On the Perception of Audified Seismograms

Lapo Boschi, Laurianne Delcor, Jean-Loic Le Carrou, Claudia Fritz, Arthur Paté, Benjamin Holtzman

On the perception of audified seismograms

Lapo Boschi[1,5], Laurianne Delcor[2,6], Jean-Loïc Le Carrou[3], Claudia Fritz[3], Arthur Paté[4], and Benjamin Holtzman[4]

[1]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Institut des Sciences de la Terre de Paris (iSTeP), 4 place Jussieu 75005 Paris, France

[2]Laboratory of Vibration and Acoustics, 69621 Villeurbanne, France

[3]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Institut Jean Le Rond d'Alembert, équipe LAM, 4 place Jussieu 75005 Paris, France

[4]Lamont-Doherty Earth Observatory, Columbia University, New York, U. S. A.

[5]ALSO AT: Sorbonne Universités, UPMC Univ Paris 06, CNRS, Institut Jean Le Rond d'Alembert, équipe LAM, 4 place Jussieu 75005 Paris, France

[6]ALSO AT: Airbus Helicopters, France

May 22, 2017

Abstract

Recordings of the Earth's oscillations made by seismometers, following earthquakes or other geophysical phenomena, can be made audible by simply accelerating and playing them through an audio reproduction system. We evaluate quantitatively the possibility of using such acoustic display of seismic data for practical applications. We first present to listeners examples of two categories of data, based on geophysical parameters (the geometry of the seismic fault; the terrain–oceanic or continental–sampled by the propagating seismic wave) that are not revealed to them. The listeners are then asked to associate each of a set of audified seismograms, that are presented to them binaurally, to either one of the two categories. After this exercise, they are asked to define the features of audified signals that helped them in completing this task. A subset of the listeners undergo a training session, before taking one of the tests for a second time. While the number of listeners is too small for a definitive statistical analysis, our results suggest that listeners are able, at least in some cases, to categorize signals according to all the geophysical parameters we had chosen. Importantly, we clearly observe that listeners' performance can be improved by training. Our work opens the way to a number of potentially fruitful applications of auditory display to seismology.

Introduction

Auditory display, or "sonification" of scientific data has been applied successfully to research topics in several disciplines [e.g. *Cowen*, 2015]. Seismic data analysis naturally lends itself to audification: a particularly simple form of sonification which consists of accelerating seismic signals (whose frequency is lower than that of audible sound) before playing them through an audio reproduction system. Auditory display of seismic data was first explored during the Cold War, when the ability to distinguish underground nuclear tests from natural earthquakes acquired a political relevance [*Speeth*, 1961; *Frantti and Leverault*, 1965; *Volmar*, 2013]. Audification was eventually discarded, in this context, in favour of seismic-array methods [*Volmar*, 2013]; in recent years, however, it has been revived by seismologists, mostly for purposes of teaching and dissemination [e.g. *Dombois and Eckel*, 2011; *Kilb et al.*, 2012; *Peng et al.*, 2012; *Holtzman et al.*, 2014; *Tang*, 2014]. Our own experiments [*Paté et al.*, 2016, 2017] have convinced us that it is a valuable and inspirational tool for the analysis of seismic data in many contexts. We suggest that it might also soon find more specific, effective research applications.

This study attempts to contribute to the quantitative analysis of the human auditory system's response to audified seismic data. As researchers peruse data via auditory display, the implicit assumption is made that they are capable of recognizing patterns and completing some related tasks by hearing. We question this assumption for the case of audified seismic data, and thus begin to evaluate what can be achieved by audification that is not already implemented through "traditional" techniques in seismic data analysis. Both the early work of *Speeth* [1961] and the recent efforts by our group [*Paté et al.*, 2016] indicate that listeners can detect meaningful clues in audified seismic signals, and thus categorize the signals according to such clues. *Paté et al.* [2016] showed that the categories formed by the listeners can be associated with several geophysical parameters, but could not entirely distinguish the effects of individual parameters (e.g., source-receiver distance, geological properties of the terrain at the receiver and between source and receiver, etc.) from one another. We present here a different approach to the analysis of audified data: listeners are asked to complete a constrained-, rather than free-categorization task, on two sets of data, each controlled by a single geophysical parameter (Earth structure in the area where the recorded seismic waves propagate; focal mechanism of the source). The listeners' performance in auditory analysis is compared with their performance in a similar task, completed via visual analysis of analogous data. We consider the visual analysis of a plot to be a "traditional" task that most individuals with some scientific background are, to some extent, familiar with. Visual analysis serves here as a reference against which results of auditory tests can be compared, and, accordingly, its results are not analyzed in as much detail. Listeners are then briefly trained, and the auditory test repeated after training, with a general improvement of test scores. Finally, listeners are asked to explain the criteria they followed to categorize the data, and their description is compared with quantitative parameters computed from the data.

Database

The work of *Paté et al.* [2016] evidenced the difficulty of disentangling the influences of different physical parameters on the seismic signal (e.g., source-receiver distance, properties of the source, geology at the receiver location, geology between source and receiver). We compiled two new audified seismic data sets, each designed to emphasize the role of one specific parameter. Both data sets only included events of magnitude between 6 and 8, with focal depths estimated by IRIS between 20 and 40 km, and recorded at epicentral distances between 4000 and 6000 km. The scale lengths under consideration are therefore different from those of *Paté et al.* [2016], who used recordings of a magnitude-5.5 event made no more than a few hundred km from the epicenter. All events contributing to either data set occurred between August 9, 2000, and April 18, 2014.

The first data set (DS1) is limited to source mechanisms of the strike-slip type, with magnitude between 6 and 7, and the propagation path (approximated by an arc of great circle) is required to lie entirely within either a continental or oceanic region. Fig. 1 shows that events in DS1 are located along the Pacific coast of Mexico and in California, while stations can be in North America (continental paths), on ocean islands throughout the Pacific ocean, in Chile or on the Alaskan coast (oceanic paths). It is well known that a seismic waveform is affected in many ways by the properties of the medium through which the wave propagates before being recorded. Based, e.g., on recent work by *Kennett and Furumura* [2013] and *Kennett et al.* [2014] on waveform differences across the Pacific Ocean, we anticipated that the bulk properties of oceanic vs. continental crust and lithosphere would result in profoundly different seismograms and audified signals. We expected this ocean/continent dichotomy to be far more important than other parameters in characterising traces in DS1, and we assumed that it would also guide the subjects' response to the corresponding audified signals.

