Canadian Journal of Experimental Psychology
2007, Vol. 61, No. 1, 54-63

DOI: 10.1037/cjep2007006

Categorization of Environmental Sounds

Catherine Guastavino, McGill University, Graduate School of Library and Information Studies and Centre for Interdisciplinary Research on Music Media and Technology

Abstract This paper investigates the way in which people categorize environmental sounds in their everyday lives. Previous research has shown that isolated environmental sounds are categorized on the basis of high-level semantic features when the sounds can be attributed to specific sound sources. However, in the presence of numerous sound sources, as occur in most real-world situations, the process of source identification is often hindered. In the present study, a free categorization task with open-ended verbal descriptions was used to investigate auditory categories for environmental sounds in complex real-world sonic environments. Two main categories emerged from the free-sort, reflecting the absence or presence of human activity in relation to hedonic judgments. At a subordinate level, subcategories were mediated by the participant's reported interactions with the environment through socialized activities. The spontaneous verbal descriptors collected were successful in discriminating categories. These findings indicate that complex environmental sounds are processed and categorized as meaningful events providing relevant information about the environment. The relevance of situational factors in categorization and the notion of auditory category in its relation to linguistic labeling are then discussed.

Résumé Cet article examine la façon dont les gens catégorisent les bruits de l'environnement dans leur vie quotidienne. Des travaux antérieurs ont montré que les bruits isolés sont catégorisés selon des attributs sémantiques de haut niveau lorsque ces bruits peuvent êtres attribués à des sources sonores. Cependant, la présence simultanée de nombreuses sources sonores rend le processus d'identification de sources difficile dans les ambiances sonores auxquelles nous sommes exposés quotidiennement. Dans la présente étude, un test de catégorisation libre avec verbalisation permet d'identifier les catégories auditives de bruit de l'environnement élaborées à partir de scènes sonores complexes du quotidien. Une première distinction émerge entre les ambiances à dominante de bruits d'activité humaine et les ambiances à dominante de bruits de circulation, en relation avec des jugements hédoniques. À un niveau subordonné, le processus de catégorisation opère selon les interactions possibles avec l'environnement au travers d'activités sociales. De plus, les verbalisations libres permettent de discriminer les catégories principales. Ces résultats indiquent que les ambiances sonores complexes sont traitées et catégorisées comme événements porteurs de sens et sources d'information concernant l'environnement. Enfin, la pertinence de facteurs situationnels pour la catégorisation ainsi que les notions de catégories auditive et linguistique sont discutées.

Research on categorization of complex objects has much evolved in the past two decades with the study of people's natural categories, which are acquired and modified in the course of learning by interacting with category members in everyday life situations (e.g., Markman & Ross, 2003; Spalding & Murphy, 1996). Furthermore, there is recent neuroscientific evidence that many visual objects are organized categorically in the visual cortex by function rather than by surface feature similarity (Ishai, Underleider, & Haxby, 2000; Thompson-Schill, Aguirre, D'Esposito, & Farah, 1999). Together, these findings provide evidence that both bottom-up and top-down processes are involved, and that cognitive attributes of mental representations, such as previous knowledge and presentation context, play a determinant role in categorization. Recent research on food categories further revealed organization principles based on situational factors. Ross and Murphy (1999) observed the salience of *script categories*, that is, categories referring to the situation or time in which the food is eaten (e.g., foods to eat at breakfast time). It was further shown that food could be cross-classified either into *taxonomic categories* on the basis of similarity (foods of the same constitutive kinds) or into *script categories* on the basis of human interactions. Script categories were used to make event-type inferences and the authors argue that it could be of importance more generally to generate plans in larger goal-oriented tasks (e.g., deciding what to eat).

Most relevant to the present investigation is previous research on environmental sounds. According to Schubert (1975), "identification of sound sources and the behaviour of those sources is the primary task of the auditory system" (see also Bregman, 1990). Gaver (1993) further introduced the distinction between musical listening and everyday listening. Musical listening focuses on perceptual attributes of the sound itself (e.g., pitch, loudness), whereas everyday listening focuses on events to gather relevant information about our environment (e.g., car approaching), that is, not about the sound itself but rather about the sound sources and the actions producing sound. Studies on human-made domestic sounds validated this view (Ballas, 1993; Guyot, Castellengo, & Fabre 1997; Susini, Misdariis, Winsberg, & McAdams, 1998; Vanderveer, 1979), which suggests that categorization operates on the basis of source identification.

