellings rated as "very positive" or "positive" were assigned to the "comfortable" category. The dwellings rated as "very negative" or "negative" were assigned to the "uncomfortable" category. Finally, the dwellings rated as "neutral" were discarded for this first set of experiments. Thus, a total of 194 were finally available.

These soundscapes were divided in a 4-fold cross-validation train-test scheme, and a logistic regressor was implemented using only the six relevant sound classes already mentioned as predictors.

For the second approach, the system predicts the acoustic satisfaction score achieved by a dwelling with an ACI using a 5-point rating scale, which emulates the Likert scale [77] used by participants of the survey to assess the global perceived quality of their surroundings (very negative (1), negative (2), neutral (3), positive (4), and very positive (5)). This ACI offers a general approach to the overall acoustic satisfaction felt by the dwellers without focusing on specific perceptual constructs such as calmness, pleasantness, or monotony, which can be subject to different cultural interpretations [78].

Kang *et al.* proposed the creation of soundscape indices (SSID) [79–81] obtained from acoustical, psychoacoustical, psychological, neural, and physiological and contextual factors as a framework to better represent soundscapes and their perception. However, most of these factors require additional information not always available. The ACI presented in this work is a simplified and minimalist version of a single SSID, which only uses the sound source type as a defining factor.

In this case, all 235 valid videos from the 2021 campaign were used. They were also divided using a 4-fold cross-validation scheme. After that, a linear regressor was used to predict the soundscape's rating. Any outcome below 1 or above 5 was rounded to avoid exceeding the rating scale margins.

This assessment gives a real number between 1 and 5 that can be optionally rounded to obtain a discrete scale of 5 points identical to the Likert scale used by participants in the survey. To study the performance of this approach, the R-squared value of the prediction is computed and the error distance between the regressor's output and the subjective assessment is calculated. Subsequently, the accuracy is evaluated on a prediction interval of ±1 points.

4 Results

In this section, the results of the designed estimator are presented. Section 4.1 conveys the results obtained by the binary assessment. Afterwards, Section 4.2 exposes the results of the ACI estimation.

4.1 Binary assessment of the acoustic comfort

Four experiments were conducted. The first two experiments (Experiments 1 and 2) used only the estimator described as Stage 2 in Figure 3 which was fed with the sounds labelled by expert annotators. The last two experiments (Experiments 3 and 4) used the complete design with

Table 1: Accuracy and *F*1-score using the real annotated sounds to predict the subjective binary assessment of the soundscapes

	Accuracy (%)	*F*1-score (%)
Exp. 1 – Manual labels and segment based	80.41	64.15
Exp. 2 – Manual labels and event based	83.51	72.41

the classifier. In this case, the classifier automatically detects the sound event present in each video and feeds them to the soundscape quality estimator.

First (Experiment 1), a segment-based approach where the prediction was based on a binary array with the annotated/detected sounds on a given video was chosen. It used binary data, i.e. the presence or absence of each sound class in any given dwelling as independent variables, without considering the exact duration of each sound event. This could be interesting if the detection of sound event is accomplished with a segment-based classifier instead of an event-based one, which normally offers better performances in polyphonic environments. Experiment 2 opted for an event-based approach trying to detect the exact time frame and duration of each sound class event. It used the relative duration of each sound event within each video as independent variables; that is the percentage of time in each audio clip where a specific sound class is spotted.

The event-based approach achieves 3.1% higher accuracy and 8.26% higher *F*1-score than the segment-based one (Table 1). Experiments 1 and 2 showed the top performance that can be achieved with this kind of estimator in the current dataset. They can only be improved with additional information or by discarding the inconsistent entries in the survey.

The next two experiments aim to achieve performances as close as possible to the ones described in Table 1. The main change between them is that instead of using the manual labels, the acoustic events are obtained using an automatic CNN-based event detector. The automatic sound event classifier, even though it achieves state-of-the-art results when working with prevalent sounds such as birds or road traffic [82], is not perfect. Therefore, a dip in the accuracy is to be expected.

A detailed analysis of the performance of each class of the classifier (when working on an event-based metric) revealed that **water** was the only class that was often mixed up with non-related categories (25.86% of the time), i.e. it was often mixed up with generally annoying noise sources (**road traffic**, **rail**, **construction**, and **leisure**). Given that **water** was one of the less relevant categories considered in the prediction process, to begin with, this classifying under-performance is high enough to consider its removal from the estimator. For that reason, Experiment 4 was conducted only with five predictors (**birds**, **road traffic**, **rail**, **construction**, and **leisure**) achieving a slightly better performance than when **water** was included.

As can be seen in Table 2, the best performance is achieved in Experiment 4, which gave the same performance as Experiment 1. That means that inaccuracies due to the automating sound event detection (ASED) had only a 3.1% impact on the global accuracy. Due to the better performance achieved by the event-based detection (Experiment 4), this study will focus on this approach from now on.

In this subsection, the results for the ACI estimators are presented. First, the results without the ASED stage will be discussed. Afterwards, the results of the two-stage estimator will be explained and compared.

Without adding the automatic sound event classifier (and, therefore, using the manually annotated labels for each audio file), the R-squared value of the regression is not high: 0.28. However, the system offers a remarkably good accuracy if we accept a prediction interval of ±1 points, which is reasonable when trying to get a first approximation of the expected acoustic comfort or discomfort. The mean absolute error distance between the global assessment of the soundscape reported (with the 5-point rating scale) and the prediction is 0.85 points. If the index is rounded, the mean absolute error decreases even more, to 0.83 points.