The second data set (DS2) is limited to continental propagation paths, but includes both strike-slip and thrust events of magnitude between 6 and 8 (Fig. 2). We expected differences between signals generated by strike-slip and thrust events to be more subtle, and harder to detect, whether visually or aurally. Again, all sources contributing to DS2 are in South-Western North America; stations are distributed throughout Canada and the United States, and, in one case, in the Caribbean. Earthquake mechanisms were obtained from the Global Centroid Moment Tensor Project (see "Data and resources" section).

Approximately 500 seismograms meeting the requirements of DS1 and DS2 were downloaded from the IRIS database (see "Data and resources" section) but only traces showing, at a visual analysis, a relatively high signal-to-noise ratio were kept. As a result, DS1 includes 23 "continental" and 23 "oceanic" signals, while DS2 includes 52 strike-slip and 52 thrust signals. No filtering or instrument-response correction was applied to the data.

The sampling rate of all downloaded seismic traces is 50 Hz. The duration of traces to be audified is 8000 s, starting 1800 s before the P-wave arrival as found in the IRIS catalog, and including the most significant seismic phases and most or all of the coda. Time is sped up by a factor of 1200, selected so that all frequencies present in the seismic traces are mapped into the audible range [*Holtzman et al.*, 2014]. Each sonified signal was normalized with respect to its maximal value. The resulting, "audified," 6-s-long signals are turned into Waveform Audio File Format (WAV) files via the Matlab function audiowrite. Their spectra show most energy between 20 and 600 Hz.

Experiments

All experiments (table 1) were conducted in an acoustically dry room (i.e., not entirely anechoic, but with very little reverberation of sound). The subjects played audified seismic signals on a laptop computer via a Matlab-based software interface, and listened to them through an audio card and closed headphones with adjustable volume. Some tests involved the visual, rather than acoustic display of the signals, which was also implemented with the same interface: seismograms were plotted in the time domain as in Fig. 3 (albeit with a longer time window, extending from $\sim$0 to $\sim$20000 s) and subjects had no way to modify the plots' size or format. We provided each subject with all necessary instructions at the beginning of the test, so that the subject would be able to take the test autonomously. The subjects knew that the signals were originated from seismograms; at the beginning of the test, they were told that all signals would belong to one and only one out of two possible "families," named A and B. By assigning "neutral" names to data families, and providing no information as to their nature, we minimize the bias that might be caused by a specialized (geophysical) knowledge/understanding of the data. After each test, subjects were asked to briefly explain the criteria they had followed in responding to it. They typed their answers on the computer used for the test.

All subjects were researchers, faculty, and graduate and undergraduate students with backgrounds in Earth sciences (referred to in the following as “geoscientists”), room or musical acoustics (“acousticians”) or applied physics/engineering (“physicists”).

Constrained categorization without training

In a first suite of experiments, families A and B were each defined by three examples, that subjects listened to or looked at before starting the test. Each of the three example audified signals could be listened to three times at most. Visual examples were plotted on the screen, and could be looked at for no more than three minutes before starting the test. All subjects were given the same examples. The subjects were then exposed to 40 unknown signals; after listening to/looking at each signal, they selected whether it belonged to family A or B; no other answer was possible. Each auditory signals could be listened to three times at most; plots were visible on the screen for 5 seconds. The subjects’ selections were recorded by the software interface.

Importantly, this approach is profoundly different from that of *Paté et al.* [2016], who asked subjects to form as many categories as they wanted according to their own criteria [*Gaillard*, 2009]. It is also different from “paired comparison,” where a subject is presented with two stimuli, and must choose which one belongs to which of two categories. We have explored the latter approach in preliminary tests with few subjects, who all obtained extremely high scores: this strengthened our hypothesis that the geophysical parameters we had selected (propagation path and orientation of the fault) do map into audible acoustic properties of the corresponding audified signals. We considered, however, that a paired-comparison test does not resemble any real task in seismic data analysis, and discared this approach in our subsequent experiments.

Auditory and visual display of DS1 (oceanic vs. continental paths)

In a first experimental session, 35 subjects (13 women, 22 men), aged between 18 and 61, took two tests involving data from DS1. The group included 18 acousticians, 9 geoscientists and 8 physicists. 40 signals were evaluated visually in one test, and their audified counterparts were listened to in another.

As explained in the "Database" section above, we made the hypothesis that data belonging to DS1 would tend to be categorized according to the terrain sampled by the propagation paths. Signals corresponding to oceanic propagation paths were presented as examples of family A, and "continental" signals as examples of B. In the following, we loosely speak of "correct" answer whenever a subject associates to family A an "oceanic" signal, or to family B a "continental" one. Exactly half of the signals in this experiment correspond to oceanic propagation paths, the other half to continental ones. The signals were the same for all subjects, but their order was random, changing at each realization of the experiment.

The average percentage of correct answers (average "score") in this first experiment amounts to 78% for the visual test, and 63% for the auditory one. All scores are summarized in the histograms of Fig. 4a and b. We suspect the very low scores of two outliers (one per test) to have been caused by a misunderstanding of the intstructions which resulted in the subjects swapping families A and B.

For the sake of comparison, we consider the case of entirely random answers, i.e., the human subject is replaced by an algorithm that generates random yes/no answers, or answers are given by tossing a coin. In this "null hypothesis," test scores are controlled by the cumulative binomial distribution [*Press et al.*, 1992, e.g.]: each signal listened to can be treated as an independent "trial", with a success probability of 50%. Fig. 4a shows that in the first auditory test about one out of three subjects scored above the 99% confidence level as defined through the cumulative binomial distribution: in other words, the probability that a subject would obtain (at least) such score by giving random answers is less than 1%. It is thus probable that some of the best-scoring subjects have identified a real difference between signals that they classified as belonging to families A and B. Given how we constructed the two families (see "Database" section), it is also reasonable to infer that the auditory clues identified by the subjects are directly related to the effects, on seismic waveforms, of wave propagation through oceanic vs. continental crust.

Our data are not numerous enough for the histograms in Fig. 4a,b to clearly suggest specific statistical distributions. By visual inspection of Fig. 4a one might speculate that the distribution of auditory test scores is bimodal, with one peak around 50% corresponding to the null hypothesis, and another peak around 70% reflecting the performance of subjects who did find meaningful clues in the signals.

Scores in the visual test (Fig. 4b) were generally quite high, and higher than for the auditory test. This indicates that, at this point, visual analysis of the data might be a more effective way to complete the task of categorizing DS1 data.

