Journal of Experimental Psychology: Applied
2012, Vol. 18, No. 1, 52–80

1076-898X/11/$12.00 DOI: 10.1037/a0026240

A Lexical Analysis of Environmental Sound Categories

Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis, and Patrick Susini
STMS-Ircam-CNRS, Paris, France

Isabel Urdapilleta
Université Paris VIII, Saint Denis

In this article we report on listener categorization of meaningful environmental sounds. A starting point for this study was the phenomenological taxonomy proposed by Gaver (1993b). In the first experimental study, 15 participants classified 60 environmental sounds and indicated the properties shared by the sounds in each class. In a second experimental study, 30 participants classified and described 56 sounds exclusively made by solid objects. The participants were required to concentrate on the actions causing the sounds independent of the sound source. The classifications were analyzed with a specific hierarchical cluster technique that accounted for possible cross-classifications, and the verbalizations were submitted to statistical lexical analyses. The results of the first study highlighted 4 main categories of sounds: solids, liquids, gases, and machines. The results of the second study indicated a distinction between discrete interactions (e.g., impacts) and continuous interactions (e.g., tearing) and suggested that actions and objects were not independent organizational principles. We propose a general structure of environmental sound categorization based on the sounds' temporal patterning, which has practical implications for the automatic classification of environmental sounds.

Keywords: environmental sounds, categorization, similarity, classification task, lexical analysis

Human participants continuously monitor their environment by identifying and interpreting environmental sounds (on occasions as diverse and common as shifting gears on a manual transmission, blending food, or hitting a tennis ball). *Environmental sounds* have been defined as "all naturally occurring sounds other than speech and music" (Gygi, Kidd, & Watson, 2007, p. 839; Gygi & Shafiro, 2007). They also are called *everyday sounds* (Ballas & Mullins, 1991). Different perceptual strategies and cognitive processes are involved when listening to environmental sounds. Categorization is an important cognitive process for understanding how listeners interpret environmental sounds and how they form categories. Experimental studies based on classification tasks have shown that participants use different types of similarity to group sounds. The taxonomy of simple sound interactions proposed by Gaver (1993b) has often been used as a starting point, but its psychological validity has only been partially tested. Therefore, our goal in this article is to explore the categorization of environmental sounds and, especially, the categories of sound events.

We first discuss these concepts, present methodological considerations. Next, we report two experiments: one focusing on a large set of environmental sounds and one focusing on a very common subcategory: sounds made by solid objects. Finally, we point out the different applications of our results concerning automatic sound classification and sound databases.

What We Are Listening to and How We Perceive Environmental Sounds

Studies reported here have shown the ability of listeners to recover some properties of sound sources, although not always accurately, indicating that context and semantic memory interact with sensory input. Truax (2001) defined *listening* as the ability to interpret information about the environment and to interact with it. He introduced different types of listening that imply an active role and involve different levels of attention: unexpected and diverted (with a global scope as a general scan of the environment) or focused on a particular sound. These types of listening are illustrated by listeners' reports of many different things when they describe what they hear. Gaver (1993b) hypothesized two modes of listening to account for this variety. He proposed that *musical listening* occurs when a listener focuses only on the acoustical properties of a sound (i.e., the listener describes perceptual qualities, such as pitch and timbre). He identified a second mode as everyday listening, which is what happens when a listener focuses on the cause of the sound. A listener can use both modes of listening for the same sound: "The distinction between everyday listening and musical listening is between experiences, not sounds" (Gaver, 1993b, p. 1).

In general, the literature suggests that listeners tend to favor everyday listening and focus on the physical phenomena that cause

This article was published Online First November 28, 2011.

Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis, and Patrick Susini, STMS-Ircam-CNRS, Paris, France; Isabel Urdapilleta, Parisian Laboratory of Social Psychology (LAPPS), Université Paris VIII, Saint Denis.

Guillaume Lemaitre is now at the Department of Psychology, Carnegie Mellon University. This work was funded by the FP6-NEST-PATH European Project No. 29085 CLOSED (closing the loop of sound evaluation and design). This work was also supported in part by National Science Foundation 0946550 to Laurie Heller. Portions of this work were presented in July 2008 at the Acoustics '08 Conference held in Paris, France and in April 2010 at the 159th joint meeting of the Acoustical Society of America/Noise-Con, Baltimore, MD.

Correspondence concerning this article should be addressed to Olivier Houix, Perception and Design Team, STMS-Ircam-CNRS, 1 Place Igor Stravinsky F-75004 Paris, France. E-mail: olivier.houix@ircam.fr

the sounds (the sound events) when interpreting them (Marcell, Borella, Greene, Kerr, & Rogers, 2000). In a previous study, Lemaitre, Houix, Misdariis, and Susini (2010) identified three properties that participants report when they describe a sound: acoustic properties (of the sound), causal properties (of the source), and semantic properties (related to the interpretation of the source). The information that listeners focus on depends on both the listener's expertise and the identifiability of the sound. Lay listeners tend to focus more spontaneously on the causal properties of the sound event. The study reported here followed this article and investigated in more detail how lay listeners organize sound events.

With regard to the perception of the causal properties of sounds, listeners may recover information both about the object(s) and the action(s) that caused a sound (Griffiths & Warren, 2004; McAdams, 1993; Michaels & Carello, 1981). For example, when listening to a small object being dropped and bouncing, listeners can recover the properties of the object (material, size, shape) or the action (the object is bouncing and not breaking). Participants can classify the same set of sounds according to either the objects or the actions (Dubois, 2000).

To classify sounds, listeners must rely on pieces of information that invariably specify these properties. Two different types of information have been introduced (invariants, to borrow a concept from ecological psychology). First, the structural invariants of an event specify the type of object and its properties (Kim, Effken, & Shaw, 1995; Pittenger & Shaw, 1975). Second, the transformational invariants specify the changes occurring in the sound source (Warren & Verbrugge, 1984). The sound event is defined as a unit of perception when both structural and transformational invariants are available for detection and evaluation (Shaw, Flascher, & Mace, 1996).

Many experimental studies have confirmed the human ability to recover both types of properties of a sound event (see Lemaitre et al., 2010, for a review): material (Giordano & McAdams, 2006; Kunkler-Peck & Turvey, 2000) or geometric (Carello, Anderson, & Kunkler-Peck, 1998; Grassi, 2005; Kunkler-Peck & Turvey, 2000; Lakatos, McAdams, & Caussé, 1997) and action (Cabe & Pittenger, 2000; Warren & Verbrugge, 1984).

However, the recovery of a sound event's properties is not always accurate and the recovery depends on whether sufficient acoustic information is available and how efficiently listeners can use it (Lutfi, 2001; Lutfi & Oh, 1997). Several studies have reported incorrect stereotypical associations between sounds and sound events; for instance, slow, loud, and low sounds are typically associated with male hand clappers (Repp, 1987) or walkers (Li, Logan, & Pastore, 1991). Lufti (2008) reviewed these phenomena. Moreover, sound events do not occur alone; they happen in a context and in conjunction with a wealth of other events perceived through different sensory modalities. The listener must isolate sound events from other concurrent events and integrate the different sensory inputs. In addition, the context influences the interpretation of sound events. Griffiths and Warren (2004) proposed a framework to understand the perception of auditory objects in a complex context. The different operational processing stages (from sound encoding to attribution of meaning) not only form a bottom-up process but also incorporate previous experiences of the auditory world and information from the other sensory modalities in a top-down process (see also the general model proposed by McAdams, 1993). As such, sound event interpretation has much in common with language perception: The context of production helps listeners to recognize the sound source (Ballas & Mullins, 1991) based on a priori knowledge about the environment (Ballas & Howard, 1987). Although the studies reported here studied the perception of sound events in isolation, it is important to remember that identification of environmental sounds is grounded not only in acoustic information from the sound event but also in the interactions of the sensory input with the semantic memory and cognitive processes. Categorization involving these different interactions could be an important cognitive process for understanding how environmental sounds are perceived in terms of categories without focusing on specific sound source properties.

Categories and Categorization

Categorization is a cognitive process that unites different entities of an equivalent status. Similarity plays a central role in the formation of categories. We present here a theoretical framework for categorization.

Following Sloutsky (2003), "Categories are defined as equivalence classes of different (i.e., discriminable) entities and categorization is the ability to form such categories and treat discriminable entities as members of an equivalence class" (p. 246).

In the case of sound perception, *semantic knowledge* (defined as a general knowledge about our world; see Tulving, 1972) allows listeners to identify the sound of a moving car as a motor vehicle, a car, or a wheeled vehicle. This inference depends on the process of the acoustic characteristics of the sound and on the context and/or the listener's expectations.

In what is one of the most influential approaches to conceptual categorization, Rosch (1978) proposed that categories do not have a stable core (i.e., they are not defined by a set of precise rules) but that members of a category are related to one another through family resemblance. Family resemblance is itself defined by the similarity to the prototypes of the categories. The context model (Medin & Schaffer, 1978) and its multidimensional extension (Nosofsky, 1986) postulate that all the members of a category are stored in memory. When context changes, different properties are engaged in the similarity between members of a category. The exemplar theory can be viewed as a generalization of the prototype theory with multiple prototypes.

However, perceptual similarity alone cannot explain category membership because, for example, context can change the properties on which the similarity is based (Tversky, 1977). Barsalou (1983) showed that ad hoc categories might be formed to achieve a specific goal (e.g., "things to pack for camping"). Categories are dependent on the context of the task. Thus, dissociations between similarity and categorization have been sometimes observed (Rips & Collins, 1993). With regard to sound perception, McAdams, Roussarie, Chaigne, and Giordano (2010) showed that different acoustic features are used for continuous judgments of similarity and categorization. Within these different approaches, similarity is a central concept that implies different properties depending on the context (Medin, Goldstone, & Gentner, 1993; Goldstone, 1994; Sloutsky, 2003). In our study, we focused on the type of similarity involved in categorization and on the categories.

Classification of Environmental Sounds and Strategies of Categorization

Classification and sorting tasks have been used to explore how people perceive environmental sounds. Categorization, implied during a classification task, allows the researcher to understand how categories are formed and which types of similarity are used. We present here the results of these experiments.

Vanderveer's (1979) seminal work showed that participants grouped sounds together that were caused by the same event (dropping a pen, a can, a piece of wood) or that shared the same acoustic properties. Marcell et al. (2000) reported the classification of 120 environmental sounds and found 27 categories corresponding to sound sources (four-legged animal, air transportation, human, tool, water/liquid), locations or contexts (kitchen, bathroom) or more abstract concepts (hygiene, sickness). Only a few classes were related to the acoustical properties. Gygi et al. (2007) reported similar results, finding 13 major categories based on 50 sounds. The most frequently used categories referred to the type of sources (animals/people, vehicles/mechanical, musical and water). In a lesser proportion, sounds were grouped by context (outdoor sports) or location (household, office, bar). Only rarely did the listeners use acoustic properties or emotional states for categorization.

These three studies did not analyze the relationships between the reported categories. However, studies by Guyot, Castellengo, and Fabre (1997) and Dubois (2000) focused on these relationships. Guyot et al. asked participants to listen to 25 domestic sounds and to classify them into different classes according to their perceptual similarity. After this classification, each participant described each category verbally. The results showed that the participants used two different strategies to classify sounds. The first strategy was based on psychoacoustic criteria (pitch, temporal evolution). The second strategy was based on the identification of the source. A close inspection of the second strategy showed two modalities of categorization, which were interpreted as the result of two different cognitive processes. One process grouped together similar sources and similar functions (e.g., opening/closing a window or a jar). The other process grouped sounds generated by the same movements or gestures (friction, etc.).

Guyot el al. (1997) authors proposed a framework for the classification of environmental sounds based on these two modalities. They used the three levels of abstraction formalized by Rosch (1978): superordinate, basic, and subordinate levels. At the basic level, listeners identified action. At the subordinate level, they identified the source. At the superordinate level, they identified abstract production, such as mechanical sounds, electronic sounds, and so forth (see Figure 1).

If we omit the few categories exclusively based on acoustical properties, these different studies showed different types of categories:

Sound sources (inanimate and animate), actions, and movements causing the sounds (Guyot et al., 1997; Gygi et al., 2007; Marcell et al., 2000; Vanderveer, 1979);

- Context, location where the event took place, or where the sound could be heard (Gygi et al., 2007; Marcell et al., 2000);
- Meaning associated with the identification of the sound sources, such as function and abstract concept (Guyot et al., 1997; Gygi et al., 2007; Marcell et al., 2000);
- Emotional responses, such as annoying, startling, or alerting (Gygi et al., 2007).

These different types of categories are not mutually exclusive and can be mixed during a classification task (across participants and/or for one participant) because an object or a sound can belong to multiple categories corresponding to alternative conceptual organizations. This cognitive process has been called *cross-classifications* (Ross & Murphy, 1999).

The studies reported above have shown that listeners use different strategies to form categories of environmental sounds. Different cognitive processes for environmental sounds may explain these strategies. Gygi et al. (2007) suggested that these results reflect a difference between categorization based on acoustic similarities and categorization based on the subject's goals and theories. Finally, these findings can also be explained by the fact that nonverbal sounds of living things and sounds produced by physical events such as tools, liquids, and dropped objects are differently processed in the brain (Lewis, 2004). For instance, Giordano, McDonnell, and McAdams (2010) showed that listeners more often used symbolic information for the sounds produced by living things and acoustic information for nonliving sounds. This is also true for the sounds caused by hand actions, such as clapping, or mouth action-related sounds, such as laughing, compared to nonaction-related sounds such as water boiling (Aglioti & Pazzaglia, 2010; Pizzamiglio et al., 2005).

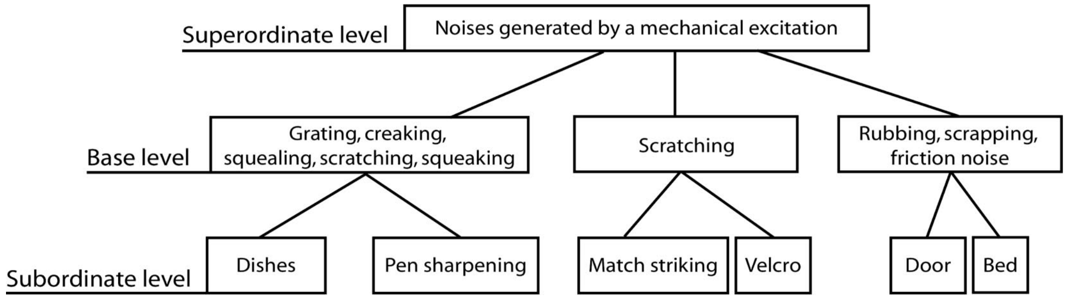


Figure 1. Hierarchical organization of domestic noises, translated from the French. Adapted from "Étude de la catégorisation d"un corpus de bruits domestiques" by F. Guyot, M. Castellengo, and B. Fabre, 1997, *Edition Krimé,* p. 43. Copyright 1997 by Edition Krimé. Adapted with permission.

The selection of the sounds for our experimental studies is a crucial choice that influences their categorization, and this selection has been taken into account with the other methodological considerations.

Methodological Considerations

This section discusses the methodological considerations that may influence the results of the studies, such as the instructions given to the participants and the choice of the statistical and lexical analyses. The previously presented experimental studies used different types of instructions, and their results were examined with different statistical and lexical analyses. First, the classification task gave participants different instructions to group sounds: based on perceptual similarity (Guyot, 1996; Guyot et al., 1997), the similarity of sounds (VanDerveer, 1979), and explicitly asking participants "to place something with other objects that have similar characteristics and are members of the same group" (Marcell et al., 2000, p. 853) or to "put together sounds that seem to belong together" (Gygi et al., 2007, p. 851). Some of the differences between the results of the aforementioned studies could be explained by the different emphasis on different aspects of the sounds or sound events.

Second, different mathematical representations were used. One technique consisted of computing a similarity matrix from the raw classification and representing it with an additive or hierarchical tree (Guyot et al., 1997) or a handmade topological representation (Vanderveer, 1979). Such statistical analyses represent results by averaged classes of individual results within a tree and can potentially mix classes based on different strategies into a single structure.

Third, once a mathematical structure accounting for the participants' classification was created, the subject's descriptions of their classification were analyzed to interpret the resulting structure. Different approaches were used to collect and analyze the descriptions of the classes. For example, Vanderveer (1979) and Guyot et al. (1997) required their subjects to describe their categories, but they did not systematically analyze these descriptions.

