Distinct categorization strategies for different types of environmental sounds

Oliver Bones
University of Salford, Acoustics Research Centre, UK.

Trevor J. Cox
University of Salford, Acoustics Research Centre, UK.

William J. Davies
University of Salford, Acoustics Research Centre, UK.

Summary

The present work involved a sound-sorting and category-labelling task that elicits rather than prescribes words used to describe sounds, allowing categorization strategies to emerge spontaneously and the interpretation of the principal dimensions of categorization using the generated descriptive words. Previous soundscape work suggests that 'everyday listening' is primarily concerned with gathering information about sound sources, and that sounds are typically categorized by perceived similarities between the sound-causing events. The present work demonstrates that this is likely to be the case when sound-sources are sufficiently differentiated for this to be a useful cognitive strategy, such as when categorizing a variety of different sound sources, or when categorizing a broad class of sounds with multiple sources such as 'water'. However, distinct strategies based upon alternative cues emerge for other types of sounds. For example, categorization of dog sounds is primarily determined by judgements relating to perceptual dimensions similar to valence ('sad'/'lonely'-'playful'/'friendly') and arousal ('bored'/'whining'-'threatening'/'vicious'), a finding that supports the circumplex model of affect as a meaningful framework for understanding human categorization of this type of sound. Categorization of engine sounds on the other hand was found to be based primarily upon explicit assessment of the acoustic signal, along dimensions which correlate strongly with the fluctuation strength ('steady'-'chugging') and sharpness ('muffled'-'jarring') of the recordings. These results demonstrate that categorization of sound is based upon different strategies depending on context and the availability of cues. It has implications for experimental methods in soundscapes that prescribe conceptual frameworks on test subjects. For instance, careful consideration should be given to the appropriateness of semantic differential scales in future perceptual soundscape work.

PACS no. 43.66.Ba

1. Introduction

Categorization describes the process by which the correlational structure of the attributes of objects in the environment is used to simplify and apply meaning to sensory experience [1, 2]. One approach to identifying the attributes with which categories are formed is the semantic differential method, whereby concepts and events (e.g. sounds) are scored on attribute rating scales. The proximity of events and concepts to one another on the underlying dimensions of the resulting data are informative of their perceived similarity, and the correlations of each dimension with individual rating scales are informative of the attributes that drive the perception of similarity. The 'core affect' framework models affective states as the linear combination of valence (a pleasure-displeasure continuum) and arousal (an alertness continuum) [3-5]. Within soundscape research the semantic differential method has identified dimensions similar to valence and arousal such as 'pleasantness' [6-8], 'preference' [9, 10], 'calmness' [11], 'relaxation' [12], 'dynamism' [12], 'vibrancy' [11], 'playfulness' [10], and 'eventfulness' [7].

An alternative method for identifying the attributes with which categories are formed is to use a sorting and category labelling task, followed by linguistic analysis of the category labels. An advantage of this method over others such as the semantic differential method is that the attributes by which sensory objects are compared are not prescribed a priori. Studies of soundscapes using this method have identified categories formed according to whether or not they contained human activity [13] and by the perceived similarity of the sound-causing events within the soundscape [13-16]. A study using a similar procedure to explore isolated domestic sounds [17] found categories that resembled those proposed by Gaver based upon the type of material and event that produced the sound [18]. Another study of isolated sounds found the similarity of non-living sounds to be predicted by their acoustic properties, but the similarity of living sounds to be predicted by their semantic meaning [19].

The present study tested the hypothesis that the cues used for category formation would differ between three different types of environmental sound: dog, engine, and water sounds. The descriptive words used to label categories from a sorting task were used as verbal correlates of sound category formation, and statistically analyzed using multinomial logit regression.

2.1. Procedure

Sound sorting experiments were performed for each of the three types of sound separately via a web interface on a website hosted by one of the authors (http://sound101.org). Sounds were represented by tiles labelled as e.g. 'Dog_1', 'Dog_2' etc. At the beginning of each study tiles were arranged in a pseudorandomized order along the left hand side of the screen in a 'sound bank'. Instructions at the top of the screen instructed participants to: click the tiles to hear the sound; group similar sounds together by dragging them from the sound bank into one of five categories; use all five categories; give each category a name describing the sounds in the category. In addition, participants were instructed not to use category names such as 'miscellaneous', 'random', or 'sounds' etc. The average amount of time taken to complete the task was approximately 20 minutes.