Auditory and visual display of DS2 (thrust vs. strike-slip faults)

Of the subjects who took part in the experiment described in the previous section, 27 (15 acousticians, 7 physicists, 5 geoscientists; 7 women, 20 men) also participated in a second session, involving 40 signals from DS2. Half of the signals were originated from the strike-slip faults, the other half from the thrust faults shown in Fig. 2b. Again, each subject took an auditory and a visual test, with average scores of 52% and 62%, respectively. The results of both auditory and visual tests are illustrated in Fig. 4c,d.

Comparison with the null hypothesis shows that the probability of achieving (at least) the average score associated with the visual test by selecting the answers randomly was relatively low (<10%); we infer that at least some subjects are likely to have found visual clues in plotted seismograms. Conversely, the probability of achieving (at least) the average score obtained in the auditory test by giving random answers was about 40%. Too high for the average observed score to be considered significant. It might be guessed that the one subject who achieved a score of 75% might have found auditory clues in the signals, but overall the test cannot be considered a success.

Constrained categorization with training

17 subjects (10 acousticians, 4 physicists, 3 geoscientists; 4 women, 13 men), who had already participated in both the constrained-categorization experiments, accepted to undergo a training session, followed by an auditory test analogous to those described above. The new exercise was conducted on data from DS2, only half of which were employed in our previous experiments. Data included in the final test had not been listened to in the course of the training session. The goal of this experiment is to determine whether performance in auditory analysis of seismic data can in principle be improved by training: this is determined below by comparison with performance in a similar task before training. It is therefore not strictly necessary to compare the results against those of visual analysis of the same data, and accordingly the visual test was not repeated.

Training

Subjects were trained [e.g. *Thorndike*, 1931; *Speeth*, 1961] by means of a software interface similar to that used in the actual tests. They first listened to three examples of each family, as before the previous test. They were then presented with up to 24 audified signals in the same way as previously. Half of these signals originated from thrust, the other half from strike-slip faults. Half had been listened to during the previous experiment, half were entirely new. The order in which the signals were presented was random. Upon hearing each signal, subjects were asked by our software interface to evaluate whether it belonged to family A or B. After giving an answer, they were immediately notified whether or not it was “correct” (i.e., consistent with our hypothesis), by the on-screen messages “you have identified the right seismological family” (*“vous avez identifié la bonne famille sismologique”*) and “that is not the right seismological family” (*“ce n’est pas la bonne famille sismologique”*), respectively. If a subject had a perfect score after listening to the first 16 sample signals, the training session would end.

Auditory display of DS2 after training

After a brief pause, all subjects who undertook the training session stayed for a final test. 36 signals were randomly picked from DS2. Half of the picked signals had to be from thrust, half from strike-slip faults. Half had to belong to the pool of signals listened to in the test of section “Auditory and visual display of DS2”.

The histogram in Fig. 5 shows that scores are generally higher now than when categorizing signals from DS2 before training (Fig. 4c). Only 4 out of 17 subjects did not improve their score at all. In the null hypothesis, with 36 trials, the probability of achieving a score of at least 69.4% (24 correct answers out of 36) is about 1%: 6 out of 17 subjects scored 70% or more, and we infer that at least some of those 6 learned to recognize relevant auditory clues in the data. Albeit small, these figures appear more significant if one considers that only one brief training session was undertaken.

Identifying audio features relevant to categorization

At the end of a test, the subject was asked to briefly explain the criteria followed to categorize the signals, via the on-screen message: "according to what criteria have you associated family A and B to the signals?" (*"sur quel(s) critère(s) avez vous attribué la famille A ou B aux signaux ?"*). The subject could answer by typing some comments through our software interface.

Given the difficulty of an exhaustive semantic study of the resulting data [*Paté et al.*, 2017], we only give here a preliminary, simplistic analysis of a subset of the recorded comments. Our goal in this endeavour is to identify some of the auditory clues that lead subjects to make their choices. We focus on the subjects whose scores were highest, as the criteria that guided them are probably related to the geophysical parameters that defined our families of signals.

Comments on DS1

We first analyze the comments made by 5 subjects (2 acousticians, 2 geoscientists, and one physicist) who all achieved scores ≥80% in discriminating audified seismograms corresponding to oceanic vs. continental paths (DS1).

Table 2 shows a number of reoccurring suggested clues, namely: the presence of what the subjects identify as "background noise," and its timbre; the duration of what is considered by the subjects to be meaningful signal;

the identification of “echos” in the signal. These features can in principle be associated to quantities calculated by seismic data analysis.

First of all, it is relatively easy to identify the onset of an earthquake recording on a seismogram (i.e., the P-wave arrival), and it is then reasonable to define as background noise the signal recorded before such arrival. In all our recordings, the first 500 samples clearly precede the arrival of the main signal and we accordingly identify them as noise. We define the beginning of the seismic signal as the first recorded sample whose amplitude is at least three times larger than the largest amplitude found within the 500 noise samples. Let n_S denote its index. The signal-to-noise ratio (SNR) in decibels can then be estimated, based on the mean amplitudes of signal and noise, by the formula

$$\mathrm{SNR} = 10 \log_{10} \left[\frac{\sum_{i=n_S}^{N} s^2[i]}{\sum_{i=1}^{500} s^2[i]} \right], \tag{1}$$

where $s[i]$ is the amplitude of the i-th sample, in a recording that consists of N samples total. We compute the SNR of all signals in DS1, and find (Fig. 6a) that continental paths tend to be associated with higher SNR values than oceanic paths. This statistical result is in qualitative agreement with the subjects comments.

We evaluate the “timbre” of background noise by taking the Fourier transform of the first 500 samples only. Fig. 6b shows the distribution of frequency values corresponding to the highest spectral peak in the resulting Fourier spectrum: whether the terrain traversed by the propagating seismic wave is oceanic or continental does not appear to affect significantly the frequency content of noise.

We next attempt to quantify the duration of meaningful seismic signal which the subjects believe to have recognized in their listening experiences: after the main, high-amplitude interval that includes body- and surface-wave arrivals, the peak amplitude of all our signals decreases until it becomes as low as the peak amplitude of background noise. For each seismogram, we find the latest sample whose peak amplitude is as large as 10% of its maximum recorded value for that seismogram; we then measure the length of the time interval that separates it from the maximum-amplitude sample, and define it as the duration of seismologically meaningful signal. Fig. 6c shows how such values are distributed for signals associated with oceanic vs. continental propagation paths, and indicates that oceanic signals are, according to our definition, longer than continental ones.