Marcell et al. (2000) developed a specific technique to analyze the descriptions, and Gygi et al. (2007) also used this technique. They required the subjects to write down descriptive labels for each category they created. The labels with similar meanings were then grouped by two independent judges (who were unaware of the purpose of the study) who assigned a label to each category. Combining the two judges' interpretations clarified the different meanings and synonyms found in the descriptions of the categories. The resulting labels were validated during a second experiment in both studies. However, this approach only highlighted general categories. The simplification of the vocabulary by the judges obscured the analysis of possible subcategories. Gygi et al. (2007) also tried to explain the structure of the categories with a multidimensional scaling representation of data from judgments of similarity between pairs of sounds, but the interpretation of the results was difficult.

We used two specific techniques to overcome what we identified as potential limitations in existing studies of environmental sound categorization. Specific techniques allowed us to use several dendrograms to represent the data to account for the possibility of different strategies (cross-classification) within and between the subjects. Furthermore, we used a systematic lexical analysis technique, based on statistical analysis of textual data, to minimize subjective bias. The methodological aspects were important for structuring our experimental studies, but our specific focus on sound event categories forced us to identify criteria for selecting the sounds.

A Map of Everyday Sounds

This section presents a taxonomy of everyday sounds that represent different classes of physical interactions (solids, liquids, gasses). Gaver (1993a, 1993b) introduced a map of everyday sounds as a starting point to study how listeners identify environmental sounds. This map is shown in Figure 2. It has a hierarchical structure (similar to a taxonomy) and is based on the physics of sound-producing events.

The first division of the hierarchical structure corresponds to the idea that each sound event involves an interaction of different types of materials: liquid, solid, or gas. The first class corresponds to the vibrating objects (solids) that are generally present in our environment, such as knocking on a door or scratching with a nail. The second class consists of aerodynamic sounds (gasses). These sounds can be produced by, for example, an exploding balloon or wind blowing through a tube. The last class groups sounds of liquids, which may be created by pouring a liquid in a glass or a drop of milk in a cup of tea. Within these three large classes, Gaver (1993b) introduced subclasses of excitation corresponding respectively to the following events:

- Solids: impacts (discrete and short inputs of energy), scrapes (continuous inputs of energy), rolling movements (involving gears or pulleys), and deformation (crumpling or crushing);
- Liquids: discrete drips or continuous pouring, splashes, or ripples;
- Gasses: explosions (sudden changes of pressure), whooshes, and winds (continuous introductions of pressure variations).

For each of these basic events, the author suggested a set of physical properties that may be perceptually relevant or may influence perception (e.g., for the sounds of scraping, the physical properties of texture, *material, speed, acceleration,* or *force*; for the sounds of wind, the *volume* and *force regularity*).

The author also assumed that different combinations of basic events produce more complex events. Temporally patterned events are composed of patterns of simple basic events. For example, crumpling is typically a patterned deformation sound. A mix of different types of basic events produces compound events. For instance, closing a door involves a scraping sound followed by an impact. The last combination, called hybrid events, is based on events involving different types of material. For instance, a drip in a container involves both the vibration of the liquid and of its container.

According to Gaver (1993b), this framework is far from exhaustive (e.g., electronic and vocalization sounds are not included), and it may be organized differently. Nevertheless, this map describes and analyzes a large part of the most common sound-producing events, and it suggests the physical attributes that may be relevant for a listener to identify these events. However, this taxonomy was motivated by only a physical analysis of the sound-producing events.

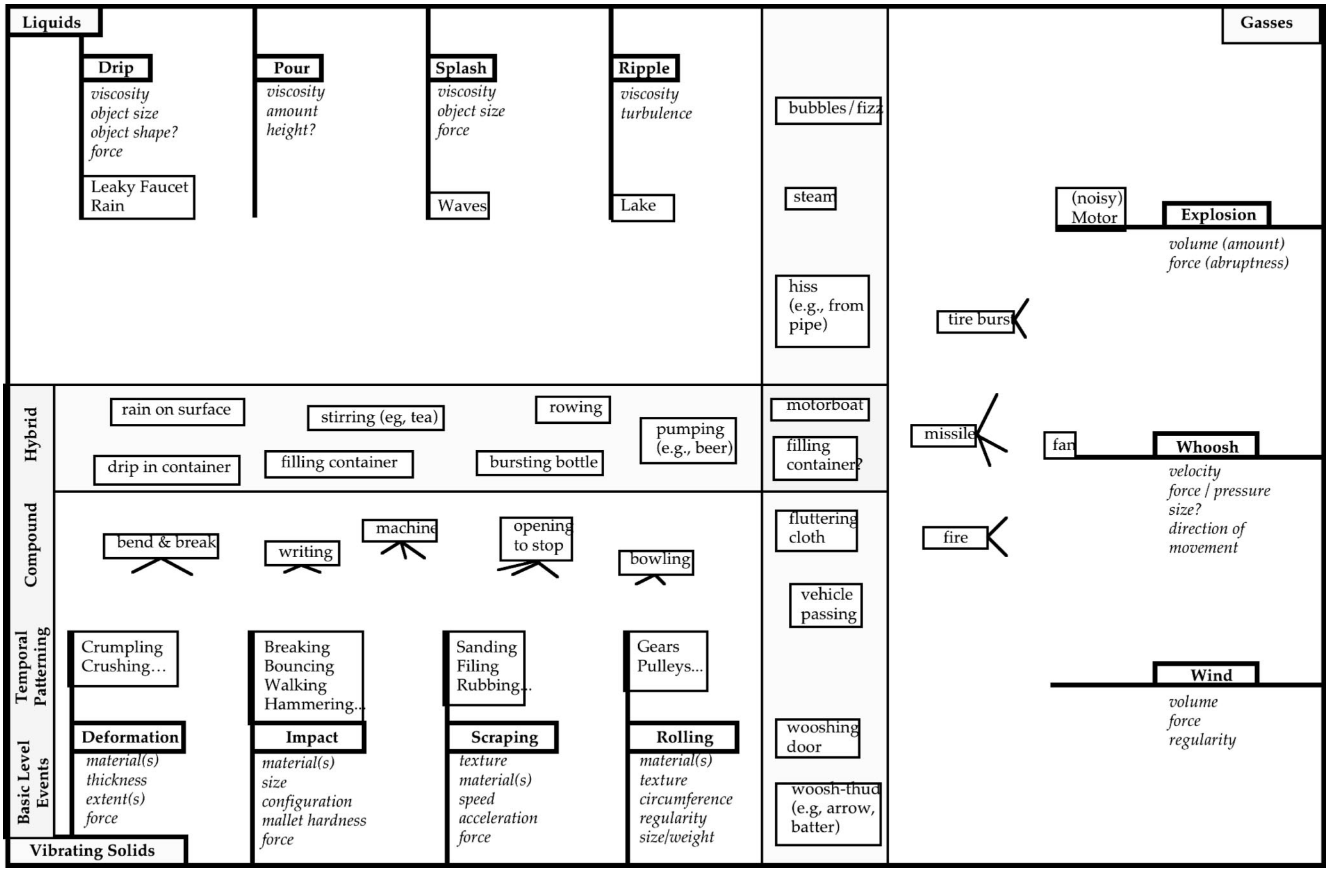


Figure 2. A map of everyday sounds. Three fundamental sources (vibrating solids, liquids, and aerodynamic sounds) are shown in the three overlapping sections. Within each section, basic sound-producing events are shown in bold, and their relevant attributes are in italics. Complexity increases toward the center of the figure, with examples showing temporally patterned, compound, and hybrid sounds. Words associated with a question mark are categories or attributes that remained questions under discussion. From "What in the World Do We Hear? An Ecological Approach to Auditory Event Perception," by W. W. Gaver, 1993, *Ecological Psychology*, *5*, p. 15. Copyright 1993 by Taylor & Francis Group, Inc. Reprinted with permission of the publisher (Taylor & Francis Group, http://www.informaworld.com).

Our aim in the study was to examine the categorization of easily identifiable sounds (especially for Experiment 2) and sounds produced by physical interactions (excluding sounds of living beings) to explore the psychological relevance of Gaver's (1993b) map of everyday sounds. Gygi et al. (2007) claimed that this map was incomplete because several categories are missing (vocalizations and electronically synthesized sounds). However, introducing sounds of living beings generally results in the participants' binary categorization of the sounds and usually hinders an analysis of potentially finer grained categories. We therefore only used sounds produced by physical interactions or by appliances (Experiment 1) and focused specifically on the sounds of solid objects (Experiment 2).

Outline of the Studies

The first experiment aimed to examine the structure and organization of the perceptual categories of sounds produced by physical interactions or appliances. The second experiment focused only on the sounds made by solid objects. We compared the resulting structure and organization with the phenomenological taxonomy proposed by Gaver (1993b) and the structure presented by Guyot et al. (1997). We used sorting tasks to emphasize the information related to the events causing the sounds (Aldrich, Hellier, & Edworthy, 2009).

Across these two experiments, our aim was as follows:

- To observe whether the classes created by the participants reflected different categories of physical events, as hypothesized by Gaver (1993b).
- To observe whether the categories of sounds made by solid objects also reflected different categories of actions.
- To explore how these two classifications could be unified in the same structure, such as the organization proposed by Guyot et al. (1997).

In the first experiment, lay participants were asked to freely sort a set of environmental sounds with different degrees of identifiability. The identifiability of the sounds was measured in Lemaitre et al.

(2010). The participants had to describe each class they made after the classification. The specific statistical analysis of the classification data and the lexical analysis of the descriptions highlighted four large categories (liquids, solids, gases, and machines).

In a second experiment, we exclusively focused on sounds produced by solid objects. We used the same protocol and analyses, but we asked the participants to group sounds produced by the same physical action, independent of the sound source.

Experiment 1: Free Classification of a Large Set of Environmental Sounds

Method

Participants. Fifteen participants (six women and nine men) volunteered as listeners and were paid for their participation. The participants were between 19 and 64 years of age (median: 32 years). All reported having normal hearing, and all participants were native French speakers.

The participants were initially selected based on questionnaires they had completed in previous experiments. From their answers, we defined nonexpert or lay participants as those who did not belong to the following categories:

- A professional musician or with a background in musical education (conservatory of music or musicological studies).
- A professional artist who regularly worked with sounds (sound installations, performances, etc.).
- A professional or semiprofessional sound engineer or recording engineer.
- A scientist working in the fields of sound perception, acoustics, or sound signal processing.

Stimuli. We selected 60 stimuli with different degrees of identification from a study by Lemaitre et al. (2010). The 60 sounds were monophonic recordings of events usually occurring in a kitchen. Our selection was guided by a list of sounds that people usually hear in their kitchen. This list was extracted from a questionnaire completed by the members of the laboratory and an exhaustive listening of different commercial sound libraries: Hollywood Edge Premiere Edition I, II and III (The Hollywwod Edge, Hollywood, USA), Sound Ideas General Series 6000 (Sound Ideas, Ontario, Canada), Soundscan v2 Vol. 61 (Ultimate Sound Bank, Paris, France) and Blue Box Audio Wav (Best Service GmbH, München, Germany). We also attempted to include the different sound examples of the taxonomy proposed by Gaver (1993b), excluding outdoor sounds such as waterfall and fireworks. We chose the context of a kitchen because of its familiarity for participants and the large variety of sound sources in a kitchen environment, such as motor, liquid, and appliance sounds. The context was provided to the participants prior to the experiment to prevent different sound recognitions due to a misinterpretation of the context inferred from the sounds. For example, a participant could recognize some sounds incorrectly if the sounds were believed to be produced in the context of sports. In contrast to Giordano et al. (2010), we excluded sounds produced directly by the bodies of living beings, such as eating, transport, or vocalization sounds generated by humans or nonhumans because the binary categorization between living and nonliving sounds could mask other categories. The stimuli included action-related sounds produced directly or indirectly (e.g., with the use of a tool) by the manipulation of an object or the interaction of several objects by an agent and nonaction sounds that do not directly involve the action of an agent, such as a gas oven sound.

The sound levels were ecologically adjusted to reproduce the expected level of sounds heard in a kitchen (Lemaitre et al., 2010). They were required to adjust individually the level of the sounds to what they should sound like in the kitchen. During the adjustments, the description of the sounds was given. The sounds had a 16-bit resolution and a sampling rate of 44.1 kHz. The labels of the sound files and their associated levels are presented in Appendix A.

Apparatus. The sounds were played by a Macintosh Mac Pro (Mac OS X v10.4 Tiger) workstation with a Motu Firewire 828 sound card (Motu audio, Massachusetts, USA). The stimuli were amplified dichotically over a pair of Yamaha MSP5 loudspeakers (Yamaha Music Europe Gmbh, Rellingen, Germany). Participants were seated in a double-walled sound isolation booth (IAC, New York, USA). Levels were calibrated using a Brüel & Kjær 2238 Mediator sound-level meter (Brüel & Kjær Sound & Vibration, Nærum, Germany; see Appendix A). The software used to run the experiment and to implement the graphical interface[1] was Matlab 7.0.41 (The MathWorks Inc., Massachusetts, USA).

Procedure. The procedure had two steps. The participants received written instructions (in French) explaining the sorting task. The participants saw a white screen on which red dots labeled from one to 60 were drawn, with each dot corresponding to a sound. The labeling was different for each participant. The participants could hear the sound by double-clicking on a dot, and they could move the dots to create classes. We asked the participants to use their own criteria to group the sounds. They were allowed to form as many classes as they wished and to put as many sounds in each class as they desired. The sound could not be interrupted while it played. Following the sorting task, in a second step, the participants were asked to type the properties shared by the sounds in each class. We asked the participants to spend about 45 min on the task.

Results

The participants created between six and 24 classes (median: nine). We used a two-step analysis to study and interpret these classes. The first step was a hierarchical cluster *analysis* that represented the classification data in two hierarchical dendrograms. In the second step, the clusters extracted from these dendrograms were interpreted by submitting the descriptions of the participants' classes to lexical analyses.

Hierarchical cluster analysis. Each individual partition was coded as a 60×60 incidence matrix (Borg & Groenen, 1997). The individual matrices were averaged to form a co-occurrence matrix of proximities. We submitted the co-occurrence matrix to a specific type of hierarchical cluster analysis using a Matlab algorithm[2] provided by Hubert, Arabie, and Meulman (2006). Whereas the classical hierarchical cluster analysis fits a single dendrogram to the classification data, the method used here fitted several dendrograms, allowing a finer-grained analysis of the participants' strat-

[1] Interface developed by Vincent Rioux.

[2] Freely available (http://cda.psych.uiuc.edu/srpm_mfiles/).

egies (see Appendix B for more detail). In our case, a recursive procedure fitted two dendrograms to the co-occurrence matrix. This procedure ensured that the primary dendrogram represented a classification strategy shared across all the participants. The secondary dendrogram, fitted to the residuals of the first analysis, accounted for possible cross-classification strategies between and within the subjects (Ross & Murphy, 1999). We used the secondary dendrogram to account for the possibility that some participants had used different strategies or criteria, which might have otherwise been masked or considered noise in the primary dendrogram. Using such a procedure ensured that the primary dendrogram was not contaminated by other marginal strategies. We report the detailed analyses of the primary dendrogram here. The secondary dendrogram is reported in Appendix C.

Figure 3 and Appendix C represent the primary and the secondary dendrograms, respectively. The variance accounted for (VAF) had a value of .92, indicating that the multiple tree representation was accurate.

To identify significant clusters in a dendrogram, it is usually cut at a given fusion level. As an alternative clustering method, in this study, we used a threshold of inconsistency. The advantage of using the inconsistency coefficient is that it emphasizes compact subclasses that would not be revealed using the fusion level. This method provided us with the clusters highlighted in Figure 3 and Appendix C.

For the primary dendrogram represented in Figure 3, we distinguished the clusters at two levels: the main clusters (identified with letters A, B, C, and D) and the subclusters (A_a to C_b).

The degree of identifiability of each sound is reported in Figure 3 and Appendix C (see the next paragraphs) and was provided by the causal uncertainty *Hcu* measured by Lemaitre et al. (2010).

To interpret the classes resulting from these analyses, we submitted the descriptions of the participants' classes to a set of lexical analyses.