2.2. Stimuli and participants

Fifty participants took part in the dog sound experiment, N=49 the engine sound experiment, and N=48 the water sound experiment. All

Table I: Demographic data of participants for each study. All values are percentages rounded up to the nearest whole percent.

		Dog	Engine	Water
Age	18-29	70	33	48
	30-39	12	49	38
	40-49	14	10	13
	50-59	4	6	2
	60-69	0	2	0
Sex	Male	46	43	38
	Female	54	57	63
Audio	Yes	18	2	10
	No	82	98	90

Table II: Variance explained by retained dimensions

	Dog		Engine		Water	
Dim	%	Cum.	%	Cum.	%	Cum.
1	27.4	27.4	17.6	17.6	22.7	22.7
2	23.1	50.5	16.2	33.8	21.5	44.2
3	11.7	62.2	7.1	40.9	16.8	61.1
4	5.7	67.9	5.7	46.6	10.4	71.4
5	3.4	71.2	5.3	51.8	4.2	75.6
6	3.1	74.3	4.3	56.1	3.1	78.7
7	2.6	76.9	3.7	59.8	2.6	81.3
8	2.2	79.1	3.2	63		
9	2	81.2	2.9	65.9		
10			2.7	68.6		

participants completed a short web form consisting of questions regarding age, sex, first language, and audio expertise (‘Are you an audio engineer, an acoustician, a proficient musician, or similar?’) prior to the onset of the experiment. Participants were screened so as to only include those aged 18 years and over and with English as a first language. Demographic data for the three studies is displayed in Table I. As can be seen, participants for each study were broadly similar, with the exception that there were more participants aged 18-29 in the dog sound study. This is addressed in the discussion section.

Dog and engine sounds were either downloaded directly from Freesound (www.freesound.org) or taken from the collection of Freesound audio files curated by ESC-50 [20]. Water sounds were downloaded from the Adobe sound effects library. Engine and water sounds were edited so as to have 5s duration. Dog sounds were selected from longer clips so as to sound like a complete dog bark (mean=5.8s, *SD*=3.05s).

2.3. **Statistical analysis of category names**

2.3.1. Multinomial logit regression of category names

Each category name for each of the three types of sounds was coded as a word-type describing either the *source-event* (referring to the inferred source of the sound), the *acoustic* signal (explicitly referring to the sound itself), or a *subjective-state* (describing an emotional response caused by the sound, or of the sound source).

Multinomial logit regression models were used to compare the likelihood of each word-type being used as a category name for each type of sound. In each case the dependent variable was the word-type (e.g. subjective-state v source-event), and the independent variable was the type of sound. Multinomial logit regression models produce log-odds coefficients (*B*) that can be expressed as an odds ratio (e^B). These describe how many times more likely a word-type is to be used relative to another word-type for one type of sound relative to another.

Table III: Percentages of different word-types used to name categories

	Dog	Engine	Water
Source	24.0	60.0	92.5
Acoustic	34.0	37.1	5.8
Subjective	42.0	2.9	1.7

2.4. **Dimensions of category names and sounds**

2.4.1. Contingency tables

Data from each participant was collected as a matrix of 1s and 0s where rows represented individual sounds and columns represented the five categories created by that participant. 1s indicated that a sound had been sorted into the category. Tables representing each sound type were then combined into a single table with rows representing sounds and each of the 5*N* columns represented a category. Category names were initially processed by: removing white space and special characters; removing the word ‘sound’ or ‘sounds’; removing numbers; converting to lower case; and correcting the spelling. Category names were then stemmed (e.g. ‘drips and ‘dripping’ were reduced to ‘drip-’) before restoring all stems to the most common pre-stemming version of that word. Following this process each combined table was then consolidated by summing columns where category names were the same or synonymous. This process reduced the number of category names from 250 to 59 for dog sounds, from 245 to 96 for engine sounds, and from 240 to 63 for water sounds. A Pearson’s Chi-squared test confirmed a dependence between sounds and category names for dog sounds ($\chi^2(2494)=3977.3$, $p<0.001$), engine sounds ($\chi^2(3705)=3915.0$, $p<0.001$), and water sounds ($\chi^2(2728)=4314.2$, $p<0.001$).