Finally, echos can be identified by visual analysis of a seismogram's envelope. We calculate the envelopes of all our audified seismograms, and take the averages of all oceanic-path and all continental-path envelopes. In analogy with *Paté et al.* [2017], the envelope is defined as suggested by *D'Orazio et al.* [2011]: starting with i coinciding with the index of the last sample in a signal, if sample $i-1$ exceeds sample i, then the value of sample $i-1$ is saved as the i-th entry of the envelope; the procedure is iterated for the preceding sample, until the entire trace is processed [*D'Orazio et al.*, 2011, figure 5]. The results of this exercise, illustrated in Fig. 7, show that (i) the amplitude of oceanic-path signal is generally larger than that of continental-path signal; (ii) the oceanic-path signal is characterized by a number of high-amplitude peaks that are not visible in the continental-path one; (iii) the large-amplitude portion of the signal lasts longer in oceanic-path than continental-path signal. While the standard deviations of both envelopes are not shown in Fig. 7 in the interest of readability, these inferences are confirmed even if the standard deviation is taken into account. We note that the standard deviation of the oceanic-path envelope is larger than the continental-path one. Observation (ii) reflects several comments made by the subjects (Table 2).

Comments on DS2

The four subjects who achieved the highest scores (> 55%) without training are combined with the four who achieved the highest scores (> 72%) after training, resulting in a group of eight subjects whose verbal comments are summarized in table 3. The group includes three geoscientists, three acousticians, and two physicists. Two of the subjects in this group were also in the group discussed in the previous section.

Table 3 shows that, despite some contradictory comments, most subjects find strike-slip-fault signals to be characterized by a relatively weak "first arrival" followed by a high-energy coda, while on the contrary they associate thrust events with a strong first arrival followed by a weaker coda. This seems to be consistent with the average envelopes of Fig. 8, where (i) the initial peak is clearly identifiable for both families and is roughly twice as high in the inverse-fault case, with respect to the strike-slip-fault one, while (ii) the later inverse-fault signal is of higher amplitude than its strike-slip-fault counterpart, with a $\sim$20% difference in their main peaks. If the envelopes' standard deviations (not shown in Fig. 8 for clarity) are taken into account, however, this observation cannot be confirmed; more tests, with a broader data set, need to be conducted to come to a definitive conclusion.

Influence of subjects' background on the results

Fig. 9 shows that test scores are not strongly affected by the background of subjects. The average score achieved by geoscientists is always (except for the auditory categorization of DS2) slightly higher than that of the other two groups, but the subjects are not numerous enough for this small difference to be considered significant.

On the other hand, our analysis of the subjects' recorded descriptions of their categorization strategy shows that acousticians have used about 20 more words than both other groups to qualify sounds. We interpret this result as a natural consequence of the acousticians' specific expertise in describing sounds, while geoscientists and physicists usually represent their data only visually. This speculation is confirmed by the study of *Paté et al.* [2017], who conducted a thorough, quantitative analysis of verbal data collected in a similar experiment (also involving audified seismic data, and subjects with similar backgrounds).

Discussion

Conclusions

In our experiments, listeners were exposed to two audified seismic data sets, each characterized by a single, binary control factor: the orientation of the fault (strike-slip or thrust) in one case, the nature of the tectonic plate through which the recorded signal had traveled prior to recording (oceanic or continental) in the other. They were then asked to split each data set into two categories, based on examples of signals associated with different values of the control factor. Purely auditory tests were compared with similar tests, where data were displayed visually rather than acoustically. Overall, listeners were able to categorize data based on audition alone. Their performance in visual tests was better, but performance in auditory categorization was significantly improved by a brief training session.

Asked to comment on the criteria they had chosen to categorize, listeners most often pointed to perception-based physical features that can be summarized as: signal-to-noise ratio (SNR); the duration of what they interpreted to be meaningful signal as opposed to background noise; the frequency content of background noise; the relative amplitude of first seismic "arrivals" with respect to coda. At least two of these features (SNR and meaningful signal duration) do correspond to quantitative parameters that we have been able to define and calculate by simple data processing; we show in Fig. 6a,c that those parameters are differently distributed depending on the value of the relevant control parameter.

In summary, human listeners are able to identify geophysically relevant features of audified seismic data, and can be trained to improve their performance at such tasks. We cannot yet predict the extent to which training can refine our skills at interpreting the data by listening, but we surmise that auditory display can be useful to a variety of endeavors in seismic data analysis.

Outlook

While the resolution and pattern recognition capabilities of the human auditory system are generally well known [e.g., *Hartmann*, 1999; *Wang and Brown*, 2006a], the seismology community does not entirely appreciate the potential of auditory display as a tool for seismic data analysis. A case in point is the interesting work of *Moni et al.* [2012] and *Moni et al.* [2013], where an algorithm designed to mimic the human auditory system (in the words of the authors, "to solve the 'cocktail party problem,' i.e., separating individual speakers in a room with multiple speakers") was successfully applied to the problem of identifying different simultaneous microseisms, and yet no attempt was made to use the human auditory system itself, simply listening to the audified data.

Besides the benefits derived from exploiting the natural skills of our auditory system, audification typically involves the acceleration of a seismic signal by a factor between $\sim 10^2$ and $\sim 10^3$, depending on the frequency content of the original data, which means that an entire day of seismic recording can be listened to in a few minutes with little or no loss of information. Being able to rapidly analyze large sets of data is important, as seismologists are faced today with large and rapidly growing databases. For instance, precisely locating the epicenters of seismic aftershocks requires solving an enormous number of inverse problems [e.g., *Valoroso et al.*, 2013]. This cannot be entirely automated if reliable results are to be otbained. All signals recorded by a seismic network can however be listened to simultaneously, by the principles of sound spatialization [e.g. *Peters et al.*, 2011], in an anechoic chamber equipped with a dense speaker network or, more simply, binaurally. The human auditory system is naturally equipped to locate the source of a sound [e.g. *Hartmann*, 1999; *Wang and Brown*, 2006b], and, through this setup, it is reasonable to hypothesize that one might be able to learn to roughly but quickly locate earthquake epicenters (global, regional or local) by listening to sets of audified seismograms. This approach would involve some important approximations (neglect of dispersion and of Earth lateral heterogeneity effects, etc.), but could be very practical because of its speed and simplicity.