Synopsis of the lexical analyses. We used two different lexical analyses to interpret the main and the subclusters in the primary dendrogram. We present here detailed lexical analyses corresponding to the primary dendrogram and a short summary for the secondary dendrogram. First, we analyzed the verbalizations to identify categories of terms that could reflect the formation of the clusters. Next, we produced a lexical portrait of each cluster. The first lexical analysis was intended to identify groups of common terms across the verbalizations, called categories of representative verbalizations (CRVs), and associate them with the clusters of the hierarchical cluster analysis. The second lexical analysis calculated the lexical occurrences in each subcluster to form their lexical portraits. In both cases, we aggregated the descriptions of the sounds included in each cluster for all 15 participants. Therefore, we did not consider the differences between each individual category and the clusters resulting from the hierarchical analysis.

First lexical analysis: Categories of representative verbalizations. We used a statistical analysis of textual data (Alceste[3] software; Reinert, 1986) to determine whether the lexical fields extracted from the verbalizations could reflect how the main clusters (A, B, C, and D) were structured. This analysis had two steps (see Appendix D). In the first step, the analysis extracted the most significant CRVs. The descriptions of the sounds associated with each significant main cluster (A, B, C, and D) were automatically split into elementary context units (ECUs). Because each sound belongs to a specific main cluster, the different ECUs are also identified as linked to a significant variable (the main clusters A, B, C, or D) called the initial context unit (ICU). The words in the ECU were reduced to simple forms (i.e., the different inflected forms of a word were grouped together and processed as a single item). A contingency table was built that indicated the presence or absence of the simple forms in each ECU. The contingency table of simple forms was submitted to a hierarchical decreasing classification analysis that resulted in a set of organized categories of terms. The categories of terms that were the most different between ECUs formed the CRVs. Each CRV was therefore characterized by a unique series of terms. In the second automatic step, the analysis computed which ICUs (the significant variables A, B, C, or D) best characterized the different CRVs by the strength of association. This second step allowed us to compare the structure between the CRV and the structure of the main clusters in the tree. Next, we interpreted the lexical fields of each CRV.

First step. The descriptions were created for 4,214 associated occurrences and 332 distinctive forms. The analysis used 176 analyzed forms (without the function words) for 2,648 occurrences. The corpus was divided into 96 ECUs, and 72% of the ECUs were considered in the analysis. The most frequently cited nouns were *water* (126 occurrences), *liquid* (77 occurrences), and *kitchen* (65 occurrences). The most frequently cited verbs were *to rub* (43 occurrences), *to crumple* (31 occurrences), and *to drain off* (29 occurrences).

This first lexical analysis identified five CRVs (I_e, II_e, III_e, IV_e, and V_e). The distribution of the ECUs in each CRV was 23.19%, 26.09%, 17.39%, 14.49%, and 18.34%, respectively. Figure 4 presents the hierarchical structure of the CRVs. This structure represents the strongest vocabulary oppositions (i.e., the terms in each CRV are the most distinctive of this CRV). The homogeneity of each CRV (and its distinctiveness from the other CRVs) was measured by chi-square values (the chi-square values measured the association strength of each terms with the CRV). Larger values of chi-square indicated that the terms were specific to this CRV. Smaller values indicated that the lexical terms of a CRV were less specific. For the both Experiments 1 and 2, we fixed arbitrarily the cut off value of chi-square equal to 15 to keep only the relevant terms (Figures 4 and 7).

To summarize, we observed across these lexical fields a clear distinction between a series of terms reflecting different physical sound sources (liquids, machines, gases, and solids) and other terms reflecting different types of actions and movements or functions (opening, closing, cutting, scratching, etc.). We also observed a few terms related to acoustic properties and appraisal judgments.

Second step. Although the CRVs were obtained by analyzing the co-occurrences of terms in the descriptions, we observed that the four main clusters of sounds (A, B, C, D) clearly overlapped with the five CRVs (I_e, II_e, III_e, IV_e, V_e):

- The verbalizations of Cluster A were associated with the CRV I_e, $\chi 2 = 25$, and II_e, $\chi 2 = 21$.

[3] Developed by Image (http://www.image-zafar.com/english/index_alceste.htm).

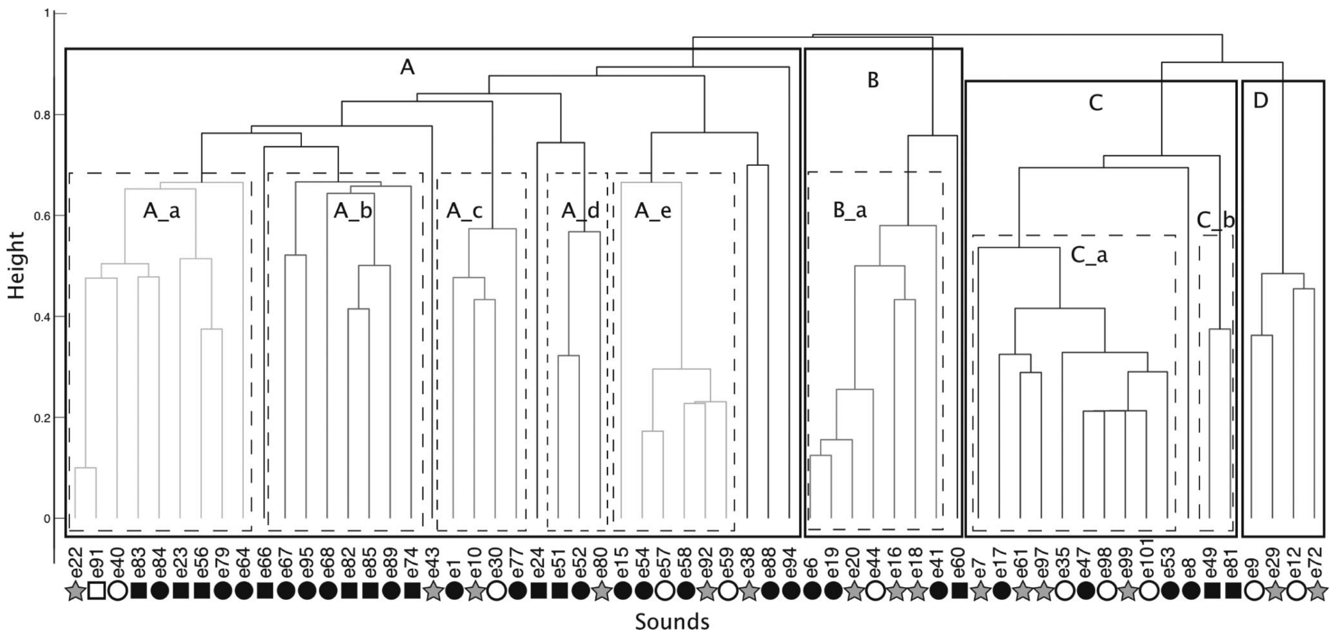


Figure 3. Experiment 1. Primary dendrogram of the hierarchical cluster analysis (Hubert, Arabie, & Meulman, 2006). The main clusters (A, B, C, and D) and their corresponding subclusters ($A_{a,}$ $A_{b,}$ $A_{c,}$ $A_{d,}$ A_{e}, B_{a}, $C_{a,}$ C_{b}, and D) are indicated by Latin letters ($A_{a,}$ $A_{b,}$ $A_{c,}$ $A_{d,}$ A_{e}, B_{a}, $C_{a,}$ C_{b}, and D). The degree of identifiability (causal uncertainty values; Lemaitre, Houix, Misdariis, & Susini, 2010) of each sound is represented by the following code: for values between {0:1} a white square, {1:2} a white circle, {2:3} a gray star, {3:4} a black circle, and {4:5} a black square. A low value of 0 indicates a perfect identification of the sound, and a value of 5 indicates a poor identification. The labels of the sound files (e1 to e60) are presented in Appendix A.

- The verbalizations of Cluster B were associated with the CRV IV_e, $\chi 2 = 69$.
- The verbalizations of Cluster C were associated with the CRV V_e, $\chi 2 = 45$.
- The verbalizations of Cluster D were associated with the CRV III_e, $\chi 2 = 37$.
- Thus, the verbalizations of each of the main clusters (B, C, D) were associated with one CRV that corresponded to one homogeneous lexical field. The verbalizations of Cluster A were associated with two CRVs and with less homogeneous lexical fields. From this analysis, we were able to propose an interpretation of the four main clusters of sounds:
- Cluster A (CRVs I_e and II_e): The descriptions of the sounds in this cluster reflected solid sound sources (e.g., containers, packaging, knife), their properties (e.g., surfaces with different properties), the actions that produced the sounds (opening/closing, compressing, cutting, rubbing), the context (to cook) and a few acoustic properties. Overall, CRVs I_e and II_e distinguished the type of objects (flexible vs. rigid) that constrained different types of actions (manipulation vs. repeated).
- Cluster B (CRV IV_e): The descriptions were related to machines (appliances) and to appraisal judgments of the sounds, but were not specifically associated with a specific event.
- Cluster C (CRV V_e): The descriptions were related to phenomena involving liquids (e.g., faucet, sink) and corresponding actions (flowing, pouring).
- Cluster D (CRV III_e): The descriptions corresponded to gases and the corresponding actions (leaking, fire).

It is important to note that we also observed some discrepancies between the hierarchical structure of the CRVs (their relationships) and the structure of the clusters (compare Figures 3 and 4). For instance, the CRV III_e (gases, associated with Cluster D) was closer to the CRV IV_e (machines, associated with Cluster B) than it was to CRV V_e (liquids, associated with Cluster C), whereas Cluster D was closer to Cluster C. We interpreted these differences based on the fact that the gas and liquid sounds shared more acoustical properties than they shared with machine sounds. However, the verbalizations did not reflect these acoustic similarities.

In sum, we found at this general level that the lexical fields describing the main clusters reflected the main categories in Gaver's (1993b) taxonomy. The same lexical analysis applied to the subclusters of the primary dendrogram did not produce stable results. Therefore, they were submitted to a second lexical analysis.

Second lexical analysis: Portraits of the subclusters. The goal of this analysis was to produce a semantic portrait of each subcluster of the primary dendrogram (Figure 3, Subclusters $A_{a,}$ $A_{b,}$ $A_{c,}$ $A_{d,}$ A_{e}, B_{a}, $C_{a,}$ C_{b}, and D), that is, the most common descriptions across participants for each subcluster. As in the previous analysis, we calculated the occurrence of each term,[4] but in this case, the descriptions were tied to each cluster. We used the same set of descriptions (and the corresponding reduced lexical

[4] We used Lexico V2, developed by CLA2T UPRES SYLED Université de la Sorbonne Nouvelle, Paris 3 (http://www.tal.univ-paris3.fr/lexico/lexico2.htm).

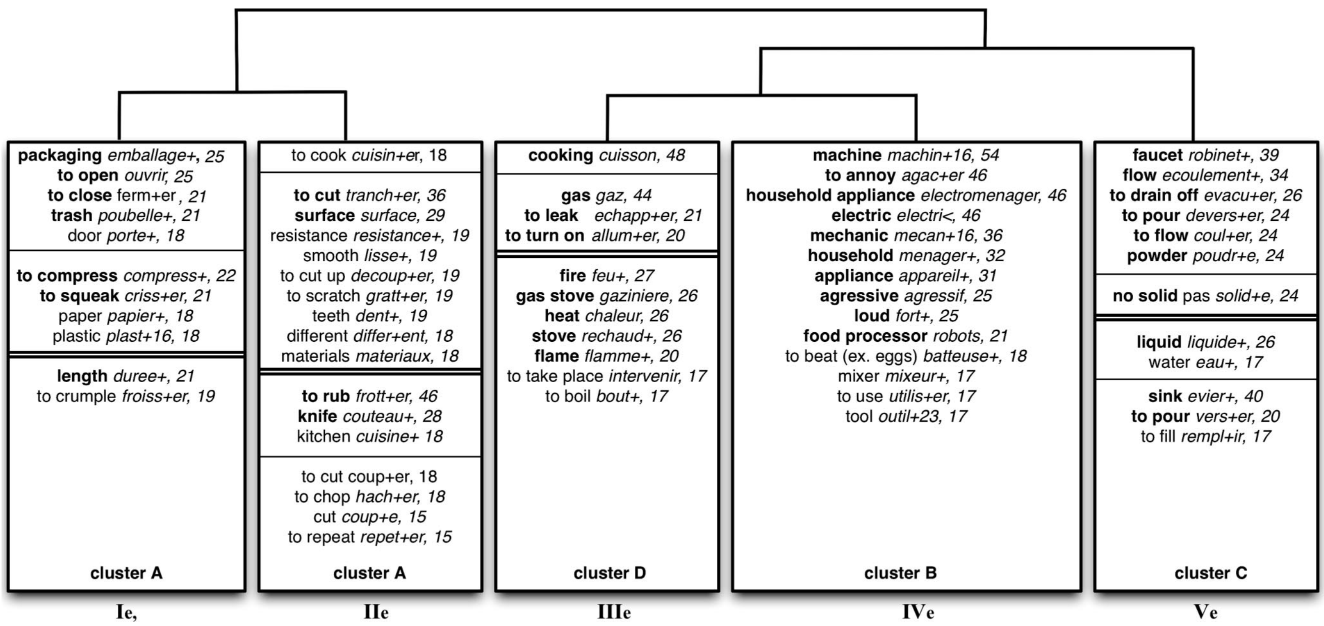


Figure 4. Synthetic view of the first lexical analysis in Experiment 1. The Roman numbers I_e, II_e, III_e, IV_e, and V_e indicate the categories of representative verbalizations (CRVs) identified by the analysis. The Latin letters A, B, C, and D correspond to the main clusters of the primary dendrogram (Figure 3). Within each CRV, the lexical forms are detailed and organized in subcategories (first level: double lines, second level: single lines). Each lexical form is associated with its original French term in italic type (reduced form) followed by the value of chi-square, indicating the importance of the words in the CRV. $\chi^2 > 20$ are in bold.

forms) as previously reported, but we excluded the descriptions of sounds that appeared to be outside stable clusters (e8, e24, e38, e43, e60, e66, e88, e94; 13.3% of the sounds) and were generally less identifiable than other sounds (based on the Hcu values, Figure 3). Their descriptions and classification appeared less reliable. The lexical forms were sorted into five semantic fields (Susini, Houix, Misdariis, Smith, & Langlois, 2009):

- Object and action involved during the production of the sound.
- *Context* of production of the sound (location, surface).
- Description of the acoustical properties of the sounds.
- *Other* types of verbalizations, including appraisal judgments or general descriptions.
- Table 1 presents the distribution of the different lexical forms in the five semantic fields. We observed many occurrences related to the object (50.87%) and fewer related to the actions (26.07%) involved in sound production, which is consistent with the analysis of the main clusters.
- Appendix Table E1 shows the verbal descriptions of each subcluster (A_a, A_b, A_c, A_d, A_e, B_a, C_a, C_b, D). The analysis of the subclusters showed that the categorization of the sounds was mainly based on the identification of the sound sources (solid, liquid, gas, and electric devices), similar to the analysis of the main clusters. This analysis also showed a second level of categorization in terms of actions or more abstract functions. In general, the descriptions appeared to be coherent with the labels of the sound files (Appendix A).

To summarize, the main Cluster A included sounds made by solid objects and their associated actions (knife and cutting in A_a, glass and shock in A_c, door and opening/closing in A_e). It also grouped more abstract functions such as opening a can or a packaging that involved an analogous goal, if not necessarily the same actions, or produced a similar sound due to the manipulation of metallic or flexible objects (A_b). One subcluster (A_d) seemed to mix a script representation (i.e., a representation of a situation organized around personal experiences in which human interaction is central; Schank & Abelson, 1977), preparation of meal, with the identification of the action of crumpling. The other main clusters, B, C, and D, grouped similar sound sources: (B) machines or electric devices; (C) liquids with the verbs *flowing*, *pouring*; and (D) gases associated with the verbs *cooking*, *leaking*. These clusters represented a lesser variety of sound events.

The analysis of the secondary dendrogram (Appendix C) indicated the use of different criteria to group sounds (see Appendix Table E2 for a synthetic view). These criteria were mixed and related to the same material (glass or metal Cluster A) or sound events (a large category of shocks, Cluster G), abstract functions (e.g., sounds from the heating of a meal, Cluster H) and script representation (e.g., the preparation of meal in a glass or container, Cluster D) and, to a lesser degree, the same acoustical properties.