4.2 ACI

As seen in Figure 5, in 86.81% of the soundscapes, the rounded assessment predicted is the same or has only 1 point of difference from the reported perception. In other words, 86.81% of soundscapes are correctly predicted inside the defined

Table 2: Accuracy and *F*1-score using automatically detected sound event to predict the subjective assessment of the soundscapes

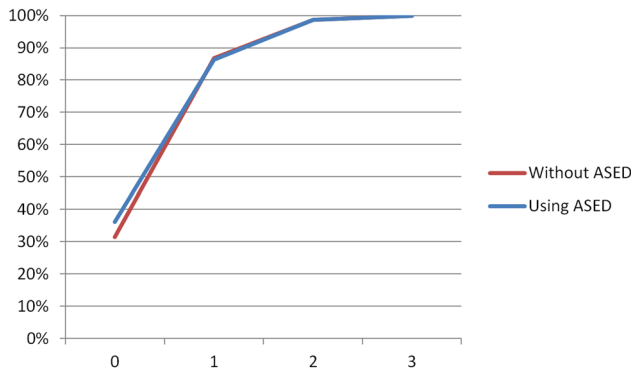|	Accuracy (%)	*F*1-score (%)
Exp. 3 – Automatic event detection and segment based	78.87	55.91
Exp. 4 – Automatic event detection and event based	80.41	64.15

Figure 5: Percentage of predictions with less than a given (absolute) error using the rounded ACI estimation.
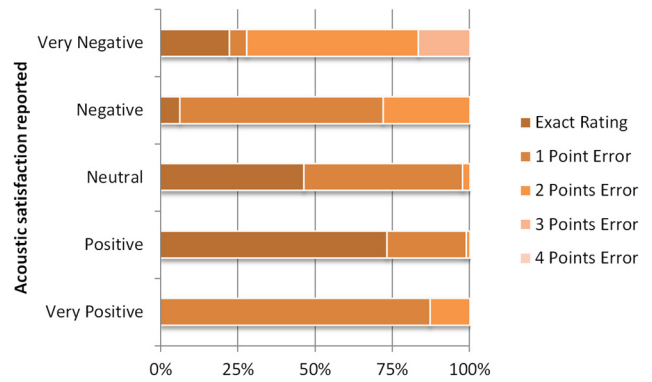

Figure 7: Prediction performance depending on the acoustic satisfaction reported.

prediction interval. Only 1.28% of the soundscapes offer a predicted index with more than 2 points of difference.

The standard deviation of the predictions is lower than the standard deviation of the reported perceptions (0.69 instead of 1.22). The system performs better in predicting middle indices and performs poorer in especially negative scenarios.

Results for the two-stage estimator are almost identical to the ones achieved with the estimator-only approach. The R-squared value is slightly diminished: 0.26. However, the mean absolute error distance between the reported assessment and the prediction is also reduced to 0.83 points (that falls even more to 0.79 points when the index is rounded). The accuracy for the prediction interval is exactly the same in both implementations (86.81%), as we can see in Figure 5. A slight improvement can be spotted in the number of perfect predictions (error = 0).

There are no significant differences in the error dispersion when using the ASED algorithm and when using manual labels. That can be further assessed in Figure 6. Median values are very close to 0 in both cases and first and third quartiles have a similar distance in both scenarios.
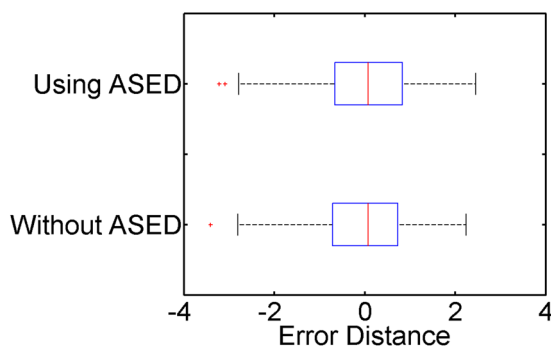
However, there is a slight asymmetry in the outliers which favour negative errors in both cases, especially in the scenario without ASED. A negative error means that the prediction describes the soundscape as less annoying that the ground truth expressed by the contributors. On the contrary, a positive error means that the prediction depicts a poorer scenario than the one stated by citizens.

The two-stage predictor performs better when predicting positive and neutral assessed soundscapes. As seen in Figure 7, when the acoustic satisfaction reported is "very negative" (the lowest), the performance is poorer. However, it must be stated that the number of soundscapes with a reported "very negative" rating is barely a 7.7% of the total (Figure 2).

The performance of the soundscape assessment depends on the size of the town/city in which it was taken. Even though the errors committed when using the 5-point rating system are smaller no matter the size of the city, Figure 8 proves that there is a vast difference depending on the rating system chosen when assessing small- and middle-sized cities (population ranging from 20,000 to 100,000). In fact, the 5-
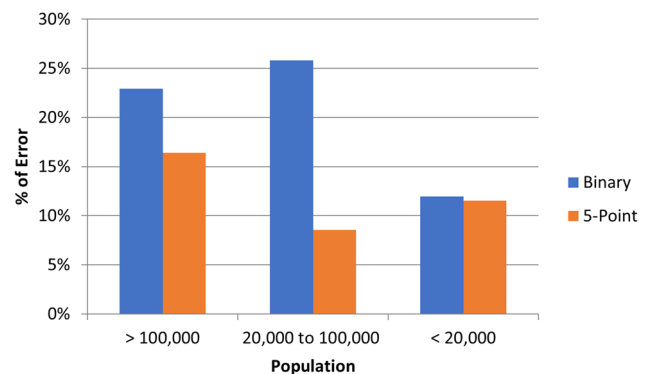

Figure 6: Comparison of the error distance for the ACI estimator with or without ASED.

Figure 8: Comparison of errors of both rating systems depending on the size of the city or town (errors in the 5-point-rated ACI are predictions outside the ±1 prediction interval).

point scale rating system stands out in small cities, even surpassing the predictions made in little towns. On the contrary, the binary rating system under-performs in this segment with an accuracy of slightly under 75%.

It is also interesting to note that the floor is inhabited by the participants, from which the videos were recorded, greatly influences the performance of the prediction. The difference is huge when using the binary assessment rating, as can be seen in Figure 9.

When considering only citizens living on floors 0 to 5 (that are approximately 85% of the participants), the accuracy rises to almost 85% with the binary assessment method. On the contrary, it drops to less than 60% for the minority living on higher floors.

5 Discussion

In this section, results are discussed and further developed. First, Section 5.1 offers a previous analysis of some relevant survey results to spot the hindrances to be considered when predicting a subjective opinion. Afterwards, a separate analysis of the accuracy according to the sounds annotated is done in Section 5.2. Finally, a detailed audit of those cases where the system failed is revealed (Section 5.3), starting with the binary assessment approach, and ending with a comparison with the 5-point scale ACI scheme.