On the linguistic side, most natural languages reflect subjective, personal conceptualizations of environmental sounds. If visual objects are often described by simple lexical devices, there are few single words on which people agree as spontaneous descriptions of sounds (David, 1997; Dubois, 2000; Dubois, Guastavino, & Raimbault, 2006; Guyot, Castellego, & Fabre, 1997). The lack of basic lexicalized terms or a priori established categories questions the relationship between words and knowledge representations and suggests broadening the linguistic material to complete statements provided by language in discourse. Moreover, hedonic judgments happen to be relevant and even discriminant in free sorting tasks of recorded sounds. Ecological psychology recently drew attention to urban soundscapes, in which noise is emitted simultaneously by a wide variety of sources, to better understand how people sort out mixtures of sounds into discrete categories in their everyday lives. Maffiolo et al. (1997) investigated memory representations of familiar urban soundscapes with open questionnaires and mental maps. The analysis of verbal and graphical descriptions suggests that soundscapes are structured into complex *script categories* integrating notions of time, location, and activities. These notions are reflected in discourse by complex prepositional phrases with multiple complements such as "*riding motorcycles at Bastille on Saturday night.*" In a similar vein, Guastavino (2006) further investigated everyday listening using an emergent theme analysis on free-format verbal descriptions of familiar urban soundscapes in a mail survey with 77 respondents. The main categories of sounds identified were human sounds, traffic noise, natural sounds, and music. They were described in relationship to hedonic judgments spontaneously evoked by respondents. Human and natural sounds gave rise to positive judgments (except when reflecting anger), whereas mechanical sounds gave rise to negative judgments. This distinction was even observed within certain categories such as Music, which gave rise to two opposite qualitative evaluations depending on whether it reflected human activity directly ("musician") or indirectly ("loud-speakers," "car radio"). In the first case, it was perceived as lively and pleasant; in the latter, it was perceived as intrusive and therefore annoying. As regards mechanical sources, only electric cars and public transportation noise gave rise to positive judgments, in relation to environmental concerns. The subjective evaluation of acoustic phenomena is therefore closely linked to the appraisal of the sound source and the meaning attributed to it, suggesting that these semantic features might play an important role in categorization.

Furthermore, results from free sorting tasks (Maffiolo, 1999) and psycholinguistic analysis of verbal descriptions (Guastavino & Cheminée, 2003) converge to highlight a distinction between *sound events*, attributed to clearly identified sources, and *ambient noise*, in which sounds blur together into collective background noise. *Sound events* are spontaneously described with reference to specific sources, by nouns referring to the object (*truck, bus*) or part of the object (*engine, muffler*) generating the noise. These metonymies – substituting the name of the source producing sound for the name of the sound itself – indicate confusions between sounds and sources producing the sound, and further suggest that the acoustic phenomenon is not abstracted from the object generating the sound. On the contrary, in the descriptions of *ambient noise,* there are few references to the object source and a majority of simple adjectives referring to the physical features of the acoustic signal (namely, temporal structure and timbre), suggesting a more abstracted conceptualization of a sound in itself. Finally, the comparison of verbal data collected in actual environments and in laboratory experiments indicates that the sense of spatial immersion contributes to the cognitive representation of urban *ambient noise* (Guastavino & Cheminée, 2003, 2004). Guastavino, Katz, Polack, Levitin, & Dubois (2005) further showed that a multichannel surround sound reproduction, providing a strong feeling of immersion, was necessary to ensure that urban noise reproduced in a laboratory setting be processed as in everyday life situations. This ecologically valid protocol was used in the present experiment.

There are two bases for the present experiment. The first is the fact that categorization of complex environmental sounds is a relatively understudied domain. Previous research on isolated sounds indicate that categorization is largely controlled by object identification processes in relation to the appraisal and linguistic labeling of the sources. How about *ambient urban noise* in which numerous sounds blur together? The second base is the fact that people cross-classify everyday life objects into either *taxonomic categories* accord-

TABLE 1
Description of the Sound Samples in Terms of Location of the Recordings, Main Sources Presented (Traffic Noise and Other Sources), and Intensity (Leq dB SPL)

Name	Location	Traffic noise	Other sources	Leq dB SPL
Alligre	Open-air market near a busy street	Cars moving, horns, distant traffic noise.	People talking at various distances, money/change	73
Bastille	Open-air market	Faint traffic	Lots of people talking, bags rustling	73
Beaubourg 1	Walled-in plaza in pedestrian area	Motorcycle, horns, distant traffic noise	Faint voices, music (portable mechanical organ), birds	73
Beaubourg 2	Walled-in plaza in pedestrian area	Faint traffic	People talking, children	75
Montorgueil	Sidewalk café-district in pedestrian area	Very faint traffic	People talking, glasses/ dishes, faint music (speakers), birds	74
Montsouris	Large park in urban area	Cars moving, distant traffic noise	Children, birds, construction work	75
Notre Dame	Square close to road	Distant traffic noise	Lots of tourists talking and walking	78
Plantes	Park close to major road	Distant traffic noise	People walking, children, birds	76
Saint-Michel 1	Square next to major road	Cars, buses, accelerations	A few people talking	78
Saint-Michel 2	Sidewalk of a major road	Vehicles stopping, dense traffic noise	A few people walking, faint voices	76
Saint-Michel 3	Sidewalk of a major road	Cars, buses, motorcycles accelerating	A few people walking, faint voices	78
Saint-Michel 4	Square close to major road	Car, buses, noisy motorcycle	A few people talking	77
Sebastopol 1	Corner of intersection	Car, buses, motorcycles	A few people walking and talking	79
Sebastopol 2	Pedestrian area next to major road	Large vehicles, motorcycles	A few people walking, no voices	76
Stravinsky 1	Walled-in square with a few trees close to road	Distant traffic noise	Water running (fountain), people talking, children, birds	73
Stravinsky 2	Walled-in square with a few trees close to road	Cars, buses, motorcycles accelerating, slowing down	Water running (fountain) faint voices, birds	78

