Auditory display of seismic data: On the use of experts' categorizations and verbal descriptions as heuristics for geoscience

Arthur Paté, Lapo Boschi, Danièle Dubois, Jean-Loic Le Carrou, Benjamin Holtzman

To cite this version:

Arthur Paté, Lapo Boschi, Danièle Dubois, Jean-Loic Le Carrou, Benjamin Holtzman. Auditory display of seismic data: On the use of experts' categorizations and verbal descriptions as heuristics for geoscience. Journal of the Acoustical Society of America, 2017, 141, pp.2143 - 2162. 10.1121/1.4978441 . hal-01522831

HAL Id: hal-01522831
https://hal.sorbonne-universite.fr/hal-01522831

Submitted on 15 May 2017

Auditory display of seismic data: On the use of experts' categorizations and verbal descriptions as heuristics for geoscience.

Arthur Paté[a)][1], Lapo Boschi[b)][2], Danièle Dubois[1], Jean-Loïc Le Carrou[1], and Benjamin Holtzman[3]

[1]Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, équipe LAM, F-75005 Paris, France
[2]Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7193, Institut des Sciences de la Terre de Paris (ISTeP), F-75005 Paris, France
[3]Lamont Doherty Earth Observatory, Columbia University, New York, USA

February 24, 2017

[a)] Arthur Paté is also at Lamont Doherty Earth Observatory, Columbia University, New York, USA

[b)] Lapo Boschi is also at Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, équipe LAM, F-75005 Paris, France

Abstract: Auditory display can complement visual representations in order to better interpret scientific data. A previous article showed that the free categorization of "audified seismic signals" operated by listeners can be explained by various geophysical parameters. The present article confirms this result and shows that cognitive representations of listeners can be used as heuristics for the characterization of seismic signals. Free sorting tests are conducted with audified seismic signals, with the earthquake/seismometer relative location, playback audification speed, and earthquake magnitude as controlled variables. The analysis is built on partitions (categories) and verbal comments (categorization criteria). Participants from different backgrounds (acousticians or geoscientists) are contrasted in order to investigate the role of the participants' expertise. Sounds resulting from different earthquake/station distances or azimuths, crustal structure and topography along the path of the seismic wave, earthquake magnitude, are found to a) be sorted into different categories, b) elicit different verbal descriptions mainly focused on the perceived number of events, frequency content, and background noise level. Building on these perceptual results, acoustic descriptors are computed and geophysical interpretations are proposed in order to match the verbal descriptions. Another result is the robustness of the categories with respect to the audification speed factor.

Keywords: free sorting; categorization; sonification; perception; auditory display; verbalization

PACS: 43.66.Lj; 43.40.Ph; 43.75.Cd

I Introduction

The development of seismology as a scientific discipline has traditionally been based upon graphical tools (through visualization of empirical data on graphs, lists, and figures[1;2]), and primarily upon the visual analysis of seismograms[3], which are representations of recordings of the oscillations of a point at the Earth's surface. The advent of digitized data acquisition and the development of modern signal processing techniques has facilitated the representation of seismic data (or of potentially any data eliciting no *a priori* modality of display) through other sensory modalities. The work presented here deals with the auditory representation ("auditory display") of seismic data.

Many instances of educational[9;10;11;12;13;14;15;16] or artistic[6;7;8] uses of seismic data "sonification" have been reported. However, to our knowledge, it has only been used twice for scientific research purposes: Speeth[4] and then Frantti and Levereault[5] accelerated seismic signals in order to shift the frequencies to the audible range, and trained people to tell "natural" earthquakes ("double-couple" sources) from explosions by listening to accelerated seismograms. This promising approach has not found practical applications, because of the development of digital seismology in the 1970s and the focus on mathematical processing with computers, largely bypassing the direct analysis of seismograms by human observers.

Auditory display as a scientific research field has grown considerably during the last decade[17]. Auditory representation of data has proven efficient for *e.g.* solar wind ion composition[18], stem-cell classification[19], recognition of patterns in stock market data[20], or in the physiological processes of trees[21]. A previous article[22] has shown that the free categorization of audified seismic signals conducted by listeners is consistent with some geophysical parameters (distance between epicenter and recording station, Earth's structure). If it seems now to be accepted that auditory display can complement visual display contributing to the interpretation of scientific data, further investigations are necessary to more precisely identify the structures of *a priori* cognitive representations that are involved when humans are exposed to auditory display, and the properties that are processed when the data (earthquake recordings in our case) are transformed into acoustic signals.

While the classification freely performed by listeners has been shown to agree with some conceptual representations of geoscientists[22] (categories made by listeners match categories based on geophysical parameters), it is now necessary to access the "intensional" definition of the categories, *i.e.* to identify the criteria used for the categorization. The aim of the present article is to show that cognitive representations of listeners, when extracted using proper testing and analysis methods, can be used as heuristics in order to identify relevant features for the discrimination and characterization of seismic signals. Thus at this stage we are not claiming general results about how humans perceive sonified seismic data (for this reason, statistical analysis would not be appropriate for this study), but rather at exploring the ability of some expert listeners (in sound *per se*, or in geophysics) to bring novel description of the data that can be used by geoscientists.

Following a first experiment[22], here referred to as T1, investigating the effect of event/station relative location, we apply the same approach in 3 further experiments to consolidate the results previously obtained and acquire more precise knowledge on auditory categories for earthquake recordings, through the investigation of other seismic parameters (magnitude in T2, audification speed factor in T3a and T3b). More importantly, while the previous study was limited to similarity measurement (only based on co-occurrences of stimuli in categories), here a thorough cognitive analysis is conducted on the verbal data collected at the end of the tests. This approach allows us to get at the relevant characteristics of the stimuli mentioned by the listeners, and therefore to guide our exploration of the seismic data. The verbal analysis, and its use to access and understand the categorization criteria, constitutes the original contribution of the present article with respect to the previous one. Following an inductive approach, the analysis of the comments associated to the categories is used to access and understand the categorization criteria, common or different between ensembles[1] of subjects (acousticians or geoscientists, the first ones being trained in listening and analyzing any acoustic signal as such, the second ones being experts in earthquakes). These criteria are further used to elicit and suggest relevant parameters for the description of the categories in terms of physical parameters.

The second contribution of this article is to take into account the expertise of the listeners. This question has been widely discussed in the literature, mostly for musical expertise and exposure to familiar sounds: Trained musicians and non-musicians were shown to have similar results on musical processing tasks[23] (although trained musicians' answers are more accurate[24]), mostly because both groups have been exposed to music on a everyday basis. Yet it is known that in sorting tasks the expertise of listeners can change the way the categories are formed[25] or the level of categorization[26]. Furthermore the focus is here on sounds any of the two ensembles of participants has never been exposed to, and the question of the use of prior knowledge (either on sound or seismic data) on such signals remains open. The previous knowledge involved in subjects' perceptual processing is investigated here through a subject-centered approach of cognition and categorization as "acts of meaning"[27;28] (*i.e.* "the nature and cultural shaping of meaning-making, as the central place it plays in human actions"[29]). For that purpose we contrasted two ensembles of subjects exposed to the "same signals": "geoscientists" — experts in visual analysis of seismograms, but not trained in processing (earthquakes as) acoustic signals — , and acousticians — experts in acoustic signal processing but without background in seismology. In this situated approach of cognition, categories resulting from individual sensory experience are not conceived as "information processing" filtered by the human senses but as a meaning-making process involving different types of knowledge[30], among which individual experience, knowhow, academic and scientific knowledge. The exploration of sensory categories cannot therefore rely only on the scientific knowledge of the world (as given by geoscience or acoustics in our case), but has to identify the categories as sets of properties making sense to the user ("*ad hoc* categories"[31;32]). Such categories as individual cognitive constructions not only include perceived physical characteristics (bottom-up processes, signal processing), but also memorized properties (top-down processes, signal interpretation) in context (*i.e.* depending on the subject's goal, cognitive orientation and attention, and expertise)[2]. Within this theoretical framework, the physical characteristics of the stimuli as defined in terms of dimensions may be psychologically meaningful only if relevant for discriminating categories. For example categories of everyday sounds are not structured along the dimension of intensity as an independent variable but in close interaction with the source identification[33], categories for soundscapes are not structured along intensity as a physical (abstracted) dimension of the acoustic signal but remain embodied within the experience, concern, and identification of the source[34].

A major empirical consequence of this theoretical positioning is that the physical description cannot be used as the reference for evaluating the human categories as deviations "errors" from the "true" representations given by physics. Therefore at this exploratory step of research, we stand apart from the psychophysical paradigm which attributes to the physical description the referential value in defining *a priori* what information is to be processed by humans[30]. We rather focus on a subject-centered paradigm, in that we explore the ability of different expert listeners to bring their specific descriptions of the data. The present approach is inductive, aiming at providing new hypotheses for future hypothetico-deductive studies, which would consolidate the hypothesis through further and more canonical experimental setting including statistical analysis. Again, statistical analysis is not relevant to our approach.

As in the previous paper[22], the audio stimuli result from a time-compression of seismic signals (this method called "audification" is a particular case of sonification[17;35]), and are presented to over headphones. Note that the seismic recordings themselves and even more so the audified signals are observable reproductions of conceptual representations elaborated from present day scientific knowledge and technologies: We implicitly assume some adequacy between these representations and the vibration of a point on the Earth's surface but it must be kept in mind that we actually deal with a specific representation of a complex phenomenon.

In the rest of the paper, Sec. II describes the database and the production of stimuli, Sec. III describes the experimental method, Sec. IV describes the analysis method for the categories and their verbal description, and Sec. V present the results as perceptual descriptions of the categories of stimuli. Building on the perceptual results, acoustic descriptors are eventually computed in order to match the verbal descriptions: they are presented in Sec. VI.

II From seismic signals to audio stimuli

A The database

The database used in this study consists of broadband recordings (sampling frequency $F_{s,0} = 40$ Hz, recording stations of the USArray experiment[36], all stations with nearly identical mechanical characteristics and spectral sensitivity) of the Earth's oscillations, made at the locations depicted in Fig. 1, of a sequence of 40 injection-triggered ("hydrofracturing"[37;38]) earthquakes (or "seismic events") in Okhlahoma that occurred in November 2011. Fig. 1 is a topographic map of the study area, showing the location of the stations and epicenters, as well as the Earth's elevation. Fig. 1 shows that central and western stations are located on relatively flat terrain, whereas northern-eastern and northern-western stations have a higher elevation. Topography differences are indicative of crustal structure heterogeneities[22;39] (mechanic properties of rocks, crustal thickness, *etc.*) so that this region is not seismologically homogeneous. In order to give a more synthetic view of the seismic events and stations, Fig. 2a presents a map of the seismic stations used in this study, and Fig. 2b presents a map of the earthquakes of the database.

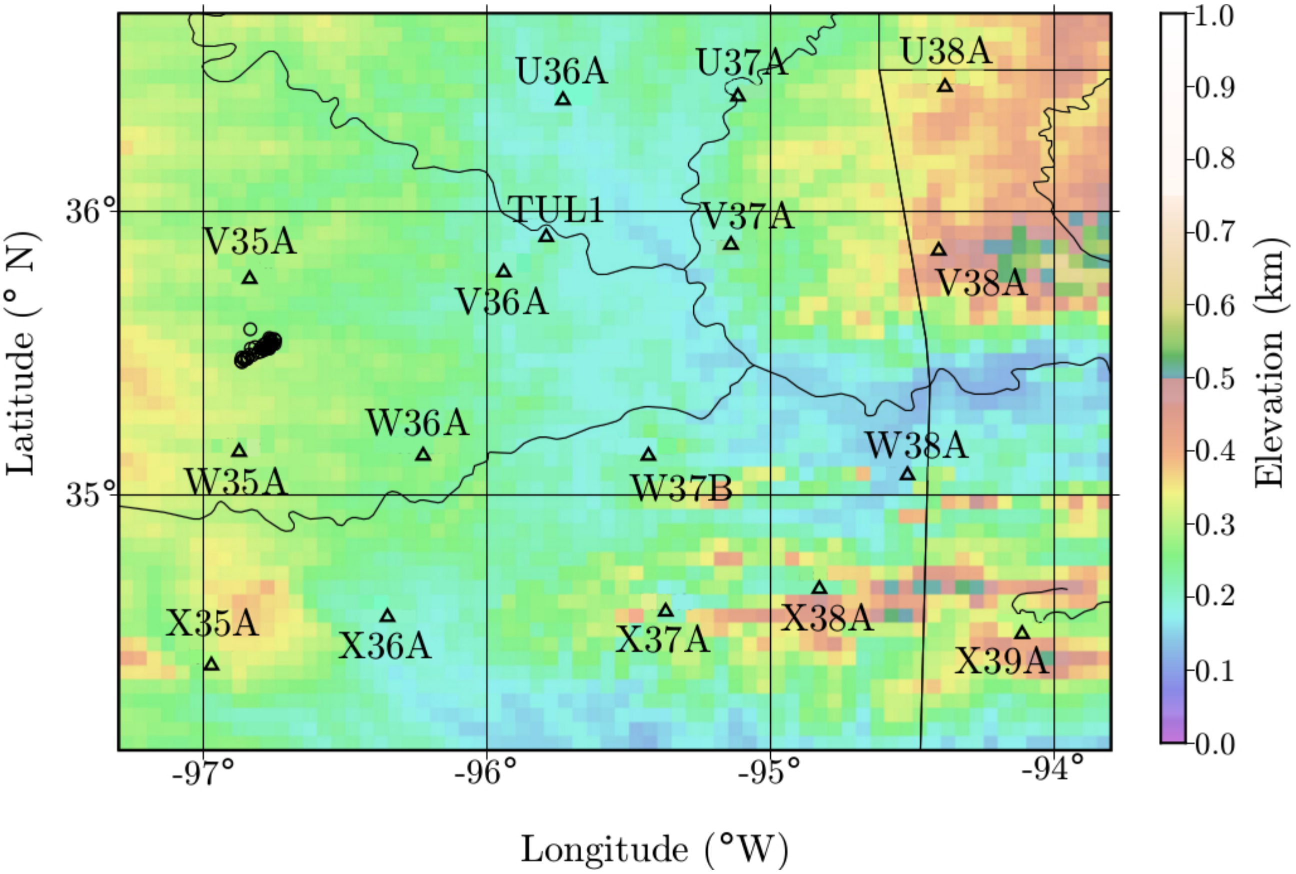

Figure 1: (color online) Topography of the study area. Black triangles denote available seismic stations, which are labelled. The color scale corresponds to the elevation of the Earth's surface with respect to sea level. Black circles on the left side denote the epicenters of the seismic events of the database.

