A systematic review of prediction models for the experience of urban soundscapes

Matteo Lionello*, Francesco Aletta*, Jian Kang*

Institute for Environmental Design and Engineering, The Bartlett, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK

Abstract

A systematic review for soundscape modelling methods is presented. The methods for developing soundscape models are hereby questioned by investigating the following aspects: data acquisition methods, indicators used as predictors of descriptors in the models, descriptors targeted as output of the models, linear rather than non-linear model fitting, and overall performances. The inclusion criteria for the reviewed studies were: models dealing with soundscape dimensions aligned with the definitions provided in the ISO 12913 series; models based on soundscape data sampled at least at two different locations and using at least two variables as indicators. The Scopus database was queried. Biases on papers selection were considered and those related to the methods are discussed in the current study. Out of 256 results from Scopus, 22 studies were selected. Two studies were included from the references among the results. The data extraction from the 24 studies includes: data collection methods, input and output for the models, and model performance. Three main data collection methods were found. Several studies focus on the different combination of indicators among physical measurements, perceptual evaluations, temporal dynamics, demographic and psychological information, context information and visual amenity. The descriptors considered across the studies include: acoustic comfort, valence, arousal, calmness, chaoticness, sound quality, tranquillity, and vibrancy. The interpretation of the results is limited by the large variety of methods, and the large number of parameters in spite of a limited amount of studies obtained from the query. However, perceptual indicators, visual and contextual indicators, as well as time dynamic embedding, overall provide a better prediction of soundscape. Finally, although the compared performance between linear and non-linear methods does not show remarkable differences, non-linear methods might still represent a more suitable choice in models where complex structure of indicators are used.

Keywords: Soundscape Modelling, Urban Soundscape, Soundscape Indices, Literature Review

1. Introduction

Soundscape is defined by the International Organization for Standardization [1] as the "acoustic environment as perceived or experienced and/or understood by a person or people, in context". The definition establishes the concept of soundscape on a subjective observation of the acoustical properties of a place [2, 3]. Different variables, including physical, psychological, and physiological factors contribute to complex interactions in the definition of a soundscape for a listener. Small variations in just a few factors could lead to similar (in terms of physical characteristics) acoustic environments being perceived differently. The complexity of these phenomena makes soundscape modelling a challenging task. In the context of this review, we refer to soundscape modelling as the ability to anticipate how acoustic environments will be perceived by people (and potentially without actually gathering data from people). This is one of the main questions soundscape researchers are debating in the field. The awareness of the lack of operational tools to deal with these aspects prompted international agencies and policy-makers [4, 5, 6] [7] to invest in research in this field.

The characterization of sound environments is traditionally addressed in terms of acoustic properties of a space. Urban acoustic models [8, 9, 10, 11, 12, 13, 14, 15, 16, 17] were initially used to assess the presence and the impact of traffic noise in urban areas. Besides sound pressure level based models, noise annoyance (as a perceptual construct) is a common descriptor used to deal with this task. However, these models aim at describing a very narrow aspect of the whole soundscape, excluding for instance, the sound sources with potentially positive contributions that the listeners may experience [18]. Nonetheless, sound pressure level has been shown not to entirely describe how urban noise can possibly affect communities' well-being and health [19]. Where traditional methods mostly focus on predicting psychoacoustic measurements, new models are structured to embed perceptual components in both descriptors and indicators. This allows to turn the study of the acoustic environment into soundscape, including a

*Corresponding author

Email addresses: `matteo.lionello.18@ucl.ac.uk` (Matteo Lionello), `f.aletta@ucl.ac.uk` (Francesco Aletta), `j.kang@ucl.ac.uk` (Jian Kang)

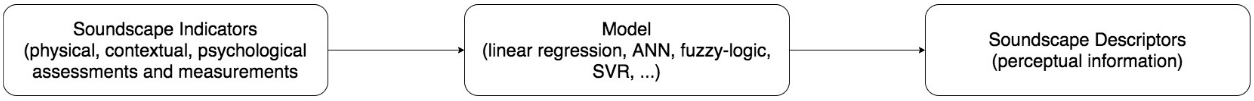

Figure 1: Soundscape model flow-chart, the indicator are the predictors input into the model which is structured according to linear or non-linear rules for predicting a set of variables named descriptors

more abstract and subjective representation of the perception of the urban sound environment.

The process for modelling soundscape (as a perceptual outcome) could be conceptualized in three components: indicators, descriptors, and the set of rules - either linear or non-linear - mapping the former to the latter (see Figure 1). Soundscape indicators and soundscape descriptors of a model are introduced [3] as "measures used to predict the value of a soundscape descriptor" and "measures of how people perceive the acoustic environment" respectively. The mapping can rely on linear regression methods or on non-linear mapping such as fuzzy-logic, Support Vector Regression Machine (SVR) and Artificial Neural Networks (ANN). Clustering methods [20, 21, 22, 23] are common strategies to find or to validate descriptors. Large soundscape datasets are commonly classified according to clusters obtained by the assessment of sound source categories presence [24, 25, 26] or according to the designed use of the site [27, 28]. The importance of sound sources [29, 2] in the categorisation of soundscapes has also been confirmed in a linguistic analysis [30]. However, sound source recognition represents the bottom level of hierarchical representations based on the attention focus as it can be extended to the following points [3]: 1) the evaluation of the presence of sound sources by categories; 2) sound sensation evaluation; and 3) overall evaluation related to the context. The subjective evaluation of point 1) is usually collected through a rating task. However, this task is mainly a soundscape classification task [31, 32] which does not address the perceptual identity of a soundscape, therefore, the studies modelling these as descriptors are excluded from the current review (see Sec. 2.2). Other approaches rely on simulating the auditory stimuli response on individuals by moving the role covered by sound source presence into event saliency recognition [33, 34, 35, 36, 37, 38], and by maintaining the cognitive and emotional factors on a higher level [32]. However, sound sources and event saliency recognition can still be considered as good indicators to fit the model. Depending on the dataset, the descriptors, and the context, other indicators can be extracted from physical contextual measurements, and cognitive and psychological evaluations.

In order to provide indicators and descriptors data for a soundscape model, in situ questionnaires, soundwalks and laboratory assessments are the most common strategies [39]. Out of the 24 references that will be presented later: 7 studies used in situ data collection approaches [27, 40, 28, 41, 21, 42, 43, 32], 2 soundwalks [44, 45], 9 laboratory experiments [46, 47, 48, 49, 50, 51, 52, 53, 54]. Six studies used a selected number of subjects assessing multiple soundscape data [21, 55, 56, 57, 58, 59]. In 2 studies [50, 57] the data corpus was taken from an online dataset of audio clips of urban contexts. Two models [49, 51] developed from laboratory experiment assessment used data collected from soundwalks, while most of laboratory assessments relied on stationary data collection. Three models [46, 21, 51] which used non-in situ data collection methods were validated with in situ datasets. One study [41] extended part of the data already used in a previous study [21] from a jury-test-like with in situ data collection.

The current review is based on the PRISMA systematic review protocol [60, 61] following, where applicable, its standardized reporting template. The PRISMA protocol provides guidelines for the writing of review manuscripts, based on hypothesis, arguments and planned reviewing methods for selecting criteria for the research papers. This review examines the state of the art with respect to the actual methods developed and implemented for soundscape prediction models. Particularly, the following point have been questioned: the quality of model in relation to a) the sample size and number of locations, b) the combination of indicators used, and c) the linearity rather than the non-linearity fitting of the model.

2. Methods

To the best of the authors' knowledge, no records of systematic review of soundscape modelling are available in the literature except for one paper in conference proceedings [62]. Some previous studies [20] worked either only on linear model structures or on indicators. Because of the exploratory nature of the current work, a standard protocol for literature review for soundscapes modelling is currently missing. Methods of the analysis and inclusion criteria were carefully considered by the authors. Peer-reviewed journal articles and full-paper conference proceedings written in English up to November 2019 were included as eligible in the database query. Conference proceedings papers have been included because soundscape modelling is a rapidly evolving field, thus excluding conference papers could have meant missing important emerging research trends.