ing to low-level features or into *script categories* according to high-level features related to the situation of use or function of the objects. The present research question then, is to what extent complex environmental sounds will be categorized on the basis of object identification, function or low-level features. The study reported here investigates the categorization of urban *ambient noise* with an emphasis on retaining ecological validity insofar as it is possible, in an attempt to bridge the gap between natural category representations and laboratory studies on novel categories.

Method

Participants

The voluntary participants were 26 students and academic staff from the Université de Paris 6, with normal hearing. The participants ranged in age from 22 to 60 years (M = 30.7; SD = 12).

Material

The sound samples were drawn from previous studies conducted to identify typical urban environments (Guastavino & Cheminée, 2003). The environments were selected from a list of places identified as representative of city noises by Maffiolo et al. (1997) and recorded at walking head height (see Table 1 for a detailed description of the sound samples). The stimuli were 16 excerpts[1] spontaneously described as "ambient noise" or "background noise" by participants in previous experiments (Guastavino & Katz, 2004; Guastavino et al., 2005). All selections chosen contained both traffic noise and human sounds (vocal sounds and footsteps), as these were identified as the most characteristic acoustic components of urban soundscapes in another study (Guastavino, 2006). The test samples were recorded with a Soundfield ST250 microphone. All samples were 13 s long. The test samples were decoded by an Ambisonics decoder and presented on six speakers regularly spaced in the horizontal plane and a subwoofer (below 100 Hz), using the full "in-phase" decoding scheme. The listening room was an acoustically isolated room (floated construction) with a reverberation time of less than 0.2 s down to 40 Hz. The recordings were played at their original level (M = 75.7 dB SPL, SD = 2.2). A graphical computer interface was designed in Matlab (The Mathworks Inc., 2001) to play the sound samples.

Procedure

Upon arriving for the experiment, each participant filled out a questionnaire for gathering background information about gender, age, and number of years spent in urban areas. The participants were then given instructions about the free sorting task. They were asked to listen to the 16 urban background noises and group them into categories according to perceived similarity. The participants could create as many categories as they wished, and they were asked to name and describe each category. The experimenter followed a written protocol to explain the graphical interface. The sound samples were represented by stars followed by a number on the computer screen. For each participant, the numbers were assigned randomly to the test samples and the stars were randomly spread out over the whole screen. Participants could play the test samples as many times as desired by clicking on them. They used the mouse to freely organize the icons on the screen and group them into categories (drag and drop). After completing this first categorization task, participants were instructed about the second part of the test. Within each category, participants were asked to subcategorize the test samples according to perceived similarity using the same interface, and to name and describe each subcategory.

Analysis

A dissimilarity matrix was generated for each participant. The value in the i^{th} row and the j^{th} column of the dissimilarity matrix Δ, denoted ∂ij, is defined as follows:

1. $\partial ij = 0$ if i and j are in the same category and the same subcategory,
2. $\partial ij = 1$ if i and j are in the same category but not in the same subcategory,
3. $\partial ij = 2$ if i and j are not in the same category.

The global dissimilarity matrix was obtained by summing the individual dissimilarity matrices. This matrix was analyzed using both QualiTree, an additive tree scaling program developed by Barthelémy and Guénoche (1991), and Statistica for the MDS (multidimensional scaling) analysis. Additive-tree representations are used in a variety of disciplinary fields, ranging from computer science to biology, as a graphical representation of dissimilarity data. In cognitive psychology, semantic memory research has been a fruitful source of inspiration for the additive tree theory. Rosch's insights (1975, 1983) initiated the development of a new theory of "psychological similarity," which can be represented by additive trees (Sattah & Tversky, 1977). Additive trees were designed to account for several empirical observations, including the typicality effect. Indeed, the traditional ultrametric tree representation, in which all items are at the same distance from the root, forces all members of a category to be equivalent. The additive tree representation, with edges of varying lengths, seems more appropriate to represent a gradient of typicality. Formally, an additive tree is a connected, nondirected, and acyclic valuated graph, together with an additive distance. The items are represented by the "leaves" (or terminal nodes) of the tree, and the observed similarity between items is represented by the distance between leaves along the edges. The goodness-of-fit is expressed in terms of both metrical crite-

[1] A preliminary study was conducted to determine the maximal number (between 12 and 25) of test samples that could be handled in a free sorting task. Due to the complex nature of our samples, we had to restrict our set of stimuli to 16 soundscapes in the present experiment.