2.4.2. Correspondence analysis

Correspondence analysis (CA) was used to identify the underlying dimensions of the data, in order to visualise the sounds and category names in the same space. CA is a method similar to the principal component analysis used to elicit dimensions of semantic differential analysis, but is suitable for use

Table IV: Results of the multinomial logit regression models. In each case the dependent variable was the word-type and the independent variables were the type of sound. Significance at <0.05 and <0.01 indicated by * and ** respectively.

		B	e^B	*SE*	*p*
Dog v Engine	Subjective v Source	3.60	36.7	0.42	<0.001**
	Acoustic v Source	0.83	2.3	0.22	<0.001**
	Acoustic v Subjective	-2.78	0.1	0.42	<0.001**
Dog v Water	Subjective v Source	3.11	22.4	0.3	<0.001**
	Acoustic v Source	4.58	97.5	0.5	<0.001**
	Acoustic v Subjective	-1.46	0.23	0.6	0.012*
Engine v Water	Subjective v Source	0.97	2.6	0.6	0.13
	Acoustic v Source	2.23	9.8	0.3	<0.001**
	Acoustic v Subjective	1.31	3.7	0.7	0.057

with categorical as opposed to continuous data [21, 22]. Each consolidated contingency table was submtited to CA using the FactoMineR package [22] in R V3.3.3. Dimensions with eignenvalues greater than would be the case were the data random were retained (see Table II).

2.4.3. Post-hoc analysis

Based upon the results of the multinomial logit regression models a post-hoc decision was taken test for correlation between the coordinates of dog category names describing subjective-states and measures of valence and arousal for those words, taken from a scored dataset of 13915 lemmas [23]. The engine sounds were tested for correlation between the coordinates of the sounds themselves and two simple acoustic features: fluctuation strength and sharpness, evaluated with dBFA software using the Zwicker and Fastl's criteria [24].

3. Results

The main purpose of this study was to test the hypothesis that category formation of different types of sounds would be based upon different attributes, using category names as a verbal correlate for category formation. The word-types used to name categories for each type of sound are presented in Table III. The series of multinomial logit regression models fitted to the word-type data are presented in Table IV. First consider the dog v engine sound models. There was 36.7 times the odds of naming a category using a word that described a subjective-state rather than the source-event when describing dog sounds compared to engine sounds, and 2.3 the odds of using a word describing the acoustic signal rather than the source-event. However, there was only 0.1 times the odds of using a word describing the acoustic signal rather than a subjective-state when describing dog sounds compared to engine sounds.

A similar pattern of results is seen in the dog v water sounds models. There was 22.4 times the odds of naming a category using a word that described a subjective-state rather than the source-event when describing dog sounds compared to water sounds, and 97.5 times the odds of using a word that described the acoustic signal rather than the source-event. However, there was only 0.23 times the odds of using a word describing the acoustic signal rather than a subjective-state.

In the engine v water sounds models, there was 9.8 times the odds of using words describing the acoustic signal rather than a source-event when naming engine sound categories compared to water sound categories. Other comparisons were not significant, although the comparison between words describing the acoustic signal and words describing a subjective-state were trending towards significance. To explore these results further, category names for dog, engine, and water sounds are plotted on the first two dimensions resulting from correspondence analysis of each consolidated contingency table in Figure 1. Note that these are the category names retained following consolidation, and therefore the ratio of word-types differs from those presented in Table III. First consider the category names for dog sounds (Fig. 1A). The majority of dog category names described subjective-states (Table III), and the odds of using this word type relative to another word type were greater for dogs than for engine and water sounds (Table IV). Category names appear to change from being broadly positive to being broadly negative along the first dimension, and from describing states of higher to lower arousal along the second dimension. This impression is confirmed by the coordinates of category names on the first dimension correlating with valence scores (Fig. 2A; $r_s(29)=-0.53$, $p<0.001$), and their coordinates on the second dimension correlating with arousal scores (Fig. 2B: $r_s(29)=-0.35$, $p=0.03$).