2.1. Scoping stage for the literature review

Soundscape is a concept that is used and interpreted differently in several scientific disciplines. A simple query on the Scopus database for "soundscape" as a search term in the title, abstract, or keywords would return more than

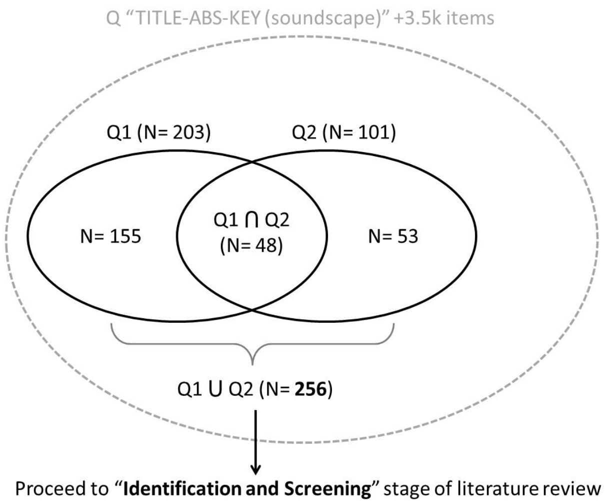

Figure 2: Schematic representation of the scoping stage of the literature review in Scopus; the first query Q1 ["TITLE-ABS-KEY (soundscape AND (model OR modeling OR modelling OR prediction OR predictive) AND (urban OR city))"] returned 203 results; the second query Q2 ["TITLE-ABS-KEY(soundscape AND (dimension OR dimensional OR valence OR arousal OR pleasant OR pleasantness OR quality) AND (model OR modelling OR modelling OR prediction OR predictive))"] returned 101 results. After removing duplicates, 256 unique results were identified.

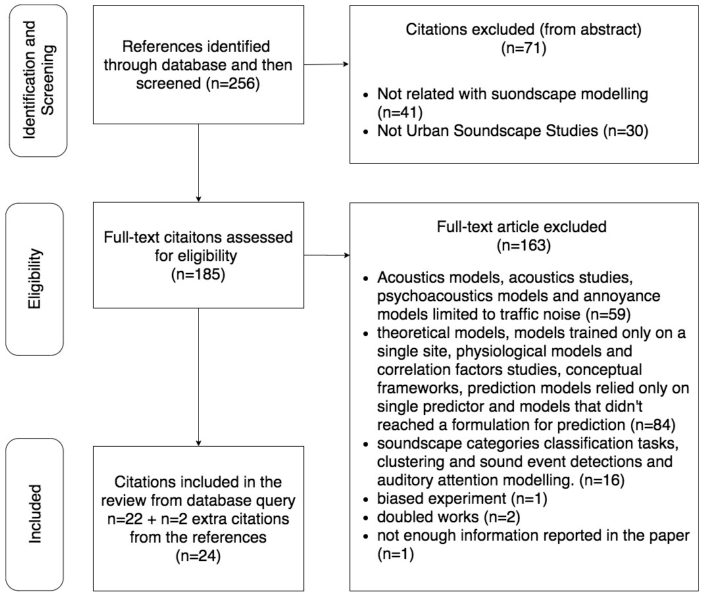

Figure 3: Flow chart for the paper search and skimming/selection process. All the references retrieved from the database query were manually assigned to one exclusion category or passed to the next step.

3,500 items to date; subject areas would vary greatly, including among others, physics and engineering, arts and humanities, medicine and healthcare, social sciences, marine ecology, gaming, noise annoyance, urban and architectural studies and so on. Because of this reason, it seemed fair to perform a preliminary search with the purpose of scoping the soundscape literature of interest; that is, it was necessary to define a search strategy so that only modelling-related urban soundscape papers in line with the definitions provided by the ISO 12913 standard series would be included. Therefore, two strings were created to query the Scopus database. The first string (Q1) focused on including studies about urban soundscape modelling: "TITLE-ABS-KEY (soundscape AND (model OR modeling OR modelling OR prediction OR predictive) AND (urban OR city))". A second string (Q2) aimed at reaching all those studies missing "urban" and "city" as keywords and including a dimensional modelling approach to soundscape: "TITLE-ABS-KEY (soundscape AND (dimension OR dimensional OR valence OR arousal OR pleasant OR pleasantness OR quality) AND (model OR modelling OR modelling OR prediction OR predictive))". This process is schematized in Figure 2. Literature that was not published in English was excluded already at this stage for the sake of practicality. Conference abstracts and book chapters were also excluded. The union of the pools of outcomes from the two main strings (N = 256) was then used as a starting point for the actual literature review and moved forward to its "Identification and Screening" stage (see also Figure 3). This dataset represents approximately 5-7% of the whole corpus of literature that Scopus reports as being mainly soundscape-related.

2.2. Search and study selection

On 1st of November 2019, Scopus returned 256 unique results from the union of the two queries. The inclusion criteria (see figure 3) for the paper selection aimed at including all the studies which developed a computational model to predict a property of urban soundscapes using more than one predictor (i.e., indicator). The term soundscape was considered as defined in the ISO 12913-1:2014 [1]. On the other hand, exclusion criteria were: models based on a single data collection site; not a predictive model, but rather, a classifier or probability density estimator such as clustering; not a soundscape study, but rather, a study on acoustics or either limited to traffic annoyance or modelling psychoacoustic features; not a urban soundscape; not related to soundscape modelling; and not predictive models but theoretical studies, meaning studies that are not in an advanced enough status to potentially be cross-validated in other contexts, or presenting only correlation analysis. Each result was assigned to a single of these rejections categories or to none. However, these rejection criteria are not mutually-exclusive as a work assigned to one rejection category could also not fit another inclusion criteria.

2.3. Data extraction protocol, data items and risk of bias

Each of the results has been fully read and information is extracted about data collection methods, indicators and descriptors used, structure of the model, and performance of each model proposed across all the selected studies. Main risks of bias may occur in individual studies and are related to the sample size of the dataset, which could be not large enough to represent real world situations; and in data collection methods affecting the choice of participants or concerning the translation of questionnaires and surveys in different languages when running the same study in multiple countries. In a laboratory setup, another bias that might affect participants' responses concerns the limited ecological validity of the experiments.

3. Results

3.1. Study selection

In accordance to the criteria stated in Section 2.2, among the 256 references, 71 (see Figure 3) were rejected after reading the abstract either because not related with soundscape modelling (e.g., head-related transfer function, auralization, health studies, virtual reality and audio gaming design..., n=41), or because not related to urban soundscapes (marine acoustics, national parks, urban forest, rural and countryside soundscape, uninhabited areas, soundscape compositions, landscape studies..., n=30). The 185 references resulting from the abstract screening were fully read and 22 of them were included in the current review. The rejected ones (n=163), are assigned to the following rejection folders: folder 1: acoustic models, acoustic studies, psychoacoustics models and annoyance models limited to traffic noise (n=59); folder 2: theoretical models, models trained only on a single site, physiological models and correlation factor studies, conceptual frameworks, prediction models relied only on a single predictor and models that did not reach a formulation for prediction (n=84); folder 3: soundscape categories classification tasks, clustering, sound event detection and auditory attention modelling (n=16); folder 4: biased experiments, smell biased experiments (n=1); List 5: doubled work, doubled journal/conference publications (n=2). One further publication has been excluded because did not report enough information for the data extraction.

Pheasant et al., 2008 [46] and Pheasant et al., 2010 [47] are separately added to the final list since 11 papers from the results are further validations of those, obtaining 24 final references included in the current review. The extracted data are reported in Table 1.