5.1 Analysis of the survey results

Some outlier opinions are bound to be almost impossible to predict without further data. In fact, data based on a citizen science project can contain incoherent assertions and inconsistencies. Therefore, a previous inspection of
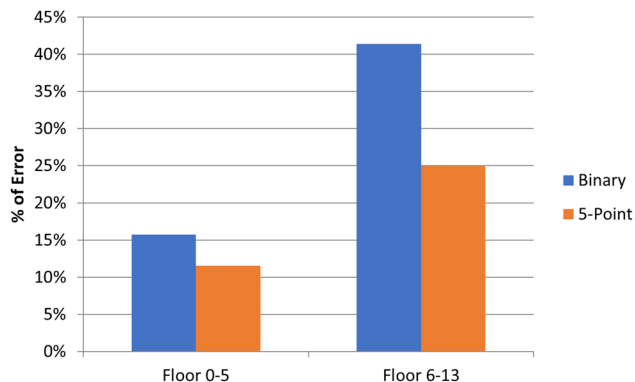
the survey results can give a clearer picture of the ceiling that can be achieved in the framework of this project.

The perception of the annoyance produced by individual predictors, the correlation or lack thereof between the annoyance reported for individual noise sources and the acoustic satisfaction reported for the dwelling, differences between reported sounds by participants and labelled sounds by annotators or the representativeness of the sound classes annotated in the videos could give valuable information to interpret the subsequent results.

5.1.1 Assessment of the perception of annoyance for individual predictors

While **birds** and **water** are almost unanimously considered as non-annoying or even pleasant sounds by participants, the assessment of the other four categories is less homogeneous. **Construction**, **leisure**, **rail**, and **road traffic** noise are normally considered annoying but there are numerous exceptions among the participants as seen in Figure 10. It has been proven that differences in impulsiveness, roughness, or tonality in some type of sound event influence the perceived annoyance [31]. This lack of consensus with the annoying noise sources betokens a higher difficulty in predicting poor quality soundscapes compared to the positive ones.

5.1.2 Correlation between the global assessment of the Dwelling's soundscape and the individual level of annoyance reported for the predictors

Inconsistencies between the subjective global assessment of the dwelling and the subjective assessment of each relevant class of noise source can affect the performance of the predictor. In most cases (89.45%), there is not a significant difference between both assessments, as seen in Table 3.

Figure 9: Comparison of errors of both rating systems depending on the floor on which the video was recorded (errors in the 5-point-rated ACI are predictions outside the ±1 prediction interval)).
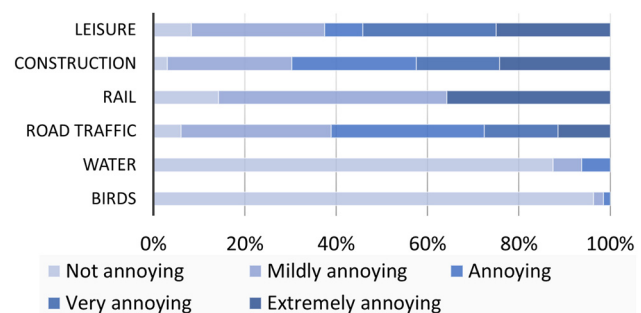

Figure 10: Individual assessment of the annoyance level for each of the six sound classes used as predictors.

Table 3: Comparison of the global negative assessment of the dwelling's soundscape provided by citizens and the mean value of the annoyance level for each of the six individual sound classes used as predictors

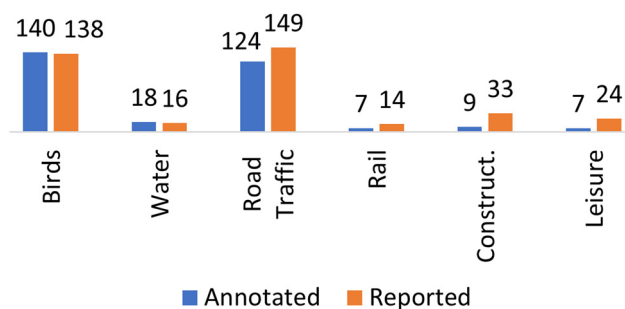|	Total answers
Perceived annoyance of individual sound classes similar to the acoustic discomfort of the soundscape	212 (89.45%)
Perceived annoyance of individual sound classes significantly higher than the global acoustic discomfort of the soundscape	10 (4.22%)
Perceived annoyance of individual sound classes significantly lower than the acoustic discomfort of the soundscape	15 (6.33%)

To be considered similar, both figures had to be less than 1.5 apart.

However, some discrepancies do exist. For 4.22% of the partakers, the perceived annoyance of the individual sound event present is significantly higher than the perceived acoustic discomfort of their residence. In some cases, this discrepancy can be explained by the low frequency of apparition of the detected noise sources. However, in other cases, the answers provided by respondents seem to be illogical or incoherent. That can be attributed to an incorrect interpretation of the questions among other causes that would be further discussed in Section 5.3.

On the other hand, 6.33% of the contributors stated that the perceived annoyance of the reported sound event present was significantly lower than the assessed acoustic discomfort of the soundscape in their dwellings. Even though a minimal part of this percentage corresponds to situations where the main noise source reported was not included in the predictors (such as neighbours or pets), some of the survey answers seem illogical, again, even after hearing the actual videos from the annotators. Further data that cannot be obtained directly from the videos could explain some of this divergence, e.g. the lack of representativeness of some of the videos sent.

5.1.3 Comparison between annotated and reported sound event

The automatic sound event classifier relies on the annotated sound event to do the training. Errors in the labelling process or discrepancies with the reported sound event by citizens affect the outcome of the classifier process and the subsequent prediction of the annoyance level. Therefore, it is interesting to compare if the annotated sounds are consistent with the reported sounds in the survey.

Figure 11 shows that some differences exist between the labelled sounds by annotators and the reported sounds by the participants in the survey. On the one hand, birds and water (not annoying sounds) were spotted in more videos by annotators than by participants even if the differences are not significant. On the other hand, construction, leisure, rail, and road traffic noise were reported in more videos by contributors than by annotators. Differences can be attributed to the quality of the recordings and to the subjective interpretation of each individual. However, some contributors may be biased in their responses as they know which sounds are normally present in their urban location, irrespective of their real apparition in the short videos sent.