Next consider the category names for engine sounds (Fig. 1B). The proportion of category names referring to the acoustic signal was larger for engine sounds than for the other sound types (Table III), and the odds of naming categories using this word type rather than others greater relative to dog and water sounds. A visual inspection of the location of category names on the first two dimensions suggests that words relating to temporal regularity are to the left of the plot (‘constant’, ‘stuttering’, ‘chugging’) and those relating to irregularity to the right (‘staccato’, ‘stuttering’, ‘chugging’). Words relating to the sound having a sharp quality appear to be located towards the top of the plot (‘jarring’, ‘drilling’, ‘piercing’), whilst words such as ‘languid’, ‘muffled’, and ‘hum’ are towards the bottom. Consistent with this observation, fluctuation strength and sharpness of the engine sounds were found to correlate with the coordinate of each sound on dimension 1 (Fig. 3A; $r_s(38)= 0.81$, $p<0.001$)

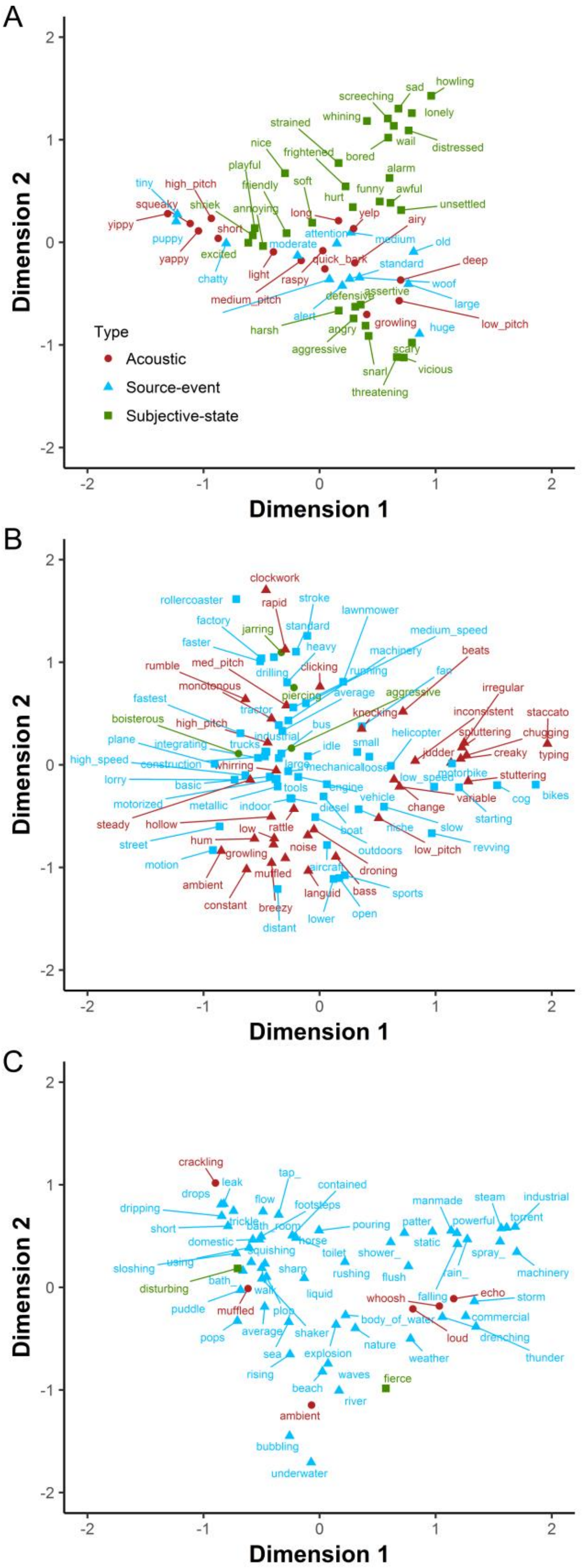

Figure 1: Category names of dog (A), engine (B), and water sounds (C) plotted on the first two dimensions