Discussion

The lexical analysis of the descriptions of the different levels of clustering in the primary dendrogram allowed us to interpret these clusters at different levels of generality and to compare the struc-

Table 1
Experiment 1: Occurrences of the Lexical Forms in the Five Semantic Fields for the Verbalizations of Each Subcluster of the Primary Dendrogram (Figure 3)

	Semantic fields					
Cluster	Object	Action	Context	Acoustic	Other	Total
A_a	88	139	55	16	47	345
A_b	46	25	20	17	6	114
A_c	35	13	10	0	10	68
A_d	6	12	0	0	0	18
A_e	93	43	10	18	6	170
B_a	100	7	6	26	7	146
C_a	214	63	0	11	19	307
C_b	0	6	0	0	0	6
D	60	21	0	0	7	88
Occ	642	329	101	88	102	1,262
%	50.87	26.07	8.00	6.97	8.08	100.00

Note. Occ = occurrences of the lexical forms.

ture of these clusters to the other structures proposed in the literature. Overall, these analyses showed that the participants used different criteria to sort the sounds. Nevertheless, we observed a large distinction between four main clusters of sounds: the sounds produced by solids, liquids, gases, and machines. This result is similar to the distinction between liquids, solids, and gases proposed by Gaver (1993b).

At a finer level, the subclusters tended to correspond to the different types of actions that produced sounds. These actions were closely tied to the different types of sound sources, and they differed from those proposed in Gaver's (1993b) taxonomy. For example, we observed only two subclusters of liquids (pouring and flowing), whereas Gaver (1993b) suggested four (dripping, pouring, splashing and rippling). This result is probably due to the limited number of liquid sounds that we used. Another interpretation is that the large variety of sounds made by radically different sound sources prevented the listeners from focusing on more precise categories. In the case of liquid sounds, we observed that the sounds drip in container and filling a container were not distinguished from the other liquid sounds, even though Gaver (1993b) would consider them hybrid sound events (liquid and solid interactions) and therefore different from all other sounds. These sounds were simply identified as liquids. Other hybrid sound events (gas and liquid interactions) such as bubbles were also clustered as liquid sounds. The subclusters of gasses were limited to what could happen in a kitchen (fire, cooking, gas leak). For instance, we did not observe the categories whoosh, explosion, and wind (Gaver, 1993b). These categories are more likely to be heard in an outdoor context. We observed many subclusters for the sounds of solid objects. However, the actions corresponding to these subclusters were very specific to the particular object on which they were executed. We did not observe clusters of actions that could generally apply to any solid object.

These results indicated a difference from the hierarchical organization proposed by Guyot et al. (1997) and Dubois (2000): listeners identified action at a basic level and at a subordinate level, the source. However, this organization did not include gas and liquid sounds. The most general level of classification in our case seemed to be related to the sound sources (sounds produced by different state of matters, objects, appliances, etc.), and the more specific levels of classification were related to the actions causing the sounds. The sounds seemed to be categorized as sound sources first and only second as actions.

Finally, we observed several other criteria for classifying the sounds: acoustical properties, appraisal judgments, and more abstract representations (function, script representation, etc.). However, these criteria were less common.

In a second experiment, we decided to investigate in greater detail the categorization of the sounds of solid objects. Many subcategories were available for the subjects to categorize these sounds.

Experiment 2: Classification of Solid Sounds

The results of the first experiment showed a general level of classification based on broad categories of sound sources (solids, liquids, gases and machines). However, the first study presented only a fragmented view of the subcategories of these broad categories, either because we had only a limited number of sounds or because the origins of the sounds were too different. In the second experiment, we decided to focus only on sounds made by solid objects to reveal these subcategories. Because the first experiment showed a specific level of categorization related to the actions or movements that produced the sounds, we decided to focus on this level.

Method

Participants. Thirty participants (15 women, 15 men) volunteered as listeners and were paid for their participation. They were between 18 and 63 years of age (median: 35 years). All reported normal hearing. They were selected using the same criteria as in the first experimental study.

Stimuli. The sounds were monophonic recordings of events exclusively produced by the interactions of solid objects or by the action of an agent on an object, with or without a tool. They were chosen from different commercial sound effect libraries: Hollywood Edge Premiere Edition I, II and III (The Hollywwod

Edge, Hollywood, USA), Sound Ideas General Series 6000 (Sound Ideas, Ontario, Canada) and Blue Box Audio Wav (Best Service GmbH, München, Germany). To perform the selection, we thoroughly listened to these different sound libraries guided by their textual descriptions. As in the first experimental study, we attempted to sample the different sound examples of Gaver's (1993b) taxonomy and added sounds that were not mentioned. We rejected hybrid sounds such as rain on a roof (solid and liquid sound). We used the same criteria to exclude the sounds of living beings (vocalization, eating, and transport sounds). Nevertheless, we included two sounds of persons walking on different floors (gravel and rubber), without a specific gait. The selected sounds were produced in the same context (indoors) to minimize the formation of large categories of sounds, such as leisure and vehicle sounds.

We asked seven members of our laboratory to listen to a restricted number of sounds, taken from a large corpus of sounds, and to note whether they could identify at least the physical actions that produced the sounds. This procedure allowed us to select 56 well-identified sounds to prevent participants from spontaneously focusing on acoustical properties. The sound levels were adjusted ecologically using the same procedure as in the first experiment but given the indoor context. The sound levels and sound labels are presented in Appendix F.

Apparatus. We used the same hardware equipment and software as in the Experiment 1.

Procedure. The experimental procedure was similar to that of the first experiment, except for the instructions. The written instructions (in French) asked the participants to sort a set of sounds produced by the interaction of solid objects, taken from different contexts. The participants were specifically required to sort sounds according to the physical actions that produced the sounds, regardless of the object involved. The results of the first study showed that, at a secondary level, the sounds were classified by different categories of actions. Therefore, we required the listeners to specifically focus on the actions producing the sounds to minimize the influence of other marginal strategies of classification.

Results

Removing outliers. The participants created between six and 28 classes (median: 15.5). We analyzed the interindividual differences between the individual proximity matrices of the 30 participants. The aim of this analysis was to observe potential groups of participants or isolated participants with different strategies. We used the R_V coefficient (Abdi, 2007) to calculate the agreement between participants. We also introduced randomly generated proximity matrices (corresponding to a random classification of the sounds in a number of classes that varied between six and 28). Figure 5 shows that the proximity matrix of Participant 17 was closer to the randomly generated matrices than to the other participants. Therefore, the results for this participant were excluded from the analyses.

Hierarchical cluster analysis. We used the same hierarchical cluster analysis as in Experiment 1. Figure 6 represents the 13

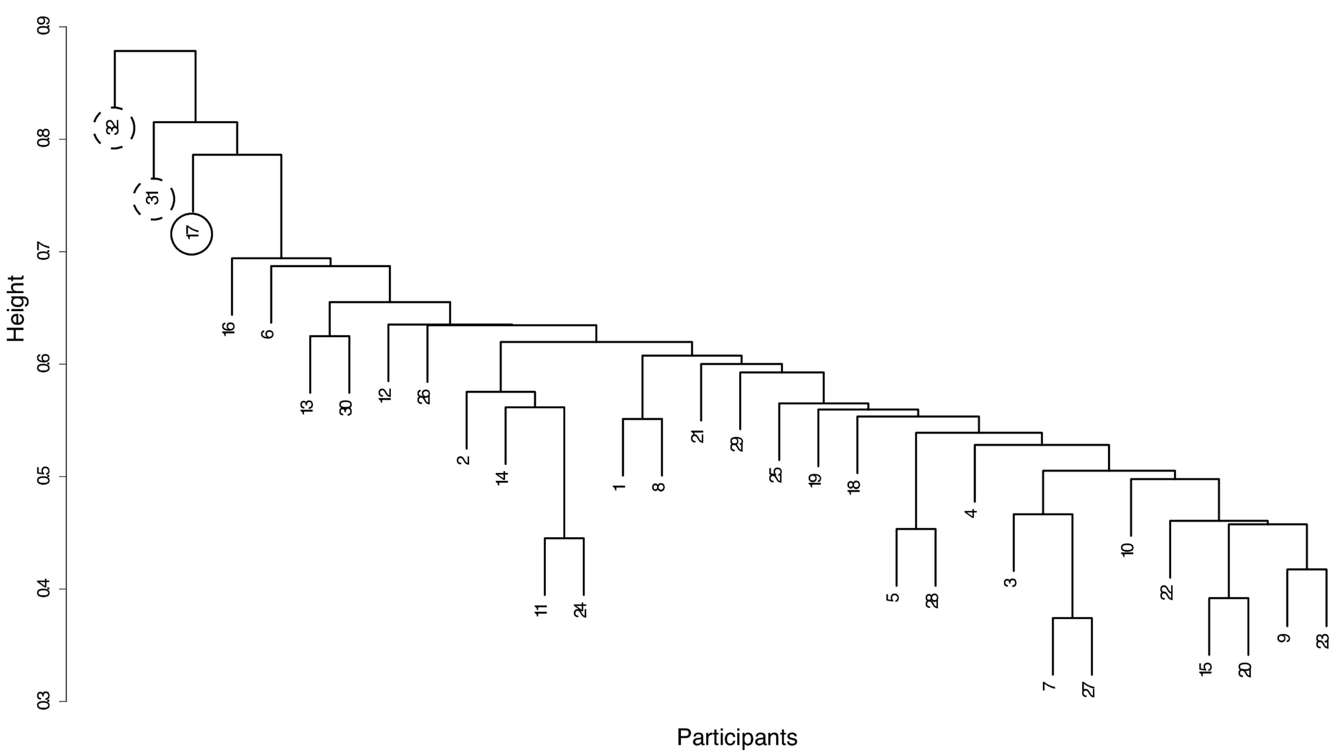


Figure 5. Hierarchical cluster analysis (average method) of the matrix of *RV* coefficients (Abdi, 2007). *RV* coefficients are computed for each pair of individual proximity matrices of the participants and calculated distances between the results of participant classifications. The Arabic numbers correspond to the 30 participants, except for numbers 31 and 32, which correspond to matrices generated randomly.

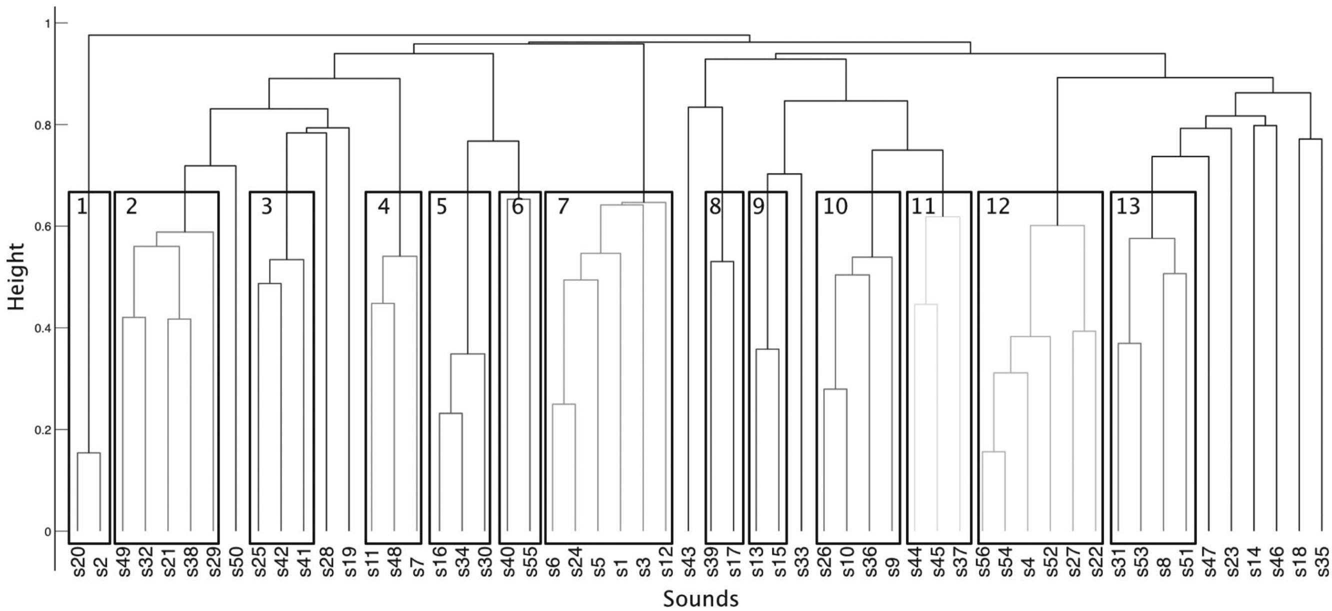


Figure 6. Experiment 2. Primary dendrogram of the hierarchical cluster analysis (Hubert, Arabie, & Meulman, 2006). The main clusters (1 to 13) are indicated by Arabic numbers. The labels of the sound files (s1 to s56) are presented in Appendix F.

main clusters (identified with Arabic numerals between 1 and 13) of the primary dendrogram (which was not separated into main clusters and subclusters) and Appendix G, the 13 main clusters (1" to 13") of the secondary dendrogram. Eleven sounds were not included in the 13 main clusters (s50, s28, s19, s43, s33, s47, s23, s14, s46, s18, s35), corresponding to 19.7% of the sounds. These sounds are not included in the lexical analyses. Again, we report only the detailed analyses of the primary dendrogram. The secondary dendrogram is reported in Appendix G.

Lexical analyses. We performed the same lexical analyses as in Experiment 1.

There were 385 distinctive forms and 4,029 associated occurrences that were accounted as verbalizations produced by the participants, and 228 analyzed forms and 3,581 occurrences that were selected for the analysis. The corpus was divided into 117 ECUs, and 90% of the ECUs were taken into account during the analysis.

The most frequently cited nouns were *object* (73 occurrences), *noise* (72 occurrences), and *impact* (63 occurrences). The most frequently cited verbs were *rubbing* (136 occurrences), *crumpling* (91 occurrences), *rolling* (78 occurrences), *tumbling* (77 occurrences), *tearing* (74 occurrences), *hitting* (58 occurrences), and *falling* (51 occurrences). This simple tally immediately showed that the verbs used to describe the sounds clearly described the actions causing the sounds.

First lexical analysis: Categories of representative verbalizations. Figure 7 presents a synthetic representation of the categories of representative verbalizations CRV. It identifies five CRVs (I_s, II_s, III_s, IV_s, V_s), with a respective distribution of the EUC of 20.95%, 14.29%, 11.43%, 29.52%, and 23.81%.

To summarize, we identified five CRVs. Considering the hierarchical structure in Figure 7, CRV I_s and II_s could be merged into a more general lexical field. CRV I_s was organized around the lexical fields of cutting, sawing, rubbing, zipping, and tearing. CRV II_s was related to the actions of crumpling, creasing, crushing, and compressing. Both of these CRVs grouped a vocabulary describing a sustained physical action.

A second general lexical field emerged from CRVs III_s, IV_s, and V_s. CRV III_s described movements of swinging producing creaking sounds as well as rotary movements; as cyclic movements. CRVs IV_s and V_s were described as hitting, knocking and slamming, as containers, pans, dishes falling or tumbling, and perhaps as stopping. Thus, the second general lexical field appeared to reflect actions or movements involving short, discrete interactions (a series of impacts).

In a second step, we analyzed the correspondence between the CRVs and the clusters of sounds. The main clusters (1 to 13) from the primary hierarchical dendrogram (see Figure 6) were related to the CRVs (I_s, II_s, III_s, IV_s, V_s; Figure 7).

- Verbalizations linked to Clusters 12, $\chi^2 = 33$, and 13, $\chi^2 = 28$, were associated with CRV I_s.
- Verbalizations associated with Clusters 10, $\chi^2 = 45$; 11, $\chi^2 = 25$; and 9, $\chi^2 = 18$; were linked to CRV II_s.
- Verbalizations linked to Clusters 5, $\chi^2 = 58$, and 6, $\chi^2 = 41$, were associated with CRV III_s.
- Verbalizations linked to Clusters 2, $\chi^2 = 23$; 4, $\chi^2 = 15$; and 3, $\chi^2 = 11$; were associated with CRV IV_s.
- Verbalizations linked to Clusters 7, $\chi^2 = 47$, and 1, $\chi^2 = 13$, were associated with CRV V_s.

The hierarchical structure of the CRVs (see Figure 7) was relatively close to that of the 13 main clusters in the primary dendrogram (see Figure 6). Some clusters presented a strong association with the CRVs, but we observed a weaker chi-square for CRV IV_s. We return to this result.