3.2. Soundscape indicators, perceptual and temporal embedding

3.2.1. Perceptual and temporal dynamic indicators

A combination of perceptual indicators and information on temporal dynamics can be found in Ricciardi et al., 2015 [21] and Aumond et al., 2017 [44] where the evaluation of overall loudness and visual amenity is accompanied by the perceived time presence of sound sources, which can also be automatically detected through kernel methods [41].

Among perceptual indicators, multiple models [27, 40, 21, 55, 56, 44, 45] used subjectively evaluated sound level as predictor. Consistently to the first layer of the descriptor hierarchy by Aletta et al., 2016 [3], [26] introduced the use of sound source identification as soundscape indicators accordingly to the inclusion criteria of the current study. Sound sources can either be evaluated by the participants to the study, or can be automatically assessed [31] and used as input layer for the models [45, 32]. The contribution of the emotional impact of an environment can be identified through two attributes [27]: the physical surrounding and the implicit attributes of social aspects by including also explicit behavioural and implicit psychological factors. Perceptual indicators can be mainly distinguished in: subjective perception of acoustics, sound sources prominence assessment [45, 32, 45, 56], and subjective preferences [40]. Higher degree of perceptual affections, such as the perceived eventfulness and the pleasantness of sounds, can also be introduced as predictors [28]. The importance of congruence perceived between soundscape and landscape has also been reported across the results [28, 55].

Sound sources can also play an important role in embedding temporal information of the soundscape. Temporal dynamic information can be encoded as the subjective evaluation of sound sources presence [21, 44], temporal standard deviation analysis features [50, 57, 58], automatic feature extraction [52, 32] and temporal derivative [44]. By using Bag-of-Frames approaches [63, 64], temporal dynamics can be extracted by either calculating standard deviation values [50, 57], the percentiles [58], or by collapsing the time domain by means of dimensional data reduction [52]. Finally, one last method to extract time dependencies consists in classifying saliency of events and so predicting the temporal density of sound sources events [32].

3.2.2. Visual and context information

The importance of visual information is proved by a significant correlation between visual quality and soundscape pleasantness [28]. Subjects rating high the visual comfort are found to be more likely to positively rate the acoustic comfort, while low visual comfort drastically rises the probability to give a negative evaluation of the acoustic comfort [40]. Moreover strong correlation is found between individual rating of sonic and visual environments [49]. Nonetheless, the exclusion of the visual predictor is observed to largely affect the explained variance of pleasantness of the model in Ricciardi et al., 2015 [21]. The percentage of blue [55] as the percentage of natural and contextual features [46, 47, 48] can be extracted from photographs as well as geometrical configuration and spatial metrics [42] of the landscape.

3.2.3. Psychoacoustic and acoustic indicators

Psychoacoustic and acoustic indicators are widely used in the retrieved results. The sound pressure level is the most common choice of indicator, either A-weighted, C-weighted, or in percentiles. Other acoustic features used by researchers include: energy, attack, spectral roll-off,

	Descriptors	Model	Indicators: Acoustic & Spectral	Psychoacoustic	Environment & Context	Sound Sources	Visual Information	Other	Perceptual Evaluation (other than s.sources)	Temporal Embedding	Study Design	#Subj.	#Sites	R^2
Pheasant et al., 2008 [46]	Tranq.	LR	L_{Amax}				NF			-	Lab	44	11	0.52
Yu and Kang, 2009 [27]	Acou. Comf.	ANN	L_{eq}.		T,W,Hum			Grouping, Movement, Age, Edu., Res, SPL_{home}	view, W, Hum, Bright., OL, LAND	-	Surveys	1840	7	0.31
Pheasant et al., 2010 [47]	Tranq.	LR	L_{Aeq}				NFC			-	Lab	102	34	0.86
Tse et al., 2012 [40]	Acou. Comf.	LR	L_{eq}		INS, B, T, W, V, Water, TR, Bike,DV			Age, Gender, AUDIT, Res, DUR	OL, $PREF_H$, $PREF_{NS}$, LAND	-	Surveys	595	4	0.60
Watts et al., 2013 [48]	Tranq.	LR	L_{Aday}		M_F		NFC			-	Surveys	252	8	0.89
Brambilla et al., 2013 [49]	Ch-Calm	LR	L_{eq}	S,R	S/H					-	Lab	32	20	0.98
Ricciardi et al., 2015 [21]	S. Quality	LR				V,TR			LAND,OL	Time presence of Sound Sources	Misc	57	102[1]	0.52
	S. Quality	LR				V,TR,B			OL	"	"	"	"	0.34
	S. Quality	LR	L_{50},L_{10}-L_{90}							-	"	"	"	0.21
Hong and Jeon, 2015 [28]	PL,EV_{SSQP}	LR[2]				H, NS, TR			PL, EV, LAND,SS_LAND	-	Surveys	631	21	χ^2/df: 4.26
Lavandier et al., 2016 [41]	PL	LR[3]						density geo-referenced variables: TR, V, Gardens		Time Ratio of Sound Sources	Surv&Misc	437[4]	89	0.68
	PL	LR[3]				V,B			OL	"	"	"	"	0.88
Fan et al., 2016 [50]	PL_{SSQP}	LR	MFCC5_{std},18,23,32_{mean}	N_{mean},S_{std}						std meas.	Lab[5]	20	125	0.57
	EV_{SSQP}	LR	Sroll,MFCC26,5_{std}2,28_{mean}							"	"	"	"	0.82
	PL_{SSQP}	SVR	98 features extracted from YAAFE software[a]							"	"	"	"	0.54
	PEV_{SSQP}	SVR	"							"	"	"	"	0.74
Herranz-Pascual et al., 2016 [55]	PL	LR	L_{10}-L_{90},L_{50},min$L_{Aeq,1s}$						Sound Source PL	-	Lab	53	9	0.18
	PL	LR	L_{10}-L_{90}			TR	%BLUE		SS_LAND	-	"	"	"	0.36
Cakır Aydin and Yilmaz, 2016 [51]	S. Quality	LR		N,R,S						-	Lab	53	27	0.77
Lundén et al., 2016 [52]	PL_{SSQP}	LR	MFCC							clustering of temp. dyn.[6]	Lab	30	77	0.74
	EV_{SSQP}	LR	MFCC							"	"	"	"	0.83
Lindborg and Friberg, 2016 [54]	PL_{SSQP}	LR		N_{10}				BigFiveTraits, funct.		-	Lab	43	12	0.49
	EV_{SSQP}	LR		N_{10}				BigFiveTraits, funct.		-	"	"	"	0.35
Puyana Romero et al., 2016 [42]	S. Quality	ANN	L_{Aeq},L_{A5}, L_{A10}, L_{A50}	R,N_5			$\%_{SEA}$, $\%_{FOOD}$, SM_F, $SM_{Sing.}$, SM_{TR}, SM_{GRDN}			-	Surveys	254	10	0.62
	S. Quality	LR	L_{Aeq}, L_{A50}	R, N_5			$\%_{SEA}$, SM_{GRDN}, SM_{TR}, SM_F, $SM_{Sing.}$			-	"	"	"	0.36
Maristany et al., 2016 [43]	S. Quality	F.L.	L_{Ceq}-L_{Aeq},L_{10}-L_{90}	S,N						-	Lab	416	12	0.88
Hong and Jeon, 2017 [56]	S. Quality	LR^7_{SLM}			H,TR,W,B					-	Misc	8	125	0.62
	S. Quality	LR^8_{GWR}			H,TR,W,B					-	Misc	8	125	0.70
Aumond et al., 2017 [44]	S. Quality	LR	$L_{50,1kHz}$,$TFSD_{500Hz}$,$TFSD_{4kHz}$							temporal derivative	Soundwalk	31	19	0.85
	S. Quality	LR				TR,V,B			OL		"	"	"	0.90
Boes et al., 2018 [32]	S. Quality	LR[9]				$SERE_{NS}$,$SERE_M$				Time Ratio of Sound Sources	Surveys	660	8	0.60
Fan et al., 2018 [57]	PL_{SSQP}	SVR	39 features from MIRToolbox[b] & YAAFE[a] soft.							std meas.	Misc[5]	1182	1213	0.62
	EV_{SSQP}	SVR	"							"	"	"	"	0.85
Kang et al., 2018 [45]	PL_{SSQP}	LR				TR,NS				-	Soundwalk	21	8	0.55
	ANN_{SSQP}	LR				TR,NS				-	"	"	"	0.52
	CM_{SSQP}	LR				TR,NS,OT				-	"	"	"	0.58
	CH_{SSQP}	LR				TR,NS,OT				-	"	"	"	0.37
Aletta and Kang, 2018 [53]	VIB_{SSQP}	LR	Fls	R,N	PEOPLE,MUSIC					-	Lab	35	46	0.76
Giannakopoulos et al., 2019 [58]	S. Quality	SVR	34 feat. extracted with pyAudioAnalysis[c]							std meas.& percentiles	Misc	10	978	acc. 42.3%; 50.2%[10]
Xichen et al., 2019 [59]	SS.Pref.	ANN	L_{Aeq}, L_{eq}	N[sone], N[phon], S						-	Misc	12	46	acc: 91.23%