Aggregating the six categories used as predictors (Table 4), it can be concluded that 88.4% of the time both labellers and contributors agree on the sound event appearing (or not appearing) in each video. By comparison, 8.23% of the sounds reported in the survey do not appear in the annotations. Finally, 3.38% of the sound event annotated were not reported by participants.

5.1.4 Representativeness of the sounds annotated in the videos

Table 5 shows the representativeness of the sounds annotated in the videos (only sounds that were both labelled by annotators and reported by contributors are being considered).

Figure 11: Comparison of the labelled sounds in the dataset and the reported sounds by the participants in the surveys.

Table 4: Crossover between labelled and reported sounds for the six predictors (aggregated)

	Not labelled (%)	Labelled (%)
Not reported	70.32	3.38
Reported	8.23	18.07

Table 5: Representativeness of the annotated and reported sound event

	Birds (%)	Water (%)	Road traffic (%)	Rail (%)	Construction (%)	Leisure (%)
Rare	3.51	55.56	13.27	0	37.5	14.29
Common	42.98	11.11	45.13	33.33	12.5	28.57
Very frequent	53.51	33.33	41.59	66.67	50	57.14

All instances of **rail** and almost all instances of **birds** are representative of the usual composition of their soundscapes. However, other sounds such as **water** and **construction** are over-represented in many videos according to the opinion of the participants. This over-representation of some noise sources may make them less relevant in the prediction of the subjective perception of the quality of some soundscapes, also affecting the performance of the estimator.

These four hindrances: heterogeneous assessment of the annoyance of individual noise sources, lack of correlation between individual noise sources and global assessment, inconsistencies in the reported and annotated sounds, and lack of representativeness of the sounds detected in the videos, limit the expected accuracy of the estimator. In Section 5, the actual errors caused by these factors are further discussed.

5.2 Effects of the type of sound source in the accuracy

As seen in Table 2, the global accuracy achieved by the two-stage implementation exceeds 80%. However, the reliability of the prediction varies based on the kind of sound present in each location. The analysis made in Section 5.1.1 with Figure 10 already hinted at this outcome. The system excels in correctly assessing the quality of the videos that only have pleasant predictors annotated (birds or water) and the videos that do not have any of the predictors because they only have other sounds labelled (such as music or pets). Table 6 shows that the performance for these videos climbs to more than 90%.

Table 6: Reliability of the prediction depending on the sound sources present at each location

Sounds present in the video	Accuracy (%)
None of the predictors present	91.67
Only pleasant sound sources present (birds, water)	90.77
Only annoying sound sources present (road traffic, rail, construction, leisure)	80.7
Both pleasant and annoying sound sources present	66.67

Performance is also slightly above average for those videos with only annoying sound sources present with an accuracy of 80.7%. However, accuracy decreases below 70% of those videos where both pleasant and annoying sound sources co-exist.

5.3 Errors analysis

It is interesting to analyse the causes of the incorrectly assessed soundscapes. Starting with the binary assessment, a total of 156 soundscapes (more than 80%) were correctly predicted. However, the system failed to match the subjective perception of the participants in 38 instances. Figure 12 shows the causes for each of these errors.

A 1.03% of the errors (Type I) were caused by inconsistencies in the survey responses. The global assessment assigned by the participants was incoherent. The reasons could be diverse and caused by the subjective nature of the project. They may be due to a misunderstanding of the questions in the survey, a lack of commitment to the veracity of the answers or an extremely outlier opinion in the assessment of the quality of the soundscape on the part of the contributor. In this present study, all these errors consisted of soundscapes reported as negative that were incorrectly predicted as positive by the estimator. This 1.03% is completely unpredictable and can only be removed by a previous screening of the survey.

A 5.15% (Type II) of the errors consisted in videos where annoying noise sources (such as road traffic) were present or even predominant but the quality of the soundscape reported was positive, nevertheless. Therefore, a negative soundscape was predicted instead of a positive one. There are several reasons that explain this situation. First, it may be possible that the presence of these particular noise sources was exceptional and not representative of the everyday soundscape. As seen in Table 5, there is a significant percentage of annoying sound event present in the videos that are rare in the studied locations. Therefore, taking this single video as an example to assess the quality of the soundscape is not appropriate. This issue can be tackled by analysing more than one video taken from the same location instead of only one. Another reason for this
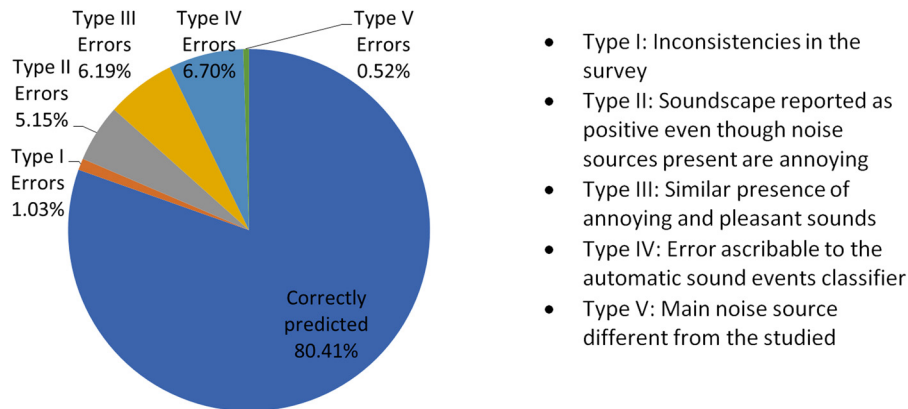
- Type I: Inconsistencies in the survey
- Type II: Soundscape reported as positive even though noise sources present are annoying
- Type III: Similar presence of annoying and pleasant sounds
- Type IV: Error ascribable to the automatic sound events classifier
- Type V: Main noise source different from the studied

Figure 12: Error analysis for the binary assessment of the soundscape.

kind of error is a different subjective appreciation of the level of annoyance of these particular noise sources (road traffic noise, leisure…) by the participants. As seen in Figure 10, even though most people consider road traffic, rail, construction, or leisure as annoying, very annoying, or even extremely annoying, there are also some participants who do not consider them annoying at all. The foreground-to-background placement of the sounds or the noise isolation of the building can explain some of these differences in appreciation. However, other factors such as socioeconomic status, demographics, time slot of occurrence, activities developed by residents, percentage of time at home, coping capacity, or expectations can also play a role. Soundscape appropriateness also has a role in positive appraisal of heavy traffic areas where road traffic is the main noise source present [83]. In any case, it should be noted that when dealing with subjective contributions, some outlier opinions are almost impossible to predict no matter the model used.