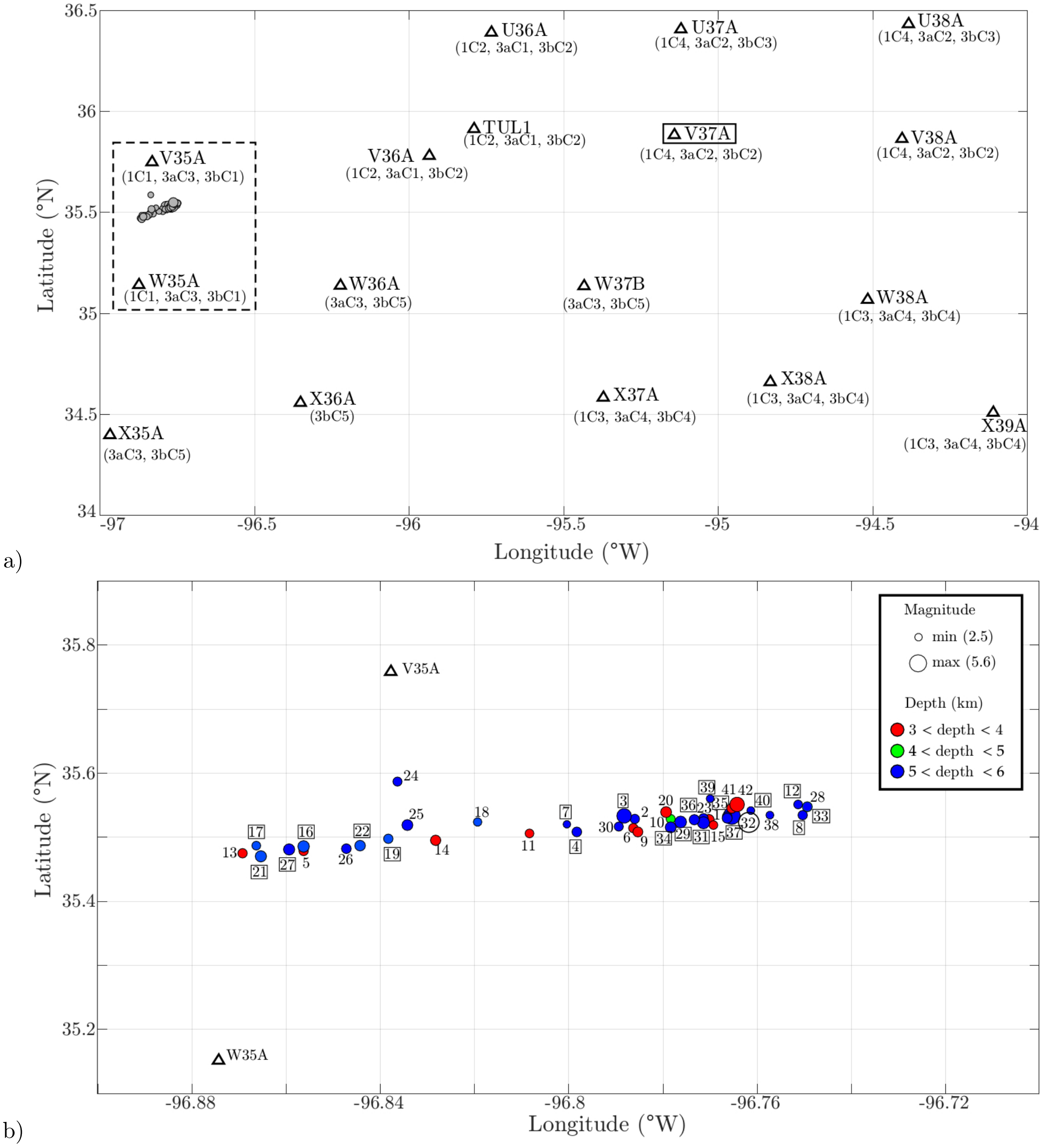

Figure 2: (color online) Map of the seismic stations (black triangles and names, followed by the consensual categories in which the resulting audio signals are put in T1, T3a and T3b, see Sec. V for more details) and the seismic events of the database (filled circles: the size is proportional to the magnitude). a) Global view, the selected station for T2 is enclosed in a solid line rectangle, the dashed line rectangle indicates the area of the epicenters, magnified below; b) Magnified view centered on the seismic event epicenters, the selected event for tests T1, T3a and T3b is circled, the selected events for T2 are enclosed in a rectangle. The events are numbered in decreasing order of appearance (1 is the latest, 42 is the earliest) during the 4 days of recording. Different colors indicate the depth of the events.

B The stimuli

The recording stations are 3-component sensors, measuring ground vibrations in the vertical (normal to the Earth's surface) and the two orthogonal horizontal directions. The audible acoustic waves are unidimensional; only the vertical component of the seismic recordings is investigated. Based on the similarity in the nature of seismic and acoustic signals (zero-mean, decreasing amplitude), the most direct sonification method is used, that is "audification". In the present case, the inaudible content of seismic recordings has to be translated to audible range. Audification then consists of playing the recorded samples at a quicker rate, which is implemented through an increase of the sampling frequency ($F_s > F_{s,0}$, the ratio $F_s/F_{s,0}$ is the speed factor). The dynamic range of seismic signals is much wider than that of audio signals, so that, in practice, signals associated with lower-magnitude events in our database would be too quiet to be heard; audified signals are therefore all normalized, each with respect to its maximum amplitude value. This means that some of the information that is contained in the signals, and that in principle could contribute to their auditive interpretation, is lost. The related issues are addressed below, in our discussion of experimental results. Four free sorting tasks are carried out. The tests are numbered T1, T2, T3a and T3b according to their order of presentation.

Stimuli for T1 The variable is the event/station relative location (distance and azimuth of the station with respect to the event), so recordings of the same event by 17 stations are used. In order to maximize the signal-to-noise ratio the seismic event with the highest magnitude is selected: Event number 32 (magnitude 5.6, depth 5.2 km, circled in Fig. 2b.) Recordings of this event from the stations plotted in Fig. 2a are audified with a speed factor of 150 ($F_s = 6,000$ Hz). All 17 signals are trimmed so as to obtain a duration of 2s for the resulting audio stimuli.

Stimuli for T2 The variable is the magnitude of the earthquakes. Twenty-two seismic events out of 42 are selected since they have the same estimated epicenter depth (5 km) and present magnitudes ranging from 2.5 to 4.8, which is typical of the database (magnitudes of the 42 events range from 2.5 to 4.8 with only one event with higher magnitude, that is 5.6). The seismic recordings used are from station *V37A*, located at an intermediate distance from the events. The seismic signals are audified with a speed factor of 150 ($F_s = 6,000$ Hz).

Stimuli for T3a The selected stations and event for T3a are the same as in T1, except that the seismic recordings are audified with a speed factor of 250 ($F_s = 10,000$ Hz).

Stimuli for T3b The selected stations and event for T3b are the same as in T1, except that the seismic recordings are audified with a speed factor of 350 ($F_s = 14,000$ Hz).

Another test variable must be made explicit here: The role of previous knowledge and expertise is investigated in all 4 tests by having 2 ensembles of subjects (geoscientists and acousticians) involved. Table 1 summarizes the effects (station location, earthquake magnitude, sonification speed factor, expertise of listeners) tested in each test, either as a variable or as a fixed effect. Tests T1, T3a and T3b, presenting the same stimuli with different levels of the speed factor as a fixed effect, can be directly compared. Additionally and as a guide for the readers, all stimuli are available online[3]. In general, they can be roughly and informally described as a gunshot-like sound with decay over a broadband background noise. This decaying part is called the "coda", and is known by geoscientists as containing all the information about wave propagation, path, scattering, attenuation.

III Experimental procedure

Classic psychophysical methods mainly rely on exclusively bottom-up models of stimulus processing, and only involve stimuli that are controlled and designed along independent physical parameters and that determine the perceptual answers, measured along dimensional indices. These methods are challenged by the two following observations. First, human perception is influenced by bottom-up (signal-driven) processes but also by top-down processes which depend on the memorized knowledge and expertise of the participants. Second, since no previous experiment dealt with the perception of audified seismic data except for our previous study[22], it is impossible to *a priori* decide what acoustical parameters are relevant for exploring the psychological processing (the study precisely aims at discovering it). The free sorting task[40;41;42;43;44;45] is chosen in this study, because it can address these two remarks. The contribution of participants' previous knowledge is explored by contrasting two ensembles of subjects.

A Theoretical background

Following Rosch's seminal work[46] on the structure of so-called "natural" categories, we aim to identify both the extensional structure of the categories as the list of their members, as well as their intensional structure as sets of properties defining the categories. Unlike well-defined categories as elaborated in scientific knowledge (with clear-cut binary membership, *i.e.* an item is either a member of the category, or is not a member), the extensional structure of natural categories is defined by similarity and distance from a prototypical exemplar. The intensional description relies on this prototype, which is defined as the exemplar gathering most of the properties of the category. The other exemplars (stimuli) are distributed along similarity ("family" resemblance) within the set of properties that they "more or less" share with one another. If a lot of psychological literature has been devoted to developing various models of categories constructed along prototype and similarity[47;48], previous research has been mainly concerned with acquired and shared established knowledge on different objects but has more rarely dealt with experiential knowledge and individually-constructed categories. Participants in this study are either acousticians or geoscientists (see Sec. B), having different education and knowledge: if acousticians are trained to describe sounds as objects *per se*, geoscientists process the "same" sounds as acoustic representations of seismic signals (as "sounds of"). The question is to identify how this difference in expertise influences the categories and the categorization criteria.

B Free sorting experiment

1 Participants

Each participant is presented with tests T1, T2 and either T3a or T3b, always in this order. 24 participants (15 acousticians, 9 geoscientists) took part in T1 and T2. From these 24 participants, 11 (8 acousticians, 3 geoscientists) in T3a and 12 (6 acousticians, 6 geoscientists) in T3b. One participant (acoustician) did not take part in T3. People in the ensemble "acousticians" are either faculty/staff of the LAM team at the d'Alembert Institute or professional sound technicians. People in the ensemble "geoscientists" are faculty/staff of the Earth Sciences Institute ISTeP. Note that the time and availability constraints did not allow us to have as many geoscientists as acousticians participating in the test. Note also that the assignment of T3a or T3b to a participant was randomized, resulting in a non-balanced number of acousticians and geoscientists for T3b. Note that at this stage of investigation the goal of the research is to find out whether such a differential approach is contrasting different ensembles of subjects is productive even with a small number of subjects, for the potential development of further research with more quantitative data allowing statistical computation.

2 Procedure

The instructions given to the participants for each test are as follows[4]:

> *Please sort the sound samples presented to you. Group the samples which seem similar to you, and put in different groups those which seem different to you. You may form as many groups as you wish.*

Each of the N stimuli has to belong to only one group. The participants are free to form as many groups as they want and can put any number of stimuli in a single group. The participants are told that the stimuli originated from seismic recordings, but no other information on their nature is given. Only the assignment of each stimulus to a group is taken into account in the analysis: The spatial arrangement of the groups and the icons within the interface area is neglected. At the end of the sorting, each participant is asked to type a comment for each group he/she made. The categorization and the verbal description provide complementary characterizations of the stimuli: descriptions allow us to identify the characterization of stimuli as properties along which the categorization has been processed.