Table 1, Tranq: Tranquillity; Acou. Comf.: Acoustic Comfort; Ch-Calm: Chaotic-Calm; S.Quality: Sound Quality; PL: Pleasant; EV: Eventful; ANN: Annoying; VIB: Vibrant; SSQP: Swedish Soundscape Protocol; SS.Pref: Soundscape Preference; T: Temperature; W: Wind; Hum: Humidity; Grouping: whether alone or in a group; Movement: whether arriving, staying or leaving; Brigth.: brightness; OL: evaluation of overall loudness; LAND: overall visual evaluation of the environment; INS: insects; T: Trees, TR: Traffic sound sources, V: Voices; AUDIT: self-reported auditory sensitivity; DUR: duration of staying; $PREF_H$: preference for anthropogenic sounds; $PREF_{NS}$: preference for natural sounds; DV: dummy variables; NF: percentage of natural features; NFC: percentage of natural and contextual features; M_f moderating factors; S: Sharpness; R: Roughness; S/H area divided by height of surrounding buildings; H: anthropogenic sound sources; NS: Natural Sound sources; SS_LAND harmony between soundscape and landscape; Sroll: Spectral Roll-off, MFCC: Mel Frequency Spectrum Coeff., N: Loudness SM: Spatial metrics; F: Fountains; Sing. single building; GRDN: Gardens; TFSD: time frequency second derivative; SERE: sound event rate estimation; M: mechanical sound sources; OT: other noise sources; Fls: Fluctuation Strength, PEOPLE: presence of people; MUSIC: Presence of music; BigFiveTraits: Extraversion, Agreeableness, Conscientiousness, Emotional stability, Openness. [1] 51 sites assessed by in winter and summer (102 objectives) by 57 subjects, each objective assessed by 20 subject; [2] structural equation based on linear regression; [3] used partial dataset from [21] including new in situ surveys from 19 locations, uses as input time ratio predicted with kernel density estimations; [4] 57 participants from the previous experiment plus 19 new locations assessed by about 20 participants each; [5] used as data corpus audio clips of urban soundscape retrieved for online databases; [6] the input features were extracted through a Gaussian Mixture Model collapsing the temporal feature over 13 components; [7] spatial lag model; [8] geographically weighted regression; [9] the sound presence perceived predicted by a SOM ANN as input; [10] 5-point Likert scale survey data reduced to 3 classes (positive, neutral, negative). For those studies reporting the correlation coefficient r, the coefficient of determination R^2 is obtained through the identity $r^2 = R^2$.

[a] available at `github.com/Yaafe/Yaafe`
[b] available at `mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox`
[c] available at `github.com/tyiannak/pyAudioAnalysis`

LINEAR MODELS

REFERENCE:	EQUATION:
Pheasant et al., 2008 [46]	Tranquillity $= 13.93 - 0.165L_{Amax} + 0.027NF$
Pheasant et al., 2008 [46]	Tranquillity $= 8.57 - 0.11L_{Aeq} + 0.036NF$
Pheasant et al., 2010 [47]	Tranquillity $= 9.68 - 0.146L_{Aeq} + 0.041NFC$
Tse et al., 2012 [40]	long formula available in the original manuscript
Watts et al., 2013 [48]	Tranquillity $= 10.55 - 0.146L_{day} + 0.41NFC + Mf$
Brambilla et al., 2013 [49]	'Chaotic-Calm' $= 10.537 - 0.129L_{Aeq} - 3.435R + 2.105S + 0.03(A/H)$
Ricciardi et al., 2015 [21]	Sound Quality $= 4.48 - 0.27OL + 0.12V + 0.52VA - 0.12T$
Ricciardi et al., 2015 [21]	Sound Quality $= 8.11 - 0.38OL + 0.20V + 0.15B - 0.15T$
Ricciardi et al., 2015 [21]	Sound Quality $= 19.08 - 0.19L_{50} - 0.06(L_{10} - L_{90})$
Hong and Jeon, 2015 [28]	long formula available in the original manuscript
Lavandier et al., 2016 [41]	Sound Pleasantness $= 8.71 - 0.74OL + 0.33V + 0.18B$
Fan et al., 2016 [50]	Valence $= 0.231 - 0.433N - 0.937S_{std} + 0.808MFCC5_{std} + 0.626MFCC18 - 2.046MFCC32 + 0.732MFCC23$
Fan et al., 2016 [50]	Arousal $= -1.441 - 0.317N + 0.556N_{std} + 4.064e - 10Sroll + 4.296MFCC26_{std} + 0.64MFCC5_{std}$ $-0.038MFCC2 - 0.604MFCC28$
Herranz-Pascual et al., 2016 [55]	N/A
Çakır Aydın and Yilmaz, 2016 [51]	Sound Quality Index $= 7.2935 - 0.05851N - 0.3723R - 0.7792S$
Lindborg and Friberg, 2016 [54]	Pleasant $= 0.893\text{Type} - 0.393N_{10} + 0.005\text{Extraversion} - 0.046\text{Agreeableness} + 0.037\text{Conscientiousness}$ $-0.111\text{EmotionalStab} - 0.053\text{Openness}$
Lindborg and Friberg, 2016 [54]	Eventful $= 0.842\text{Type} + 0.325N_{10} + 0.112\text{Extraversion} - 0.065\text{Agreeableness} + 0.044\text{Conscientiousness}$ $-0.112\text{EmotionalStab} - 0.005\text{Openness}$
Puyana Romero et al., 2016 [42]	S. Quality $= 0.166L_{Aeq} - 0.033R - 0.207L_{A50} - 0.086N_5 + 0.027\%\text{\_Sea} + 0.037\text{SM\_Fountain}$ $+0.045\text{SM\_Singular} - 0.027\text{SM\_Garden} + 0.048\text{SM\_Traffic}$
Hong and Jeon, 2017 [56]	Spatially lagged and geographically weighted regressions available in the original manuscript
Aumond et al., 2017 [44]	Pleasantness $= 9.70 - 0.47OL - 0.21T + 0.12V + 0.09B$
Aumond et al., 2017 [44]	Pleasantness $= 16.48 - 0.25L_{50,1Khz} - 15.82TSFD_{500Hz} + 16.82TFSD_{5KHz}$
Kang et al., 2018 [45]	Pleasant $= -0.577T + 0.252NS$
Kang et al., 2018 [45]	Annoying $= 0.64T - 0.144NS$
Kang et al., 2018 [45]	Chaotic $= 0.437T + 0.223OT - 0.152NS$
Kang et al., 2018 [45]	Calm $= -0.582T + 0.24NS - 0.11OT$
Boes et al., 2018 [32]	Sound Quality = 6.65 + 0.2739NS - 0.1726 MS
Aletta and Kang, 2018 [53]	Vibrancy $= 0.682R + 0.436PEOPLE + 0.383Fls - 0.579N + 0.272MUSIC$