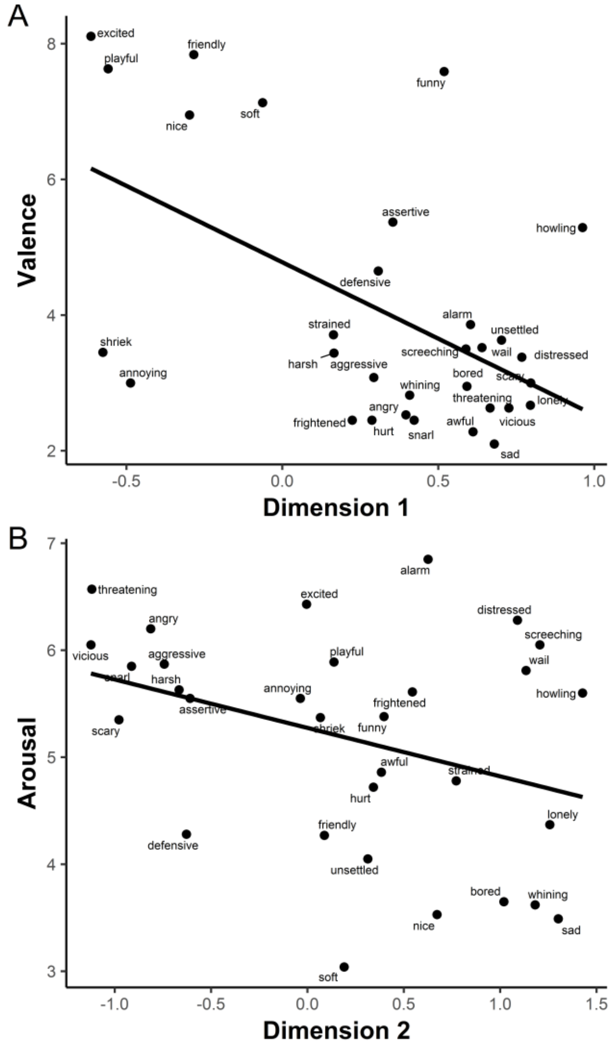

Figure 2: Valence (A) and Arousal (B) scores of dog category names plotted on the first and second dimensions respectively

and dimension 2 (Fig. 3B; r (38)=0.83, p<0.001) respectively.

The largest proportion of word type used to name water sound categories was source-event, and the odds of using this word-type rather than another was greater for this sound type than for dog and engine sounds. Accordingly, 'bath', 'bathroom' and 'domestic' are located close together on the first two dimensions of water sounds (Fig. 1C), as are 'beach', 'river', 'waves', and 'nature'. However, since the principal organizing principle for water sounds was source-event, an attempt was not made to find correlates for the first two dimensions in the same was as for dog and engine sounds.

4. Discussion

Previous work suggests that everyday listening is primarily concerned with gathering information about sound-sources [18, 25], and that sounds are therefore typically categorized by perceived similarities between sound sources rather than by abstracted acoustic features [14, 16-18, 26]. In the present study, this appears to have been the case for water sounds. However, using verbal correlates of categorization the present study suggests that dog sounds were categorized based primarily upon similarities in subjective-state, whilst engine sounds were categorized based upon explicit assessment of the acoustic signal. These results are consistent with source-event identification being the primary method for categorizing environmental sounds, and that this method was sufficient to categorize water sounds but not the other types of sound: in the case of water, it may be that participants were able to identify a sufficient variety of source-events with which to perform the categorization task, whereas this was not possible for dog and engine sounds. The first two dimensions of the engine sound data strongly correlated with fluctuation strength and sharpness, suggesting that these were strong organizing principles for these sounds. The two-dimensional plots presented here represent a mapping between the acoustic correlates and their subjective meaning: for example, it is possible to infer that as fluctuation strength increases the engine sounds here were perceived as being more 'chugging' and 'judder'-like.