The auditory properties of audified seismograms have also been shown to be indicative of several specific seismic processes, including mainshock/aftershock sequences, earthquake swarms that accompany volcanic eruptions, or deep non-volcanic tremors [*Kilb et al.*, 2012; *Peng et al.*, 2012]. Audification is likely to find other potentially important applications in seismology, wherever large datasets are to be investigated, and unknown/unexpected patterns recognized. Examples include the analysis of the Earth's seismic background signal [e.g. *Boschi and Weemstra*, 2015] with implications for monitoring of natural hazards [e.g. *Wegler and Sens-Schonfelder*, 2007; *Brenguier et al.*, 2008], and the problem of determining the evolution of a seismic rupture in space and time from the analysis of seismic data [e.g., *Ide*, 2007; *Mai* *et al.*, 2016]. The study of large sets of audified data can further benefit from the possibilities offered by crowd-sourcing platforms: if the sounds are short and meaningful enough, if the listeners' task is simple enough, and if the data set is correctly distributed among listeners (each sound is given to at least one listener, some are given to several listeners for verification and variability assessment), then a large data set can be effectively explored by the "collaborative" work of a number of listeners.

Data and resources

The image and audio files that were presented to subjects in all the experiments described here are available online at http://hestia.lgs.jussieu.fr/ boschil/downloads.html.

The Global Centroid Moment Tensor Project database was searched using www.globalcmt.org/CMTsearch.html (last accessed September 2016).

The IRIS database was searched via the Wilber interface at http://ds.iris.edu/wilber3/find_eve (last accessed September 2016).

Figs. 1 and 2 were made using the Generic Mapping Tools version 5.2.1 [*Wessel and Smith*, 1991, www.soest.hawaii.edu/gmt].

Acknowledgements

We acknowledge financial support from INSU-CNRS, and from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 641943 (ITN WAVES). Danièle Dubois and Piero Poli have contributed to our work through many fruitful discussions. Zhigang Peng, Debi Kilb and one anonymous reviewer provided careful reviews of our original manuscript. We are grateful to the volunteers who kindly accepted to take our tests.

References

Boschi, L., and C. Weemstra, Stationary-phase integrals in the cross-correlation of ambient noise, *Rev. Geophys.*, *53*, doi:10.1002/2014RG000,455, 2015.

Brenguier, F., N. M. Shapiro, M. Campillo, V. Ferrazzini, Z. Duputel, O. Coutant, and A. Nercessian, Towards forecasting volcanic eruptions using seismic noise, *Nat. Geosci.*, *1*, 126–130, 2008.

Cowen, R., Sound bytes, *Scientific American*, *312*, 4447, 2015.

Dombois, F., and G. Eckel, Audification, in *The Sonification Handbook*, edited by T. Hermann, A. Hunt, and J. G. Neuhoff, pp. 301–324, Berlin: Logos Publishing House, 2011.

D'Orazio, D., S. De Cesaris, and M. Garai, A comparison of methods to compute the effective duration of the autocorrelation function and an alternative proposal, *J. Acoust. Soc. Am.*, *130*, 19541961, 2011.

Ekström, G., M. Nettles, and A. M. Dziewoński, The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, *200-201*, 1–9, 2012. doi:10.1016/j.pepi.2012.04.002 , 2012.

Frantti, G. E., and L. A. Leverault, Auditory discrimination of seismic signals from earthquakes and explosions, *Bull. Seism. Soc. Am.*, *55*, 1–25, 1965.

Gaillard, P., Laissez-nous trier ! TCL-LabX et les tâches de catégorisation libre de sons, in *Le Sentir et le Dire*, edited by D. Dubois, pp. 189–210, L'harmattan, Paris, France, 2009.

Hartmann, W. M., How we localize sound, *Phys. Today*, *11*, 24–29, 1999.

Holtzman, B., 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*, edited by M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, pp. 161–174, Springer International Publishing, 2014.

Ide, S., Slip inversion, in *Treatise of Geophysics, Vol. 4*, edited by H. Kanamori, pp. 193–223, Elsevier, Amsterdam, 2007.

Kennett, B. L. N., and T. Furumura, High-frequency Po/So guided waves in the oceanic lithosphere: I-long distance propagation, *Geophys. J. Int.*, *195*, 1862–1877, doi: 10.1093/gji/ggt344, 2013.

Kennett, B. L. N., T. Furumura, and Y. Zhao, High-frequency Po/So guided waves in the oceanic lithosphere: II-heterogeneity and attenuation, *Geophys. J. Int.*, *199*, 614–630, doi: 10.1093/gji/ggu286, 2014.

Kilb, D., Z. Peng, D. Simpson, A. Michael, and M. Fisher, Listen, watch, learn: SeisSound video products, *Seismol. Res. Lett.*, *83*, 281–286, doi:10.1785/gssrl.83.2.281, 2012.

Mai, P. M., et al., The earthquake-source inversion validation (SIV) project, *Seism. Res. Lett.*, *87*, 690–708, doi:10.1785/0220150,231, 2016.

Moni, A., , C. J. Bean, I. Lokmer, and S. Rickard, Source separation on seismic data, *IEEE Signal Process Mag.*, *29*, 16–28, doi:10.1109/MSP.2012.2184,229, 2012.

Moni, A., D. Craig, and C. J. Bean, Separation and location of microseism sources, *Geophys. Res. Lett.*, *40*, 3118–3122, doi:10.1002/grl.50,566, 2013.

Paté, A., L. Boschi, J. L. le Carrou, and B. Holtzman, Categorization of seismic sources by auditory display: a blind test, *International Journal of Human-Computer Studies*, *85*, 57–67, doi:10.1016/j.ijhcs.2015.08.002, 2016.

Paté, A., L. Boschi, D. Dubois, B. Holtzman, and J. L. le Carrou, Auditory display of seismic data: Expert categorization and verbal description as heuristics for geoscience, *J. Acoust. Soc. Am.*, *accepted*, 2017.

Peng, Z., C. Aiken, D. Kilb, D. Shelly, and B. Enescu, Listening to the 2011 magnitude 9.0 Tohoku-Oku, Japan earthquake, *Seismol. Res. Lett.*, *83*, 287–293, doi:10.1785/gssrl.83.2.287, 2012.

Peters, N., G. Marentakis, and S. McAdams, Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis, *Computer Music Journal*, *35*, 10–27, 2011.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77*, Cambridge University Press, 1992.

Speeth, S. D., Seismometer sounds, *J. Acoust. Soc. Am.*, *33*, 909–916, doi:10.1121/1.1908,843, 1961.

Tang, Y., Data sonification with the seismic signature of ocean surf, *The Leading Edge*, *33*, doi:10.1190/tle33101,128.1, 2014.