NON-LINEAR MODELS

REFERENCE:	MODELS AND PARAMETERS:
Yu and Kang, 2009 [27]	Acoustic comfort <- ANN: 16 input dimension, 2 hidden layers 6 nodes each
Fan et al., 2016 [50]	Valence and Arousal <- SVR: sequential minimal optimization algorithm & polynomial kernels
Lundén et al., 2016 [52]	Valence and Arousa l<- SVR: N/A
Puyana Romero et al., 2016 [42]	Sound Quality <- ANN: 15 input dimension, 1 hidden layer with 15 nodes
Maristany et al., 2016 [43]	Sound Quality <- Fuzzy-logic: 11 rules over 4 input conditions (full rule system available in the original manuscript)
Fan et al., 2018 [57]	Valence and Arousal <- SVR: Radial basis function kernel & grid search method to find the parameters C and γ
Giannakopoulos et al., 2019 [58]	Sound Quality <- SVR: Radial basis function kernel
Xichen et al., 2019 [59]	Soundscape preference <- ANN: Radial Basis Function Neural Network

Table 2: Top: Equations of the linear models provided across the results. NF: percentage of natural features; NFC: natural and contextual features; Mf: moderating factor for urban decay; R: roughness; S: sharpness; A/H surface area divided by height of surrounding building; OL: overall loudness; VA: visual amenity; T: traffic presence; V: voice presence; B: birds presence; N: Loudness; TFSD: time frequency second derivative; NS: presence of natural sounds; OT: other sounds, PEOPLE: presence of people; MUSIC: presence of music; %_Sea, SM_Fountain, SM_Singular, SM_Garden, SM_Traffic: Spatial metric parameters. Bottom: Non-linear model structure and parameters

Mel-frequency cepstral coefficients (MFCC), spectral flatness, spectral flux, spectral slope, spectral variations, spectral gravity centre, time and frequency second derivatives, kurtosis and fluctuation. Loudness is the most recurrent choice among the psychoacoustic parameters and it is implemented in 7 publications; 4 publications use sharpness as indicator, whilst other 4 use roughness. One study [51] implemented a model based on psychoacoustics only, while a larger set of models are based on acoustic and spectral indicators only [50, 52, 54, 44]. Three studies [58, 57, 50] use large dimensional feature vectors automatically calculated through software packages (MIRToolbox, YAAFE, pyAudioLab).

Other indicators that can be found in the results include: personality traits [54], demographic data [27, 40], and environmental and physical measurements [21, 28], such as the ratio between surface area and height of surrounding buildings of an observed space.

3.3. Soundscape descriptors

In order to achieve the perceptual modelling, the descriptors from points 2) and 3) of the hierarchical modelling paradigm introduced in Section 1, are commonly identified through dimensional decomposition across semantic differentials. The Swedish Soundscape Quality Protocol (SSQP) [65] provides 8 adjectives ranging in a circumplex model [66, 67], namely: pleasant, unpleasant, eventful, uneventful, exciting, monotonous, calm and chaotic. Other studies found correspondence in the pairs vibrant-calm and chaotic-calm [68, 49] dimensions. Notwithstanding, the essence of soundscapes components laying in dimensions rather than in categories [69, 70] is an open debate [71, 72] which concerns the existence of relationships between the abstract structure of subjects' mind and the bi-dimensional representation onto the circumplex model. Five studies reported in this review used descriptors within the SSQP [50, 57, 52, 45, 53, 28, 54], whilst most of the studies reported sound quality [51, 42, 58, 56, 43, 21, 32] and soundscape pleasantness [44, 55, 41] used as descriptor. Three studies [46, 47, 48] aimed at modelling tranquillity [73]. Other descriptors used in the modelling include: acoustic comfort evaluation [27, 40], chaotic-calm dichotomy [49] and soundscape preference [59]. However, the mentioned descriptors are not linearly independent, for instance, a strong correlation (r=0.73) is found between the acoustic comfort and sound quality [55].

3.4. Soundscape models: linear and non-linear mapping

Besides location-based models, soundscape regression models can also be based on individual response [27, 40, 54]. These latter ones include psychological and personal traits as predictors. While the former, despite possibly being validated by individual responses [50], predict overall statistics of a soundscape. In order to predict the individual responses, models need to deal with a higher degree of complexity of information embedded within the variables. The data used for this purpose must comprehend features that characterise single participants or a set of them. To reach this degree of information psychological demographic and well-being-related features [74] are usually needed. Most of the selected studies implement models exclusively relying on overall statistics of the responses across locations. Soundscape topological information can be encoded [41, 56] through kernel-based density models or developed by using interpolation algorithms [45] to evaluate soundscape maps. Other topological conditions are introduced in a limited number of studies such as spatial dependencies matrix, and Gaussian-kernel density estimation [56], unsupervised neural network self-organized map-like model [32] and radial basis function neural network [59]. When the temporal domain of the data is substantially large, a lower-level model can be used to collapse the temporal domain across multiple clusters by using unsupervised methods upon which to compute linear regression [52]. Only two studies [42, 50] directly compared the results from the same dataset across multiple models, while other studies which run comparisons across different models are done on a subset or a different dataset than the one used for the proposed modelling. A large majority of the results uses linear regression based models which are reported in Table 2. Only a few models implemented non-linear structures comprehending neural networks, support vector machines and fuzzy-logic models (see Table 1). Among the linear regression models the following further techniques are used: logistic regression [40], structural equation model [28, 40], and density kernel estimations to predict time ratio presence as input of the linear regression [41]. Descriptor values are usually computed over the average of responses, except for a few cases where the values are computed through a ranking sorting task [57] or by splitting the values according to some ranges such as splitting the 5-point Likert scale in 3 classes [58].

4. Discussion

The performance of the models is measured in terms of correlation between its predictions and some validation data and it depends on several aspects. Some of these aspects have been discussed in the previous sections, namely: data collection methods, sample size of participants for the study, geographical and cultural context, choice of soundscape descriptor, physical measurements, scaling and experiments method, the choice of the type and the hyper-parameters of the model, and feature extraction and feature selection methods. The change of even just one of these parameters makes the comparison between two models very difficult. In this section some general considerations are pointed out with respect to separate models proposed within- and between-studies in relation to their performance. Moreover, many studies report the performance of linear regressions over the whole dataset and do not consider a validation set. Non-linear models mostly use 10-fold cross validation techniques [75] or separate test and validation datasets on which they measure the performance of the model, which is the performance reported in the current study. However, this discrepancy still allows for a good overall comparison.

4.1. Sample size and number of locations

The sample size of participants in the data collection, and the amount of locations across the data are the first aspects to discuss. The models are split with respect to the number of participants involved ($n < 35$, $35 \leq n < 100$, $100 \leq n$) and to the performance of the model ($R^2 \geq 0.7$, $0.5 \leq R^2 < 0.7$ and $R^2 < 0.5$). Given a high performance of the models ($R^2 \geq 0.7$) studies can be distinguished in strongly performing models which require many participants [47, 48, 43, 57, 41], and models easier to implement in spite of a weaker reliability as they could potentially be affected by over-fitting [49, 50, 52, 56, 44, 53]. A trade-off between these two conditions is provided by models still performing well ($R^2 > 0.5$) and involving a moderate number of participants ($35 \leq n < 100$) [46, 21, 51]. Similar considerations can be addressed with respect to the amount of different locations involved in the studies. It is possible to distinguish well-performing models ($R^2 > 0.5$) based on limited range of locations ($n \leq 12$) [46, 48, 45, 40, 42, 32, 43], and models based on a wider range of locations ($n > 40$) [21, 41, 50, 52, 56, 57, 53]. It can be noted that a few models [47, 41] obtained a significantly higher performance by extending the number of locations and participants compared to previous models.