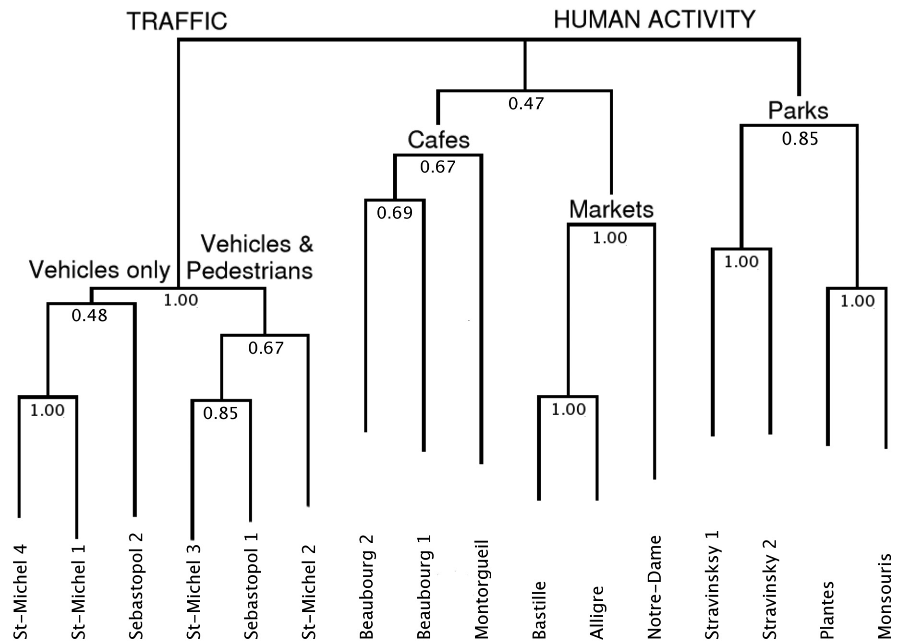

Figure 1. Additive tree representation of the dissimilarity matrix with verbal descriptors of the main categories and the subordinate categories given by the participants.

ria, based on edge lengths, and topological criteria, based on the tree topology (Guénoche & Garreta, 2001).

The verbal data consisted of spontaneous descriptions of the categories and the subcategories. A total of 412 phrasings were classified in semantic categories emerging from the spontaneous descriptions of the categories and subcategories. The verbal data was lemmatized, that is, inflectional and variant forms of a word are reduced to their lemma (i.e., their base form). Synonyms were grouped together, as well as linguistic devices constructed on the same stem (e.g., "life," "liveliness," "lively"). Lexical devices belonging to the same semantic field as indicated in a French thesaurus (Péchoin, 1992), were grouped into semantic themes. Semantic themes with fewer than three occurrences were excluded from the analysis. Two coders independently combined semantic themes into larger semantic categories, with an interrater agreement of 92%. Finally, all occurrences in each category were counted. A G-test (log-likelihood ration for goodness-of-fit) with Williams correction (for small sample sizes) was used to test the significance of the difference in frequency distribution of verbal descriptors between the two main categories.

As regards acoustic analysis, the sound pressure level of each sample was measured in the room at the position of the listener with a sound analyzer CEL 328, over 500-ms time periods. The third-octave band Leq (defined as the equivalent continuous sound pressure level for each third-octave over the duration of the sound samples) was derived and used to compute the spectral centroid and the loudness, using Zwicker's model (Zwicker, 1984) implemented in Matlab (The Mathworks Inc., 2001). Percentage sound pressure levels were also computed, namely, the 5, 10, 50, and 95 percentiles (L_5, L_{10} , L_{50}, and L_{95}). The L_5 (or L_{10}, L_{50}) is the sound level exceeded for 5% (or 10%, 50%) of the duration of the sound sample. Percentile sound pressure levels are commonly used in environmental acoustics to measure the sound level of emerging events. The L_{95} is the sound level exceeded for 95% of the time. It is commonly used to measure ambient sound level.

Results

Dissimilarity Analysis

The additive tree representing the data of the present experiment is presented in Figure 1. The dissimilarity between two objects is proportional to the length of the edge path connecting them in the tree. The number between 0 and 1 shown at each node is a topological indicator of goodness-of-fit of a given edge. The greater this number, the more reliable the grouping between the corresponding nodes.