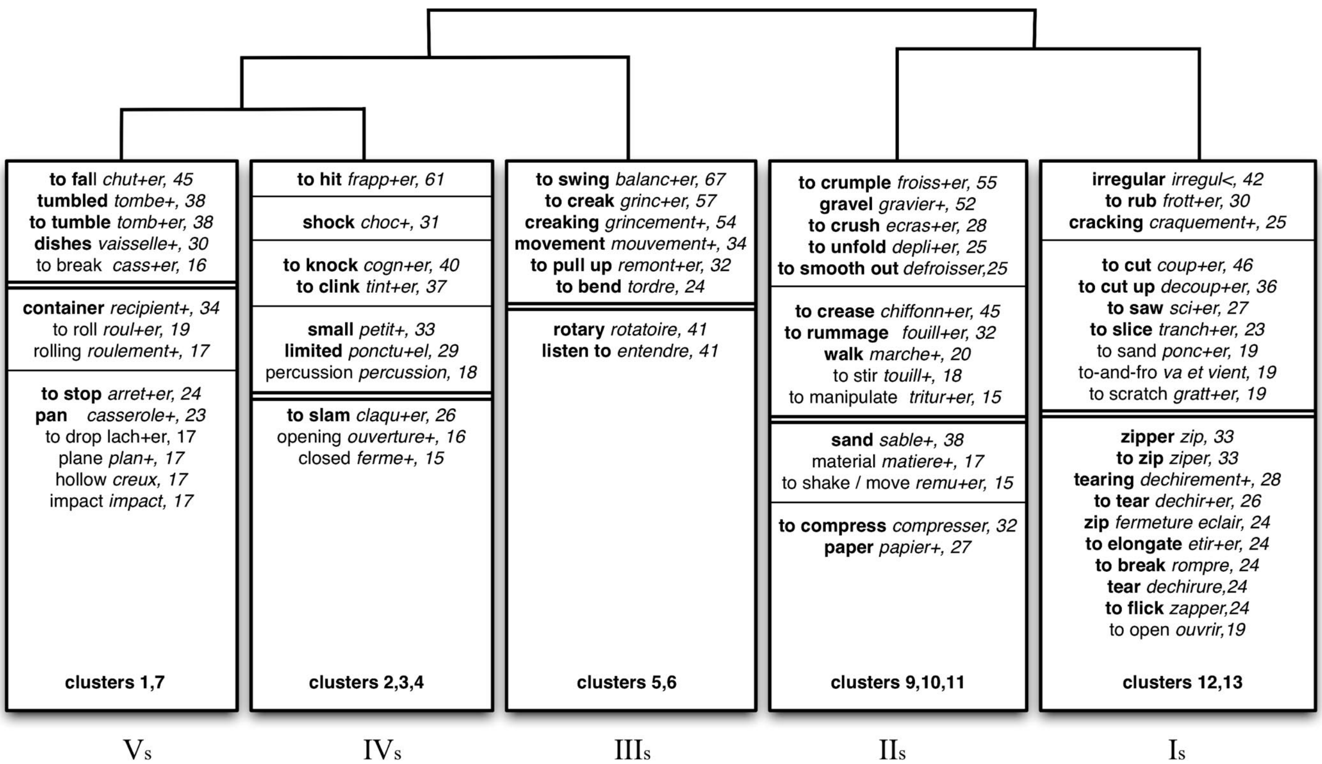


Figure 7. Experiment 2. Synthetic view of the first lexical analysis. The Roman numbers I_s, II_s, III_s, IV_s, and V_s indicate the categories of representative verbalizations (CRVs) identified by the analysis. The Arabic numbers 1 to 13 correspond to the main clusters of the primary dendrogram (see Figure 6). Within CRVs, the lexical forms are detailed and organized in subcategories (first level: double lines, second level: single lines). Each lexical form is associated with its original French term in italic type (reduced form) followed by the value of chi-square, indicating the importance of the words in the CRV. $\chi^2 > 20$ are in bold face.

We also observed some discrepancies between the two structures. Clusters 1 and 7 were different but were associated with the same CRV (V_s) because of the concomitance of the action of falling with the action of rolling in the representation of the event and the associated verbalizations, such as compound events (Gaver, 1993b). The action of falling precedes the action of rolling. We also observed that Cluster 8 was not connected to a CRV.

Second lexical analysis: Portraits of the clusters. Using the same analysis as in Experiment 1, we calculated the occurrence of each term. In this case, however, the descriptions were associated with each respective cluster.

Table 2 presents the raw results of the analysis. This table shows that the distribution of the lexical forms was primarily related to the actions (61.16%) and was related to a lesser degree to the objects (19.51%) producing the sounds. Appendix Table H1 reports the details of the analysis.

To summarize, the 13 clusters were described mainly by the identification of the physical action producing the sounds. The sound sources were sometimes identified, generally at a generic level (participants were explicitly required to focus on the action and not to consider the object).

If we consider the primary dendrogram (see Figure 6), we observe two large clusters regrouping Clusters 2, 3, 4, 5, 6, and 7, on the one hand, and 8, 9, 10, 11, 12, and 13, on the other hand. The organization of the CRVs (see Figure 7) reflects this structure, with the exception of cluster 1, which was regrouped with Cluster 7 but separated from all the other clusters.

The distinction between these two large clusters is related to the distinction between two types of physical actions to produce sounds. Clusters 2, 3, 4, 5, 6, and 7 grouped discrete interactions (short contact time), whereas Clusters 8, 9, 10, 11, 12, and 13 grouped physical actions related to a continuous and sustained contact, generally implying noise. Within each large cluster, the clusters were organized around similar physical actions that were constrained by the physical properties of the objects (one part or several parts, shape, material).

For example, Clusters 2, 3, and 4 grouped sounds produced by simple impacts referring to different types of objects: small parts of a mechanism, glass, and metal objects, and large mechanisms. Clusters 5 and 6 distinguished rotary mechanisms such as gears (Cluster 6) and rotary mechanisms with objects returning to their initial position (Cluster 5). The presence of gears (impacts with rotary patterning) could explain why this cluster was tied to Clusters 2, 3, and 4. Cluster 7 seemed to be associated with physical actions including multiple impacts.

Clusters 9, 10, and 11 were related to the deformation of objects made of different materials: small rigid objects (Cluster 9), soft objects (Cluster 10), and rigid objects (Cluster 11). Similarly, Clusters 12 and 13 reflected the separation of soft materials (Cluster 12), deformable material (Cluster 13), and friction (Cluster 13).

Table 2
Experiment 2: Occurrences of the Lexical Forms in the Five Semantic Fields for the Verbalizations of Each Main Cluster of the Primary Dendrogram (Figure 6)

	Semantic fields					
Cluster	Object	Action	Context	Acoustic	Other	Total
1	45	0	0	0	8	53
2	22	27	0	18	13	80
3	13	24	11	6	7	61
4	18	41	0	0	6	65
5	25	8	0	30	0	63
6	6	16	0	0	6	28
7	48	181	12	9	13	263
8	7	25	0	0	0	32
9	0	15	7	6	0	28
10	16	61	0	8	0	85
11	0	27	0	0	6	33
12	9	131	0	10	0	150
13	0	99	20	11	0	130
Occ	209	655	50	98	59	1,071
%	19.51	61.16	4.67	9.15	5.51	100.00

Note. Occ = occurrences of the lexical forms.

Cluster 8 represented the action of shaking small rigid objects. This cluster was close to Clusters 9, 10, and 11.

Cluster 1 was distant from the other two large clusters (2, 3, 4, 5, 6, and 7, and 8, 9, 10, 11, 12, and 13), but the general analysis of verbalizations showed that this cluster shared a lexical field with Cluster 7. The distance in the primary dendrogram could be explained by unpleasant breaking events sounding completely different from other sounds. Warren and Verbrugge (1984) pointed out an auditory distinction between breaking and bouncing sounds.

The analysis of the secondary dendrogram (Appendix G) summarized in Appendix Table H2 and its related verbalizations indicated the use of marginal types of similarity, sometimes mixed, to group the sounds of acoustical properties (e.g., crumpling and walking sounds, Cluster 1), specific categories of sound events (e.g., shaking small marbles, Cluster 10) or abstract functions (e.g., walking sounds, Cluster 1).

Discussion

The goal of Experiment 2 was to observe whether the categorization of the sounds made by solid objects would result in categories corresponding to the different actions generating the sounds, as assumed by Gaver (1993b). Figure 8 summarizes the two lexical analyses of the verbalizations. The first lexical analysis produced the hierarchical structure of the lexical fields related to clusters. The second analysis identified each cluster of sounds and indicated how well the participants identified the sounds.

The first general lexical analysis resulted in two large lexical fields: one related to continuous interactions and the other related to discrete interactions. We identified more specific fields within these two large lexical fields. The continuous interactions included cutting, sawing, and rubbing; zipping and tearing; and crumpling, creasing, crushing, and compressing. The discrete interactions included hitting, knocking, and slamming; falling or tumbling; and rotary or swinging movements. The structure of the different CRVs (see Figure 7) was consistent with the organization of the clusters in the primary dendrogram (see Figure 6), although Cluster 8 was not linked to any CRV (see Figure 7), possibly due to terms that were not specific to the verbalizations associated with this cluster.

The different clusters emerging from the primary dendrogram indicated that the listeners identified the different physical actions that produced the sounds. The second (and more specific) analysis of each cluster also resulted in the distinction between discrete interactions (simple and multiple impacts, rotary movements) and continuous interactions (shaking, crushing small objects, creasing soft object, crushing rigid object, tearing, cutting, and friction).

As a comparison, four classes of solid interactions were described in Gaver's (1993b) taxonomy: (a) deformation (crumpling, crushing, etc.), (b) impact (breaking, bouncing, walking, hammering, etc.), (c) scraping (sanding, filing, rubbing, etc.) and (d) rolling (gears, pulleys, etc.). Following Gaver's terminology, compound events mixed these different basic interactions.

Therefore, the clusters highlighted by our analyses were similar to these four classes. We identified a cluster of impact sounds (simple or multiple), a cluster related to the action of rotation (rotary mechanisms with or without gears), another grouping sounds categorized as actions of deformation (small sets of objects, soft materials, deformable objects) and, finally, a cluster of actions producing sounds of physical separation (with tools, mixed with friction, or by force). The cluster rotation could be considered similar to the rolling class, and the scraping class was close to the cluster of separating with friction. The cluster shaking could be understood as compound sound events mixing multiple impact and deformation actions.

If we consider the structure of the cluster of sound events, our results highlight a similarity with the three levels proposed by Guyot et al. (1997) and Dubois (2000). In the category of sounds generated by solid interactions, the most general level is related to the movements or actions generating sounds, implying a different

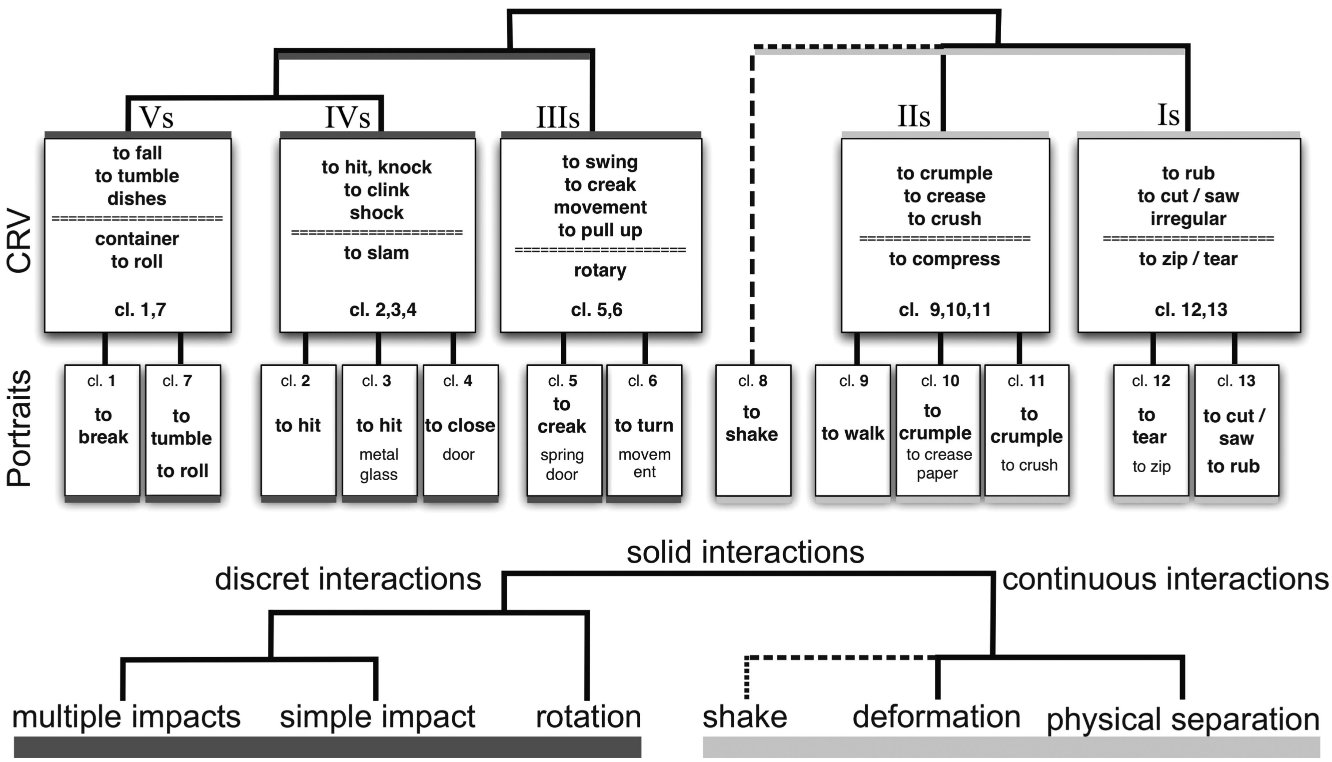


Figure 8. Experiment 2. Summary of the first and second lexical analyses. The organization between the different categories of representative verbalizations (CRVs; first analysis) and the different main categories and their semantic portraits (second analysis) is depicted with lines. The Roman numbers I_s, II_s, III_s, IV_s, and V_s indicate the categories of representative verbalization CRVs identified by the analysis. The Arabic numbers 1 to 13 correspond to the main clusters from the primary dendrogram (see Figure 6). Within each CRV, the lexical forms are summarized and organized in subcategories (first level: double lines, second level: single lines). Category 8 is marked with a dotted line because it was not associated with a CRV during the first analysis.

type of temporal patterning. These objects are only found at a more specific level.

Finally, our results stress that temporal patterning is a very important cue to identify and categorize sound events and reflects the physical action generating the sound.

General Discussion

Our aim in this research was to study how listeners categorize environmental sounds and whether their categorization corresponds with the taxonomy of sound production hypothesized by Gaver (1993b). First, we studied the categorization of a set of everyday sounds. Next, we only focused on sounds made by solid objects.

A survey of previously reported studies of sound categorization highlighted two issues: the coexistence of several classification strategies in the mathematical representations of the data and the difficulty of analyzing the participants' verbalizations. To address these issues, we used a technique to mathematically represent the data (two dendrograms). The verbalizations were submitted to a set of systematic linguistic analyses.

The study of the categorization of a set of everyday sounds (Experiment 1) highlighted an organization based on the causes of the sounds. The categorization occurred first at a general level of sound sources, such as solids, liquids, gases, and machines. These categories were similar to the categories of sound events proposed by Gaver (1993b). At a second level of categorization, the subcategories were related to the actions generating the sounds. These actions were specific to each sound source. Overall, we found fewer categories than Gaver (1993b) did. This was probably because the other categories do not usually occur in a kitchen. The sounds of liquids seemed to have a specific perceptual status: they were systematically categorized as liquid sounds even if they could be labeled as hybrid sound events (liquid and solid interactions). We hypothesize that these sounds had their own acoustical properties that distinguished them from the other sounds (Gygi et al., 2007).

Following the results of the first experiment, we focused on the sounds produced exclusively by solid interactions. We asked the participants to classify sounds by physical actions. We used the same protocol to analyze the data (hierarchical cluster and statistical analysis of the verbalizations). The first lexical analysis showed five different categories of representative verbalizations that were organized into two large semantic categories. The structure of the semantic categories overlapped the organization of the main clusters of the primary dendrogram (Experiment 2), indicating a correspondence between the verbalizations and the categorization of the sounds. The second lexical analysis of the verbalizations provided a semantic portrait of each cluster of the primary dendrogram. The labels of the sound files were used to understand how the sounds were identified and categorized.