A significant part of these Type II errors appears in videos recorded from higher floors, which partly explains the differences revealed in Figure 9. On higher floors, the annoying noise sources still are present and detected by the ASED algorithm, but they are not as annoying to the inhabitants. Accurate measuring of the L_{Aeq} in the studied floor could be helpful in improving the prediction performance in this case. However, as it was already stated, this situation only affects a small part of the population (15% in the *Sons al Balcó* sample), as most people live on lower floors.

A 6.19% of the errors (Type III) occurred because a similar presence of annoying noise sources and pleasant sound event were equally present in the spot. The subjective appreciation of this situation is especially variable, and the algorithm will miss about a third of the situations (as seen in Table 6) if the only criteria to assess the quality is the detection of the sound classes present. Errors can go both ways: a negative soundscape is predicted as positive or otherwise. To improve the performance, the assessment of these soundscapes should be complemented with other data such as the L_{Aeq} measured in the location (when it is available). If noise levels are not available, another valid alternative could be using psychoacoustic metrics extracted from the sounds detected.

Table 7: Comparison of errors made by both rating systems

Type of error	Errors in the binary rating	Errors in the ACI	Improvement
Type I: Inconsistencies in the survey	2	2	0%
Type II: Soundscape reported as positive even though noise sources are annoying	10	2	80%
Type III: Similar presence of pleasant and annoying sound sources	12	8 (+1 not previously existing)	33.33%
Type IV: Errors ascribable to the ASED algorithm	13	7	46.15%
Type V: Noise source different from the studied	1	1	0%
Type VI: Errors in the assessment of extreme ratings (1 or 5) predicted as middle	—	11	Not applicable

Types I, II, and III errors are exceedingly difficult to improve using only this approach to make the decision, marking an accuracy ceiling of 87.63%.

There are two more kinds of errors. On the one hand, a 6.7% of the videos were incorrectly predicted (Type IV errors) due to mistakes committed by the classifier (incorrect detection or classification of some sound event). A detailed analysis of these errors showed that in four audio clips, water was incorrectly detected as road traffic noise leading to a negative assessment of a positive soundscape. Moreover, even though the algorithm performs exceptionally well in detecting rail noises, a slight confusion to other sounds considered pleasant leads to a positive assessment of a negative soundscape. As ASED algorithms are continuously improving their accuracy, it is conceivable that these errors could diminish, especially if the classifier can be trained with more extensive data.

On the other hand, 0.52% of the errors (Type V) were caused by the presence of a noise source different from the ones used as predictors. As a result, the soundscape was predicted as positive although it had poor quality. There were not enough samples in the *Sons al Balcó* project to include this noise source (Industry) as a predictor with positive results. However, a broader collection of data in future campaigns would solve this specific issue.

In order to compare the performance of both ratings (binary and ACI), it can be considered that an error of 2 or more points in the rounded ACI is a poor assessment. Following this criterion, 86.38% of the soundscapes assessed with the ACI can be considered good predictions (inside the ±1 prediction interval) and 13.62% can be considered poor predictions. This outperforms the binary assessment accuracy of 80.41% even though there are more videos assessed using the ACI than using the binary one (235 to 194).

To better understand the improvement of the ACI over the binary one, Table 7 compares the errors of the latter with the poor predictions of the former.

Errors derived from inconsistencies in the survey or from noise sources different from the predictors cannot be solved with the 5-point rating system. On the contrary, the other types of errors described in Figure 12 are significantly improved, especially Type II errors. A binary categorization of this kind of location was not the best suited. Even some of the errors ascribable to the ASED algorithm can be avoided when using the 5-point scale rating. In fact, all the situations where the rail was detected slightly mixed with pleasant sounds are correctly predicted as negative locations with a the 5-point rating.

However, it must be noted that the rating underperforms with extreme appraisals ("very negative" or "very positive"). It struggles especially with some locations rated as "very negative" by citizens that are upgraded two points by the predictor.

6 Conclusion

This article presents a system capable of predicting the acoustic satisfaction level for a dwelling based only on the ASED of a short video. Although both a segment-based approach and an event-based approach have been tested, the prediction based on events is preferable. The improved accuracy of the ASED algorithm in segment-based metrics does not make up for the less information obtained with that approach.

Accuracies obtained are good and encouraging (topping 80% or even better depending on the rating system used). However, for an even higher reliability, it is recommended to add additional information of the noise exposure when available, such as the mean L_{Aeq} level measured on the spot. The reason is that the system has a ceiling of accuracy that can hardly be surpassed without further information, mainly due to several hindrances related to the subjective nature of the study and the representativeness of the videos used.

The binary assessment alternative achieved particularly remarkable accuracies when assessing soundscapes from floors 0 to 5 reaching almost 85%. That figure is especially relevant considering that only a small fraction of the Catalan population has its residence in the upper floors. The performance would probably take a dip in regions cluttered with skyscrapers, making it less suitable. The population of the studied city or town also affects the performance of both assessments. For small- and medium-sized cities the 5-point scale ACI approach is recommended.

This implementation works comparatively better in predicting pleasant scenarios (with less annoying noise exposure) than in predicting deteriorated soundscapes. That was to be expected as opinions on the annoyance (or lack thereof) of pleasant sounds such as birds or water are homogeneous. However, there is a clear lack of consensus in the assessment of the level of annoyance for other less agreeable noise sources: road traffic, train, leisure, and construction that give place to outlier opinions difficult to predict.

In general, the ACI approach makes even better predictions, offering a more nuanced assessment, taking into account that the prediction should be interpreted inside a ±1 interval. However, it tends to be conservative in the assessment of extreme soundscapes ("very negative" or "very positive") with a reduced standard deviation and a bias towards a neutral assessment.

As it takes a little extra effort to extract both ratings, it is advised to use both of them to have a more precise assessment. The ACI option is especially valid and error-free (within the sample of the study) when the rail is detected. However, the number of occurrences of rail in the videos collected is insufficient to make a general assertion.