The additive tree shown in Figure 1 can be consid-

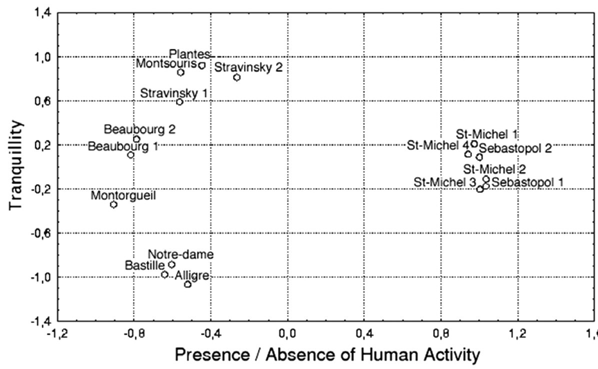

Figure 2. Two-dimensional MDS representation of the dissimilarity matrix.

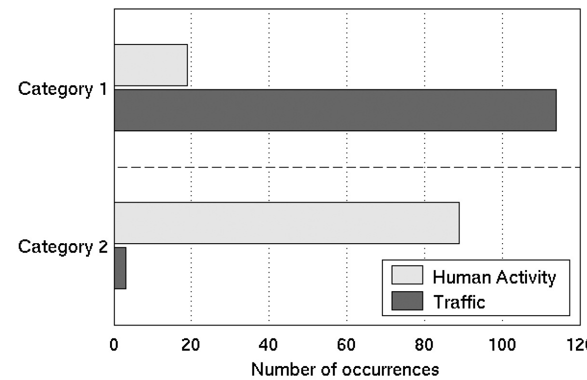

Figure 3. Distribution of verbal descriptors for the two main categories. The verbal data are classified into descriptions of human activity and descriptions of traffic.

ered as a good representation of the data in terms of both metrical criteria (stress = 0.06; 93% of variance explained) and topological criteria (89% of well represented quadruplets, arboricity = 0.85), according to Guénoche and Garreta (2001). The tree can be subdivided into two main categories. The first category, on the left side of Figure 1, consists of six items very close to each other, which correspond to recordings of two boulevards at different times of the day. This category can be subdivided into two subcategories of three items each. The other main category, on the right side of Figure 1, can also be subdivided into two subcategories. The first subcategory consists of recordings of commercial areas (outdoor markets, sidewalk café district), whereas the second subcategory consists mostly of recordings of parks.

The two-dimensional representation of the data derived from MDS analysis is presented in Figure 2 (stress = 0.049, 97.5% of variance explained). The dissimilarity between two objects is represented by the euclidean distance between the two corresponding points in space. Again, two main categories clearly emerge at a generic level along the first dimension, which accounts for 82% of the variance. The items of the first category (to the right) are very clustered whereas the items of the second category (to the left) are scattered along the second dimension, which accounts for 15.5% of the variance. A linguistic analysis of the verbal descriptors and an acoustic analysis of the sound samples were carried out to interpret these dimensions.

Linguistic Analysis

Two main categories were derived from the semantic analysis: traffic (117 occurrences) and human activity (121 occurrences). The *traffic* semantic category includes the generic term "traffic" itself (36 occurrences), as well as more specific descriptions in terms of vehicles ("cars," "truck," 34 occurrences), road type ("street," "boulevard," 19 occurrences), parts of the vehicles ("engines," 14 occurrences), and action of vehicles producing the noise ("acceleration," 14 occurrences). The *human activity* semantic category includes descriptions of human presence either explicitly ("pedestrians," "people," 38 occurrences) or by means of sounds reflecting human presence ("voice," "people speaking," 36 occurrences), as well as descriptors indicating activity ("liveliness," "animation," 32 occurrences) or human contacts ("encounter," "meet someone," 15 occurrences). These two semantic categories were compared with the groupings of sequences produced in the free sorting tasks at a generic level. Figure 3 shows the distribution of verbal descriptors for the two main categories elaborated at a generic level.

It can be seen that the verbal descriptors are very successful for discriminating the two main categories elaborated in the free sorting task. Indeed, the G-test with Williams correction (Sokal & Rolph, 1995) indicates a significant difference ($G = 7{,}056$, $p < 0.001$) between the distribution of the verbal data, classified into traffic and human activity descriptors, for both categories of sound samples. At a generic level, the presence or absence of human sounds is therefore discriminant for the process of categorization. The first dimension of the two-dimensional representation of the data (Figure 2) can thus be interpreted as the presence/absence of human activity. As regards qualitative evaluation, the human activity category was spontaneously described as pleasant (8 occurrences), whereas the traffic noise category was described as unpleasant (17 occurrences). A finer grain categorization distinguishes subcategories within each of these two categories. Half of the participants subcategorized traffic samples on the basis of the presence of human sounds in relation to the judgments of pleasantness: "unbearable" in the absence of human sounds versus "unpleasant" when few human sounds can be heard. This distinction can be seen on the additive tree representation (Figure 1). Twenty-three percent of the participants elaborated subcategories according to the type of vehicles ("bus," "heavy vehicles").