The different clusters reflected categorization at the level of the action or movement generating sound. The structure in the primary dendrogram showed two large clusters that distinguished clusters related to different types of physical actions, one type related to discrete interactions and the other type associated with continuous interactions. Within each large cluster, subclusters were organized around different physical actions across different objects. The category of discrete interactions was divided into impact sounds (simple or multiple) and rotation sounds (rotary mechanisms with or without gears). The second category, related to continuous interactions, was organized around the action of deformation (small sets of objects, soft materials, deformable objects) and the action of physical separation (with instruments, mixed with friction, or by force). A significant contribution of these analyses is the proposal of a perceptual organization of the different clusters proposed in Figure 8. These categories resulted from listeners' categorizations, in contrast to Gaver's (1993b) taxonomy, which was created by analyzing the physics of sound production. Similar to Gaver's (1993b) taxonomy, we found a compound category mixing multiple impacts and deformation. These compound actions generated sounds such as shaking. However, our results highlighted a clear distinction between discrete and continuous interactions, which reflects a difference in temporal patterning. The participants may mentally represent how the sounds are physically produced by manipulating objects with different physical properties, directly or indirectly.

This last result is similar to the results of Gygi et al. (2007), who used acoustical similarity judgments of 50 environmental sounds and a multidimensional scaling analysis (MDS). They found categories of continuous sounds, discrete impact sounds, and harmonic sounds. They also identified categories for nonvocal sounds and vocalizations. A hierarchical cluster analysis of the multidimensional scaling solution further showed a distinction between single and multiple impacts, on the one hand, and a specific class of scraping sounds separate from the other continuous nonharmonic sounds, on the other hand. Combining our results with those found by Gygi et al. (2007) suggested that listeners were able to represent the physical cause of the sounds by focusing on the temporal aspects of the sounds (patterns, etc.).

Notably, our results suggest that action and object are not independent principles for organizing the representation of sound. For instance, we did not find a set of general actions that could be executed on any kind of solid object. On the contrary, we found different actions constrained by the objects on which they were executed. This is a major difference from Gaver's (1993b) view.

Finally, the results of the two experiments reported here allow us to propose a general structure of environmental sound categories in three levels. This structure extends the taxonomy proposed by Guyot et al. (1997). A first general level distinguishes large categories of sound sources: solids, gases, liquids, and machines. At a second level, the structure is based on the physical actions that produce sounds and is indexed by different temporal patterns (discrete and continuous interactions). At the third level, each temporal patterning action category is further divided into specific actions. The present results support this structure for the sounds of solids, but it remains to be tested for liquids and gases. However, fewer actions seemed to be available to produce sounds with gases and liquids.

Our stimuli did not include all environmental sounds, such as living sounds and synthetic sounds. The stimuli were selected within specific contexts of sound production: the kitchen for the first experiment and the indoors for the second experiment. These specific contexts may limit the formation of large semantic categories (e.g., the transport category) or more conceptual categories (, e.g., the leisure category) involving semantic relations (*part of*, *hyponyme*, *hypernyme*, etc.), such as those implemented in the Wordnet network (Fellbaum, 1998). This structure is more concerned with sound event categories, specifically, the physical production of the sounds.

From a methodological point of view, we developed an original approach to analyze the partitions produced by participants and their associated verbalizations. Our statistical analyses were based on a particular hierarchical cluster analysis producing two dendrograms (Hubert et al., 2006), allowing us to focus on more homogeneous categories (i.e., minimizing different types of similarity). In our study, the lexical analyses of the secondary dendrograms indicated marginal criteria for grouping sounds, such as acoustical similarity, abstract function, and script representation. Considering a wide corpus of environmental sounds with a large variety of categories (e.g., living and nonliving sounds, action and nonaction sounds, and synthetic sounds that can imply a complex set of semantic relations), an analysis with multiple dendrograms could be useful to highlight complex relations between categories. The goal of these experiments was not to test the internal structure of the category (e.g., typicality between sounds within a category) or to test different theoretical approaches (e.g., comparing the prototypical view and the exemplar view). Therefore, we used a hierarchical cluster analysis rather than an additive tree to produce clusters with clear boundaries. Different categories with clear boundaries and structures allowed us to perform a textual data analysis rather than an analysis of the verbalizations in combination with additive tree representations, which requires more interpretation by the person analyzing the data. Our analysis of verbalizations was twofold, permitting us to relate semantic categories and the structure of the clusters (the dendrogram) and producing a semantic portrait of each cluster. This protocol of analysis proved to be a useful contribution to the analysis of classification tasks.

The verbalizations were produced in French following the sorting task. The convergence between the structure of the clusters (hierarchical cluster analysis of the sorting tasks) and the structure of the lexical categories (lexical analysis of the verbalizations) can be understood by two assumptions. The first assumption is a weak influence of the language (French) on the categorization of these sounds. We explain this weak influence based on the categories (sound events) grounded in the physical production of the sounds, which therefore constrains their acoustical properties. Listeners use more acoustic information for nonliving sounds (Giordano et al., 2010). In a contrasting assumption, we propose that French influences the categorization because listening to sounds involves similar neuronal networks as language if people name the sound or its properties when listening (Cummings et al., 2006; Lebrun et al., 2001). Our results cannot disentangle these different assumptions. An interesting study would consist in testing subjects with different native languages.

These results have practical applications. For instance, sound-database organization and the perceptual approach to synthesizing physical sounds have both benefited from the recent development of environmental sound studies (see Gygi & Shafiro, 2007, for an overview).

A recent and major improvement in sound synthesis techniques involved the introduction of new algorithms based on physical modeling. These algorithms are powerful and are based on the simulation of the physical phenomena. However, they are difficult to control and parameterize, particularly because their control parameters are unintuitive. Obtaining the precise sound that the sound designer or musician has in mind requires a mastery of the underlying algorithms. Therefore, the major challenge of this technique is to develop intuitive control interfaces. Drioli et al. (2009) developed an experimental tool to control sound synthesis based on a simplified auditory representation of reference sounds in a two-dimensional space. The sound designer can navigate in this 2-D space in which reference sounds are located and may choose a location to generate a new sound with new synthesis parameters. These parameters are interpolated from synthesis parameters associated with these sonic landmarks. One approach could consist of mapping a set of synthesis parameters to specific meaningful categories of sounds (e.g., the material or the type of interaction). The knowledge of the categories of sound-producing events could allow developers to achieve this goal.

Our results (and, specifically, our representation of the physical actions producing sounds) are an interesting contribution to the automatic classification of sounds, especially because of their relation with temporal patterning. For example, Gygi and Shafiro (2010) proposed an automatic classification of environmental sounds related to the work of Gaver (1993b). This automatic classification was based on their work on spectro-temporal acoustical factors in the identification of environmental sounds (Gygi, Kidd, & Watson, 2004). Our structure of categories may be useful for supervised classification based on acoustical descriptors, like the work by Misdariis et al. (2010) proposing a model for classifying sounds according to three pre-established classes (impact sounds, instrument-like, and motors) based on psychoacoustic descriptors and the work by Roma et al. (2010) based on Gaver's (1993b) taxonomy. In supervised classification, the categories and their associated sounds are used by the system to learn to classify sounds. For this purpose, new acoustical descriptors related to the spectro-temporal profiles of sounds should be developed, as in the work by Peeters and Deruty (2010). Our categories of physical actions generating sounds may serve as reference classes to develop these new acoustical descriptors.

Searching based exclusively on acoustical properties of the sounds is now effective (Misdariis, Smith, Pressnitzer, Susini, & McAdams, 1998). Search requests based on textual descriptions by commercial sound libraries or by textual tags facilitate searching within the databases and relating different semantic categories (Cano, Koppenberger, Herrera, & Celma, 2004). Internet Web sites, such as Findsounds (http://www.findsounds.com) or Soundfisher (http://www.soundfisher.com/), provide such complex search engines to find sounds. Generally, however, these sites use a simple hierarchical structure to navigate sounds. For example, Soundfisher, using the Musclefish technology (Wold, Blum, Keislar, & Wheaten, 2002), has developed tools to search sounds based on similarity (fine acoustic descriptors). It also uses basic navigation processes based on the semantic descriptions of sounds in commercial sound databases such as Sound Ideas (http://www.sound-ideas.com/). The sounds belong to classes that are hierarchically organized based on these semantic descriptions. The navigation operates through these classes.

As shown by the review of classification studies, a hierarchical structure is probably too simple to account for the human categorization of sounds. Therefore, some technologies use more complex structures. Audioclas, the technology behind Freesound (http://www.freesound.org/), is an interesting project of online sound databases using the semantic network Wordnet (Fellbaum, 1998). This project uses the different relations of this network to allow fuzzy queries using synonyms and to relate different categories of sounds to their semantic descriptions (Cano, Koppenberger, Herrera, & Celma, 2004). For example, *piano* has two sets of meanings: the musical attributes or the musical instrument made of different parts (a keyboard, a pedal, etc.). Freesound uses all these different meanings and semantic relations to link sounds. Audioclas technology also developed an experimental function that was not implemented in Freesound: Wordnet is used to label new sounds (Cano, Koppenberger, Le Groux, et al., 2004). When a new sound is imported into the database, the sound is compared with other sounds based on acoustical similarity. When a sound is acoustically similar to the sounds of a category, the semantic descriptions of the category are associated with this new sound, which then belongs to the taxonomy. However, if sounds share similar acoustical properties but can belong to different semantic categories (a situation exploited by Foley artists for sound design, such as using a can opener to produce a gun sound; Ament, 2009), frequent errors of classification can occur. The relationship between acoustical descriptors of the sounds and their textual descriptions is difficult to establish because of the abstraction of the properties shared by items of a semantic category.

These two types of organization (one based on a simple hierarchical structure and the other on a complex semantic network) are either too simple to fit our perception of environmental sounds or too complex and broad to be effective (e.g., in the field of postproduction, a sound designer should be able to efficiently search sounds in sound databases). Another proposed approach is a structure for organizing sound databases that reflects how human listeners categorize sound events. For example, the ability to find sounds made by the same action without considering a particular source may be very effective for the postproduction of sound effects for motion pictures. Our results could be a useful contribution toward an intermediate level of searching grounded in the physical production of the sounds that relates acoustical descriptors specific to these categories and the lexical descriptions. These three levels of relation (acoustical, physical production of the sounds, semantic descriptions) can be a powerful tool for searching and organizing sounds.

References

Abdi, H. (2007). The RV coefficient and the congruence coefficient. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 849–853). Thousand Oaks, CA: Sage.

Aglioti, S. M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research, 14*(3), 99–110. doi: 10.1007/s00221-010-2344-x

Aldrich, K. M., Hellier, E. J., & Edworthy, J. (2009). What determines auditory similarity? The effect of stimulus group and methodology. *Quarterly Journal of Experimental Psychology, 62,* 63–83. doi: 10.1080/17470210701814451

Ament, V. T. (2009). *The Foley grail: The art of performing sound for film, games, and animation.* Oxford, England: Focal Press.

Ballas, J. A., & Howard, J. H. (1987). Interpreting the language of environmental sounds. *Environment & Behavior, 19*(1), 91–114. doi: 10.1177/0013916587191005

Ballas, J. A., & Mullins, R. T. (1991). Effects of context on the identification of everyday sounds. *Human Performance, 4,* 199–219. doi: 10.1207/s15327043hup0403_3

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11,* 211–227. doi: 10.3758/BF03196968

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling. Theory and applications.* New York, NY: Springer.

Cabe, P. A., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 313–324. doi: 10.1037/0096-1523.26.1.313

Cano, P., Koppenberger, M., Herrera, P., & Celma, O. (2004). Sound effect taxonomy management in production environments. In *Proceedings of 25th international AES conference* (pp. T3–2). London, England: Audio Engineering Society.

Cano, P., Koppenberger, M., Le Groux, S., Ricard, J., Wack, N., & Herrera, P. (2004, May). *Nearest-neighbor generic sound classification with a wordnet-based taxonomy.* Paper presented at the Audio Engineering Society Convention 116, Berlin, Germany.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of the length by sound. *Psychological Science, 9,* 211–214. doi: 10.1111/1467-9280.00040

Cummings, A., Èeponienë, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research, 1115,* 92–107. doi: 10.1016/j.brainres.2006.07.050

Drioli, C., Polotti, P., Rocchesso, D., Delle Monache, S., Adiloglu, K., Robert, A., & Obermayer, K. (2009). Auditory representations as landmarks in the sound design space. In *Sixth proceedings of sound and music computing conference* (pp. 315–20). Porto, Portugal: SMC.

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

Fellbaum, C. (1998). *Wordnet: An electronic lexical database.* Cambridge, MA: MIT Press. Retrieved from http://isbn.nu/9780262061971

Gaver, W. W. (1993a). How do we hear in the world ? Explorations in ecological acoustics. *Ecological Psychology, 5,* 285–313. doi: 10.1207/s15326969eco0504_2

Gaver, W. W. (1993b). What is the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology, 5,* 1–29. doi: 10.1207/s15326969eco0501_1

Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America, 119,* 1171–1881. doi: 10.1121/1.2149839

Giordano, B. L., McDonnell, J., & McAdams, S. (2010). Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds. *Brain and Cognition, 73,* 7–19. doi: 10.1016/j.bandc.2010.01.005

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52,* 125–157. doi: 10.1016/0010-0277(94)90065-5

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception and Psychophysics, 67,* 274–284. doi: 10.3758/BF03206491

Gaver, W. W. (1993). What in the World Do We Hear? An Ecological Approach to Auditory Event Perception, Copyright 1993 by Taylor & Francis Group, Inc., *Ecological Psychology, 5,* 15. http://www.informaworld.com

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5,* 887–892. doi: 10.1038/nrn1538

Guyot, F. (1996). *Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative: Une approche par la catégorisation* [A study of sound perception, in terms of recognition and qualitative appraisal]. Université du Maine, Le Mans.

Guyot, F., Castellengo, M., & Fabre, B. (1997). *Étude de la catégorisation d'un corpus de bruits domestiques* [A study of the categorization of an everyday sound set]. Catégorisation et Cognition: De la Perception au Discours (pp. 41–58). Paris, France: Éditions Kimé.

Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America, 115,* 1252–1265. doi: 10.1121/1.1635840

Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics, 69,* 839–855. doi: 10.3758/BF03193921

Gygi, B., & Shafiro, V. (2007). General functions and specific applications of environmental sound research. *Frontiers in Bioscience, 12,* 3152–3166. doi: 10.2741/2303

Gygi, B., & Shafiro, V. (2010). Development of the database for environmental sound research and application (DESRA): Design, functionality, and retrieval considerations. *EURASIP Journal on Audio, Speech, and Music Processing, 2010*(Article ID 654914). doi: 10.1155/2010/654914

Hubert, L. J., Arabie, P., & Meulman, J. (2006). *The structural representation of proximity matrices with MATLAB. ASA-SIAM Series on Statistics and Applied Probability.* Philadelphia, PA: Society for Industrial and Applied Mathematics, American Statistical Association.

Kim, N. G., Effken, J. A., & Shaw, R. E. (1995). Perceiving persistence under change and over structure. *Ecological Psychology, 7,* 217–256. doi: 10.1207/s15326969eco0703_3

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shapes. *Journal of Experimental Psychology: Human Perception & Performance, 26,* 279–294. doi: 10.1037/0096-1523.26.1.279

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception and Psychophysics, 59,* 1180–1190. doi: 10.3758/BF03214206

Lebrun, N., Clochon, P., Etevenon, P., Lambert, J., Baron, J. C., & Eustache, F. (2001). An ERD mapping study of the neurocognitive processes involved in the perceptual and semantic analysis of environmental sounds and words. *Cognitive Brain Research, 11,* 235–248. doi: 10.1016/S0926-6410(00)00078-1

Lemaitre, G., Houix, O., Misdariis, N., & Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied, 16,* 16–32. doi: 10.1037/a0018762

Lewis, J. W. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex, 14,* 1008–1021. doi: 10.1093/cercor/bhh061

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America, 90,* 3036–3049. doi: 10.1121/1.401778

Lutfi, R. A. (2001). Auditory detection of hollowness. *Journal of the Acoustical Society of America, 110,* 1010–1019. doi: 10.1121/1.1385903

Lutfi, R. A. (2008). Human sound source identification. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources, Springer handbook of auditory research* (pp. 13–42). New York, NY: Springer.

Lutfi, R. A., & Oh, E. (1997). Auditory discrimination of material changes in a struck-clamped bar. *Journal of the Acoustical Society of America, 102,* 3647–3656. doi: 10.1121/1.420151

Marcell, M. M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000).

Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology, 22,* 830–864. doi: 10.1076/jcen.22.6.830.949

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). New York, NY: Oxford University Press.

McAdams, S., Roussarie, V., Chaigne, A., & Giordano, B. L. (2010). The psychomechanics of simulated sound sources: Material properties of impacted thin plates. *Journal of the Acoustical Society of America, 128,* 1401–1413. doi: 10.1121/1.3466867

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278. doi: 10.1037/0033-295X.100.2.254

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238. doi: 10.1037/0033-295X.85.3.207

Michaels, C. F., & Carello, C. (1981). *Direct perception.* New York, NY: Prentice-Hall.

Misdariis, N., Minard, A., Susini, P., Lemaitre, G., McAdams, S., & Parizet, E. (2010). Environmental sound perception: Metadescription and modeling based on independent primary studies. *EURASIP Journal on Audio, Speech, and Music Processing, 2010*(Article ID 362013). doi: 10.1155/2010/362013

Misdariis, N., Smith, B., Pressnitzer, D., Susini, P., & McAdams, S. (1998). Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres. *The Journal of the Acoustical Society of America, 103,* 3005. doi: 10.1121/1.421751

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57. doi: 10.1037/0096-3445.115.1.39

Peeters, G., & Deruty, E. (2010). Sound indexing using morphological description. *IEEE Transactions on Audio, Speech, and Language Processing, 18,* 675–687. doi: 10.1109/TASL.2009.2038809

Pittenger, J., & Shaw, R. (1975). Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 374–382. doi: 10.1037/0096-1523.1.4.374

Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'Amico, S., & Di Russo, F. (2005). Separate neural systems for processing action-or non-action-related sounds. *NeuroImage, 24,* 852–861. doi: 10.1016/j.neuroimage.2004.09.025

Reinert, M. (1986). Un logiciel d'analyse lexicale: ALCESTE [Software for lexical analysis: ALCESTE]. *Les Cahiers de l'Analyse des Données, 11,* 471–484.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America, 81,* 1100–1109. doi: 10.1121/1.394630

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122,* 468–486. doi: 10.1037/0096-3445.122.4.468

Roma, G., Janer, J., Kersten, S., Schirosa, M., Herrera, P., & Serra, X. (2010). Ecological acoustics perspective for content-based retrieval of environmental sounds. *EURASIP Journal on Audio, Speech, and Music Processing, 2010,* 1–11. doi: 10.1155/2010/960863

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38,* 495–553. doi:10.1006/cogp.1998.0712

Sauvageot, F., Urdapilleta, I., & Peyron, D. (2006). Within and between variations of texts elicited from nine wine experts. *Food Quality and Preference, 17,* 429–444. doi: 10.1016/j.foodqual.2005.05.007

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Shakespeare, W. (1992). The Tragedy of Hamlet, Prince of Denmark (B.A. Mowat & P. Werstine, Eds.). New York: Washington Square-Pocket.

Shaw, R., Flascher, O., & Mace, W. (1996). Dimensions of event perception. *Handbook of Perception and Action, 1,* 345–395. doi: 10.1016/S1874-5822(96)80013-2

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7,* 246–251. doi: 10.1016/S1364-6613(03)00109-8

Susini, P., Houix, O., Misdariis, N., Smith, B., & Langlois, S. (2009). Instruction's effect on semantic scale ratings of interior car sounds. *Applied Acoustics, 70,* 389–403. doi: 10.1016/j.apacoust.2008.06.008

Truax, B. (2001). *Acoustic communication* (2nd ed). Westport, CT: Ablex Publishing Corp.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organisation of memory* (pp. 381–402). London, England: Academic Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi: 10.1037/0033-295X.84.4.327

Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds.* Ithaca, NY: Cornell University.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and boucing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704–712. doi: 10.1037/0096-1523.10.5.704

Wold, E., Blum, T., Keislar, D., & Wheaten, J. (2002). Content-based classification, search, and retrieval of audio. *Multimedia IEEE, 3,* 27–36. doi: 10.1109/93.556537

Appendix A

List of Sounds Used in Experiment 1

No.	Description (translated from French)	Maximum level (dB)	Duration (s)
e1	Ice clinks in glass without liquid	67	1.0
e6	Dishwasher running	66	3.0
e7	Coffee perking	62	6.9
e8	Boiling pot	56	3.9
e9	Gas stove, gas grill on, with gas release	66	4.1
e10	Glass ding, crystal champagne flute toast	59	1.0
e12	Match strike and ignition	62	1.1
e15	Toaster release	68	1.9
e16	Microwave beeps, three times	55	3.0
e17	Wade through water	66	2.4
e18	Microwave on	55	3.8
e19	Food processor	72	3.5
e20	Blender	72	2.8
e22	Knife scraping plate	66	1.0
e23	Knife cutting food	65	2.9
e24	Metal pan, scraping	67	0.6
e29	Gas stove turning on	42	3.6
e30	Glass bowl and spoon been placed on table	59	1.4
e35	Sink draining	68	5.0
e38	Bottle top	58	0.7
e40	Bread cutting	61	1.2
e41	Coffee pot whistling	67	10.0
e43	Cork	63	1.5
e44	Fridge	64	3.4
e47	Pouring water into a metal kettle	67	3.8
e49	Cooking with fat	64	2.4
e51	Blowing up paper bag	68	2.5
e52	Opening a new plastic bag	67	4.4
e53	Bubbles, water cooler	62	4.1
e54	Folding a wood chair	75	1.7
e56	Plastic container, unscrewing cap	58	3.2
e57	Door, a cupboard closing	72	1.4
e58	Door, lock turning	73	1.0
e59	Drawer, opening on track	71	1.7
e60	Blind, Venetian lowering down	68	2.7
e61	Pouring beer in into glass	61	6.6
e64	Replacing a screw lid on a bottle	60	2.5
e66	Single squeeze of near empty	62	1.4
e67	Opening beer can	70	1.3
e68	Crushing tin metallic can	65	1.2
e72	Spraying polish on table	68	3.7
e74	Removing lid from plastic container	59	0.5
e77	Stirring coffee in mug	70	2.8
e79	Beating eggs in bowl with whisk	71	6.1
e80	Pouring cereal into bowl	68	5.1
e81	Pouring milk on cereal	57	5.5
e82	Crunching eggs shells	62	2.2
e83	Grating carrots with hand grater	56	4.4
e84	Slicing celery on cutting board	67	6.2
e85	Pulling tops off a bunch of carrots	62	1.8
e88	Dropping metallic lid on ground	72	1.0
e89	Salt grinder, single grind	62	0.5
e91	Sharpening a knife	66	0.5
e92	Closing microwave door	72	1.5
e94	Pulling and tearing paper towel from a holder	67	2.3
e95	Switch a lamp	66	0.6
e97	Water dripping into a container	64	3.6
e98	Water running in porcelain sink	67	5.0
e99	Water filling porcelain sink	70	6.0
e101	Sink flowing and stopping	69	4.1

(*Appendices continue*)

Appendix B

Hierarchical Cluster Analysis With Several Dendrograms

For the hierarchical cluster analyses, we used the freely available algorithms provided at http://cda.psych.uiuc.edu/srpm_mfiles/ (Hubert et al., 2006). These algorithms in Matlab generate the best-fitting ultrametric distances, minimizing the least square criterion, L_2norm (Equation 1), between distances in a proximity matrix and ultrametric distances (see Hubert et al., 2006, for the detail of the algorithm). These algorithms use a heuristic search strategy using iterative projection to locate the best-fitting ultrametric distance in the L_2norm. This L_2norm corresponds to

$$L_2norm = \sum_{i<j}(p_{ij} - u_{ij}^*)^2, \quad (1)$$

where p_{ij} is the input proximity between the objects i and j, and u_{ij} is the corresponding ultrametric distance.

The variance accounted for the ultrametric distance matrix (VAF, Equation 2) representing the proximity matrix is given by

$$VAF = 1 - \frac{\sum_{i<j}(p_{ij} - u_{ij}^*)^2}{\sum_{i<j}(p_{ij} - \bar{p})^2}, \quad (2)$$

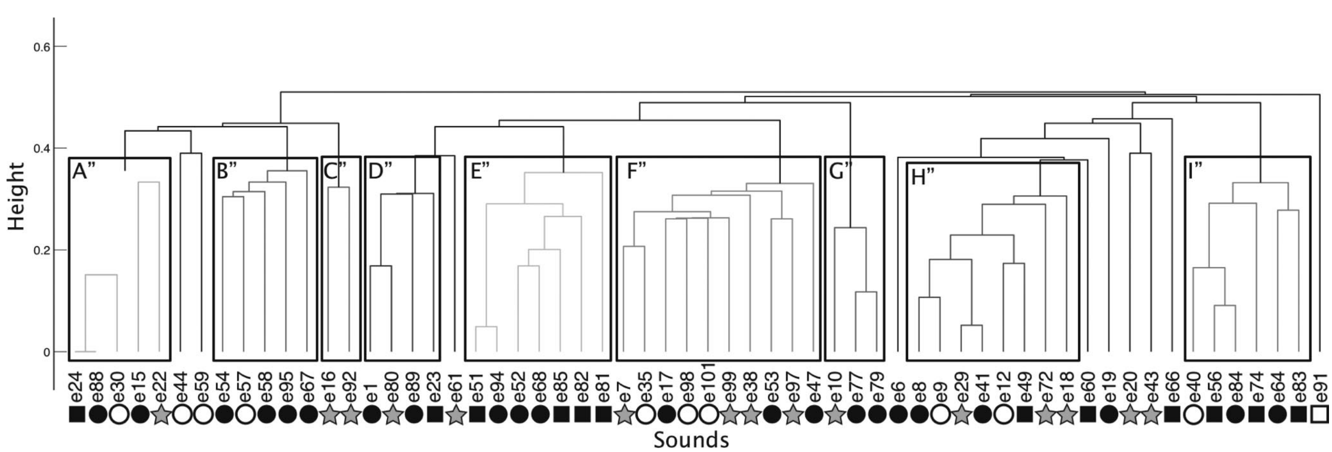
where $\bar{p}$ is the mean off-diagonal proximity, and u_{ij}^* is the best-fitting ultrametric distance for a pair of objects i and j. The variation of the VAF is between 0 and 1, with 1 for a perfect fit.

A first matrix is fitted to the proximity matrix using the L_2norm, and a first residual matrix is calculated. A second matrix is then fitted to this first residual matrix, producing a second residual matrix. Iterating, the second fitted matrix is now subtracted from the original proximity matrix, and a first (re)fitted matrix is obtained. This first (re)fitted matrix, in turn, is subtracted from the original proximity matrix, and a new (re)fitted second matrix is fitted. This process continues until the VAF by the sum of both fitted matrices no longer changes by a given value (Hubert et al., 2006).

Appendix C

Experiment 1

Secondary dendrogram. Hierarchical cluster analysis (Hubert, Arabie, & Meulman, 2006). The main clusters are indicated by Latin letters (A”, B”, C”, D”, E”, F”, G”, H”, and I”). The degree of identifiability (causal uncertainty values; Lemaitre, Houix, Misdariis, & Susini, 2010) of each sound is represented by the following code: for values between {0:1} a white square, {1:2} a white circle, {2:3} a gray star, {3:4} a black circle, and {4:5} a black square. A low value of 0 indicates a perfect identification of the sound, and a value of 5 indicates a poor identification. The labels of the sound files (e1 to e60) are presented in Appendix A.



(Appendices continue)

Appendix D

Introduction to Lexical Analysis

The goal of the statistical analysis performed by Alceste (Reinert, 1986) is to quantify a text to extract the most significant structures, categories of representative verbalizations (CRV), based on the hypothesis that these structures are closely linked to the distribution of the words in a text and that this distribution is not random. The method consists of modeling the distribution of words in a description and identifying the language patterns that are most frequently used by the participants.

The analysis has four steps. For detailed explanations, see Reinert (1986); Sauvageot, Urdapilleta, and Peyron (2006) and the tutorial illustrating the analysis of Shakespeare's *Hamlet* (1992).[5]

1. The first stage consists of dividing the text into context units (CU). This stage consists of two parts. First, the text is tagged with initial context units (ICUs) by the experimenters (i.e., the different acts and the dialogue associated with each character for Shakespeare's *Hamlet*). These ICUs correspond to the experimental variables of the study. For our study, these variables are the verbalizations of the main clusters, for example, the main Clusters A, B, C, D (Figure 3) or the main Clusters 1 to 13 (Figure 7) extracted from the hierarchical cluster analysis and their associated verbalizations. The analysis then divides the corpus into segments, the elementary context units (ECUs, approximately 10–20 words). This segmentation is automatic and iterative and is linked to the Step 3.

2. The second step concerns the lemmatization of the corpus. The words of the ECUs are reduced to simple forms based on a dictionary of morphosyntactic forms. For example, the words *players*, *player,* and *play* are reduced to *play*+. The sign + is a reduction indicator. The reduced forms are divided into two groups: the nouns, verbs, adjectives, and adverbs (analyzable forms) and the function words such as prepositions and conjunctions (illustrative forms).

3. The contingency table is built during the third step. The lines of the matrix correspond to the ECUs, and the columns correspond to the reduced forms (analyzable forms) extracted from the previous step. The analysis crosses the ECUs and the presence or absence (co-occurrence) of the reduced forms in a square matrix. It then calculates the value of the chi-square attached to each co-occurrence value. A hierarchical decreasing classification is applied to this new matrix through an iterative process to obtain CRVs.

4. The fourth step is the representation and the description of each CRV. Each CRV is described by a list of the most significant words with their associated frequencies and the chi-square measure of the importance of the words in the class. Words from a CRV are those that distinguish this class from others.

The analysis also searches for the most representative IUC, the experimental variables in our case, the main clusters (e.g., Clusters A, B, C, and D for Experiment 1) that characterize the category of representative verbalizations. For Shakespeare's *Hamlet*, the analysis can determine whether the different CRVs (the lexical field of the action or related to the concept of possession, to use examples from the tutorial) are related to specific act(s) or specific character(s) (for example, the King and the Ghost, two characters in the play).

[5] The lexical analysis of Shakespeare's *Hamlet* is available as a tutorial (http://www.image-zafar.com/PLAQGB.ZIP).

(*Appendices continue*)

Appendix E

Table E1

Experiment 1. Portraits of the Subclusters. Occurrences of the Lexical Forms in the Five Semantic Fields Associated With Each Subcluster of the Primary Dendrogram (see Figure 3)

	Cluster A_a	(e22, e91, e40, e83, e84, e23, e56, e79, e64)	Cluster A_b	(e67, e95, e68, e82, e85, e89, e74)
Semantic fields	%	(terms, occ)	%	(terms, occ)
Object	25.51	(knife, utensil, 41; canned food, can, 27)	40.35	(canned food, can, trash, packaging, 28; knife, 11)
Action	40.29	(to cut, to carve, to chop, 45; manual gesture, to make, action, 39; preparation, to cook, 28; rubbing, 21)	21.03	(to open, to close, 14; preparation, 11)
Context	15.04	(over, surface, 35; kitchen, 20)	17.54	(kitchen, 11; over, 9)
Acoustic	4.64	(sound, 10; repeated, 6)	14.91	(sound, 11; high, 6)
Other	13.62	6	5.26	6
	Cluster A_c	(e1, e10, e30, e77)	Cluster A_d	(e51, e52, e80)
	%	(terms, occ)	%	(terms, occ)
Object	58.33	(glass, containers, 28)	33.33	(cereal, 6)
Action	21.67	(shock, 7; preparation, 6)	66.67	(to crumple, 6; preparation, 6)
Context	10.00	(in, 6)	0.00	
Acoustic	0.00		0.00	
Other	10.00	(two, 6)	0.00	
	Cluster A_e	(e15, e54, e57, e58, e92, e59)	Cluster B_a	(e6, e19, e20, e44, e16, e18, e41)
	%	(terms, occ)	%	(terms, occ)
Object	54.71	(door, drawer, cupboard, microwave, furniture, 65; trash, packaging, 16)	68.49	(machine, electric device, 57; household appliance, food processors, mixer, refrigerator, 37)
Action	25.29	(to close, to open, 25; impact, slam, 12)	4.79	(to make, 7)
Context	5.88	(kitchen, 10)	4.11	(kitchen, 6)
Acoustic	10.59	(sound, 6; noise, 6; low, 6)	17.81	(noise, 19; sound, 7)
Other	3.53	6	4.79	(annoying, 7)
	Cluster C_a	(e7, e17, e61, e97, e35, e47, e98, e99, e101, e53)	Cluster C_b	(e49, e81)
	%	(terms, occ)	%	(terms, occ)
Object	69.71	(liquid, water, tea, 158; faucet, sink, 26; powder, 10; not solid, 10)	0.00	
Action	20.52	(to flow, to drain off, to pour, 39; to pour, to fill, 14)	100.00	(cooking, 6)
Context	0.00		0.00	
Acoustic	3.58	(sound, 11)	0.00	
Other	6.19	19	0.00	
	Cluster D	(e9, e29, e12, e72)		
	%	(terms, occ)		
Object	68.18	(gas, 29; fire, heat, 22; water, 9)		
Action	23.86	(cooking, 14; to leak, 7)		
Context	0.00			
Acoustic	0.00			
Other	7.95	7		

Note. For each subcluster and each semantic field, we counted the occurrences of lexical terms with close meanings together (terms, occ). For each cluster, the associated sounds are indicated (Appendix A). Occ = occurrences.