3 Experimental setup

The tests are run on a laptop equipped with a *RME Fireface UCX* soundcard. The stimuli are played back through *Sennheiser HD380 Pro* headphones. Audio stimuli are monophonic, each ear being exposed to the same signal, in phase. The participants can set and change the sound level in the headphones at any time during the test (but no participant did it). The *TCL-LabX*[49] software is used for the free sorting interface. The graphic interface displays each stimulus as a small square icon. Illustrations of the graphic interface and of the test setup are given in the previous study[22]. The N icons for a test with N stimuli are randomly numbered from 1 to N. A double click on an icon launches the stimulus playback, and the icon can be moved within the entire interface area with a click-and-drag operation. Each stimulus can be played back as many times as wanted.

4 Output and duration

The output of each test, for each participant, is referred to in the following as a "partition;" it consists of a suite of groups of stimuli, accompanied by a verbal description of each group in this partition.

The mean duration for test T1 was 13.6±9.1 minutes, and for T2, T3a and T3b respectively 9.3±3.7 min., 5.3±1.5 min. and 5.3±2.1 min.

IV Method of data analysis

The data analyzed are of two kinds, and as such impose different types of processing in order to evaluate their robustness and interpret them within the theoretical framework discussed in Sec. A: Individual partitions resulting from the sorting task (Sec. A), and b) verbal comments for each individual partition (Sec. B).

A Categories of stimuli

Partitions reflect similarities and differences between stimuli as evaluated by the subjects: Stimuli within a category are more similar to one another than stimuli sorted into different categories. Individual partitions are added up, with the number of subjects grouping together a certain pair of stimuli functioning as a metric of the similarity between those two stimuli. It is very important to note that this measure of similarity a) relies on the consensus between subjects, and b) processes stimuli as whole and indivisible items. In other words we get a representation "in extension" of the categories (*i.e.* an explicit list of its members). Consequently a projection of the stimuli on a multidimensional space is not necessary relevant, before further investigations of the "intensional" description of the stimuli (*i.e.* as sets of dimensional properties or other characteristics). It is therefore not adequate to use statistics relying on distributions of values on dimensions and Gaussian assumptions, and we prefer other mathematical metrics developed through the close collaboration between mathematicians and psychologists in order to account for classification analysis[50;51;52].

Only a concise description of the mathematical method of analysis is given here: A more detailed description can be found in the literature[22;41;45;50;53;54;55]. For each test, a perceptual distance between stimuli is defined as follows:

1. A co-occurrence matrix M^k is defined and computed for each participant (k=1,…,N_s, where N_s is the number of participants). M^k is a square matrix of size N, where N is the number of stimuli:
    - $M^k_{ij} = 1$ if stimuli i and j are in the same group according to participant k,

- $M_{ij}^k = 0$ if stimuli i and j are in different groups according to participant k;

2. The total co-occurrence matrix is computed: $M_{ij} = \sum_{k=1}^{k=N_s} M_{ij}^k$ (the more often stimuli i and j are grouped together, *i.e.* the more subjects having grouped them together, the larger M_{ij});
3. The distance matrix D is defined as: $D_{ij} = 1 - M_{ij}/N$ (the more often stimuli i and j are grouped together, the smaller D_{ij}; $0 \leq D_{ij} \leq 1$).

The values in D are "consensual" measures of perceptual distance between stimuli, *i.e.* they represent a consensus between the participants. They can be represented by an additive tree[50]: the length of the branches (connecting the leaves, or vertices, representing the stimuli) is proportional to the perceptual distance between stimuli. Branches aggregate at "nodes", enabling to consider categories at different levels of generality/inclusion. The orientation of the branches is arbitrary, only the distance along branches matters. The distances in D are fitted to an additive tree distance by means of the *Addtree* software[56].

The resulting trees are represented in Figs. 3 through 6 (in Sec. V) for each test. They take account of data from all participants (acousticians and geoscientists). On the trees, the consensual categories are identified visually as the most compact clusters of leaves/stimuli. These identified consensual categories are circled in Figs. 3 through 6, and numbered for clarity (numbers are arbitrary). Note that the identification of these clusters depends on the experimenter[22;41;45] and might slightly change for another experimenter, however we believe the visual identification of consensual categories to be robust enough for our purpose. The same analysis is conducted separately for acousticians and geoscientists (the trees from separated ensembles of participants are not shown here for brevity, but are available as supplementary material[5]. There is no major differences between the categories of both ensembles of participants in the structural properties of the categories, but there are differences in their verbal descriptions, interestingly showing that participants process along the same bottom-up constraints but conceptualized in a different way that the verbal analysis will make explicit (see Sec. A).

B Verbal comments

In this section the verbal descriptions of the consensual categories identified on the trees in Sec. A are analyzed. The method has already been applied in the literature[22;45;57;58]. Different ensembles of participants are formed and independently studied: all participants ("all"), "acousticians" and "geoscientists". In the two latter cases, the consensual categories described are not those shown in Figs. 3 to 6 but the ones computed from the total co-occurrence matrix of all acousticians and the total co-occurrence matrix of all geoscientists respectively.

As far as the verbal comments are concerned, our analysis relies on a "differential" conception of lexical semantics, that considers that the meaning of a word (as a "lexical form" or "significant") is not given by its referential value (a label on pre-existing things it refers to in the physical world), but mainly relies on a consensus between speakers to attribute its form to something or to a concept. For example in scientific discourses, scientific concepts are named by "terms", *i.e.* words whose meaning is an explicit rectification of a common sense word negotiated and accepted in the scientific community. Same words used in different communities may have different meanings[59], and individual variations in the meaning assigned to words are a common and well-known phenomenon. In other words, variations in meaning attributed by different individuals cannot be considered as errors with respect to a "true" meaning, but rather as data the analysis has to account for. It follows that it is, at least, problematic, to apply classical statistics (*e.g.* averaging, test of significant differences) to our data, and we refrain from doing so in this study. Furthermore, there is too little verbal data to undertake any statistical analysis.

In verbalization tasks, expert participants (*e.g.* expert guitarists talking about how they perceive their instrument[45]) have a specific use of lexicon, assigning to words different meanings than when they are used by non-experts or in a generic context. In those cases, it is necessary to undertake a linguistic analysis in order to identify the semantics of the words, using linguistic clues such as reformulations, definitions, *etc.*, present in the participants' discourse when they are invited to describe their sensory experience of the stimuli. In the present case, the instructions are oriented on the "objectivity" of the stimuli (participants are told that the stimuli originated from seismic recordings) that encourage a more straightforward naming by the use of simple and common words. Furthermore the constraints associated with typing prevent participants from producing long sentences they might utter if speaking[60]. Even in lack of substantial discourses as it is the case here it is worth keeping in mind the different background of the two ensembles of subjects. Due to the constraints of the task discussed above, geoscientists and acousticians are expected to produce short statements using simple, everyday-life words, with meanings slightly differing from the "common sense" meaning, as given by a dictionary. Geoscientists are expected to use common sense meaning for words referring to the sound itself but technical terms for referring to the sound as the "sound of an earthquake", whereas acousticians are expected to use technical terms for the sound itself and common sense words when referring to the sound as the sound of an earthquake. In other words, the subjects all share the same language and culture, but slight differences in educational background and training may change the way they conceptualize and therefore verbally describe sounds.

The analysis of verbal comments[6] is carried out by first organizing them according to the aspect of the stimuli they refer to. Words are grouped into semantic classes, which are labelled by a word picked from the corresponding class[7]. The comments can be split into those referring to: number of perceived *impacts*[8] in the stimuli (semantic class IMPACTS), frequency content of these impacts (semantic class FREQUENCIES[9]); *duration* or *speed* of the stimulus (semantic class DURATION / SPEED); distance from the presumed source of the stimulus (semantic class DISTANCE). Other aspects are identified and split into the classes REVERBERATION[10] (references to the part of the audio stimuli after the impacts), PERCUSSIVE (*sharp* and *clear impacts*), BOUNCING SOUND (chirp-like sound sometimes occurring after the impact), BASS BACKGROUND (*large amount of very low frequencies* before the *impact*; note that this does not describe the impacts, hence it is separated from the class FREQUENCIES), DEEP (referring to the supposed, perceived depth of the seismic event), AGGRESSIVE (words sometimes referring to a very strong high-frequency content, but always bearing a judgment of unpleasantness), CRACKLING NOISE (crackings added to the background noise), and VOLUME (the perceived loudness of the stimulus). The semantic classes as well as the words assigned to these classes are presented in Tabs. 2, 3, 4 and 5 (original French in normal font, lexical units separated by commas, English translation in italic font).

Verbal descriptions of the consensual categories defined in Sec. A are next formed by associating with them the comments made by each participant back to her/his own actual groups of stimuli. As individual groups often slightly differ from the consensual categories, a threshold has been defined: We gather in the verbal description of a consensual category the comments associated with each individual group sharing more than half of its stimuli with the consensual category. For example, if 2 out of 3 (or 3 out of 4, 3 out of 5, *etc.*) stimuli are common to an individual group and to the consensual category, the comments of this individual group are added to the verbal description of the consensual category.

The third and last step of our analysis consists of summarizing the lists of words formed above into a synthetic verbal description for each consensual category[45;57]. For this purpose words in each semantic class are gathered into subclasses labelled by simple lexical units: *e.g.* in the semantic class IMPACTS, words *impacts*, *strokes* and *shots* are assumed to refer to the same aspect of sound, summarized by the label IMPACTS; in the semantic class FREQUENCIES *bass*, *low-frequency* and *dull* are summarized by the label "*bass*". The number of comments under each subclass is then counted: A positive number (+1) is assigned to a comment showing the presence of the corresponding label, or sound aspect (*e.g. with reverberation* for the label/subclass REVERBERATION); a negative number (-1) is assigned to a comment showing the absence of the corresponding label (*e.g. without reverberation* for the label/subclass REVERBERATION). These numbers are then added together for each label. Note that absolutely no value judgement is made by the authors when using positive and negative numbers: this only depends on the arbitrary choice of the label word. Tabs. 6, 7, 8 and 9 show the presence or absence of the various sound aspects evaluated for tests T1, T2, T3a and T3b respectively. A positive (resp. negative) number indicates that the sound aspect in question is present (resp. absent) according to the majority of evaluations. The categories resulting from the comments of all participants are numbered for ease of reading.

V Results

This section first provides a general comparison of acousticians' and geoscientists' categories. Next, the results of the consensual categories and the verbal descriptions for each test are presented. The section ends with a discussion about the playback speed of the stimuli.

A How acousticians and geoscientists differ

One first result is that that acousticians and geoscientists differ both in the words they use and in the objects they refer to. This is evidence for the fact that listeners make use of their experience, memory and knowledge in their interpretation of the stimuli. Acousticians interpret, describe and name the stimuli as acoustic objects, which they are used to listening to analytically (searching for "acoustical similarities"[25]); whereas geoscientists interpret, describe and name the stimuli as cues referring to geophysical processes that make sense for them based on their knowledge (searching for "causal similarities"[25]). This difference in conceptualization can be inferred from the following results:

Number of words used Acousticians use more words than geoscientists (average of 105 and 49 words per ensemble). This indicates that acousticians are trained to describe acoustic signals and have richer vocabularies available (not only common words they share with geoscientists such as *echo* but also technical acoustic terms such as *reverberation*). Similar observations can be found in the literature[25].

Object described It is interesting to note that when giving a description of the stimuli, acousticians tend to use the word *sound* in plural rather than in singular form (62.2% of the occurrences of *sound* are in plural form), whereas geoscientists prefer to use the singular form (only 17% of the occurrences of *sound* are in plural form). This means that acousticians identify differents sounds in the one stimulus, and process in an analytical mode; whereas geoscientists process the stimulus as the sound representation of one seismic event, a meaningful unitary object for them.

Words used Geoscientists use a different vocabulary than acousticians, who are experts in describing sounds. For instance, geoscientists describe the impacts as *impact* or *sound*, as opposed to the expressions *impact*, *stroke*, *attack*, *wave*, *shot* employed by acousticians. According to their expertise, working customs and training, acousticians are more precise in their description of the spectral content, distinguishing medium frequencies from low-medium and high-medium, which geoscientists do not do. In the description of the BASS BACKGROUND, acousticians show a more accurate ability to describe the noise (*extremely low frequencies*, *sub basses*, *constant bass*, *continuous low-frequency sound*, *humming*, *<100Hz*, *throbbing*, *rumble*, *rolling*, *dull / deep / bass / low background*, *low-frequency noise*), while geoscientists are more succinct (*low-frequency background*, *rumble*, *vibrations*).