Meanwhile, human activity samples were subcategorized into *script categories* by 75% of the participants on the basis of the type of activities performed ("do the groceries," "have a drink," "take a walk"), ranging on a second dimension related to the degree of tranquillity (from "busy" open markets to "quiet" parks), as seen on the MDS representation (Figure 2). The traffic noise samples are located in the middle of this second dimension, since they reflect less animation than busy markets but more activity than the quiet parks. Human activity samples were described mostly by nouns referring to the type of locations ("market," "café" or "restaurant," "park") and identified sound sources ("vendors," "music," "birds"). Furthermore spatial attributes played an important role at a subordinate level, although all recordings were carried out in outdoor environments. One-third of the participants distinguished reverberant spaces (described as "reverberant," "semi-closed," "hall," "shopping mall") and open spaces ("open," "large squares," "outdoors") within the human activity category.

Acoustic Analysis

Several psychoacoustic parameters were computed and used for statistical analysis to predict category membership from physical parameters of the sound samples. Both logistic regression and discriminant analysis were carried out. The acoustic parameters investigated were the spectral centroid, the sound pressure level in dB Lin and dB(A)2 the sound pressure level in dB_{SPL} and dB(A)[2] in third octave bands. No significant interaction between psychoacoustic parameters and category membership was observed. Together with the observed lack of spontaneous descriptions of physical properties.[3] These results suggest that people categorize the sound samples on the basis of semantic features rather than perceptual ones.

[2] (A) refers to a weight function that approximates an equating loudness across frequency for low intensities.

[3] Only 8 descriptors of physical properties were collected, all describing traffic sequences: 4 referred to the sound pressure level ("loud") and 4 referred to pitch (as either "high" or "low").

Discussion

Two main categories were derived from the free sorting task at a generic level, on the basis of absence or presence of human activity in relation to judgments of pleasantness. Soundscapes in which mechanical sounds dominate were tightly grouped together. They were subcategorized according to either, again, the absence or presence of human activity, or the type of vehicles involved. Soundscapes in which human sounds dominate subdivide at a subordinate level into subcategories related to the different types of activities along a second dimension corresponding to tranquillity and ranging from busy markets to quiet parks.

The results reported here provide evidence for categorical perception of complex auditory scenes, with a clear distinction between ambient noise consisting of predominantly human sounds as opposed to predominantly traffic noise. Given the complex structure of the urban soundscapes, in which noise stems simultaneously from a variety of sources, categorical structure was not tested in a strict sense with stimuli ranging along a physical continuum. However, all sound samples contained both traffic noise and human sounds but were still classified into two qualitatively distinct categories reliably across participants. The strongly categorical structure derived from the free sorting task in this experiment confirms the findings of previous research on auditory categories for domestic human-made sounds (Dubois, 2000; Guyot, Castellengo, & Fabre, 1997) and urban soundscapes (Dubois, Guastavino, & Raimbault, 2006; Guastavino & Cheminée, 2003; Maffiolo, 1999), and further extends the results to ambient urban noise. Complex acoustic phenomena, as experienced in our everyday life, appear to be sorted out into discrete categories elaborated on the basis on semantic features (source identification, meaning attributed to the source), even in the presence of numerous sources.

Our findings provide further evidence that in the case of everyday life situations, acoustic phenomena are processed as meaningful events, in which case semantic attributes exert their influence over perceptual ones. It has often been reported that people sort auditory stimuli along a single dimension, specifically, along loudness, which often dominates other perceptual attributes in sound quality assessment, leading many researchers to equalize sound samples along these parameters (Susini, McAdams, & Winsberg, 1999). However, in the present experiment, semantic features overpowered perceptual attributes, and no effect of loudness was observed. This finding suggests that acoustic features are less relevant to sort out familiar everyday life auditory scenes than they are for abstracted stimuli.

Although using a different methodology, the results of our free sorting task are consistent with the linguistic analysis of spontaneous descriptions collected in the absence of auditory stimulation (Guastavino, 2006; Guastavino & Cheminée, 2003). The main two categories of sounds (predominantly human sounds vs. predominantly traffic noise) were also observed in the analysis of memory representation of familiar soundscapes in relationship to hedonic judgments. Soundscapes where human sounds are predominant tend to be perceived as more pleasant than soundscapes consisting of mechanical sounds predominantly. These findings provide empirical evidence for the sound classifications proposed by Schafer (1977). Schafer introduced the idea that soundscapes reflect human activities, and proposed four main categories of environmental sounds, namely, mechanical sounds (traffic noise in an urban context), human sounds (voices, footsteps), collective sounds (resulting from social activites), and sounds conveying information about the environment (e.g., spatial effects). The results reported here highlight the distinction between the first two categories, namely, traffic noise and human sounds, and suggest that categorization further operates according to the other two, namely, type of human activities (e.g., walks) and auditory information about the environment (open market vs. narrow streets) at a subordinate level. The salience of human voice has also been demonstrated using functional magnetic resonance imaging. Belin, Zatorre, Lafaille, Ahad, and Pike (2000) found voice-selective areas in human auditory cortex. Regions along the upper bank of the superior temporal sulcus showed greater neuronal activity when participants listened to vocal sounds than to nonvocal environmental sounds (including human nonvocal sounds). Hence, there is converging behavioural and neuroscientific evidence for the distinction between human sounds and mechanical sounds.