(*Appendices continue*)

Table E2

Experiment 1. Portraits of the Main Clusters. Occurrences of the Lexical Forms in the Five Semantic Fields Associated With Each Main Cluster of the Secondary Dendrogram (see Appendix C)

	Cluster A"	(e24, e88, e30, e15, e22)	Cluster B"	(e54, e57, e58, e95, e67)
Semantic fields	%	(terms, occ)	%	(terms, occ)
Object	63.16	(metal, glass, 15; trash, 8; knife, 7; door, 6)	63.46	(door, 21; trash, 6; drawer, 6)
Action	12.28	(preparation, 7)	23.08	(to close, 12)
Context	24.56	(over, 7; kitchen, 7)	0.00	
Acoustic	0.00		13.46	(sound, 7)
Other	0.00		0.00	
	Cluster C"	(e16, e92)	Cluster D"	(e1, e80, e89, e23)
	%	(terms, occ)	%	(terms, occ)
Object	0.00		33.33	(glass, container, 14)
Action	0.00		19.05	(preparation, 8)
Context	0.00		47.62	(over, in, 13; kitchen, 7)
Acoustic	0.00		0.00	
Other	100.00	10	0.00	
	Cluster E"	(e51, e94, e52, e68, e85, e82, e81)	Cluster F"	(e7, e35, e17, e98, e101, e99, e38, e53, e97, e47)
	%	(terms, occ)	%	(terms, occ)
Object	21.43	(cereal, 6; water, 6)	71.13	(liquid, water, 145; faucet, sink, 27; powder, 10; not solid, 10)
Action	42.86	(preparation, 12; to crumple, 12)	21.48	(to flow, to drain off, to pour, 40; to pour, to fill, 12; gesture, 9)
Context	12.50	(kitchen, 7)	0.00	
Acoustic	12.50	(sound, 7)	4.58	(sound, 13)
Other	10.71	6	2.82	8
	Cluster G"	(e10, e77, e79)	Cluster H"	(e8, e9, e29, e41, e12, e49, e72, e18)
	%	(terms, occ)	%	(terms, occ)
Object	63.16	(glass, 12)	61.41	(fire, heat, 36; gas, 31; water, 15; air, breath, 12; food processors, 7)
Action	36.84	(shock, 7)	34.24	(cooking, to boil, to defrost, to reheat, 49; to leak, 7)
Context	0.00		0.00	
Acoustic	0.00		4.35	(sounds, 8)
Other	0.00		0.00	
	Cluster I"	(e40, e56, e84, e74, e64, e83)		
	%	(terms, occ)		
Object	30.50	(knife, 25; canned, can, 12; water, 9)		
Action	26.95	(preparation, 16; to rub, 16; to cut, 6)		
Context	19.86	(over, 15; kitchen, 13)		
Acoustic	9.93	(sound, 8; repeated, 6)		
Other	12.77	18		

Note. For each main cluster and each semantic field, we counted the occurrences of lexical terms with close meanings together (terms, occ). For each cluster, the associated sounds are indicated (Appendix A). Occ = occurrences.

(*Appendices continue*)

Appendix F

List of Sounds Used in Experiment 2

Table F1

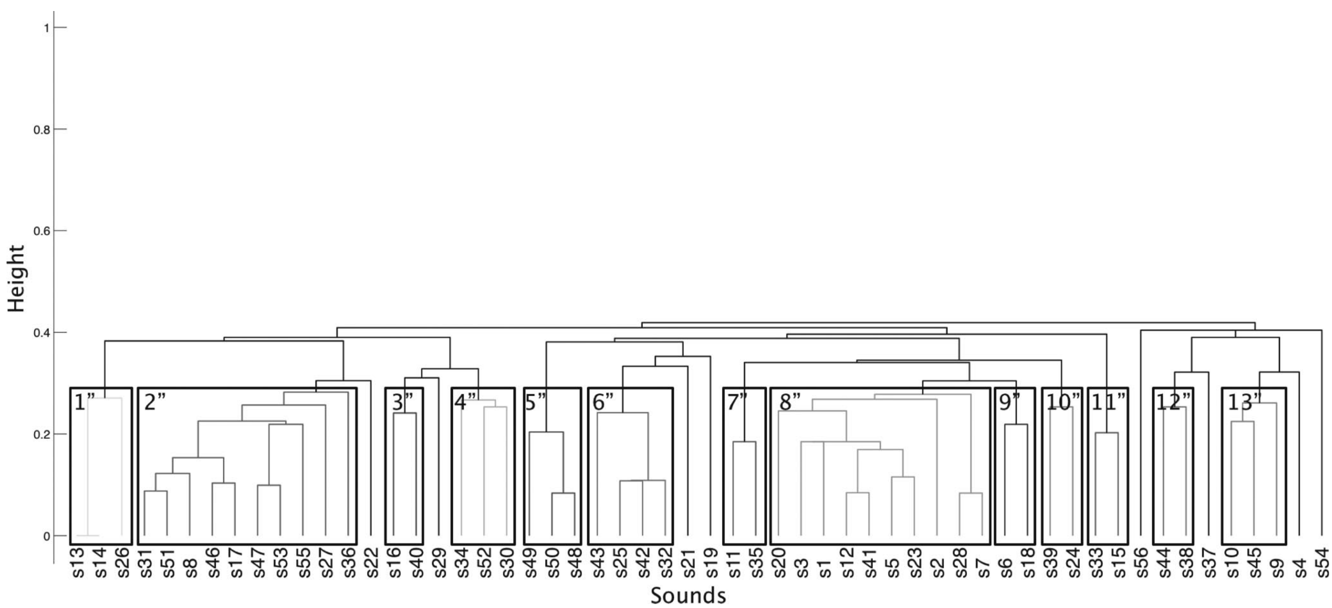
No.	Description (translated from French)	Maximum level (dB)	Duration (s)
s1	Coin in a glass	61	6.0
s2	Glass breaking	62	1.2
s3	Coins falling	52	1.0
s4	Tearing paper	57	1.2
s5	Hollow object falling	61	1.6
s6	Hollow object rolling	64	6.2
s7	Garbage closing	58	0.7
s8	Cutting a slice of bread	51	4.7
s9	Handling paper	50	1.4
s10	Crumpling papers	53	3.2
s11	Closing an old door	58	1.6
s12	Falling stone	59	1.2
s13	Walking on gravel	58	4.1
s14	Walking with rubber soles	54	2.7
s15	Tossing screws in a box	48	2.5
s16	Door squeal	63	2.0
s17	Shaking matches	52	1.3
s18	Glass that is moved	54	0.6
s19	Knocking	68	1.4
s20	Pottery breaking	70	3.8
s21	Clicking with a mouse	42	0.6
s22	Zipping	48	1.4
s23	Scratching interior iron pot	54	1.7
s24	Ball turning in a casino wheel	57	6.9
s25	Hammer on an anvil	57	1.6
s26	Opening a plastic bag	53	4.7
s27	Rubbing a finger on balloon	59	1.9
s28	Putting an empty bucket on the floor	60	1.5
s29	Opening the latch of a suitcase	50	1.3
s30	Rocking a rocking chair	56	3.8
s31	Sanding with sandpaper	58	1.7
s32	Opening a zippo lighter	47	0.8
s33	Opening a shower curtain	57	2.4
s34	Opening a screen door	56	3.9
s35	Opening a drawer	58	1.8
s36	Placing a ski glove	50	3.6
s37	Crushing a metal can	55	1.7
s38	Replacing the lid of an aerosol can	48	0.8
s39	Stirring an aerosol paint	53	2.9
s40	Winding an old clock	57	4.4
s41	Taking a bowl from stack	54	1.7
s42	Glass ding, crystal champagne flute toast	53	1.2
s43	Stirring coffee in a mug	51	2.4
s44	Crushing egg shells	42	2.3
s45	Pulling tops off a bunch of carrots	46	2.0
s46	Salt grinder, single grind	49	1.9
s47	Sharpening a knife	55	0.6
s48	Closing the door of a microwave	56	1.7
s49	Small breaker switch	49	0.6
s50	Toaster release	55	1.1
s51	Cutting paper with scissors	52	4.0
s52	Unwinding adhesive tape	46	1.7
s53	Writing with a pencil	49	1.3
s54	Tearing cloth	59	0.8
s55	Small pulleys of metal turning	55	3.1
s56	Detaching Velcro	57	2.6

(*Appendices continue*)

Appendix G

Experiment 2

Secondary dendrogram of the hierarchical cluster analysis (Hubert, Arabie, & Meulman, 2006), Experiment 2. The main clusters (1” to 13”) are indicated by Arabic numbers. The labels of the sound files (s1 to s56) are presented in Appendix Table F1.



(*Appendices continue*)

Appendix H

Table H1

Experiment 2. Portraits of the Main Clusters. Occurrences of the Lexical Forms in the Five Semantic Fields Associated With Each Main Cluster of the Primary Dendrogram (see Figure 6)

	Cluster 1	(s20, s2)	Cluster 2	(s49, s32, s21, s38, s29)	Cluster 3	(s25, s42, s41)
Semantic fields	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	0.00		27.50	(thing, object, 16; mechanism, 6)	21.31	(metal, glass, 13)
Action	84.91	(to break, to break down, 36; to fall, 9)	33.75	(to hit, shock, 27)	39.34	(to hit, shock, 24)
Context	0.00		0.00		18.03	(over, 11)
Acoustic	0.00		22.50	(sound, 10; to tinkle, 8)	9.84	(to tinkle, 6)
Other	15.09	8	16.25	13	11.48	7

	Cluster 4	(s11, s48, s7)	Cluster 5	(s16, s34, s30)	Cluster 6	(s40, s55)
	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	27.69	(door, 10; thing, 8)	39.68	(spring, 9; door, 8; object, 8)	21.43	(thing, 6)
Action	63.08	(to close, 23; shock, 6; to fall, 6)	12.70	(movement, 8)	57.14	(to turn, 9; movement, 7)
Context	0.00		0.00		0.00	
Acoustic	0.00		47.62	(to creak, 30)	0.00	
Other	9.23	6	0.00		21.43	6

	Cluster 7	(s6, s24, s5, s1, s3, s12)	Cluster 8	(s39, s17)	Cluster 9	(s13, s15)
	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	18.25	(object, thing, 28; pan, container, 14)	21.88	(thing, 7)	0.00	
Action	68.82	(to tumble, to fall, 76; to roll, 62; to make, 19; to stop, 6; shock, 6; movement, 6; to leave, 6)	78.13	(to shake, 25)	53.37	(to walk, 15)
Context	4.56	(in, 6; surface, 6)	0.00		25.00	(in, 7)
Acoustic	3.42	(sound, 9)	0.00		21.43	(sound, 6)
Other	4.94	13	0.00		0.00	

	Cluster 10	(s26, s10, s36, s9)	Cluster 11	(s44, s45, s37)	Cluster 12	(s56, s54, s4, s52, s27, s22)
	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	18.82	(paper, 16)	0.00		6.00	(zipper, 9)
Action	71.76	(to crumple, to crease, 54; to turn, 7)	81.82	(to crumple, 20; to crush, 7)	87.33	(to tear, 81; to zip, 17; to flick, 12; to rub, 8; to open, 7; to elongate, 6)
Context	0.00		0.00		0.00	
Acoustic	9.41	(noise, 8)	0.00		6.67	(noise, 10)
Other	0.00		18.18	6	0.00	

	Cluster 13	(s31, s53, s8, s51)
	%	(terms, occ)
Object	0.00	
Action	76.15	(to cut, to saw, 49; to rub, 42; to-and-fro, 8)
Context	15.38	(surface, 20)
Acoustic	8.46	(noise, 11)
Other	0.00	

Note. For each main cluster and each semantic field, we counted the occurrences of lexical terms with close meanings together (terms, occ). For each cluster, the associated sounds are indicated (Appendix F). Occ = occurrences.

(*Appendices continue*)

Table H2
Experiment 2. Portraits of the Main Clusters. Occurrences of the Lexical Forms in the Five Semantic Fields Associated With Each Main Cluster of the Secondary Dendrogram (see Figure G)

	Cluster 1”	(s13, s14, s26)	Cluster 2”	(s31, s51, s8, s46, s17, s47, s53, s55, s27, s36)	Cluster 3”	(s16, s40)
Semantic fields	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	0.00		5.07	(thing, object, 18)	0.00	
Action	67.35	(to walk, 22; to crumple, 11)	64.23	(to rub, 85; to cut, to saw, 77; to shake, 10; to-and-fro, 10; to scratch, 9; to crumple, 8; to turn, 8; to manipulate, 8; to sand, 7; to sharpen, 6)	53.33	(to turn, 9; movement, 7)
Context	12.24	(in, 6)	11.27	(surface, 27; over, 13)	0.00	
Acoustic	20.41	(noise, 10)	16.90	(irregular, repeated, rhythm, 24; noise, 20; crackling, 8; sound, 8)	0.00	
Other	0.00		2.54	9	46.67	14
	Cluster 4”	**(s34, s52, s30)**	**Cluster 5”**	**(s49, s50, s48)**	**Cluster 6”**	**(s43, s25, s42, s32)**
	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	24.53	(object, 7; spring, 6)	12.00	(door, 6)	25.29	(metal, glass, 14; thing, 6)
Action	30.19	(to tear, 10; movement, 6; to fall, 6)	48.00	(shock, to hit, 14; to close, 10)	28.74	(shock, to hit, 25)
Context	11.32	(over, 6)	0.00		12.64	(over, 11)
Acoustic	33.96	(to creak, 18)	12.00	(sound, 6)	17.24	(to tinkle, 8; sound, 7)
Other	0.00		28.00	14	16.09	14
	Cluster 7”	**(s11, s35)**	**Cluster 8”**	**(s20, s3, s1, s12, s41, s5, s23, s2, s28, s7)**	**Cluster 9”**	**(s6, s18)**
	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	0.00		20.44	(thing, object, 44; pan, container, dishes, 22; metal, glass, 18)	0.00	
Action	100.00	(to close, 6)	62.53	(to fall, 106; to break, 45; to hit, shock, 42; rolling, 23; to make, 16; to close, 7; to throw, 6; to leave, 6; to knock over, 6)	64.29	(to roll, 18)
Context	0.00		2.43	(over, 10)	0.00	
Acoustic	0.00		3.16	(sound, 13)	0.00	
Other	0.00		11.44	47	35.71	10
	Cluster 10”	**(s39, s24)**	**Cluster 11”**	**(s33, s15)**	**Cluster 12”**	**(s44, s38)**
	%	(terms, occ)	%	(terms, occ)	%	(terms, occ)
Object	16.22	(thing, 6)	0.00		0.00	
Action	64.86	(to shake, 12; rolling, 6; to fall, 6)	25.00	(to shake, 6)	100.00	(to crumple, 6)
Context	0.00		25.00	(in, 6)	0.00	
Acoustic	0.00		0.00		0.00	
Other	18.92	7	50.00	12	0.00	

(*Appendices continue*)

Table H2 (*continued*)

	Cluster 13”	(s10, s45, s9)
	%	(terms, occ)
Object	18.18	(paper, 12)
Action	62.12	(to crumple, 34; to turn, 7)
Context	0.00	
Acoustic	0.00	
Other	19.70	13

Note. For each main cluster and each semantic field, we counted the occurrences of lexical terms with close meanings together (terms, occ). For each cluster, the associated sounds are indicated (Appendix F). Occ = occurrences.

Received January 5, 2011
Revision received August 10, 2011
Accepted September 6, 2011 ■