Aspects of sound Geoscientists and acousticians focus on different aspects of the stimuli. On the one hand, geoscientists do not make use of the word *balanced* (spectral balance between the frequencies), and do not write about the DISTANCE or the DEPTH of the earthquakes (presumably because those words refer to precise parameters of seismic events, which geoscientists felt that they were not able to estimate from the presented stimuli). Furthermore, the chirp-like sound sometimes occurring after the impacts (BOUNCING SOUND) has not been mentioned by the geoscientists. On the other hand, acousticians do not use the loudness of the stimuli (class VOLUME) as a categorization criterion.

Selection of evaluation criteria While acousticians evaluate systematically the same sound aspects for each test (apart from the DURATION / SPEED in T3b), geoscientists adjust their evaluation criteria depending on the test and therefore geophysical relevant parameters (BASS BACKGROUND evaluated only in T2 and

T3b, PERCUSSIVE only in T3a).

The above analysis is complicated by the sample size, which is small from a statistical standpoint, and by the slight disparity in the size of the "acousticians" and "geoscientists" ensembles. Nevertheless both ensembles produced comments that are relevant for distinguishing the consensual categories of stimuli, and hence attempt to reconstruct the categorization criteria. The next paragraphs summarize the differences between consensual categories for each test, focusing not on the similarities but on the differences, in particular when categories are contrasted according to a criterion.

B T1 (variable: event/station relative location, speed factor: 150)

All identified consensual categories in the tree in Fig. 3 group together stations which are close to one another[22]. Information about the categorization criteria is given by the verbal description associated with each consensual category (Tab. 6).

Consensual categories are first formed according to the perceived number of *impacts* in the stimuli. The physical interpretation of this number of *impacts* is quite straightforward: the longer the propagation distance between the event and the station, the more the P- and S- wave (travelling at different speeds) are temporally separated[22]. To a lesser extent the categorization relies on the event/station azimuth (*i.e.*, the propagation path between the event and the station). More in details, category 1C1 (stations closest to the event) includes stimuli where only one *impact* is perceived. Category 1C2 (grouping stations located at an intermediate distance, North-East from the epicenter) includes stimuli where 2 *impacts* are distinguished but found to be close or very close one to another. Categories 1C3 and 1C4, consisting of stations further away, respectively located South-East and North-East from the epicenter, include stimuli where 2 *impacts* are perceived to be clearly separated.

A second criterion on which categorization relies is the frequency content, which can be related to the event/station azimuth. The spectrum of stimuli of categories 1C2 and 1C4 (North-East from the event) are perceived to have more *treble* and *medium frequencies*, whereas the frequency contents of categories 1C1 (close to the event) and 1C3 (South-East from the event) are respectively more in the *medium* to *low frequencies.*

A third criterion is related to the mention of the perceived *speed* and *distance* from the sound event(s), which can be related to the event/station azimuth, just as the evaluations of the FREQUENCIES. Sounds from categories 1C2 and 1C4 (both categories North-East from the event) are perceived as *fast* and *near*, whereas stimuli from category 1C3 (South-East from the event) are perceived as *slow* and *far.* It is possible that the perceived *speed* and *distance* refer to different frequency contents: further studies dealing with more complete verbalizations may make this point clearer.

Other criteria are not relevant to distinguish categories, but do contribute to the description of specific categories. For stations North-East from the event, stimuli are described as more *percussive* when the event/station distance increases (see the *percussive* evaluation of 1C1, 1C2 and 1C4). No straightforward geophysical interpretation can be provided for now for the fact that categories 1C2 and 1C3 are often described as having a *bass background* (this aspect is evaluated by acousticians only, so it might not be a relevant parameter, geophysically speaking), while categories 1C3 and 1C4 are characterized, among other things, as having *reverberation*.

It could have been hypothesized that the event/station distance is related to the perceived intensity or loudness of the stimuli. However, the stimuli had to be normalized according to amplitude, presumably making loudness differences so subtle that they could not be used as a relevant and discriminative criterion.

It should be noted that the categories of acousticians only, geoscientists only, and all participants are quite similar. Geoscientists and acousticians are able to produce relevant (*i.e.* allowing us to discriminate between categories) evaluations of the number of impacts and the spectral content, but the other aspects are explicitely considered by acousticians only.

C T2 (variable: magnitude, speed factor: 150)

The tree in Fig. 4 does not show clusters as clearly as the tree of T1 (Fig. 3), indicating a weaker consensus (all listeners generally tend to follow a broader range of different criteria than in T1, and acousticians and geoscientists are less consistent with each other, as shown in Tab. 7). Note that the categories identified from the data of all participants are more similar to the categories made by acousticians, but simply because acousticians form the most numerous ensemble of participants.

A few consensual categories can be identified however on the tree in Fig. 4. We observe a tendency to group together stimuli coming from seismic events close to one another: Category 2C1 includes stimuli associated with seismic events East of latitude -96.8°, *E3* and *E34* (2C2) are at latitude -96.78°, *E22* and *E27* (2C3) are between latitudes -96.86°and -96.84°, *E4* and *E29* (2C5) are between -96.8°and -96.78°(see Fig. 2b). Exceptions are categories 2C4 and 2C6, which include seismic events occurring on both sides of the cluster of seismic events. In general, the categories are not related to the magnitude, with the exception of 2C1, grouping the two lowest-magnitude seismic events, E7 and E40, as well as two other seismic events of relatively small magnitude (<3.1).

The majority of the comments focuses on the presence/absence and level of the *bass background*, and stimuli in 2C1 have clearly a high level of *bass background*, whereas in all other categories the *bass background* is either very *low* or absent. Two interpretations can be proposed for the mention of the presence and level of this *bass background*. First, each audified seismogram is normalized with respect to its maximum amplitude: As a result the background level is raised for lower-amplitude events (with lower magnitude). Second, the *(bass) background*, as opposed to the number of *impacts*, depends on the medium through which the elastic waves travel. If one considers *e.g* 2C1, the former explanation can apply because the two lowest-magnitude events (E7, magnitude 2.5; E40, magnitude 2.6, see Fig. 4) are included in this category, but the latter explanation can apply as well, since the events in 2C1 have an epicenter very close one to another.

Clearly in T2 the number of *impacts* is no longer relevant for discriminating stimuli. There are indeed only a few comments about the number of impacts. Furthermore, there is no clear consensus about the number of *impacts* in the stimuli of a category (*e.g.* it is not sure whether 2C2 has stimuli with 1 or 2 *impacts*, nor is it certain whether stimuli in 2C2 or 2C6 have clearly separated or very close *impacts*). This is not surprising to us, as here the variable is magnitude, and all stimuli are associated with one cluster of events very close to one another, and one (always the same) station.

Some differences can be seen in the frequency content of the stimuli: stimuli in 2C1 and 2C6 are perceived to have more *treble* and *bass*, stimuli in 2C4 and 2C5 are perceived to have less *bass*, stimuli in 2C3 are perceived to have more *treble* and no *bass*.

An occasionally perceived and mentioned *bouncing sound* allows us to make a difference between the categories: 2C2, 2C3, 2C4 and 2C6 have stimuli with this *bouncing sound*, whereas 2C1 does not have. The geophysical origin of this *bouncing sound* is however unclear.

D T3a (variable: event/station relative location, speed factor: 250)

Tests T3a and T3b are aimed at determining whether perception changes when the audification frequency (speed factor) changes. This section and the following one (Sec. E) show the results obtained for tests with the same stimuli as T1 but with different speed factors, and Sec. F compares the results of T1, T3a and T3b in order to identify the influence of the speed factor.

As in the case of T1, each consensual category includes stimuli associated with stations located close to one another (Fig. 5). The relevance of the criterion IMPACTS (see Tab. 8) and the mapping between event/station distance and category is not as clear as in T1. Yet the event/station distance and azimuth remain the main criteria for categorization (in decreasing order of importance). More in details, there is no consensus between participants about the number of *impacts* heard in the stimuli of category 3aC1 (stations at an intermediate distance, North-East from the event), and for 3aC2 (stations at intermediate to long distance, North-East from the event) and 3aC3 (stations at close to intermediate distance from the event, spanning around it) the consensus is weak. Only stimuli of 3aC4 (stations far, South-East from the event), are perceived as certainly having 2 (*clearly separated*) *impacts*. This difficulty at identifying the number of *impacts*, and therefore at grouping according to the number of *impacts*, may be explained by the change in the stimuli playback speed: Increasing the playback speed makes the *impacts* temporally much closer, making them harder to resolve.

The descriptions of the FREQUENCIES, although hard to interpret, are quite close to those provided in T1. Sounds in 3aC2 (intermediate to long distance, North-East from the event) and 3aC3 (short to intermediate distance) have a spectrum with *treble* frequencies and no *bass*, whereas stimuli in 3aC4 (far, South-East from the event) have more *bass*.

The categories also differ according to the classes CRACKLING NOISE, BASS BACKGROUND, AGGRESSIVE. Further investigations would be necessary to precisely interpret these observations, but this goes beyond the scope of the present study. Furthermore, the relation between the event/station distance and the *percussive* aspect of sound, as identified in T1, is no longer observed.

Like in T1, the consensual categories of acousticians only and geoscientists only are quite similar to the consensual categories identified when taking into account all participants. Like in T1 again, both geoscientists and acousticians provide evaluations of the number of *impacts* and the spectral content, but the other aspects (except PERCUSSIVE) are evaluated by acousticians only.

E T3b (variable: event/station relative location, speed factor: 350)

Just like in T1 and T3a, the consensual categories group together stimuli from stations located close to one another (Fig. 6). Tab. 9 shows that the consensus about the number of impacts is also weaker than in T1.

Just like in T1 and T3a, the participants have focused first on the event/station distance, related to the "number of impacts" criterion; and then on the event/station azimuth in order to group the stimuli, related to the other criteria, as discussed below. Sounds in 3bC1 (smallest event/station distance) are perceived as having one impact or two very close one to another, stimuli in 3bC2 and 3bC5 (intermediate event/station distance) as having 2 impacts quite close one to another, stimuli in 3bC3 and 3bC4 (stations far from the event) are perceived as having two impacts. Category 3bC2 shows a weaker consensus about the perceived temporal distance between the impacts. Two sub-categories can be indeed identified in 3bC2: stations *V37A* and *V38A* lie at a greater distance from the epicenter than stations *U36A*, *V36A* and *TUL1*.

Here, again, categories differ by the perceived spectral content of their stimuli. Quite like T1 and T3a, stimuli in 3bC4 (far from the event, South-East from it) have more *bass*, stimuli in 3bC2 and 3bC3 (intermediate to long distance from the event, North-East from it) are perceived to have more *treble* and *medium* frequencies, stimuli from 3bC1 and 3bC5 (short to intermediate distance from the event) have more *treble* frequencies.

The categories also differ according to the classes VOLUME, REVERBERATION, CRACKLING NOISE, BASS BACKGROUND: these classes require further investigations.

Like in T1 and T3a, the consensual categories of acousticians only and geoscientists only are similar to the consensual categories when grouping all the participants. But contrarily to the previous tests, the geoscientists mention all criteria: Increasing even more the playback speed may have made some aspects of sound more salient and noticeable, even for participants not trained to analyze sound[11].

F Effects of the playback speed (T1, T3a, T3b)

Categories are robust with respect to the playback speed. The effect of the playback speed is indeed small in comparison with the event/station relative location, as shown in Tab. 10 by the similarity of categories in T1, T3a and T3b. In particular, subcategories {*TUL1, U36A, V36A*}, {*U38A, U37A*}, {*V38A, V37A*}, {*V35A, W35A*} and {*X37A, X38A, X39A*} remain unchanged with speed factor variations. Signals from the stations in each of these subcategories may share some strong similarities that remain to be interpreted in terms of geophysical parameters. Increasing the speed factor with respect to T1 (T3a and T3b) pushes forward the grouping of stimuli {*W36A, W37A, X35A, X36A*}. The speed factor applied in T3a favors the grouping of stimuli {*V35A, W36A, W37B, X35A*}, whereas the speed factor of T3b favors the grouping of stimuli {*W36A, W37B, X35A, X36A*} and {*TUL1, U36A, V36A, V37A, V38A*}. A higher speed factor enhances similarities between stimuli, while a lower speed factor enhances the differences between stimuli. Again, these similarities and differences have to be interpreted in terms of geophysical parameters: Different speed factors highlight different aspects of the signals, that are translated in terms of perceived similarities and differences, so possible developments of audition-based seismic data analysis methods may adapt the speed factor depending on the feature of interest in the signals.

Additionally, it can be remarked that the consensus about the number of *impacts* is weaker in T3a and T3b than in T1. This is most probably due to the speed factor (playback of the seismic time series) which is increased from T1 to T3b: this necessarily reduces the time interval between the *impacts*, making them harder to resolve/discriminate.