Thorndike, E. L., *Human Learning*, Century, New York, 1931.

Valoroso, L., L. Chiaraluce, D. Piccinini, R. D. Stefano, D. Schaff, and F. Waldhauser, Radiography of a normal fault system by 64,000 high-precision earthquake locations: The 2009L'Aquila (central Italy) case study, *J. Geophys. Res.*, *118*, 1156–1176 doi:10.1002/jgrb.50,130, 2013.

Volmar, A., Listening to the Cold War: The Nuclear Test Ban negotiations, seismology, and psychoacoustics, 1958-1963, *Osiris*, *28*, 80–102, 2013.

Wang, D., and G. J. Brown, Fundamentals of computational auditory scene analysis, in *Computational Auditory Scene Analysis*, edited by D. Wang and G. J. Brown, pp. 1–44, Wiley & Sons, Hoboken, N. J., 2006a.

Wang, D., and G. J. Brown, *Computational Auditory Scene Analysis*, Wiley & Sons, Hoboken, N. J., 2006b.

Wegler, U., and C. Sens-Schonfelder, Fault zone monitoring with passive image interferometry, *Geophys. J. Int.*, *168*, 1029–1033, 2007.

Wessel, P., and W. H. F. Smith, Free software helps map and display data, *EOS Trans. Am. Geophys. Union*, *72*, 445–446, 1991.

Table 1: Summary of listening experiments.

data set	number of waveforms	subjects			audio	visual	training
		A	G	P			
DS1	40	18	9	8	yes	yes	no
DS2	40	15	5	7	yes	yes	no
DS2	36	10	3	4	yes	no	yes

The first two columns to the left indicate how many signals from which data set were presented to the subjects. The letters A, G and P stand for "acousticians," "geoscientists" and "physicists," respectively; "audio" and "visual" indicate which type(s) of data were provided to the subjects; "training" refers to whether subjects were trained before taking the test.

Table 2: Listeners' comments on DS1.

Family A (oceanic paths)	Family B (continental paths)
second shock very close to the first	echo of the first impact's sound
with an echo / rebound	small rebounds
a lot of background noise	little background noise
high-pitched background noise	low-pitched background noise
background noise	shorter and duller sound
longer signal	sharper and shorter
	rising perceived frequency
	faster arrival
	buzz or intense reverberation after the explosion

Summary of written, verbal explanations given by 5 subjects (scoring $\geq$80%) concerning their auditory cateogorization of DS1. All text was originally in French and has been translated into English as literally as possible.

Table 3: Listeners' comments on DS2.

Family A (strike-slip events)	Family B (thrust events)
first shock weaker than second one	louder low frequencies
wave of rising frequency louder than the first heard shock	first shock louder than the second one
after the detonation, sound decays more slowly	first shock louder than the wave
faster attack and decay	more powerful and present sound
significant intensity even after a long time	sound decays quickly after the detonation
lower-frequency shock	slower decay
duller signal	
higher frequencies	

Summary of written, verbal explanations given by 8 subjects (scoring ≥80%) for the auditory cateogorization of DS2 before (4 subjects scoring > 55%) and after training (4 subjects scoring > 72%). Again, the original French text was translated into English.

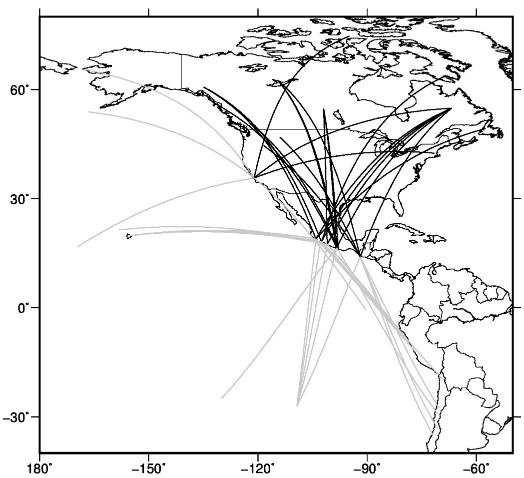
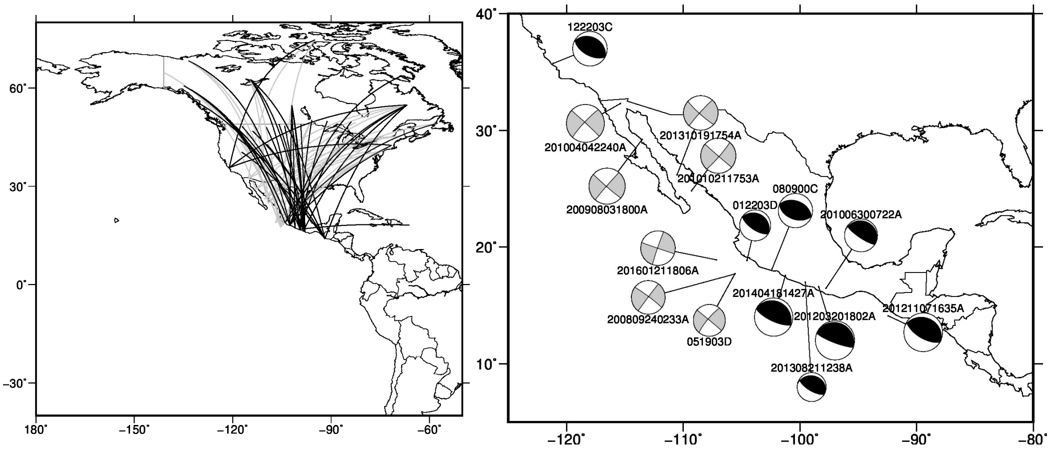


Figure 1: Surface projections of ray paths associated with audified data set DS1. DS1 consists of recordings of events occurring along the west coast of Mexico, made at stations at epicentral distances of ∼4000 to 6000 km; recordings made at north American stations correspond to ray paths only traversing continental terrain (black lines), while stations along the Pacific coast or on ocean islands result in purely oceanic paths (grey lines).



(a)

(b)

Figure 2: (a) Same as Fig. 1, but for data set DS2, which only includes recordings made at stations within the north American continent, of either strike-slip (grey ray path curves) or thrust (black) events. Their epicenters and focal mechanisms [*Ekström et al.*, 2012] are shown in (b) using the same color code.