For dog sounds the present study suggests that a third cue for categorization was used. In this case participants spontaneously employed an evaluation of the subjective-state of the sound source, or the emotional response that it caused. This finding, and the finding that the principal dimensions of the dog data correlated with valence and arousal, lends support to the circumplex model of affect [3-5] as a meaningful framework for understanding human categorization of some environmental sounds. Using the two-dimensional plot it can be said that dog sounds that elicit a strong valence response are perceived as 'excited' and 'playful', and those that elicit a large arousal response are perceived as 'vicious' and 'snarling'.

As noted previously a larger proportion of younger participants took part in the dog sound study thank in the engine and water sound study. As such it cannot be ruled out the use of subjective-state cues in the dog study was an effect of age. However, it is contended that it is more likely that

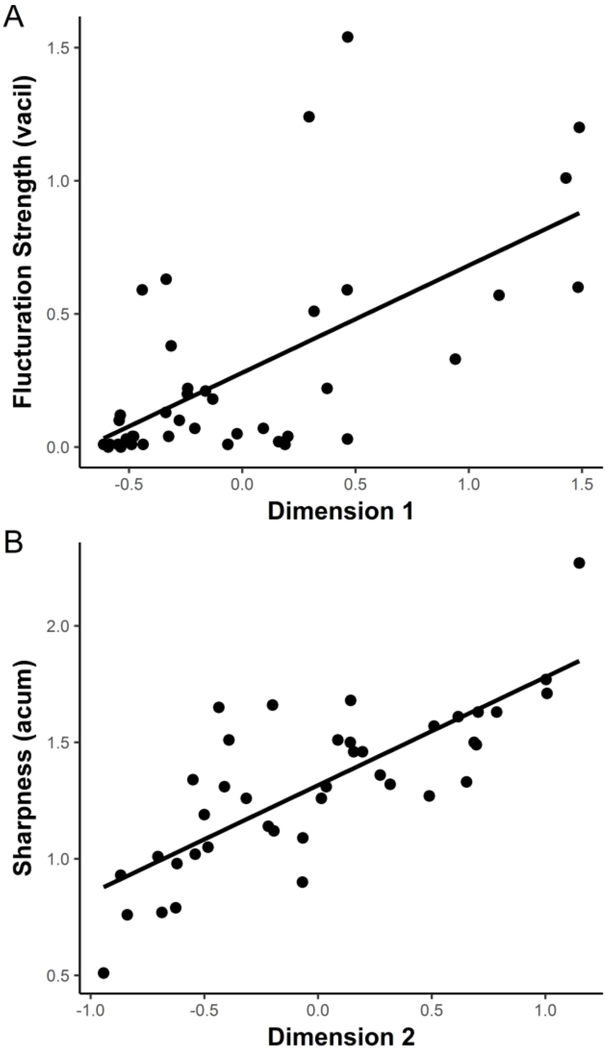

Figure 3: Fluctuation Strength (A) and Sharpness (B) of engine sounds plotted on the first and second dimensions respectively

the results represents the greater availability and utility of strategies for categorization. It should be also be noted that the categorization cue most used for engine sounds was that of an explicit assessment of the acoustic signal, despite there being fewer audio experts in the engine study.

In summary, the results here are consistent with categorization of sound being based upon different strategies depending on context and the availability of cues. It has implications for experimental methods in soundscapes that prescribe conceptual frameworks on test subjects. For example, a number of soundscape studies have reported principal dimensions related to subjective-states [6, 10-12]: however, careful consideration should be given to the appropriateness of prescribed semantic differential scales in future perceptual soundscape work.

Acknowledgement

This project has been funded by the EPSRC [EP/N014111/1] of the United Kingdon.