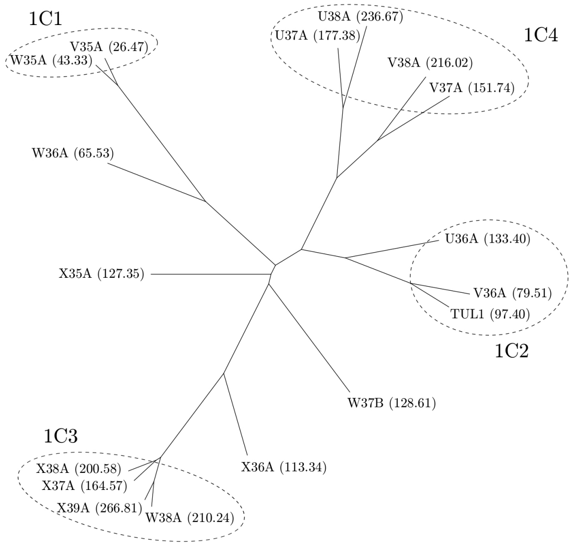

Figure 3: Additive tree for T1 (variable: event/station relative location, speed factor: 150), all participants. The identified consensual categories are numbered from 1C1 to 1C4. Leaves are labelled with the stimulus name, and the event/station distance in km in parentheses.

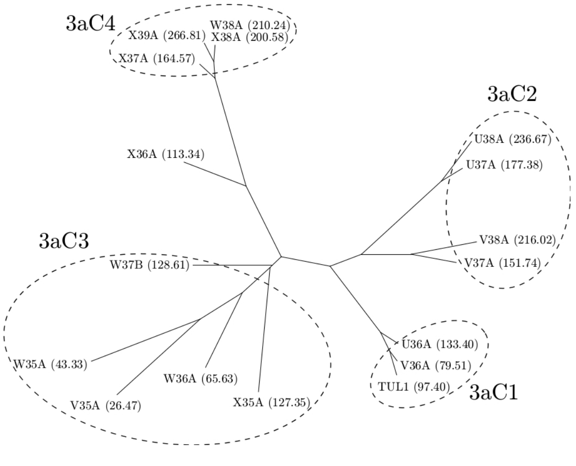

Figure 5: Additive tree for T3a (variable: event/station relative location, speed factor: 250), all participants. The identified consensual categories are numbered from 3aC1 to 3aC4. Leaves are labelled with the stimulus name, and the event/station distance in km in parentheses.

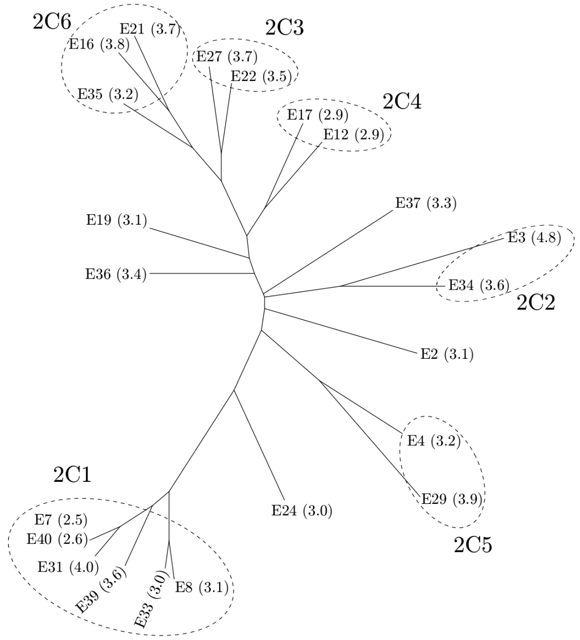

Figure 4: Additive tree for T2 (variable: magnitude, speed factor: 150), all participants. The identified consensual categories are numbered from 2C1 to 2C6. Leaves are labelled with the event number, and the magnitude in parentheses.

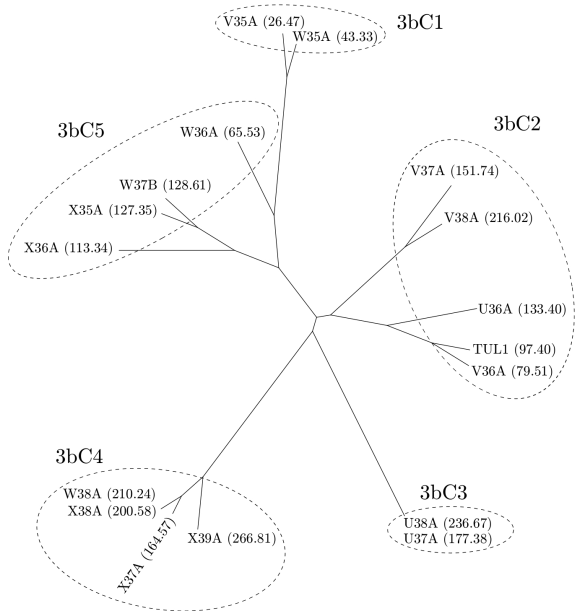

Figure 6: Additive tree for T3b (variable: event/station relative location, speed factor: 350), all participants. The identified consensual categories are numbered from 3bC1 to 3bC5. Leaves are labelled with the stimulus name, and the event/station distance in km in parentheses.

VI Acoustic descriptors

Some sound aspects have been shown in Sec. V to be particularly relevant for the perceptually- and cognitively-based categorization of stimuli. These aspects are: the number of IMPACTS (relevant for T1, T3a and T3b), the FREQUENCIES (relevant for T1, T3a and T3b), and the amount of BASS BACKGROUND (relevant for T2). Building on these observations, we searched for acoustic descriptors matching these perceptually relevant sound aspects. These descriptors are presented here. Results about the number of IMPACTS (Sec. A) and the FREQUENCIES are shown only for stimuli of T1 (similar results are found for stimuli of T3a and T3b, not shown here for the sake of brevity), and results about the BASS BACKGROUND are shown only for stimuli of T2. Note that these descriptors have been chosen after the linguistic analysis. These descriptors have also been chosen to be as simple as possible, hence they are classical and well-known acoustic descriptors. Importantly, the linguistic analysis has been conducted without any *a priori* on the nature of these acoustic descriptors (or on the fact that there was going to be any acoustic descriptors to derive).

A Number of IMPACTS and the temporal envelope

A simple way to visualize the number of *impacts* (note that we keep using the participants' wording) in the stimuli is the computation of the temporal envelope of the stimuli. The method used here is described in an article by D'Orazio *et al.*[62] (computation with 500 iterations). Fig. 7 shows the envelopes of all stimuli of T1. On the one hand, stimuli in category 1C1, perceived as having one *impact* only, exhibit a rather smooth envelope with a main *impact* (higher value). On the other hand, stimuli in categories 1C3 and 1C4, perceived as having 2 *impacts*, have a more irregular envelope with a main *impact* preceded by a lower-amplitude event. Stimuli of category 1C2, for which the number of perceived *impacts* is not clear, have intermediate envelopes: The preceding lower-amplitude event does not clearly stand out from the main *impact*. Note that the envelopes are sketchy and somewhat "quantized" due to the envelope computation algorithm and to our parameters, but this "simplified" aspect is adequate for qualitative interpretation.

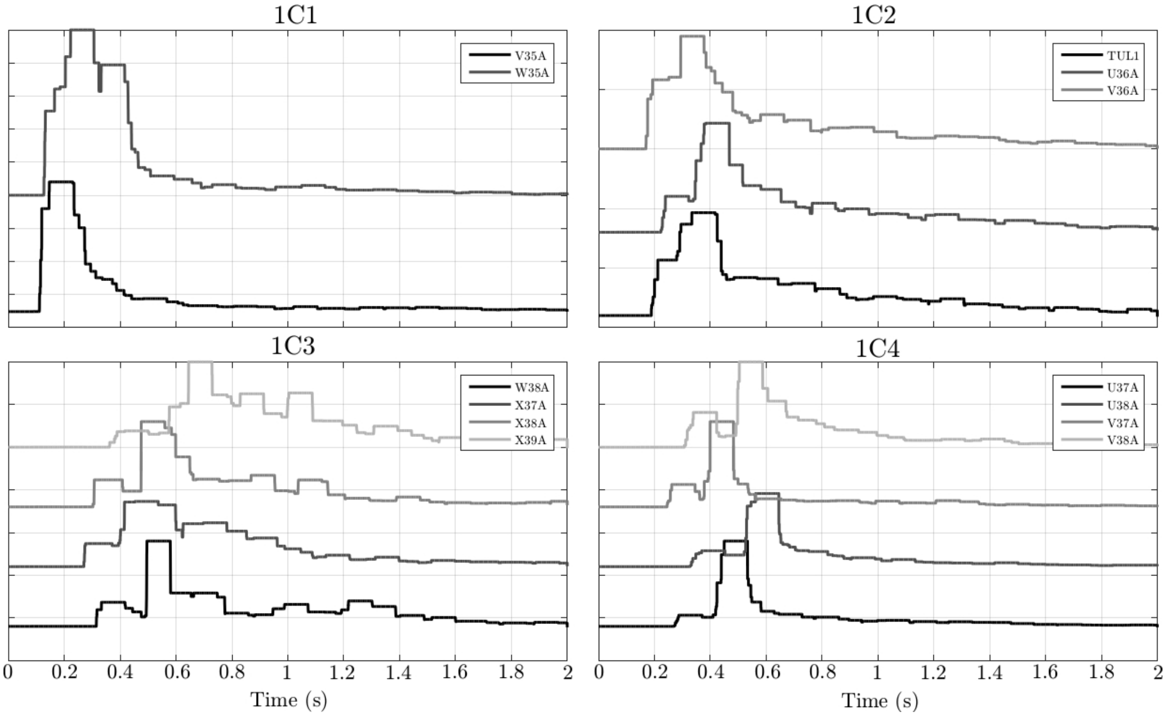

Figure 7: Envelopes of all stimuli of T1. Stimuli are grouped by consensual categories. For each category, the envelopes are shifted vertically for ease of reading.

B FREQUENCIES and the spectral centroid

Psychoacousticians usually describe the frequency content of sounds with the concept of spectral distribution of energy[63], which is classically illustrated by the spectral centroid. The spectral centroid is the “center of gravity” of the spectrum and is defined[64] as:

$$SC = \frac{\sum_{k=1}^{N} f_k a_k}{\sum_{k=1}^{N} a_k}, \tag{1}$$

where $f(k)$ and $a(k)$ are respectively the frequency and amplitude in bin k. Thus more energy in the low (resp. high) frequencies gives a lower (resp. higher) spectral centroid. Fig. 8 shows the spectral centroid computed on each stimulus of T1. Stimuli of category 1C3 have a lower spectral centroid, which is consistent with the verbalisations describing them as having more *bass*. Stimuli in categories 1C2 and 1C4 have higher spectral centroids and are judged as having more *treble* and *medium* frequencies. Stimuli in 1C1, perceived as having more *bass*, have intermediate spectral centroids.

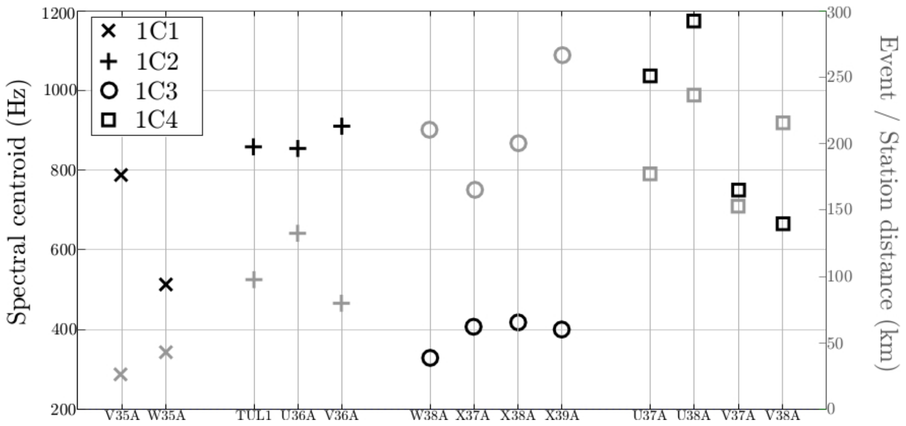

Figure 8: Spectral centroid (black) and event/station distance (gray) for all stimuli of T1. Stimuli are grouped by consensual categories.

C Bass background and the signal-to-noise ratio

The most natural descriptor of *background* noise is the signal-to-noise ratio SNR. For each stimulus of T2, the maximum value of the first 500 points of the audio signal is computed. The "noise part" ends and the "signal part" starts when the audio signal first exceeds 3 times this maximum value. Then the SNR is computed as:

$$SNR = 10log_{10}\left(\frac{\int Signal^2(t)}{\int Noise^2(t)}\right) \tag{2}$$

Fig. 9 shows the computed SNR value for each stimulus of T2. While stimuli of categories 2C2, 2C3, 2C4 and 2C5 have similar SNR values, stimuli in category 2C1 clearly have lower SNR values. This is consistent with the perception: Only stimuli of 2C1 are perceived as having a large amount of *bass background*. Furthermore, Fig. 9 confirms that higher-amplitude events correspond to signals with a higher signal-to-noise ratio.