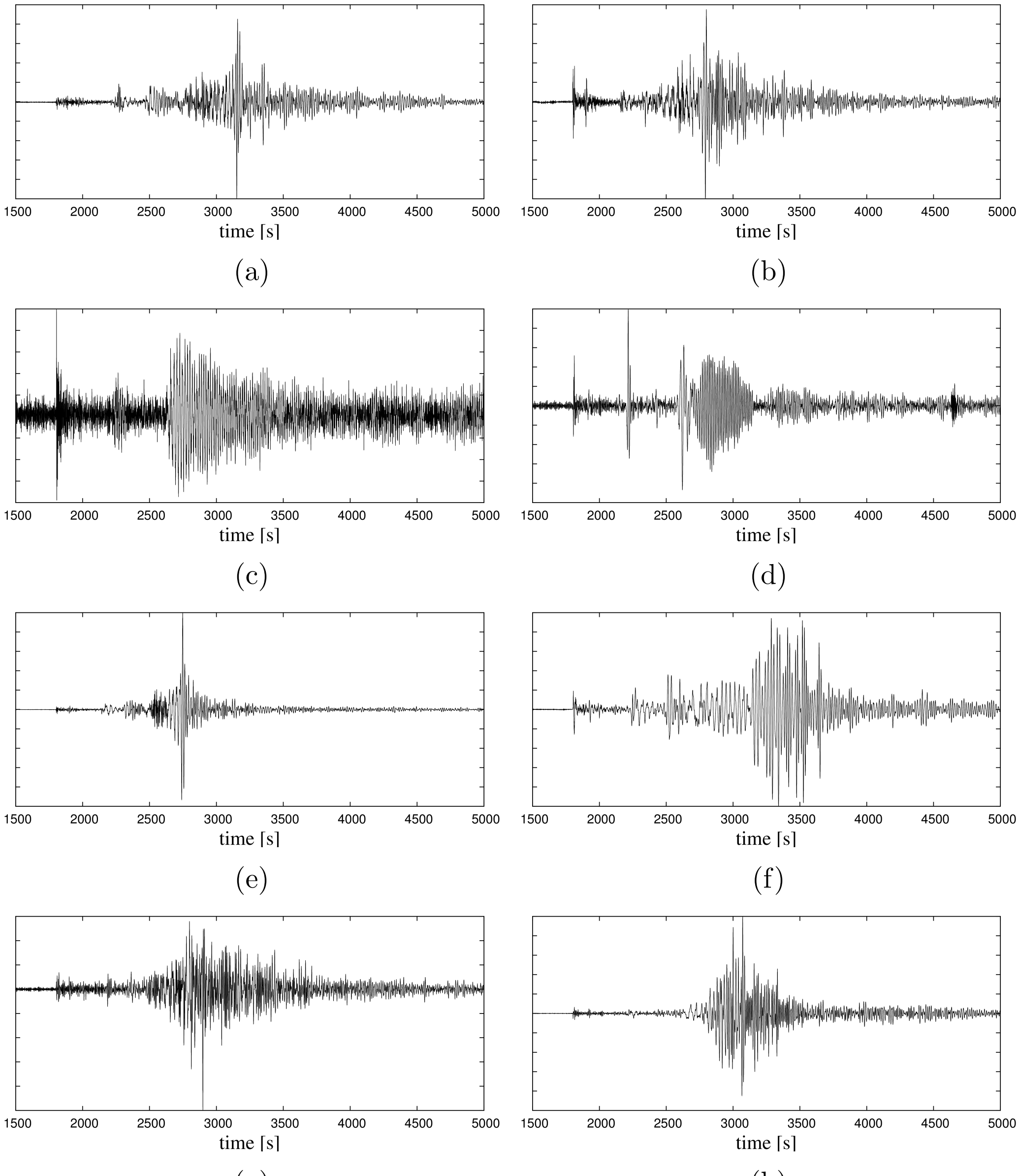
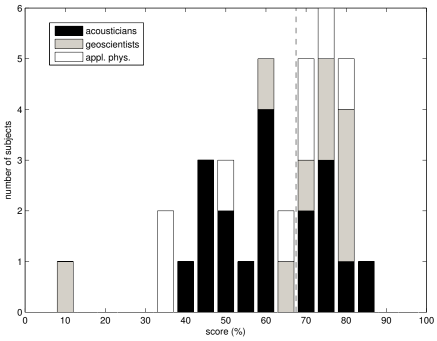
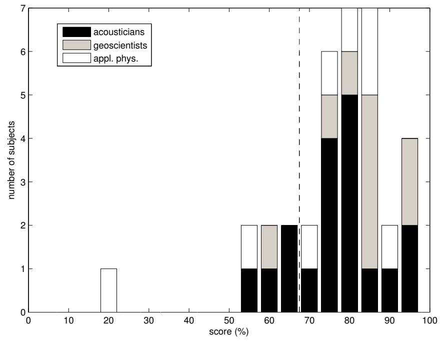


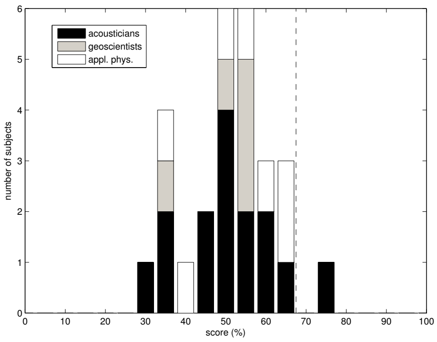
Figure 3: Examples of seismograms used in our study. (a) and (b): DS1, continental paths. (c) and (d) DS1: oceanic paths. (e) and (f): DS2, thrust faults. (g) and (h) DS2, strike-slip faults. The vertical axis is not labeled as we systematically normalize all seismograms (both visual and audio). In our visualization experiments, the horizontal axis was less exaggerated and the time span much longer, so that in principle the exact same information was provided to subjects in visualization and listening tests. The images files used in experiments are available online (see "Data and resources" section).



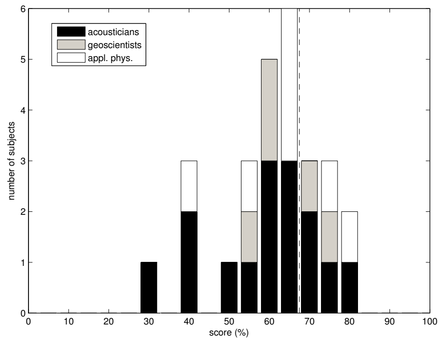
(a) DS1, auditory



(b) DS1, visual



(c) DS2, auditory



(d) DS2, visual

Figure 4: These histograms summarize the results of the constrained categorization experiments conducted before training on audified seismograms from data sets DS1 (panels (a) and (b)) and DS2 ((c) and (d)). Scores achieved in auditory tests are shown in panels (a) and (c); scores achieved in visual tests are shown in (b) and (d). The vertical dashed line marks the "99% confidence level," i.e. the probability of achieving at least that score by categorizing the signal at random is less than 1%. Colors correspond to the different background of subjects, as explained in the inset.