The videos collected for the *Sons al Balcó* campaigns only include daytime soundscapes, which is a recurrent problem in this kind of citizen science project. It would be interesting to also collect videos depicting night-time scenarios to better evaluate the impact of night recreational activities on the acoustic comfort of the dwelling.

The approach proposed can be very useful to make a first approximation of the perceived acoustic comfort in urban areas without needing expensive dedicated equipment to do so. As short videos can be recorded with a mobile phone, everyone is able to easily upload or send the video without technical expertise. It can also be used by municipal technical staff complementary to other noise surveillance techniques (such as sound meters) to more accurately map the subjective noise exposure in a city (or town).

References

[1] Agency EE. Environmental noise in Europe, 2020. European Environment Agency; 2020.

[2] Khomenko S, Cirach M, Barrera-Gómez J, Pereira-Barboza E, Iungman T, Mueller N, et al. Impact of road traffic noise on annoyance and preventable mortality in European cities: A health impact assessment. Environ Int. 2022;162:107160.

[3] Petri D, Licitra G, Vigotti MA, Fredianelli L. Effects of exposure to road, railway, airport and recreational noise on blood pressure and hypertension. Int J Environ Res Public Health. 2021;18(17):9145.

[4] Thacher JD, Poulsen AH, Raaschou-Nielsen O, Hvidtfeldt UA, Brandt J, Christensen JH, et al. Exposure to transportation noise and risk for cardiovascular disease in a nationwide cohort study from Denmark. Environ Res. 2022;211:113106.

[5] Pyko A, Roswall N, Ögren M, Oudin A, Rosengren A, Eriksson C, et al. Long-term exposure to transportation noise and ischemic heart disease: A pooled analysis of nine Scandinavian cohorts. Environ Health Perspectives. 2023;131(1):017003.

[6] Smith MG, Cordoza M, Basner M. Environmental noise and effects on sleep: an update to the WHO systematic review and meta-analysis. Environ Health Perspectives. 2022;130(7):076001.

[7] Taoussi AA, Yassine AsA, Malloum MSM, Assi C, Fotclossou T, Ali YA. Effects of noise exposure among industrial workers in power plants of the National Electricity Company in N'Djamena, Chad. The Egypt J Otolaryngol. 2022;38(1):63.

[8] van Kamp I, Davies H. Environmental noise and mental health: Five year review and future directions. In: 9th International Congress on Noise as Public Health Problem (ICBEN) - Foxwoods, CT; 2008.

[9] Tortorella A, Menculini G, Moretti P, Attademo L, Balducci PM, Bernardini F, et al. New determinants of mental health: The role of noise pollution. A narrative review. Int Rev Psychiatry. 2022;34(7–8):783–96.

[10] Rossi L, Prato A, Lesina L, Schiavi A. Effects of low-frequency noise on human cognitive performances in laboratory. Building Acoustics. 2018;25(1):17–33.

[11] Vukiccc L, Mihanoviccc V, Fredianelli L, Plazibat V. Seafarers' Perception and attitudes towards noise emission on board ships. Int J Environ Res Public Health. 2021;18(12):6671.

[12] Minichilli F, Gorini F, Ascari E, Bianchi F, Coi A, Fredianelli L, et al. Annoyance judgment and measurements of environmental noise: a focus on Italian secondary schools. Int J Environ Res Public Health. 2018;15(2):208.

[13] Thompson R, Smith RB, Karim YB, Shen C, Drummond K, Teng C, et al. Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence. Environ Int. 2022;158:106905.

[14] Preisendörfer P, Liebe U, Enzler HB, Diekmann A. Annoyance due to residential road traffic and aircraft noise: Empirical evidence from two European cities. Environ Res. 2022;206:112269.

[15] Wang Q, Hongwei W, Cai J, Zhang L. The multi-dimensional perceptions of office staff and non-office staff about metro noise in commercial spaces. Acta Acustica. 2022;6:15.

[16] Radun J, Maula H, Rajala V, Scheinin M, Hongisto V. Acute stress effects of impulsive noise during mental work. J Environ Psychol. 2022;81:101819.

[17] Pal J, Taywade M, Pal R, Sethi D. Noise pollution in intensive care unit: a hidden enemy affecting the physical and mental health of patients and caregivers. Noise Health. 2022;24(114):130.

[18] Désiré SSM, Ngum TM, Mambo AD, Lawrence F, Fogam BN, Landry NSJ. Effects of noise pollution on learning in schools of Bamenda II municipality, Northwest region of Cameroon. In: Mambo AD, Gueye A, Bassioni G, editors. Innovations and Interdisciplinary Solutions for Underserved Areas (InterSol 2022). Cham, Switzerland: Springer; 2022. p. 3–15.

[19] Tomek R, Urhahne D. Effects of student noise on student teachers' stress experiences, concentration and error-correction performance. Educat Psychol. 2022;42(1):64–82.

[20] Nagarnaik PB, Mohitkar V, Parbat DK. Evaluation of noise pollution annoyance at uninterrupted traffic flow condition. 2011 Fourth International Conference on Emerging Trends in Engineering & Technology; 2011 Nov 18–20; Port Louis, Mauritius. IEEE, 2012. p. 156–63.

[21] Lemaitre G, Aubin F, Lambourg C, Lavandier C. How does the train background noise affect passengers' activities? Determining thresholds of noise levels ensuring a good comfort for passengers.

World Congress on Railway Research (WCRR); 2022 Jun 6–10; Birmingham, UK.

[22] Cardoso M, Quintas M, Tavares D. Observational study on the influence of noise pollution on the quality of sleep of Porto residents compared with that of rural communities. Territorium. 2023;(30 (I)):107–14.

[23] Bonet-Solà D, Vidaña-Vila E, Alsina-Pagès RM. Predicting the perceptual rating of a soundscape using artificial intelligence. In: 25th International Conference of the Catalan Association for Artificial Intelligence (CCIA 2023); 2023 Oct 25–27; Barcelona, Spain. p. 287–8.

[24] Mitchell A, Oberman T, Aletta F, Kachlicka M, Lionello M, Erfanian M, et al. Investigating urban soundscapes of the COVID-19 lockdown: A predictive soundscape modelling approach. J Acoustic Soc America. 2021;150(6):4474–88.