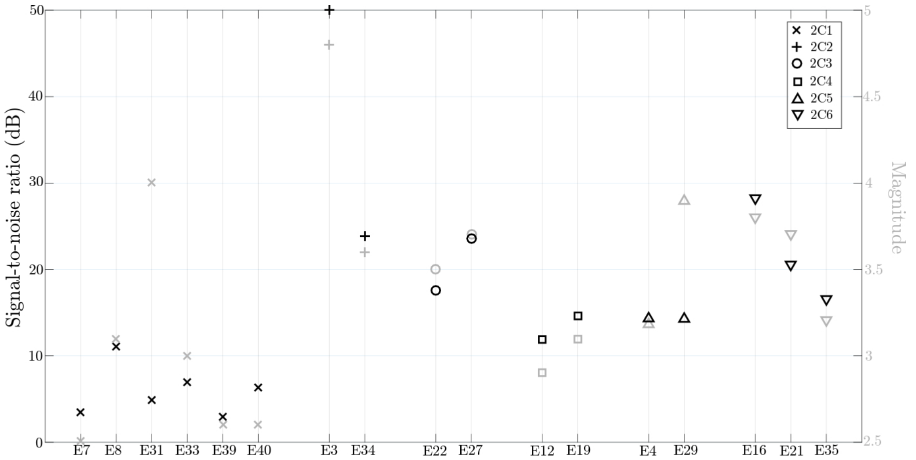

Figure 9: Signal-to-noise ratio (black) and magnitude (gray) for all stimuli/corresponding events of T2. Stimuli are grouped by consensual categories.

VII Conclusion

Expert human categorization of audified seismic signals is found to match geophysical parameters (event/station distance and azimuth), confirming previous results[22]. With respect to earlier work, limited to one of the four tests presented here, the present study contributes a thorough analysis of the categorization criteria used by participants. With the help of basic acoustical features, that have been derived and selected after the linguistic analysis, and that are shown to correspond to the verbal descriptions of stimuli, this makes the link between geophysical parameters and psychological responses clearer, and will facilitate future applications of auditory display as a didactic, data analysis and possibly research tool in seismology. Among the categorization criteria, one can mention the following important ones:

Number of *impacts* Participants primarily sort stimuli according to event/station distance: The number of *impacts* and the time difference between the *impacts* is directly related to the difference in arrival time between P- and S- waves[22]. The computed temporal envelope of the signal can help visualize, detect and confirm this temporal distance between seismic phases.

Frequency content The frequency content of the stimuli can be related to the medium through which seismic waves travel from the event to the station. The North-East and South-East regions of the investigated area differ in their elevation, local phase velocity variations, crustal structure and composition of ground, *etc.*, inducing different scattering, dissipation, attenuation, dispersion behaviors, that may act on the seismic or sonic waves as filters. As a result audified signals differ in their frequency content depending on whether they result from recordings made in one of these two regions: Signals from North-East are perceived to have more *treble* and *medium* frequencies, whereas stimuli from South-East are perceived as having more *low* frequencies; for similar event/station distances. It has been shown that the spectral centroid is a good indicator of the perceived frequency content: Stimuli perceived as having more *bass* (resp. *treble*) have lower (resp. higher) spectral centroids.

***Background* noise** T2 showed that evaluating the loudness/volume balance between the *background noise* and the main *impact* can give clues about the magnitude of the seismic event. As mentioned in Sec. B, signals were normalized according to maximal amplitude, so that the same level of background noise ends up sounding louder for smaller-magnitude events than for higher-magnitude ones. The computed signal-to-noise ratio (SNR) has been shown to be a good indicator of the perceived *background noise* level. The issue of amplitude of the sonified vs. seismic signal is further complicated by the frequency-dependence of loudness as perceived by the human ear. Frequencies that carry important seismological information might systematically be underestimated or neglected by the auditory system. In future work, a frequency-dependent amplitude correction (equalization) of sonified data that accounts for this effect is envisaged.

Other sound aspects have been pointed out by participants, and may prove relevant for the interpretation in terms of geophysical parameters, but they remain to be investigated further: In T1 stimuli from North-East (resp. South-East) stations are perceived as *fast* and *near* (resp. *slow* and *far*); acousticians notice a *bouncing sound* after the *impacts*, which probably derives fom some properties of the coda of the seismic recordings; *crackling noises* are heard in T3a and T3b, maybe related to some geophysical features. Note that in order to reduce the dynamic range of seismic signals, all stimuli had to be normalized (the solution here was to normalize each signal separately according to its maximum amplitude), and as a result loudness differences may have been much reduced, preventing the subjects to use loudness as an informative and discriminative criterion.

As stated in the introduction, this study is aimed at grasping new ideas for making new hypotheses in joining researchers' expertise from different domains: acoustics and geophysics but also psycholinguistics. Future more systematic tests will assess and generalize the listeners' sensitivity to the identified acoustic parameters and their supposed relationship to the geophysical parameters.

A logical continuation of this test is to investigate more systematically how geophysical parameters are translated into acoustical parameters and then into perceptual evaluations, using controlled-source experiments, *i.e.* using signals from laboratory experiments where the geophysical parameters (propagation medium, direction of failure, fault geometry) are known and controlled.

The categories are quite robust with respect to the playback speeds investigated here, but further tests should focus on performing a more complete parametric test of the playback speed, in order to identify whether certain playback speeds facilitate the resolution of impacts, the evaluation of spectral features, or of any other relevant features.

The investigation of the effect of training is also a good candidate for future research: participants may be trained beforehand to recognize some features, in order to get enhanced results in listening tasks. This may be applied to discrimination (*e.g.* threshold regarding the number of *impacts*) or recognition tasks (*e.g.* tell *impacts* from *echoes* resulting from reverberations on different Earth layers).

In future work, we also plan to address the issue of normalization. Normalization is needed if earthquakes of different magnitudes are investigated, due to the greater dynamics in seismic signals than in audio signals. It has been shown here that normalizing each signal according its maximum amplitude can help for the identification of some parameters (magnitude, by changing the signal-to-noise ratio), but on the other hand it can remove auditory clues for the identification of other parameters (distance, by reducing the loudness differences). Other normalization methods have to be tested (*e.g.* based on RMS value of the signal, or mapping a given seismic amplitude range to a given loudness range), depending on which seismic parameters are investigated.

The study presented here, developing a procedure aiming at employing user perception as heuristics in acoustics and geophysics, is a step towards the use of calibrated auditory display devices as complementary to or independent from visual devices. A further idea for future research is to compare human perception in two contexts, visual display (*e.g.* plot of the seismic wave) and auditory display, *e.g.* audification of the same signal: How can these two approaches to data display complement each other for the identification of some geophysical parameters?

We expect that the findings presented here will open the way to numerous applications and further developments. The verbal description of stimuli may bring new ideas for the automated analysis of seismic data. This study (and previous ones) focused on the earthquakes themselves. Further studies may focus on the "background noise", *i.e.* the soft seismic activity occurring before and after majors earthquakes. Higher speed factors (up to 1,000 or even higher) may be considered for the audification of this background noise, making auditory analysis much faster than visual analysis: one day of data can be monitored in a couple of minutes only. It is possible that auditory display techniques help understanding aftershocks, seismic swarms, or even possible earthquake precursors.

Acknowledgements

We gratefully acknowledge financial support from INSU-CNRS which made our work possible. We would like to thank François Baudin, Nicolas Bellahsen, Etienne Caylou, Delphine Chadefaux, Elia d'Acremont, Hugo Dujourdy, Augustin Ernoult, Claudia Fritz, Clément Gariel, Hugues Genevois, Ambroise Harivel, Philippe Huchon, Elsa Jauffret, Emanuel Kaestle, Félicie Korostelev, Alexis Plunder, Laurent Quartier, Alain Rabaute, Alexandre Roy, Alexandre Schubnel, Pauline Thierry, Camille Vauthrin, and Jean-Louis Waflart for taking part in the psychoacoustic tests. Thanks to Nolan Lem and Pascal Gaillard for fruitful discussion. Thanks also to Nolan Lem, Meredith Nettles, Hannah Rabinowitz, and Heather Savage, for discussions and participation in a preliminary test at the Lamont-Doherty Earth Observatory, and to Katie Keranen, Patty Lin and Nolan Lem for data and other assistance. Matthew Vaughan is acknowledged for his work in setting up an early predecessor to this study at LDEO. Lapo Boschi is grateful to Florian Dombois and Olivier Warusfel for some very interesting discussions, that inspired part of this study.

Notes

[1] Throughout the manuscript, the following words — usually synonyms — will be used depending on what is being described: "ensemble" will denote a group of subjects of similar expertise (the ensemble of acousticians and the ensemble of geoscientists); "group" will denote a group of stimuli as produced by a subject; "category" will denote a group of stimuli, when added up over the subjects; and "class" will denote a group of words of similar meaning, that is a semantic class.

[2] Signal processing refers to an object-centered conception of perception as "data driven", which is mainly a bottom-up approach of perception as information processing extracted from the stimulus; whereas signal interpretation refers to a subject-centered conception in which it is the subject who is giving meaning to the stimulation along knowledge-based, top-down processing, "down to" the selection of the relevant characteristic to be perceived and to which it is concerned.

[3] `http://www.lam.jussieu.fr/Membres/Pate/Fichiers/Sounds_ArticleEQ/`

[4] This is an English translation of the original French instructions *Nous vous demandons de procéder à un tri des extraits sonores qui vous sont présentés. Pourriez-vous regrouper les extraits qui se ressemblent et placer dans des groupes différents ceux qui vous semblent différents ? Vous faites autant de groupes que vous le souhaitez.*

[5] See supplementary material at ...

[6] This analysis has been carried out by the first author.

[7] In the following, words are italicized when they originate from the verbal comments of the participants, or capitalized when they are labels of semantic classes.

[8] In this article the words used by the participants are written in italic font; the labels of the semantic classes are written in uppercase

[9] Because this article aims at interpreting the participant's wording in terms of physics, the word *frequencies*, borrowed from the lexicon of acoustics, is preferred to the other words of the category. "*Frequencies*", as an acoustical concept, stops being a word and becomes a "term" (word with a semantic constructed by the speakers sharing a specific expertise)[61].

[10] For the same reason, the technical term "*Reverberation*" used in acoustics is chosen here to denote the semantic category.

[11] Geoscientists may have learnt and got trained in processing acoustics signals after the four experimental sessions (the type of stimulus becoming somehow more familiar and therefore being processed in a more discriminative way). This hypothesis, to be further tested, could be of interest for training geoscientists in acoustics signal processing in order to complement their previous training in reading graphical representations of seismographs

References

[1] B. Latour, "Visualisation and Cognition: Thinking with Eyes and Hands," in *Knowledge and Society Studies in the Sociology of Culture Past and Present*, H. Kuklick (editor), Jai Press vol. 6, pp. 1–40 (1986). Reprinting and revision in *Representation in Scientific Activity*, Michael Lynch and Steve Woolgar (editors), MIT Press, Cambridge Mass, pp. 19–68 (1990).

[2] B. Latour, "Visualisation and Cognition: Drawing Things Together," in *Knowledge and Society Studies in the Sociology of Culture Past and Present* 6:1–40, H. Kuklick (Ed.), JAI Press, (2003).

[3] E. von Rebeur-Paschwitz, "The Earthquake of Tokio, April 18, 1889," Nature **40**:294–295 (1889).

[4] S. D. Speeth, "Seismometer sounds," Journal of the acoustical society of America **33**(7):909–916 (1961).

[5] G. E. Frantti and L. A. Levereault, "Auditory discrimination of seismic signals from earthquakes and explosions," Bulletin of the seismological society of America, **55**(1):1–25 (1965).

[6] M.S. Karney "Natural radio (news, comments and letters about natural radio)," December 2006, available online at http://naturalradiolab.com/wp-content/uploads/2015/11/NR2006_12.pdf, retrieved on December 22nd 2016.

[7] F. Dombois, "Using audification in planetary seismology," in *Proceedings of the international conference on auditory display* (2001).

[8] F. Dombois, "Auditory seismology – on free oscillations, focal mechanisms, explosions and synthetic seismograms," in *Proceedings of the international conference on auditory display* (2002).

[9] B. Holtzman, J. Candler, M. Turk, and D. Peter, "Seismic sound lab: Sights, sounds and perception of the Earth as an acoustic space," in *Sound, Music, and Motion*, pp.161–174, Springer International Publishing (2014).