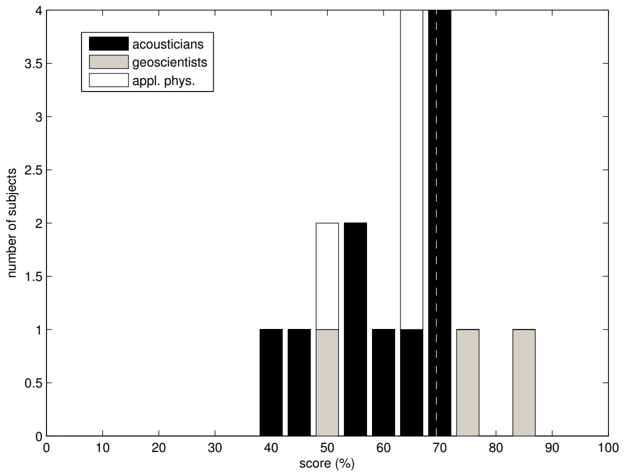


Figure 5: Histogram summarizing the results of the constrained categorization experiment conducted (on DS2) after training. The vertical dotted line, corresponding to a score of 69.4%, marks the 99% confidence level; all scores in the 67.5%-to-72.5% bin actually fall to its right.

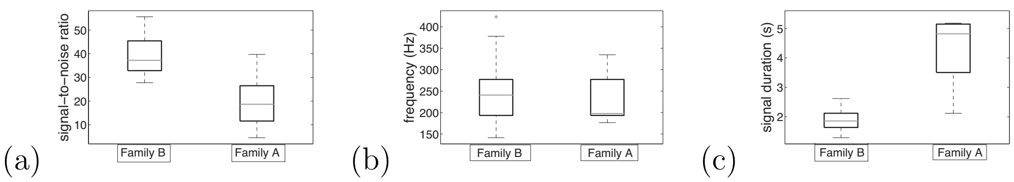


Figure 6: Distributions, shown as box-plots, of three physical parameters, corresponding to properties of the signal that subjects tend to describe as important: (a) SNR; (b) dominant frequency of background noise; (c) duration of meaningful signal. For each parameter, the distributions of parameter values for oceanic-path ("Family A) and continental-path ("Family B) signal are shown separately. Distributions are summarized by their median (thick grey segments), first and third quartiles (upper and lower sides of boxes), and minimum and maximum values (endpoints of dashed lines). Values that we neglect as outliers (their absolute value is more than 1.5 times the interquartile distance) are denoted by grey crosses.

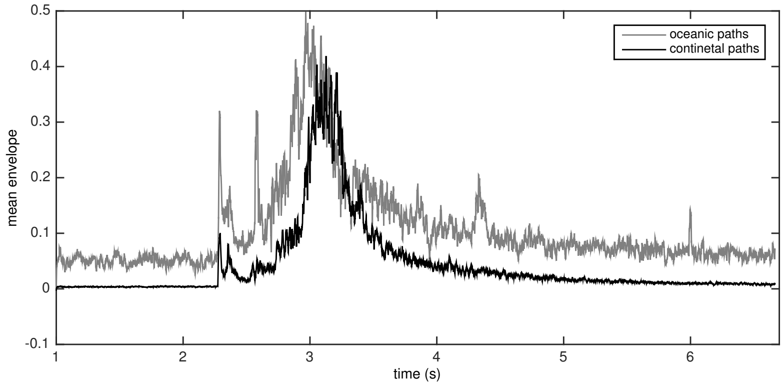


Figure 7: Signal envelope averaged over all DS1 audified seismograms corresponding to continental (black line) vs. oceanic (grey) paths.

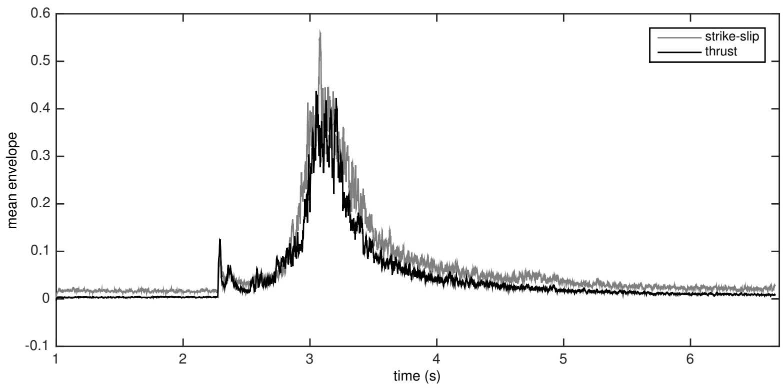


Figure 8: Same as Fig. 7, but envelopes are averaged over all DS2 signals originated from thrust (black line) vs. strike-slip (grey) events.

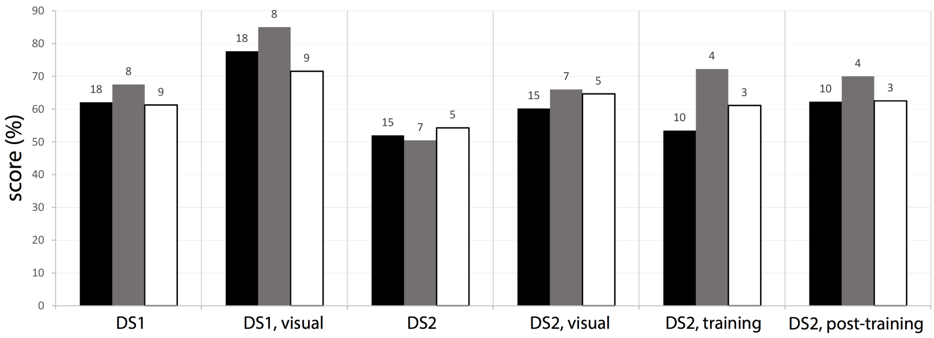


Figure 9: Average scores by test (left to right, as indicated under each bar) and by subjects' background group. Black, grey and white bars are associated with acousticians, geoscientists and physicists, respectively. The number of subjects participating to a test is shown above the corresponding bar. The label "post-training" refers to the auditory test of DS2 conducted after the training session; "training" refers to answers that were given *during* the aforementioned training session.