[25] Sadeghian M, Shekarizadeh S, Abbasi M, Mousavi SM, Yazdanirad S. The use of artificial neural networks to predict tonal sound annoyance based on noise metrics and psychoacoustics parameters. Noise Control Eng J. 2022;70(4):309–22.

[26] Aletta F, Axelsson Ö, Kang J. Towards acoustic indicators for soundscape design. Forum Acusticum; 2014 Sep 7–12; Kraków, Poland. European Acoustics Association, 2014.

[27] Vardaxis NG, Bard D, Persson Waye K. Review of acoustic comfort evaluation in dwellings-part I: Associations of acoustic field data to subjective responses from building surveys. Building Acoustics. 2018;25(2):151–70.

[28] Alsina-Pagès RM, Orga F, Mallol R, Freixes M, Baño X, Foraster M. Sons al balcó: soundscape map of the confinement in Catalonia. Eng Proc. 2020;2(1):77.

[29] Baño X, Bergadà P, Bonet-Solà D, Egea A, Foraster M, Freixes M, et al. Sons al Balcó, a citizen science approach to map the soundscape of Catalonia. Eng Proc. 2021;10(1):54.

[30] Bravo-Moncayo L, Lucio-Naranjo J, Chávez M, Pavón-García I, Garzón C. A machine learning approach for traffic-noise annoyance assessment. Appl Acoustic. 2019;156:262–70.

[31] Orga F, Mitchell A, Freixes M, Aletta F, Alsina-Pageees RM, Foraster M. Multilevel annoyance modelling of short environmental sound recordings. Sustainability. 2021;13(11):914–21.

[32] Kang J, Aletta F, Margaritis E, Yang M. A model for implementing soundscape maps in smart cities. Noise Mapping. 2018;5(1):46–59.

[33] Titu AM, Boroiu AA, Mihailescu S, Pop AB, Boroiu A. Assessment of road noise pollution in urban residential areas-a case study in Pitesssti, Romania. Appl Sci. 2022;12(8):4053.

[34] Gheibi M, Karrabi M, Latifi P, Fathollahi-Fard AM. Evaluation of traffic noise pollution using geographic information system and descriptive statistical method: a case study in Mashhad, Iran. Environ Sci Pollution Res. 2022;29:1–14.

[35] Kalisa E, Irankunda E, Rugengamanzi E, Amani M. Noise levels associated with urban land use types in Kigali, Rwanda. Heliyon. 2022;8(9):e10653.

[36] Rendón J, Gómez DMM, Colorado HA. Useful tools for integrating noise maps about noises other than those of transport, infrastructures, and industrial plants in developing countries: Casework of the Aburra Valley, Colombia. J Environ Manag. 2022;313:114953.

[37] Ahmed S, Gadelmoula A. Industrial noise monitoring using noise mapping technique: a case study on a concrete block-making factory. Int J Environ Sci Technol. 2022;19(2):851–62.

[38] Baffoe PE, Duker AA, Senkyire-Kwarteng EV. Assessment of health impacts of noise pollution in the Tarkwa Mining Community of

Ghana using noise mapping techniques. Global Health J. 2022;6(1):19–29.

[39] de Souza TB, Alberto KC, Barbosa SA. Evaluation of noise pollution related to human perception in a university campus in Brazil. Appl Acoustics. 2020;157:107023.

[40] Alsina-Pagès RM, Freixes M, Orga F, Foraster M, Labairu-Trenchs A. Perceptual evaluation of the citizenas acoustic environment from classic noise monitoring. Cities Health. 2021;5(1–2):145–9.

[41] Di G, Wang Y, Yao Y, Ma J, Wu J. Influencing Factors Identification and Prediction of Noise Annoyance-A Case Study on Substation Noise. Int J Environ Res Public Health. 2022;19(14):8394.

[42] Abbaszadeh MJ, Madani R, Ghaffari A. Effects of non-acoustic factors on noise annoyance in apartment buildings (case study: Aseman-E Tabriz residential complex). Iran Univ Sci Technol. 2022;32(1):1–12.

[43] Ouis D. Annoyance from road traffic noise: a review. J Environ Psychol. 2001;21(1):101–20.

[44] Wang J, Wang X, Yuan M, Hu W, Hu X, Lu K. Deep Learning-Based Road Traffic Noise Annoyance Assessment. Int J Environ Res Public Health. 2023;20(6):5199.

[45] Maschke C, Niemann H. Health effects of annoyance induced by neighbour noise. Noise Control Eng J. 2007;55(3):348–56.

[46] Jahangeer F. Acoustic comfort in the living environment and its association with noise representation: A systematic review. SHS Web Confer. 2019;64:03012.

[47] Dümen AS, Rasmussen B. Neighbour noise in multi-storey housing with poor sound insulation-facts and occupants' viewpoints. Forum Acusticum; 2023 Sep 11–15; Turin, Italy. European Acoustics Association, 2023.

[48] Ottoz E, Rizzi L, Nastasi F. Recreational noise: impact and costs for annoyed residents in Milan and Turin. Appl Acoustics. 2018;133:173–81.

[49] Feder K, Marro L, Portnuff C. Leisure noise exposure and hearing outcomes among Canadians aged 6 to 79 years. Int J Audiol. 2022:1–17.

[50] Lee SC, Hong JY, Jeon JY. Effects of acoustic characteristics of combined construction noise on annoyance. Building Environ. 2015;92:657–67.

[51] Van Kamp I, van Kempen E, Simon S, Baliatsas C. Review of evidence relating to environmental noise exposure and annoyance, sleep disturbance, cardio-vascular and metabolic health outcomes in the context of the interdepartmental group on costs and benefits noise subject group (IGCB (N)). The Netherlands: National Institute for Public Health and the Environment; 2020.

[52] Wang Y, Wang G, Li H, Gong L, Wu Z. Mapping and analyzing the construction noise pollution in China using social media platforms. Environ Impact Assessment Review. 2022;97:106863.

[53] Mendonça C, Arruda A, Mesquita C, Couto R, Sousa V, Dogs barking and babies crying: the effect of environmental noise on physiological state and cognitive performance. SSRN Elsevier. Available at SSRN 4218179.