References

1. Dubois, D., *Categories as Acts of Meaning: The Case of Categories in Olfaction and Audition.* Cognitive Science Quarterly, 2000. **1**: p. 35-68.
2. Rosch, E.H., *Principles of categorization*, in *Cognition and categorization*, E.H. Rosch and B.B. Lloyd, Editors. 1978, Lawrence Erlbaum Associates.: Hillsdale, NJ.
3. Posner, J., J.A. Russell, and B.S. Peterson, *The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology.* Development and Psychopathology, 2005. **17**(3): p. 715-734.
4. Russell, A.R., *Core affect and the psychological construction of emotion.* Psychological Review, 2003. **100**(1): p. 145-172.
5. Russell, A.R., *A circumplex model of affect.* Journal of Personality and Social Psychology, 1980. **39**(6): p. 1161-1178.
6. Payne, S.R., P. Devine-Wright, and K.N. Irvine. *People's perceptions and classifications of sounds heard in urban parks: semantics, affect and restoration*. in *Inter-Noise*. 2007. Istanbul, Turkey: Proceedings of Inter-Noise 2007.
7. Axelsson, O., M.E. Nilsson, and B. Berglund, *A principal components model of soundscape perception.* J Acoust Soc Am, 2010. **128**(5): p. 2836-46.
8. Hong, J.Y. and J.Y. Jeon, *Influence of urban contexts on soundscape perceptions: A structural equation modeling approach.* Landscape and Urban Planning, 2015. **141**: p. 78-87.
9. Kawai, K., et al., *Personal evaluation structure of environmental sounds: experiments of subjective evaluation using subjects' own terms.* Journal of Sound and Vibration, 2004. **277**(3): p. 523-533.
10. Yu, B., J. Kang, and H. Ma, *Development of indicators for the soundscape in urban shopping streets.* Acta Acustica united with Acustica, 2016. **102**: p. 462-473.
11. Cain, R., P. Jennings, and J. Poxon, *The development and application of the emotional*

dimensions of a soundscape. Applied Acoustics, 2013. **74**(2): p. 232-239.

12. Kang, J. and M. Zhang, *Semantic differential analysis of the soundscape in urban open public spaces.* Building and Environment, 2010. **45**(1): p. 150-157.
13. Guastavino, C., *Categorization of environmental sounds.* Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 2007. **61**(1): p. 54-63.
14. Dubois, D., C. Guastavino, and M. Raimbault, *A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories.* . Acta acustica united with acustica, 2006. **92**(6): p. 865-874.
15. Morel, J., et al., *Road Traffic in Urban Areas: A Perceptual and Cognitive Typology of Pass-By Noises.* Acta Acustica united with Acustica, 2012. **98**(1): p. 166-178.
16. Guastavino, C., *The ideal urban soundscape: investigating the sound quality of French cities.* Acta Acustica united with Acustica, 2006. **92**: p. 945-951.
17. Houix, O., et al., *A lexical analysis of environmental sound categories.* J Exp Psychol Appl, 2012. **18**(1): p. 52-80.
18. Gaver, W.W., *What in the world do we hear?: an ecological approach to auditory event perception.* Ecological Psychology, 1993. **5**(1): p. 1-29.
19. Giordano, B.L., J. McDonnell, and S. McAdams, *Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environmental sounds.* Brain Cogn, 2010. **73**(1): p. 7-19.
20. Piczak, K.J. *ESC: Dataset for Environmental Sound Classification*. in *23rd ACM international conference on Multimedia*. 2015. ACM.
21. Greenacre, M., *Theory and application of correspondence analysis.* . 1984, London: Academic Press.
22. Lê, S., J. Josse, and F. Husson, *FactoMineR: An R package for multivariate analysis.* Journal of statistical software, 2008. **25**(1): p. 1-18.
23. Warriner, A.B., V. Kuperman, and M. Brysbaert, *Norms of valence, arousal, and dominance for 13,915 English lemmas.* Behavior Research Methods, 2013. **45**: p. 1191-1207.
24. Zwicker, E. and H. Fastl, *Psychoacoustics: Facts and models*. 2013: Springer Science & Business Media.
25. Schubert, E.D., *The role of auditory perception in language processing.* , in *Reading, perception and language: Papers from the World Congress on Dyslexia*, D.D. Duane and M.B. Rawson, Editors. 1975, York: Oxford, England.
26. Marcell, M.M., et al., *Confrontation naming of environmental sounds.* J Clin Exp Neuropsychol, 2000. **22**(6): p. 830-64.