[10] D. Kilb, Z. Peng, D. Simpson, A. Michael, M. Fisher, and D. Rohrlick, "Listen, Watch, Learn: SeisSound Video Products," Seismological Research Letters: Electronic Seismologist **83**(2):281–286 (2012).

[11] Z. Peng, C. Aiken, D. Kilb, David. R. Shelly, and B. Enescu, "Listening to the 2011 magnitude 9.0 Tohoku-Oki, Japan, earthquake," Seismological Research Letters, **83**(2):287–293 (2012).

[12] C. Hayward, "Listening to the earth sing," in Gregory Kramer, editor, *Auditory Display – Sonification, audification, and auditory interfaces*, chap. 15, pp. 369–404. Addison–Wesley, Reading, MS (1994).

[13] M. Meier and A. Saranti, "Sonic exploration with earthquake data," in *Proceedings of the international conference on auditory display* (2008).

[14] A. J. Michael, "Earthquake sounds," in H. Gupta, editor, *Encyclopedia of Solid Earth Geophysics*, Springer (2011).

[15] N. Barrett and K. Mair, "Sonification for geoscience: Listening to faults from the inside," in *EGU General Assembly Conference Abstracts 16*, p. 4489 (2014).

[16] N. Barrett and K. Mair, "Aftershock: A scienceâĂŞart collaboration through sonification", Organised Sound **19**(1):4–16 (2014).

[17] T. Hermann, A. Hunt, and J. G. Neuhoff, *The sonification handbook*, Logos Verlag, Berlin (2011).

[18] E. Landi, R. L. Alexander, J. R. Gruesbeck, J. A. Gilbert, S. T. Lepri, W. B. Manchester, and T. H. Zurbuchen, "Carbon ionization stages as a diagnostic of the solar wind," The Astrophysical Journal **744**(2):100 (2012).

[19] D. Vicinanza, R. Stables, G. Clemens, and M. Baker, "Assisted differentiated stem cell classification in infrared spectroscopy using auditory feedback," in *Proceedings of the international conference on auditory display* (2014).

[20] K. V. Nesbitt and S. Barrass "Finding trading patterns in stock market data," *IEEE transactions on Computer graphics and applications* 24 (2004).

[21] M. Maeder and R. Zweifel "Downy oak: rendering ecophysiological processes in plants audible," in *Proceedings of Sound and Music Computing* (2013).

[22] A. Paté, L. Boschi, B. Holtzman, and J.-L. Le Carrou, "Categorization of seismic sources by auditory display: a blind test," International journal of human-computer studies **85**:57–67 (2016).

[23] E. Bigand and B. Poulin-Charronnat, "Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training," Cognition **100**:100–130 (2006).

[24] S. McAdams, S. Winsberg, S. Donadieu, G. de Soete, and J. Krimphoff, “Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes,” Psychological Research **58**:177–192 (1995).

[25] G. Lemaitre, O. Houix, N. Misdariis, and P. Susini, “Listener expertise and sound identification influence the categorization of environmental sounds,” Journal of Experimental Psychology: applied **16**(1):16–32 (2010).

[26] J. W. Tanaka and M. Taylor, “Object categories and expertise: Is the basic level in the eye of the beholder?,” Cognitive Psychology **23**:457–482 (1991).

[27] D. Dubois, “Categories as acts of meaning: the case of categories in olfaction and audition,” Cognitive science Quaterly **1**:35–68 (2000).

[28] D. Dubois, “From psychophysics to semiophysics: Categories as Acts of Meaning : a case study from olfaction and audition, back to colors,” in *Speaking of colors and odors. An interdisciplinary approach to cognitive and linguistic categorization of color vision and olfaction*, M. Plumacher and P. Holz (Eds.), Benjamins, Amsterdam, the Netherlands, pp. 45–119 (2006).

[29] J. Bruner, *Acts of meanings*, Harvard University press, Cambridge, MA (1990).

[30] D. Dubois, M. Coler, and H. Wörtche, “Knowledge, sensory experience and sensor technology,” *WSPC Proceedings of the CHESS II interaction Conference*, Saskatoon, Canada (2013).

[31] L. W. Barsalou, “Ad hoc categories,” Memory and Cognition **11**:211–27 (1983).

[32] L. W. Barsalou, “Grounded cognition,” Annual Review of Psychology **59**:617–645.

[33] V. Maffiolo, D. Dubois, S. David, M. Castellengo, and J.-D. Polack, “Loudness and pleasantness in structuration of urban soundscapes,” Proceedings of *InterNoise*, Christchurch, New Zealand (1998).

[34] D. Dubois, C. Guastavino, and M. Raimbault, “A cognitive approach to soundscape: Using verbal data to access everyday life auditory categories,” Acta Acustica united with Acustica, **92**(6):865-874 (2006).

[35] G. Kramer, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Perseus Publishing (1993).

[36] R. Kerr, “Geophysical exploration linking deep earth and backyard geology,” Science **340**:1283–1285 (2013).

[37] K. Keranen, M. Weingarten, G. A.. Abers, B. Bekins, and S. Ge, “Sharp increase in central oklahoma seismicity since 2008 induced by massive wastewater injection,” Science **345**:448–451 (2014).

[38] N. J. van der Elst, H. M. Savage, K. M. Keranen, and G. A. Abers,“Enhanced remote earthquake triggering at fluid-injection sites in the midwestern united states,” Science, **341**(6142):164–167 (2013).

[39] G. Ekström, “Love and Rayleigh phase-velocity maps, 5-40 s, of the western and central USA from USArray data,” Earth and Planetary Science Letters (2013).

[40] D. Dubois, *Le Sentir et le Dire : concepts et méthodes en psychologie et linguistique cognitives (Feel and Say: concepts and methods in cognitive psychology and linguistics)*, L’Harmattan, Paris, France (2009).

[41] P. Gaillard, “Etude de la perception des transitoires d’attaques des sons de steeldrums : particularités acoustiques, transformation par synthèse et catégorisation (Study of the perception of attack transients of steeldrums’ sounds: acoustic characteristics, synthesis and categorization),” Ph.D. Thesis, Université de Toulouse II - Le Mirail (2000).

[42] E. Parizet and V. Koehl, “Application of free sorting tasks to sound quality experiments,” Applied Acoustics, **73**:61–65 (2011).

[43] C. Guastavino, “Categorisation of environmental sounds,” Canadian journal of experimental psychology, **61**(1):54–63 (2007).

[44] J. Morel, C. Marquis-Favre, D. Dubois, and M. Pierrette, “Road traffic in urban areas: a perceptual and cognitive typology of pass-by noise,” Acta acustica united with acustica, **98**:166–178 (2012).

[45] A. Paté, J.-L. Le Carrou, B. Navarret, D. Dubois, and B. Fabre, “Influence of the electric guitar’s fingerboard wood on guitarist’s perception,” Acta Acustica united with Acustica, **101**(2):347–359 (2014).

[46] E. Rosch and B. B. Lloyd, *Cognition and categorization*, Lawrence Erlbaum Associates, Hillsdale, NJ (1978).

[47] D. Dubois, *Sémantique et cognition (Semantics and cognition)*, Éditions du CNRS, Paris, France (1991).

[48] J. Hampton and D. Dubois, “Psychological Models of Concepts,” in *Categories and Concepts, Theoretical views and inductive data analysis*, Van Mechelen *et al.* (Eds), Academic Press, London, UK, pp. 11–33 (1993).

[49] P. Gaillard, Website for “TCL-LabX”: http://petra.univ-tlse2.fr/tcl-labx/, retrieved on December 22nd 2016.

[50] J.-P. Barthélémy and A. Guénoche, *Trees and proximity representations*, J. Wiley, Chichester (New York) (1991).

[51] D. Dubois, D. Fleury, “From classification to cognitive categorization : the road lexicon”, in *New approaches in classification and data analysis*, E. Diday *et al.* (Eds), Springer Verlag, Berlin, Germany, pp. 25–35 (1994).

[52] A. Guénoche, “Consensus of partitions: a constructive approach”, Advances in Data Analysis and Classification **5**(3):215–229 (2010).

[53] C. Guastavino, *Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement urbain (Semantic and acoustical study of the perception of low frequencies in the urban environment)*, Ph.D. Thesis, Université Pierre et Marie Curie (2003).

[54] F. Guyot, M. Castellengo, and B. Fabre, "Etude de la catégorisation d'un corpus de bruits domestiques (Study of the categorization of a household noises corpus)," in *Catégorisation et cognition : de la perception au discours (Categorization and cognition: from perception to discourse)*, D. Dubois (ed), Kimé Paris, France, pp. 41–58 (1997).

[55] F. Guyot, *Etude de la perception sonore en termes de reconnaissance et d'apprÃľciation qualitative : une approche par la catégorisation (Study of the sonic perception in terms of recognition and qualitative assessment: a categorization approach)*, Ph.D. Thesis, Université du Maine (1996).

[56] J. Poitevineau, Website for "Addtree": http://petra.univ-tlse2.fr/tcl-labx/, retrieved on December 22nd 2016.

[57] A. Paté, *Lutherie de la guitare électrique solid body: aspects mécanique et perceptif (Lutherie of the solid body electric guitar: mechanical and perceptual aspects)*, Ph.D. Thesis, Université Pierre et Marie Curie (2014).

[58] J. Morel, C. Marquis-Favre, D. Dubois, and M. Pierrette, "A perceptive and cognitive typology of vehicle pass-by noises," in *Proceedings of Internoise 2010*, Lisbon, Portugal (2010).

[59] M. Niessen, T. Cruys, C. Cance, and D. Dubois, "Sound and Noise in Sonic Environmental Studies: Comparing Word Meaning in Discourses of Community Noise and Soundscape Research," Acta Acustica united with Acustica **99**(6):853-862, (2013).

[60] C. Cance and D. Dubois, "Dire notre expéerience du sonore: nomination et dénomination (To say our sonic experience: designation and denomination)," Langue Française **188**:15–32 (2015).

[61] D. Dubois and A. Giboreau, "Descriptors: attributes? labels? terms? names? a contribution of psycholinguistics to sensory evaluation," in H. MacFie and C. Kuesten *Developing, comparing and using consumer and sensory vocabularies*, Food Quality and Preference **17**:669–672 (2006).

[62] D. D'Orazio, S. De Cesaris, and M. Garai, "A comparison of methods to compute the effective duration of the autocorrelation function and an alternative proposal," Journal of the acoustical society of America **130**(4):1954–1961 (2011).

[63] J. M. Grey, "Multidimensional scaling of musical timbres," Journal of the Acoustical Society of America **61**(5):1270–1277 (1977).

[64] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The Timbre Toolbox: Extracting audio descriptors from musical signals," Journal of the Acoustical Society of America **130**(5):2902–2016 (2011).

Variable / test	**T1**	**T2**	**T3a**	**T3b**
Station location	variable	fixed	variable	variable
Magnitude	fixed	variable	fixed	fixed
Speed factor	fixed (150)	fixed (150)	fixed (250)	fixed (350)
Expertise of listeners	variable	variable	variable	variable

Table 1: Summary of the variable and fixed effects over the 4 tests.

	acousticians		geoscientists	
IMPACTS	*2, 1, only one, impact, strokes, waves, shots, temporal distance, close, near, separated by an intermediate duration, distinct*	2, 1, 1 seul, impact, coups, ondes, détonations, distance temporelle, proches, rapprochés, séparés par une durée intermédiaire, distincts	*2, 1, impact, near, separated*	2, 1, impact, rapprochés, séparés
FREQUENCIES	*high frequencies, treble, clear, high medium, treble medium, medium, low medium, low frequencies, low, dull, sub bass*	hautes fréquences, aigus, clair, haut médium, médiums aigus, médiums, bas médium, basses fréquences, graves, sourds, sub basses	*high frequencies, treble, clear, medium, low frequencies, dull*	hautes fréquences, aigu, clair, médium, basses fréquences, sourd
	balanced	équilibré		
DURATION	*short*	court	*short*	court
/ SPEED	*accelerated, quick, slowness, slow*	accéléré, rapide, lenteur, lent		
DISTANCE	*proximity, close, far, remoteness, remote, away*	proximité, proche, lointain, éloignement, loin, distants		
REVERBERATION	*delay, reverberation, echo*	delay, réverbération, écho	*echo, resonant*	écho, résonant
PERCUSSIVE	*percussive, abrupt, sharp, dry*	percussif, cassant, pointu, sec	*net, cutting*	net, tranchant
BOUNCING SOUND	*bouncing sound*	rebonds		
BASS BACKGROUND	*rumble, low-frequency noise, sub bass*	grondement, bruit basses fréquences, infra-basse		
DEEP	*deep*	profond		
AGGRESSIVE	*aggressive*	agressif	*aggressive*	agressif

Table 2: Words used during T1 (variable: event/station relative location, speed factor: 150), sorted by semantic classes and listener ensemble. In normal font the French original words, in italic font the English translation. Words used by the two ensembles of subjects are underlined.