[54] Jégh-Czinege N, Faragó T, Pongrácz P. A bark of its own kind - the acoustics of 'annoying' dog barks suggests a specific attention-evoking effect for humans. Bioacoustics. 2020;29(2):210–25.

[55] Koffi E. Infant cry annoyance scale and indexes. Linguistic Portfolios. 2023;12:3.

[56] Jeon JY, Ryu JK, Lee PJ. A quantification model of overall dissatisfaction with indoor noise environment in residential buildings. Appl Acoustics. 2010;71(10):914–21.

[57] Aletta F, Kang J, Axelsson Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape Urban Plann. 2016;149:65–74.

[58] Kim S, Kim J, Lee S, Song H, Song M, Ryu J. Effect of temporal pattern of impact sound on annoyance: children's impact sounds on the floor. Building Environ. 2022;208:108609.

[59] Hoeger R, Schreckenberg D, Felscher-Suhr U, Griefahn B. Night-time noise annoyance: state of the art. Noise Health. 2002;4(15):19–25.

[60] Pennig S, Schady A. Railway noise annoyance: exposure-response relationships and testing a theoretical model by structural equation analysis. Noise Health. 2014;16(73):388–99.

[61] Li HN, Chau CK, Tang SK. Can surrounding greenery reduce noise annoyance at home? Sci Total Environ. 2010;408(20):4376–84.

[62] Gidlöf-Gunnarsson A, Öhrström E. Noise and well-being in urban residential environments: the potential role of perceived availability to nearby green areas. Landscape Urban Plan. 2007;83(2):115–26.

[63] Eggenschwiler K, Heutschi K, Taghipour A, Pieren R, Gisladottir A, Schäffer B. Urban design of inner courtyards and road traffic noise: influence of façade characteristics and building orientation on perceived noise annoyance. Building Environ. 2022;224:109526.

[64] Von Szombathely M, Albrecht M, Augustin J, Bechtel B, Dwinger I, Gaffron P, et al. Relation between observed and perceived traffic noise and socio-economic status in urban blocks of different characteristics. Urban Sci. 2018;2(1):20.

[65] Meijer H, Knipschild P, Sallé H. Road traffic noise annoyance in Amsterdam. Int Archives Occupat Environ Health. 1985;56(4):285–97.

[66] De Muer T, Botteldooren D, De Coensel B, Berglund B, Nilsson M, Lercher P. A model for noise annoyance based on notice-events. INTER-NOISE 2005: Environmental Noise Control; 2005 Aug 7–10; Rio de Janeiro, Brazil.

[67] González DM, Morillas JMB, Rey-Gozalo G. Effects of noise on pedestrians in urban environments where road traffic is the main source of sound. Sci Total Environ. 2023;857:159406.

[68] Font L, Gómez A, Oliveras L, Realp E, Borrell C. Soroll Ambiental i Salut a la Ciutat de Barcelona. Barcelona, Spain: Agència de Salut Pública de Barcelona; 2022.

[69] Namba S, Kuwano S, Schick A, Accclar A, Florentine M, Rui ZD. A cross-cultural study on noise problems: comparison of the results obtained in Japan, West Germany, the U.S.A., China and Turkey. J Sound Vibrat. 1991;151(3):471–7.

[70] Bonet-Solà D, Vidaña-Vila E, Alsina-Pagès RM. Analysis and acoustic event classification of environmental data collected in a citizen science project. Int J Environ Res Public Health. 2023;20(4):3683.

[71] Alsina-Pagès R, Bergadà P, Martínez-Suquía C. Sounds in Girona during the COVID Lockdown. J Acoustic Soc America. 2021;149:3416.

[72] Bonet-Solà D, Martínez-Suquía C, Alsina-Pagès RM, Bergadà P. The soundscape of the COVID-19 lockdown: Barcelona noise monitoring network case study. Int J Environ Res Public Health. 2021;18(11).

[73] Valero X, Alías F. Gammatone Cepstral coefficients: biologically inspired features for non-speech audio classification. IEEE Trans Multimedia. 2012;14(6):1684–9.

[74] Bonet-Solà D, Alsina-Pagès R. A comparative survey of feature extraction and machine learning methods in diverse acoustic environments. Sensors. 2021;21(4):1274.

[75] Li Z, Liu F, Yang W, Peng S. A survey of convolutional neural networks: analysis, applications, and prospects. 2020. arXiv:200402806.

[76] Ajuntament de Barcelona. Barcelona Resilience Plan Diagnosis; 2020.

[77] Likert R. A technique for the measurement of attitudes. Archives of Psychology. New York: American Physchological Association; 1932.

[78] Aletta F, Oberman T, Mitchell A, Kang J, Consortium S. Preliminary results of the soundscape attributes translation project (SATP): lessons learned and next steps. Forum Acusticum; 2023 Sep 11–15; Turin, Italy. European Acoustics Association, 2023.

[79] Kang J, Aletta F, Oberman T, Erfanian M, Kachlicka M, Lionello M, et al. Towards soundscape indices. In: Proceedings of the International Congress on Acoustics; 2019 Sep 9–13; Aachen, Germany. p. 2488–95.

[80] Mitchell A, Oberman T, Aletta F, Erfanian M, Kachlicka M, Lionello M, et al. The soundscape indices (SSID) protocol: a method for urban soundscape surveys-questionnaires with acoustical and contextual information. Appl Sci. 2020;10(7):2397.

[81] Kang J, Aletta F, Oberman T, Mitchell A, Erfanian M. Subjective evaluation of environmental sounds in context-towards Soundscape Indices (SSID). Forum Acusticum; 2023 Sep 11–15; Turin, Italy. European Acoustics Association, 2023.

[82] Bonet-Solà D, Vidaña-Vila E, Alsina-Pagès RM. Analysis and acoustic event classification of environmental data collected in Sons al Balcó Project. Forum Acusticum; 2023 Sep 11–15; Turin, Italy. European Acoustics Association, 2023.

[83] Tan JKA, Lau SK, Hasegawa Y. The effects of aural and visual factors on appropriateness ratings of residential spaces in an urban city. INTER-NOISE and NOISE-CON Congress and Conference Proceedings; 2021 Washington DC, USA. Institute of Noise Control Engineering, 2021. p. 5314–26.