Distinctions between different types of human activities were observed in our study at a subordinate level. These findings are consistent with the *script categories* reported by Maffiolo et al. (1997) in their study of memory representations of urban soundscapes. Their psycholinguistic analysis of spontaneous verbal descriptions indicated that soundscapes give rise to complex semantic categories integrating notions of time, location, and activities. In our study, participants primarily referred to locations ("market," "parks") and activities ("take a walk," "have a drink") when describing the sound samples. Notions of time of the day were not explicitly evoked but were implicitly included in the descriptions of the activities. Similar evidence that environmental sounds are processed as meaningful events providing relevant information about possible interactions with the environment were reported in the context of auditory comfort evaluation inside trains (Mzali, Dubois, Polack, Letourneaux, & Poisson, 2001). Passengers' auditory judgments were collected during a train ride using open-ended questions. The respondents evaluated sounds in relationship to specific activities they were involved in, and the judgment of physical properties of sounds were modulated by the activities of respondents. For instance, the same sound can be judged as "quiet enough to sleep" but "too loud to have a discussion." These findings suggest that the elaboration of auditory categories is mediated by the participant's interactions with his/her environment through socialized activities, as in the case of food categories (Ross & Murphy, 1999). This leads us to consider relations between individual sensory experiences and collective representations of categories elaborated within a community through common practise or activities. We concur with Dubois (2000) and Gaver (1993) in thinking that familiar acoustic phenomena produce categories of *events* or *effects* of objects of the world, rather than "knowledge or theory-based" categories of objects, described within the theoretical frame of scientific conceptualizations.

The predominance of cognitive factors, namely, source and situation identification, is closely linked to linguistic constraints. Given the lack of simple lexicalized linguistic forms available in French to describe auditory phenomena (with the exception of music and speech), participants make use of complex phrasings, which are mostly constructed on denominations of sources. Numerous linguistic studies have stressed the diversity of linguistic forms across languages to describe cognitive categories and pointed to the interaction between linguistic constraints and concept elaboration (e.g., Wierzbicka, 1992 for colour perception and Waxman, 1999 for developmental psychology). Further research is required to investigate the complex relationship between linguistic and cognitive categories, more specifically, to better understand how individual sensory experiences are objectivized in language when elaborating shared cognitive categories. The results of Crawford, Regier, and Huttenlocher (2000) on spatial categories suggest that there is not necessarily a direct correspondence between categories labeled by simple lexicalized terms (linguistic categories) and categories for which there is no simple linguistic labeling (nonlinguistic categories). However, their results also indicate that the two systems are not independent. Linguistic and nonlinguistic spatial categories both rely on a common underlying structure, namely, the cardinal axes. These axes appear to play different roles in linguistic and nonlinguistic categorization of space. Indeed, the vertical axis, which serves as a category *prototype* in spatial language, serves as a *category boundary* in nonlinguistic organization of space. The study on nonlinguistic categories is of particular interest as it provides insights on how categories, for which there are no obvious "basic terms," are processed. More specifically, further research is required to investigate whether non-linguistic categories share the same cognitive status as linguistic categories (Waxman, 1999), since it is known that category labels indicate stable properties of categories, even in young children (Markman, 1989).

In the absence of established shared knowledge, categorization principles rely mostly on experiential knowledge (Barsalou, 1983), grounded in shared and socialized participants' activities (Dubois, 2000). Barsalou (1991) demonstrated the importance of situational factors in the construction of novel goal-derived *script categories* (e.g., places to go on vacations). Our findings, together with those of Ross and Murphy (1999) and Dubois (2000), suggest that some *script categories* are well established in memory and shared across participants. Future directions include investigating how the situational factors on which these categories rely can be used to derive inferences and predictions in order to guide action in everyday life goal-derived tasks.

This study was conducted at the Laboratoire d'Acoustique Musicale, Université Paris 6/CNRS (UMR 7604), 11 rue de Lourmel, 75015 Paris, France. The author is grateful to Daniel Levitin for helpful comments on previous versions of this manuscript, to Danièle Dubois, Pascale Cheminée, Jean-Dominque Polack, and Jacques Poitevineau for fruitful discussions, and to Vincent Rioux for the design of the graphical interface. This work was funded by the French Ministries of Research (funding for the author's doctoral work conducted at the Université Paris 6) and Environment (equipment). The writing of this paper was supported by CFI and FQRNT grants to the author.

References

Ballas, J. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology, 19*, 250-267.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*, 211-227.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1-64). New York: Academic Press.

Barthélémy, J. P., & Guénoche, A. (1991). *Trees and proximity representations*. New York: Wiley.

Belin, P., Zatorre, R., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309-312.