	acousticians		geoscientists	
IMPACTS	*2, double, 1, impact, stroke, separated, near*	2, double, 1, impact, coup, séparés, rapprochés	*2 or 3, 2, double, 1st, impact, separated*	2 voire 3, 2, double, 1er, impact, séparés
FREQUENCIES	*high frequencies, high medium, medium, low frequencies, low, bass*	hautes fréquences, haut-médium, médiums, basses fréquences, graves, basses	*high frequencies, medium, low frequencies, dull*	hautes fréquences, médiums, basses fréquences, sourd
DISTANCE	*far*	lointain		
REVERBERATION	*echo, delay*	écho, delay		
PERCUSSIVE	*clear*	clair		
BOUNCING SOUND	*bouncing sound, sweep, swept sine*	rebond, sweep, sinus glissant		
BASS BACKGROUND	*sub bass background, continuous low-frequency sound, constant bass, humming noise, wobbly noise, <100Hz, low-frequency throbbing, bass background, extreme low-frequency, low-frequency background, rumble, rolling*	fond sonore sub, son basse fréquence continu, constante grave, bourdonnement, basses tremblantes, <100Hz, vrombissement basses fréquences, bruit de fond grave, extrême grave, bruit de fond basses fréquences, grondement, roulements	*low-frequency background, vibrations, rumble*	bruit de fond basses fréquences, vibrations, grondement
AGGRESSIVE	*aggressive*	agressif		

Table 3: Words used during T2 (variable: magnitude, speed factor: 150), sorted by semantic classes and listener ensemble. In normal font the French original words, in italic font the English translation. Words used by the two ensembles of subjects are underlined.

	acousticians		geoscientists	
IMPACTS	*2, 1, only 1, impact, attack, distinct, spaced, near, close*	2, 1, 1 seul, impact, attaque, distinct, espacés, rapprochés, proches	*2, 1, only 1, impact, sound, near*	2, 1, 1 seul, impact, son, rapprochés
FREQUENCIES	*high frequencies, low frequencies, pastel, soft*	hautes fréquences, basses fréquences, pastel, doux	*high frequencies, medium, low frequencies*	hautes fréquences, médium, basses fréquences
DURATION / SPEED	*short*	courts		
DISTANCE	*remote*	lointains		
REVERBERATION	*reverberating, resonance*	réverbérant, résonance		
PERCUSSIVE	*twangy, percussive, dry, abrupt, sharp*	claquant, percussif, secs, cassant, pointu	*net*	net
BOUNCING SOUND	*sweep, bouncing noise*	sweep, rebond		
BASS BACKGROUND	*bass / low / deep/ dull background*	bruit de fond basses fréquences / grave/ profond / sourd		
DEEP	*deep*	profond		
AGGRESSIVE	*aggressive, hissing*	agressif, stridents		
CRACKLING NOISE	*crackling noise*	grésillement		

Table 4: Words used during T3a (variable: event/station relative location, speed factor: 250), sorted by semantic classes and listener ensemble. In normal font the French original words, in italic font the English translation. Words used by the two ensembles of subjects are underlined.

	acousticians		geoscientists	
IMPACTS	*2, only 1, impact, near*	2, 1 seul, impact, rapprochés	*2, double, 1, impact, separated, near, close*	2, double, 1, impact, séparés, rapprochés, proches
FREQUENCIES	*treble, high medium, low frequencies, dull, low*	aigu, médium aigu, médium, basses fréquences, sourd, bas	*high frequencies, low frequencies, dull*	hautes fréquences, basses fréquences, sourd
	balanced – équilibré			
DURATION / SPEED			*short*	court
			quick	rapide
REVERBERATION	*echo, delay*	écho, delay	*tail*	tail
BOUNCING SOUND	*sweep*	sweep		
BASS BACKGROUND	*extremely low frequencies*	composantes extrêmement graves	*low-frequency background*	bruit de fond basses fréquences
CRACKLING NOISE	*small high-pitched crackings*	petits claquements aigus	*cracklings*	crépitements
VOLUME			*amplitude, low, volume, high*	amplitude, faible, volume, fort

Table 5: Words used during T3b (variable: event/station relative location, speed factor: 350), sorted by semantic classes and listener ensemble. In normal font the French original words, in italic font the English translation. Words used by the two ensembles of subjects are underlined.

		acousticians					geoscientists					ALL			
												1C1	1C2	1C3	1C4
		V35A W35A W36A	TUL1 U36A V36A	W38A X36A X37A X38A X39A	U37A U38A V37A V38A	W37B X35A	V35A W35A	TUL1 V36A W36A	W38A X37A X38A X39A	U37A U38A V37A V38A	W37B X35A X36A	V35A W35A	TUL1 U36A V36A	W38A X37A X38A X39A	U37A U38A V37A V38A
IMPACTS	2 clearly/well separated			2	5				2	1	1			7	6
	2 impacts		4	5	4	2	1	2	4	5	5	1	7	9	9
	2 close to one another	1	1		1			3			1	1	3		1
	2 very close to one another	1	4			2						1	4		
	1 or 2 very close to one another		1				2	2				2	3		
	1 impact	9	1				4					10	1		
	have different pitch				1										1
	2nd impact louder									1	1				1
FREQUENCIES	treble		5	-1	2	1	1	3	-1	3	1	1	8	-2	5
	high medium		2		1								2		1
	medium	3	1		3	1		2		2	1	3	3		5
	low medium	1										1			
	bass	-1	2	11	-2				5		2	-1	2	16	-2
	balanced		1								1		1		
DURATION / SPEED	short	1								1		1			1
	fast		2	-2	2	1							2	-2	2
DISTANCE	far			-1	3	-1							-1	3	-1
OTHERS	Reverberation	-1		1		3	1	2	1	2	1			2	2
	Percussive	2	2	1	3	2				2		1	2		5
	Bouncing sound	1	-1			1						1	-1		
	Aggressive		2										2		
	Bass background		2	2									2	2	
	Deep			1										1	

Table 6: T1 (variable: event/station relative location): Summary of the verbal description of the consensual categories of stimuli, grouped by semantic classes. The numbers indicate the cumulative number of evaluation for each verbal descriptor (line). A positive (resp. negative) number indicates the presence (resp. absence) of the sound aspect. The categories resulting from the comments of all subjects are numbered from 1C1 to 1C4.

		acousticians					geoscientists					ALL					
												2C1	2C2	2C3	2C4	2C5	2C6
		E7 E8 E31 E33 E39 E40	E3 E34	E12 E17 E19 E36	E16 E21 E22 E27 E35	E29 E4 E37	E7 E8 E31 E33 E39 E40	E3 E16	E22 E36	E2 E19 E34 E37	E21 E35	E7 E8 E31 E33 E39 E40	E3 E34	E22 E27	E12 E17	E4 E29	E16 E21 E35
IMPACTS	2 clearly/well separated	1		1	1		2			1		3		1			1
	2 impacts					2	3	3	2	3	3	3	1	2		1	3
	2 very close to one another		1		1								1	1			1
	1 impact			1									1				
	1st impact shorter								1	1	1						1
	1st impact weaker							1									
FREQUENCIES	treble	1			1		2		2	2	1	3		1			2
	high medium					1											
	medium			1						1							
	bass	2		2		-2	1	1				3		-1	-1	-1	1
DISTANCE	Far					1										1	
OTHERS	Bouncing sound	-1	1	5	7	-1						-1	2	6	5		7
	Percussive	2				1						2				1	
	Reverberation				1	-1								1		-1	1
	Bass background	13	-5	-6	-6	-5	9	-1		-3	-1	22	-7	-4	-5	-7	-4
	Aggressive					1											

Table 7: T2 (variable: magnitude): Summary of the verbal description of the consensual categories of stimuli, grouped by semantic classes. The numbers indicate the cumulative number of evaluation for each verbal descriptor (line). A positive (resp. negative) number indicates the presence (resp. absence) of the sound aspect. The categories resulting from the comments of all subjects are numbered from 2C1 to 2C6.

		acousticians					geoscientists				ALL			
											3aC1	3aC2	3aC3	3aC4
		TUL1 U36A V36A	U37A U38A V37A V38A	V35A W35A	W36A W37B X35A X36A	W38B X37A X38A X39A	TUL1 U36A V36A	U38A U37A V38A V37A	V35A W35A W36A	W38A X36A X37A X38A X39A	TUL1 U36A V36A	U38A U37A V38A V37A	V35A W35A W36A W37B X35A	W38A X37A X38A X39A
IMPACTS	2 clearly separated		1			2						1		2
	2 impacts		1	2	2	1		2	5	3	1	3	7	4
	2 close one to another			2	1						1		2	
	2 very close one to another		3		2			1			1	4		
	1 or 2 very close one to another			1			2				2		1	
	1 impact	2		1	1		1	1			1	1	1	
FREQUENCIES	treble		4	4	2	1		1	3		1	5	7	1
	medium							1	1			1	1	
	bass		-2	-2		3	1		1	3	1	-2	-1	7
DURATION / SPEED	short				1								1	
DISTANCE	far			1									1	
OTHERS	bass background			-2	3								-2	
	deep			-1		1					1		-1	1
	aggressive			2									2	
	crackling noise	1	1	1	1	-1					1	1	1	-1
	bouncing sound			1	2	1					1		1	1
	Percussive		2	5					1	1		2	6	1
	reverberation					2								2

Table 8: T3a (variable: event/station relative location, speed factor: 250): Summary of the verbal description of the consensual categories of stimuli, grouped by semantic classes. The numbers indicate the cumulative number of evaluation for each verbal descriptor (line). A positive (resp. negative) number indicates the presence (resp. absence) of the sound aspect. The categories resulting from the comments of all subjects are numbered from 3aC1 to 3aC4.

		acousticians						geoscientists					ALL				
													3bC1	3bC2	3bC3	3bC4	3bC5
		V35A W35A W36A	TUL1 U36A V36A	V37A V38A	U37A U38A	W38A X37A X38A X39A	W37B X35A X36A	V35A W35A	TUL1 V36A W37B X35A	U36A V37A V38A	U37A U38A	W38A X37A X38A X39A	V35A W35A	TUL1 U36A V36A V37A V38A	U37A U38A	W38A X37A X38A X39A	W36A W37B X35A X36A
IMPACTS	2 clearly separated									1	1	2		1	1	2	
	2 impacts		1	1	2	1		1	2	3	3	3	1	2	5	4	3
	2 close to one another								1								1
	2 very close to one another		1				1	1	1			1	1	1			2
	1 or 2 very close	1					1										1
	1 impact	2	1					2					4				
FREQUENCIES	treble		1		3			2	3	3	2	1	2	2	5	1	3
	high medium		1	1		-1	-1							1		-1	-1
	medium	1	3	2	1									1	1		
	bass		1	-1	-1	3						4		1	-1	7	-1
	balanced		1				1							1			1
DURATION / SPEED	short							1	1	1	1	1	1		1	1	1
	fast							1	1	1			1	1			1
OTHERS	bouncing sound	1															
	reverberation	-1	3	3		3	-1	-1	1	1	1	1	-1	2	1	4	
	bass background					1			1	1	-1				-1	1	1
	crackling noise		2	-1	2	-1	2	1	1	1	1	-1		2	3	-2	2
	volume							-1	2	1	1	-1	-1	1	1	-1	2

Table 9: T3b (variable: event/station relative location, speed factor: 350): Summary of the verbal description of the consensual categories of stimuli, grouped by semantic classes. The numbers indicate the cumulative number of evaluation for each verbal descriptor (line). A positive (resp. negative) number indicates the presence (resp. absence) of the sound aspect. The categories resulting from the comments of all subjects are numbered from 3bC1 to 3bC5.

U36A	T1, T3a, T3b															
U37A																
U38A			T1, T3a, T3b													
V35A																
V36A	T1, T3a, T3b	T1, T3a, T3b														
V37A	T3b	T3b	T1, T3a	T1, T3a		T3b										
V38A	T3b	T3b	T1, T3a	T1, T3a		T3b	T1, T3a, T3b									
W35A					T1, T3a, T3b											
W36A					T3a				T3a							
W37B					T3a				T3a	T3a, T3b						
W38A																
X35A					T3a				T3a	T3a, T3b	T3a, T3b					
X36A										T3b	T3b		T3b			
X37A												T1, T3a, T3b				
X38A												T1, T3a, T3b			T1, T3a, T3b	
X39A												T1, T3a, T3b			T1, T3a, T3b	T1, T3a, T3b
	TUL1	U36A	U37A	U38A	V35A	V36A	V37A	V38A	W35A	W36A	W37B	W38A	X35A	X36A	X37A	X38A	X39A

Table 10: Comparison of T1, T3a and T3b. For each pair of stimuli we give in the corresponding cell the tests during which they are put in the same consensual category.