4.2. Indicators and Models Performance

In order to study the impact of the combination of the indicators in the performance of the models, the different models collected across the results are sorted in four categories [76]: models reporting substantial performance with $R^2 \geq 0.7$, those with a moderate performance with $0.5 \leq R^2 < 0.7$, those with a weak performance in their results $0.25 \leq R^2 < 0.5$ and those with low performance $R^2 < 0.25$.

4.2.1. Low performance model indicators

The results reported below the bottom limit $R^2 = 0.25$ are performed by models using a limited amount of information across the indicators or by context constraints. The choice of indicators based only on acoustic measurements does not provide enough information for the prediction of soundscape descriptors based on perceptual dimensions. Models relying only on acoustic indicators are shown to perform worse ($R^2 = 0.21$ [21], $R^2 = 0.18$ [55]) than the other proposed models in the same studies. On the other hand, the introduction of higher level of complexity of indicators might not be enough to provide a well-performing model. Using only sound source presence has been reported [45] to affect the prediction of eventful, uneventful, vibrant and monotonous descriptors worse (coefficients of determination between 0.03 and 0.09) than using the same indicator for predicting pleasantness, annoyance, calmness and chaoticness. These phenomena have been discussed by the authors indicating that it is plausible that not all the soundscape dimensions might emerge in every place.

4.2.2. Weak performance model indicators

The information provided by acoustic variables alone can be extended by introducing perceptual and contextual features. The use of perceptual indicators instead of acoustic measurements [21] shows an increase of the performance (from $R^2 = 0.21$ to $R^2 = 0.34$). By introducing contextual features, models initially based only on acoustic paramters show improvements in their performance [55]. However, even the model using also contextual variables show a weak ($R^2 = 0.36$) performance in predicting the soundscape quality descriptor [55]. Chaotic model [45] showed a better predictability ($R^2 = 0.37$) than eventful, uneventful, vibrant and monotonous descriptors based on sound sources presence. A similar performance ($R^2 = 0.31$) can also be noted in the prediction of individual responses [27]. However, one possible reason for this weak result might relate with the large variance in the number of participants per location, causing the model to be unstable.

4.2.3. Moderate performance model indicators

The importance of accompanying acoustic measurements with visual features is proved [46] showing linear dependency between L_{Aeq} and the proportion of natural visual factors in predicting the tranquillity descriptor. Extending the perceptual indicators with visual evaluations suggests [21] a better performance of the model ($R^2 = 0.34$ to $R^2 = 0.52$ and validated in another country $r = 0.62$). Moreover, significant correlation between visual quality and soundscape pleasantness showed [40] that subjects rating high visual comfort were 2.2 times more likely to positively rate the acoustic comfort, while low visual comfort rises 7.6 times the probability to give a negative evaluation of the acoustic comfort. Pleasant, annoying and calm models are shown [45] to be those with better performance, suggesting that these descriptors might more commonly understood (i.e., there is consensus on the interpretation), compared to the rest of the attributes composing the SSQP, or to be better explained depending on sound source presence.

4.2.4. Substantial performance model indicators

The embedding of high-level indicators such as time dynamics, the employment of perceptual-based [41, 53] and topological information, shows to provide the best options for accurate soundscape prediction. Topology explained with kernel density and matrix distance information showed [56] to predict accurately the soundscape quality ($R^2 = 0.62$, $R^2 = 0.7$). The importance of the population sampling and particularly of the geographical context has proved [47] to enhance the performance in modelling tranquillity (from $R^2 = 0.52$ [46] to $R^2 = 0.89$ [48] by introducing contextual information to the last model.

Contrarily to the above results obtained to predict SSQP attributes [45], a better performance [50, 57, 52] is achieved by introducing MFCC indicators, to model eventfulness (respectively $R^2 = 0.82$, $R^2 = 0.85$ and $R^2 = 0.83$) rather than the modelling of pleasantness (respectively $R^2 = 0.57$, $R^2 = 0.62$ and $R^2 = 0.74$). Among these results, the use of indicators of temporal patterns detected from MFCC analysis only shows [52] better outcomes overall. A comparison [50] between linear regression and support vector regression machine shows an easier implementation of the former assuming that the second method over-fitted with respect to the dataset size. The support vector regression machine has been used in a later study [57] with a larger dataset returning better results. Models based only on psychoacoustic indicators [51] can still achieve a relatively good performance ($R^2 = 0.77$) However, better results ($R^2 = 0.98$ [49], $R^2 = 0.79$ [42])

are obtained by extending the models with visual information. The use of non-linear model over only acoustic and psychoacoustic indicators shows [43] an improvement ($R^2 = 0.88$) compared to the average performance of linear methods based on the same set of indicators. In opposition to only acoustic measurement performances [21], acoustic measurements can still be used with good results ($R^2 = 0.85$) if augmented with time dynamics embedding [44]. Although, even better results ($R^2 = 0.90$) [44] are obtained by using models based on perceptual indicators.

4.3. Linear and non-linear model performances

Finally, due to the small amount of studies implementing non-linear models, no conclusion can be drawn regarding the most suitable approach between linear and non-linear models. Since non-linear methods generally provide greater accuracy than linear regression methods, the challenge of implementing them in spite of the good results is not always a preferable choice for the researchers. However, non-linear models are still a good strategy to fit complex information encoded within and between the indicators, such as temporal dynamics. Moreover, more effectiveness can be achieved when implementing machine learning and kernel methods in the data processing and feature extraction in the creation of the inputs [52, 32, 41].

5. Conclusion

In the context of establishing prediction models for soundscape descriptors, using physical indicators as predictors, collection methods cover an essential role dictating the direction towards which the modelling task can proceed. A first comparison between linear and non-linear methods suggests that the former provide a strategy that is easier to implement, while the latter provide better results at the cost of a more difficult definition of the model, feature extraction, and analysis. The use of only acoustic indicators results to be far more difficult to fit the demand of soundscape models. The information provided by these show the need to be extended, as a first step, with visual and contextual information. Indeed, the review found a strong relationship between the hierarchical level of abstraction of the indicators and the performance of the model. In particular, it can be observed that the embedding of time dynamics, the use of subjects' evaluation factors, and contextual features play an important role in increasing the quality of the model. Topological dependencies across data points are also an important factor which could be included into the model structure. The contextual information is mostly used as an additional indicator and only few studies refer to the relationship between soundscape and its context as a descriptor. With regards to the directions of the research questions above stated (see Section 1), the main conclusions of this literature review are:

- data collection methods and the amount of data points used to fit the model define the complexity and strength of the model: for well-performing models the larger the dataset, the more stable and complex the model is and, in a complementary way, the reduction of datapoints provides simpler model in terms of implementation but, at the same time, a model more susceptible to over-fitting;
- the degree of subjective and perceptual information encoded in the indicators proved to be a great contributor leading to better performance in predicting soundscapes (as perceived) compared to combination of acoustic and psychoacoustics indicators;
- non-linear methods provide more accurate tools compared to linear methods to predict soundscapes in spite of a more difficult implementation by the researchers.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 740696, project title: soundscape indices - SSID).

References

[1] Acoustics — soundscape — part 1: Definition and conceptual framework, Standard, International Organization for Standardization, Geneva, CH (2014).

[2] A. L. Brown, J. Kang, T. Gjestland, Towards standardization in soundscape preference assessment, Applied Acoustics 72 (6) (2011) 387 – 392.

[3] F. Aletta, J. Kang, O. Axelsson, Soundscape descriptors and a conceptual framework for developing predictive soundscape models, Landscape and Urban Planning 149 (2016) 65–74. doi:10.1016/j.landurbplan.2016.02.001.