Bregman, A. (1990). *Auditory scene analysis: The perceptual organizations of sound*. Cambridge, MA: MIT Press.

Crawford, L. E., Regier, T., & Huttenlocher, J. (2000). Linguistic and non-linguistic spatial categories. *Cognition, 75*, 209-235.

David, S. (1997). Représentation d'objets sensoriels et marques de la personne. In D. Dubois (Ed.), *Catégorisation et cognition: de la perception au discours* (p. 211-242). Paris: Kimé.

Dubois, D. (2000). Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, *1*, 35-68.

Dubois, D., Guastavino, C., & Raimbault, M. (2006). A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories. *Acta Acustica united with Acustica, 92*(6), 865-874.

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*(1), 1-29.

Guastavino C. (2006). The ideal urban soundscape: Investigating the sound quality of French cities. *Acta Acustica united with Acustica, 92*(6), 945-951.

Guastavino, C., & Cheminée, P. (2003). Une approche psycholinguistique de la perception des basses fréquences : conceptualisations en langue, représentations cognitives et validité écologique. *Psychologie Francaise, 48*(4), 91-101.

Guastavino, C., & Cheminée, P. (2004). A psycholinguistic approach to the ecological validity of experimental settings. *Food Quality and Preference*, *15,* 884-886.

Guastavino, C., & Katz, B. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *Journal of the Acoustical Society of America. 116*(2), 1105-1115.

Guastavino, C., Katz, B., Polack, J.-D., Levitin, D., & Dubois, D. (2005). Ecological validity of sound reproduction systems. *Acustica united with Acustica, 91*(2), 333-341.

Guénoche A., & Garetta L. (2001). Can we have confidence in a tree representation? Proceedings of JOBIM'2000. *Lecture Notes in Computer Sciences, 2066*, 43-53.

Guyot, F., Castellengo, M., & Fabre, B. (1997). Étude de la catégorisation d'un corpus de bruits domestiques. In D. Dubois (Ed.), *Catégorisation et cognition: de la perception au discours*, (pp. 41-58). Paris: Kimé.

Ishai, A., Underleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron, 28,* 979-990.

Maffiolo, V. (1999). De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement sonore urbain. PhD thesis, Université du Maine, Le Mans, France.

Maffiolo, V., David, S., Dubois, D., Vogel, C., Castellengo, M., & Polack, J.-D. (1997). *Sound characterization of* *urban environment*. Proceedings of Internoise 97, Budapest.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592-613.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Mzali, M., Dubois, D., Polack, J.-D., Letourneaux, F., & Poisson, F. (2001, August). Mental representation of auditory comfort inside trains: Methodological and theoretical issues. *Proceedings of Internoise 2001*, The Hague.

Péchoin, D. (1992). *Thésaurus Larousse; des idées aux mots, des mots aux idées*. Paris: Larousse.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-233.

Rosch, E. (1983). Prototype classification and logical classification: The two systems. In E. K. Scholnik (Ed.), *New trends in conceptual representations*. Hillsdale, NJ: Lawrence Erlbaum.

Ross, B. H., & Murphy, G. L. (1999). Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*, 495-553.

Sattah, S., Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319-45.

Schafer, R. M. (1977). *The tuning of the world*. New York: Knopf.

Schubert, E. D. (1975). The role of auditory perception in language processing. In D. D. Duane & M. B. Rawson (Eds.), *Reading, perception, and language*, (pp. 97-130). Baltimore, MD: York Press.

Sokal, R. R., & Rohlf, F. J. (1995). *Biometry* (3rd ed.). New York: W. H. Freeman and Co.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 525-538.

Susini, P., McAdams, S., & Winsberg, S. (1999). A multidimensional technique for sound quality assessment. *Acta Acustica united with Acustica*, *85*(5), 650-656.

Susini, P., Misdariis, N., Winsberg, S., & McAdams, S. (1998). Caractérisation perceptive de bruits. *Acoustique et Techniques, 13*, 11-15.

The MathWorks Inc. (2001). *Matlab version 6.1*. The Mathworks Inc., Natick, MA. www.mathworks.com

Thompson-Schill, S. L., Aguirre, G. K., D'Esposito, M., & Farah, M. J. (1999). A neural basis for category and modality specific of semantic knowledge. *Neuropsychologia, 37*, 671-676.

Vanderveer, N. J. (1979). Ecological acoustics: Human perception of environmental sounds. *Dissertation Abstract International, 40*; 4543 B (University Microfilms number 8004001).

Waxman, S. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In D. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233-284). Cambridge, MA: MIT Press.

Wierzbicka, A. (1992). *Semantics, culture and cognition: Universal human concepts in culture-specific configurations*. Oxford: Oxford University Press.

Zwicker, E. (1984). BASIC-Program for calculating the loudness of sounds from their 1/3-octave band spectra according to ISO 532-B. *Acustica, 55*, 63-67.