[4] Directive 2002/49/ec relating to the assessment and management of environmental noise, Tech. rep., Brussels: Publications Office of the European Union (2002).

[5] Noise and soundscape action plan 2018-2023, Tech. rep. (2018).

[6] Environmental noise guidelines for the european region, Tech. rep., Regional Office for Europe (2018).

[7] J. Kang, From dba to soundscape indices: Managing our sound environment, Frontiers of Engineering Management 4 (2) (2017) 184. doi:10.15302/J-FEM-2017026.

[8] M. A. Burgess, Noise prediction for urban traffic conditions - related to measurements in the sydney metropolitan area, Apllied Acoustics 10 (1977) 1–7.

[9] R. Josse, Notions d'acoustique, Ed. Eyrolles, Paris. France, 1972.

[10] C. Fagotti, A. Poggi, Traffic noise abatement strategies. the analysis of real case not really effective, in: 18th International Congress for Noise Abatement, Bologna, Italy, 1995, pp. 223–233.

[11] D. Bertoni, A. Franchini, M. Magnoni, Il Rumore Urbano e l'Organizzazione de Territorio, Pitagora Editrice, Bologna, Italy, 1987, pp. 45–69.

[12] D. Cvetković, M. Praščević, V. Stojanović, Naiss model for traffic noise prediction, Facta Universitatis Series: Working and living environment protection 1 (1997) 73–81.

[13] M. Prascevic, D. Cvetkovic, D. Mihajlov, Z. Petrović, B. Radičević, Verification of naiss model for road traffic noise prediction in urban areas, Electronics and Electrical Engineering 19. doi:10.5755/j01.eee.19.6.1294.

[14] F. Mietlicki, C. Mietlicki, C. Ribeiro, P. G. Bruitparif, B. Vincent, V. Gissinger, www.noiseineu.eu: New tools to inform the public about environmental noise in cities and to assist decision-making, in: EuroNoise, Maastricht, 2015.

[15] N. Genaro, A. Torija, A. Ramos-Ridao, I. Requena, D. P. Ruiz, M. Zamorano, A neural network based model for urban noise prediction, The Journal of the Acoustical Society of America 128 (4) (2010) 1738–1746.

[16] G. Cammarata, S. Cavalieri, A. Fichera, A neural network architecture for noise prediction, Neural Networks 8 (6) (1995) 963 – 973. doi:10.1016/0893-6080(95)00016-S.

[17] Y. Avsar, A. Saral, M. T. Gonullu, E. Arslankaya, U. Kurt, Neural network modelling of outdoor noise levels in a pilot area, Turkish Journal of Engineering and Environmental Sciences 28 (2004) 149–155.

[18] F. Aletta, T. Oberman, J. Kang, Positive health-related effects of perceiving urban soundscapes: a systematic review, The Lancet 392 (2018) S3. doi:10.1016/S0140-6736(18)32044-0.

[19] K. Persson, R. Rylander, Disturbance from low-frequency noise in the environment: A survey among the local environmental health authorities in sweden, Journal of Sound and Vibration 121 (2) (1988) 339 – 345. doi:10.1016/S0022-460X(88)80034-8.

[20] F. Aletta, J. Kang, Descriptors and indicators for soundscape design : vibrancy as an example, 2016.

[21] P. Ricciardi, P. Delaitre, C. Lavandier, F. Torchia, P. Aumond, Sound quality indicators for urban places in paris cross-validated by milan data, Journal of the Acoustical Society of America 138 (4) (2015) 2337–2348.

[22] A. J. Torija, D. P. Ruiz, Ángel F. Ramos-Ridao, A tool for urban soundscape evaluation applying support vector machines for developing a soundscape classification model, Science of The Total Environment 482-483 (2014) 440 – 451. doi:https://doi.org/10.1016/j.scitotenv.2013.07.108.

[23] B. D. Coensel, D. Botteldooren, K. Debacq, M. E. Nilsson, B. Berglund, Soundscape classifying ants, 2007.

[24] L. Yu, J. Kang, Ann models for sound preference evaluations, in: International Congress on Acoustics, ICA, 2010.

[25] L. Yu, J. Kang, Using ann to study sound preference evaluation in urban open spaces, Journal of Environmental Engineering and Landscape Management 23 (2015) 163–171. doi:10.3846/16486897.2015.1050399.

[26] O. Axelsson, M. E. Nilsson, B. Berglund, A swedish instrument for measuring soundscape quality, in: Proceedings of the Euronoise 2009 Conference, Edinburgh, 2009.

[27] L. Yu, J. Kang, Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach, Journal of the Acoustical Society of America 126 (3) (2009) 1163–1174, cited By :61. doi:10.1121/1.3183377.

[28] J. Y. Hong, J. Y. Jeon, Influence of urban contexts on soundscape perceptions: A structural equation modeling approach, Landscape and Urban Planning 141 (2015) 78 – 87. doi:10.1016/j.landurbplan.2015.05.004.

[29] R. M. Schafer, The Soundscape: Our Sonic Environment and the Tuning of the World, 1977.

[30] M. Niessen, C. Cance, D. Dubois, Categories for soundscape: Toward a hybrid classification, Internoise 2010 (2010) 5816–5829.

[31] W. Yang, J. Kang, Acoustic comfort evaluation in urban open public spaces, Applied Acoustics 66 (2) (2005) 211 – 229, urban Acoustics. doi:https://doi.org/10.1016/j.apacoust.2004.07.011.

[32] M. Boes, K. Filipan, B. De Coensel, D. Botteldooren, Machine listening for park soundscape quality assessment, Acta Acustica united with Acustica 104 (1) (2018) 121–130. doi:doi:10.3813/AAA.919152.

[33] B. De Coensel, D. Botteldooren, Modeling auditory attention focusing in multisource environments, The Journal of the Acoustical Society of America 123 (2008) 3394. doi:10.1121/1.2934070.

[34] B. D. Coensel, D. Botteldooren, Models for soundscape perception and their use in planning, 2007.

[35] B. D. Coensel, D. Botteldooren, A model of saliency-based auditory attention to environmental sound, 2010.

[36] D. Oldoni, B. De Coensel, M. Rademaker, T. Van Renterghem, D. Botteldooren, B. De Baets, Imitating human auditory processing for urban soundscape measurement, in: Proceeding of the Institute of Acoustics, Vol. 32, Institute of Acoustics, 2010, pp. 71–78.

[37] D. Oldoni, B. D. Coensel, M. Boes, T. V. Renterghem, D. Botteldooren, A computational auditory attention model for urban soundscape design, 2012.

[38] D. Oldoni, B. De Coensel, M. Boes, M. Rademaker, B. De Baets, T. Van Renterghem, D. Botteldooren, A computational model of auditory attention for use in soundscape research, The Journal of the Acoustical Society of America 134 (1) (2013) 852–861. doi:10.1121/1.4807798.

[39] M. Engel, A. Fiebig, C. Pfaffenbach, J. Fels, A review of socio-acoustic surveys for soundscape studies, Current Pollution Reports 4. doi:10.1007/s40726-018-0094-8.

[40] M. S. Tse, C. K. Chau, Y. S. Choy, W. K. Tsui, C. N. Chan, S. K. Tang, Perception of urban park soundscape, The Journal of the Acoustical Society of America 131 (4) (2012) 2762–2771. doi:10.1121/1.3693644.

[41] C. Lavandier, P. Aumond, S. Gomez, C. Dominguès, Urban soundscape maps modelled with geo-referenced data, Noise Mapping 3. doi:10.1515/noise-2016-0020.

[42] V. Puyana Romero, L. Maffei, G. Brambilla, G. Ciaburro, Modelling the soundscape quality of urban waterfronts by artificial neural networks, Applied Acoustics 111 (2016) 121–128.

[43] A. Maristany, M. R. López, C. A. Rivera, Soundscape quality analysis by fuzzy logic: A field study in cordoba, argentina, Applied Acoustics 111 (2016) 106 – 115. doi:10.1016/j.apacoust.2016.04.013.

[44] P. Aumond, A. Can, B. De Coensel, D. Botteldooren, C. Ribeiro, C. Lavandier, Modeling soundscape pleasantness using perceptual assessments and acoustic measurements along paths in urban context, Acta Acustica united with Acustica 103 (3) (2017) 430–443.

[45] J. Kang, F. Aletta, E. Margaritis, M. Yang, A model for implementing soundscape maps in smart cities, Noise Mapping 5. doi:10.1515/noise-2018-0004.

[46] R. Pheasant, K. Horoshenkov, G. Watts, B. Barrett, The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places?, The Journal of the Acoustical Society of America 123 (2008) 1446–57. doi:10.1121/1.2831735.

[47] R. Pheasant, K. Horoshenkov, G. Watts, Tranquillity rating prediction tool (trapt), Acoustics Bulletin 35 (2010) 18–24.

[48] G. Watts, A. Miah, R. Pheasant, Tranquillity and soundscapes in urban green spaces- predicted and actual assessments from a questionnaire survey, Environment and Planning B: Planning and Design 40 (1) (2013) 170–181.

[49] G. Brambilla, L. Maffei, M. Gabriele, V. Gallo, Merging physical parameters and laboratory subjective ratings for the soundscape assessment of urban squares, The Journal of the Acoustical Society of America 134 (2013) 782–90. doi:10.1121/1.4768792.

[50] J. Fan, M. Thorogood, P. Pasquier, Automatic soundscape affect recognition using a dimensional approach, AES: Journal of the Audio Engineering Society 64 (9) (2016) 646–653.

[51] D. Çakır Aydın, S. Yilmaz, Assessment of sound environment pleasantness by sound quality metrics in urban spaces, A—Z ITU Journal of the Faculty of Architecture 13. doi:10.5505/itujfa.2016.75547.

[52] P. Lundén, O. Axelsson, M. Hurtig, On urban soundscape mapping: A computer can predict the outcome of soundscape assessments, in: Proceedings of the INTER-NOISE 2016 - 45th International Congress and Exposition on Noise Control Engi-

neering: Towards a Quieter Future, 2016, pp. 4725–4732.

[53] F. Aletta, J. Kang, Towards an urban vibrancy model: A soundscape approach, International Journal of Environmental Research and Public Health 15 (2018) 1712. doi:10.3390/ijerph15081712.

[54] P. Lindborg, A. Friberg, Personality traits bias the perceived quality of sonic environments, Applied Sciences 6 (2016) 405. doi:10.3390/app6120405.

[55] K. Herranz-Pascual, I. García, I. Aspuru, I. Díez, A. Santander, Progress in the understanding of soundscape: objective variables and objectifiable criteria that predict acoustic comfort in urban places., Vol. 3, 2016. doi:10.1515/noise-2016-0017.

[56] J. Y. Hong, J. Y. Jeon, Exploring spatial relationships among soundscape variables in urban areas: A spatial statistical modelling approach, Landscape and Urban Planning 157 (2017) 352 – 364. doi:10.1016/j.landurbplan.2016.08.006.

[57] J. Fan, M. Thorogood, P. Pasquier, Emo-soundscapes: A dataset for soundscape emotion recognition, in: 2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017, Vol. 2018-January, 2018, pp. 196–201.

[58] T. Giannakopoulos, M. Orfanidi, S. Perantonis, Athens Urban Soundscape (ATHUS): A Dataset for Urban Soundscape Quality Recognition, Vol. 11295 LNCS of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.

[59] H. Xichen, J. Yu, W. Shuting, Z. Linying, L. Siren, Study on evaluation model of soundscape in urban park based on radial basis function neural network: A case study of shiba park and kamogawa park, japan, in: IOP Conf. Series: Earth and Evnironemntal Sceince, Vol. 300, 2019.

[60] D. Moher, A. Liberati, J. Tetzlaff, D. Altman, Preferred reporting items for systematic reviews and meta-analyses: The prisma statement, PLoS Medicine 7. doi:10.1371/journal.pmed1000097.

[61] A. Liberati, D. Altman, J. Tetzlaff, C. Mulrow, P. Gøtzsche, J. Ioannidis, M. Clarke, P. Devereaux, J. Kleijnen, D. Moher, The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration, Journal of clinical epidemiology 62 (2009) e1–34. doi:10.1016/j.jclinepi.2009.06.006.

[62] M. Yang, A review of regression analysis methods: Establishing the quantitative relationships between subjective soundscape assessment and multiple factors, Proceedings of the ICA 2019 and EAA Euroregio : 23rd International Congress on Acoustics integrating 4th EAA Euroregio 2019 : 9-13 September 2019 (2019) 9 Sep 2019–13 Sep 2019; Aachen (2019). doi:10.18154/RWTH-CONV-239497.

[63] D. Barchiesi, G. [email protected, D. Stowell, M. Plumbley, Acoustic scene classification: Classifying environments from the sounds they produce, Signal Processing Magazine, IEEE 32 (2015) 16–34. doi:10.1109/MSP.2014.2326181.

[64] L. Su, C.-C. M. Yeh, J.-Y. Liu, J.-C. Wang, y.-h. Yang, A systematic evaluation of the bag-of-frames representation for music information retrieval, Multimedia, IEEE Transactions on 16 (2014) 1188–1200. doi:10.1109/TMM.2014.2311016.

[65] O. Axelsson, M. Nilsson, B. Berglund, A principal components model of soundscape perception, The Journal of the Acoustical Society of America 128 (2010) 2836–46. doi:10.1121/1.3493436.

[66] J. Russell, A circumplex model of affect, Journal of Personality and Social Psychology 39 (6) (1980) 1161–1178.

[67] J. Posner, J. Russell, B. Peterson, The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology, Development and psychopathology 17 (3) (2005) 715–34.

[68] R. Cain, P. Jennings, J. Poxon, The development and application of the emotional dimensions of a soundscape, Applied Acoustics 74 (2) (2013) 232 – 239, applied Soundscapes: Recent Advances in Soundscape Research. doi:https://doi.org/10.1016/j.apacoust.2011.11.006.

[69] L. M. Aiello, S. Rossano, D. Quercia, F. Aletta, Chatty maps: constructing sound maps of urban areas from social media data, The Roayl Society Open Sciencedoi:10.1098/rsos.150690.

[70] K. Sun, B. D. Coensel, K. Filipan, F. Aletta, T. V. Renterghem, T. D. Pessemier, W. Joseph, D. Botteldooren, Classification of soundscapes of urban public open spaces, Landscape and Urban Planning 189 (2019) 139 – 155. doi:https://doi.org/10.1016/j.landurbplan.2019.04.016.

[71] V. Maffiolo, M. Castellengo, D. Dubois, Is pleasantness for soundscapes dimensional or categorical?, Journal of The Acoustical Society of America 105(2).

[72] M. Lionello, F. Aletta, J. Kang, On the dimension and scaling analysis of soundscape assessment tools: A casestudy about the "method a" of iso/ts 12913-2:2018, in: The International Conference on Acoustics, 2019.

[73] T. R. Herzog, G. J. Barnes, Tranquillity and preference revisited, Journal of Environmental Psychology (1999) 171–181.

[74] F. Aletta, T. Oberman, A. Mitchell, M. Erfanian, M. Lionello, M. Kachlicka, J. Kang, Associations between soundscape experience and self-reported wellbeing in open public urban spaces: a field study, The Lancet 394. doi:10.1016/S0140-6736(19)32814-4.

[75] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, Berlin, Heidelberg, 2006.

[76] J. Henseler, C. Ringle, R. Sinkovics, The use of partial least squares path modeling in international marketing, Advances in International Marketing (AIM) 20 (2009) 277–320.