

A COMPARISON OF AUDIO & VISUAL ANALYSIS OF COMPLEX TIME-SERIES DATA SETS

Sandra Pauletto & Andy Hunt

Media Engineering Group
Electronics Dept., University of York
Heslington, York, YO10 5DD, U.K.
{sp148,adh} @ ohm.york.ac.uk

ABSTRACT

This paper describes an experiment to compare user understanding of complex data sets presented in two different modalities, a) in a visual spectrogram, and b) via audification. Many complex time-series data sets taken from helicopter flight recordings were presented to the test subjects in both modalities separately. The aim was to see if a key set of attributes (noise, repetitive elements, regular oscillations, discontinuities, and signal power) were discernable to the same degree in the different modalities. Statistically significant correlations were found for all attributes, which shows that audification can be used as an alternative to spectrograms for this type of analysis.

1. CONTEXT AND BACKGROUND

This paper describes an experiment to verify that sound can be used as an alternative to graphs in the analysis of complex signals. We have compared a visual and an audio display of the same data sets in order to confirm that certain key attributes are *at least as discernable* from a complex data set by sonification as by visualization.

This verification is important to those projects which aim to use sound representation for data analysis. The world is currently dominated by visual techniques, and many people need to be convinced that information will not somehow be 'lost' by representing it as sound. Once that has been established, it becomes a lot easier to stress the advantages of using sound.

1.1. Previous work on audio / visual comparisons

Visual representations of data have been used for a lot longer than auditory representations. In fact, visual displays can be said to be the norm, and particular visual displays (graphs, diagrams, spectrograms) are widely understood. It is therefore natural when evaluating new auditory displays that we compare their efficacy in portraying information to that of a somewhat equivalent visual display. In the literature there are various studies which compare audio and visual displays. Nesbitt & Barass [1] compared a sonification of stock-market data with a visual display of the same data and with the combined display (audio-visual). Brown & Brewster [2] designed an experiment to study the understanding of sonified line graphs. Peres and Lane [3] evaluated different ways of representing statistical graphs (box plots) with sound. Valenzuela et al [4] compared the sonification of impact-echo signals (a method for non-destructive testing of concrete and masonry structures) with a visual display of the signal. Fitch and Kramer [5] compared the

efficacy of an auditory display of physiological data with a visual display by asking the subjects (who play the role of anesthesiologists) to try to keep alive a 'digital patient' by monitoring his status with each display.

The evaluation methods used in the above examples are dependent on the type of data, the type of auditory display and the context in which the displays are used. These examples show how important it is to compare auditory displays with visual ones for their evaluation, but their results are specific to the type of data, their complexity and the sonification used.

In this paper the sonification method used is audification, i.e. where data are appropriately scaled and used as sound samples. There are some studies in the literature about the efficacy of audification of complex data. Audification is often used for the sonification of data that are produced by physical systems. Hayward [6] describes audification techniques of seismic data. He finds that audification is a very useful sonification method for such data, but he stresses that proper evaluation and comparisons with visual methods are needed. Dombois [7, 8] presents more evidence of the efficacy of audification of seismic data which appears to complement the visual representations.

Rangayyan et al [9] describe the use of audification to represent data related to the rubbing of knee-joint surfaces. In this case though the audification is compared to other sonification techniques (not to a visual display) and it is not found to be the best at showing the difference between normal and abnormal signals.

In all these studies on audification of data, the scaling of the data is informed by an a priori knowledge of the basic properties of the data to be represented.

The novel slant of the experiment presented here is that no assumption is made on the characteristics of the data.

1.2. So why use sound anyway?

This work is a small part of a larger project to work with professionals who use data analysis on a day-to-day basis, but are finding visual analysis techniques inadequate for the task. We have built an interactive sonification toolkit [10] to allow the human analyst to interact with the recorded data as sound, in order to spot unusual patterns to aid in the diagnosis of system faults. The power of a human interacting in a closed loop with sonic feedback is described in [11], and in the IEEE Multimedia special issue on Interactive Sonification [12].

The use of sound is particularly good way of portraying time-series data, because the time-base is preserved in sound playback. The eye tends to scan a picture at its own speed, yet

sound is heard as it is revealed. This yields a particularly natural portrayal of the dynamics of a complex data set. Complex frequency responses in the data are often perceived holistically as timbral differences. Large amounts of data can be rendered rapidly, yet the microstructure is still manifest as timbral artifacts. However, the purpose of this experiment is to determine if some basic attributes of the data are lost by moving from a visual representation to a sonic one.

In our project, we are working specifically with two groups of professionals who need to analyze large quantities of complex data which emanate from sensors connected to the subject being studied.

Physiotherapists at the University of Teesside, UK, record the complex bursts of activity from several EMG sensors attached to the surface of a patient's skin. From these signals the therapists hope to build up a mental image of how the patient's muscles and joints are working, and what is perhaps going wrong in a particular case. We are working with them in sound as it appears to portray the dynamic response of the muscles in a more natural way than by looking at traces on a graph (which is the established, conventional technique). However, our second collaborators have provided us with much more complex data, the analysis of which is the focus of this paper.

1.3. Helicopter flight analysis

We are working with flight analysis engineers at Westland Helicopters, UK. These engineers are routinely required to handle flight data and analyze it to solve problems in the prototyping process. As we have reported in [10] flight data is gathered from pilot controls and many sensors around the aircraft. The many large data sets that are collected are currently examined off-line using visual inspection of graphs. Printouts of the graphs are laid across an open floor and engineers walk around this paper display looking for anomalous values and discontinuities in the signal. The paper is considered more useful than the limited display on a computer monitor.

The current project aims to improve the analysis technique by providing a sonic rendition of the data which can be heard rapidly, and therefore will save valuable technician time and speed up the analysis process. Sound representation also provides the added benefit of allowing the presentation of several time-series data sets together, for dynamic comparison of two (or many more) signals. We are currently also working on methods of portraying many tens of complex parameters together to give a picture of the whole helicopter's flight data. [<Reference to MMViz paper to come later for the camera copy>](#)

The flight engineers are often given the task of analyzing this data because a pilot has reported something wrong in a test flight. The analysts now have a huge amount of data to sift through in order to look for unusual events in the data. These unusual events could be, for instance:

- unwanted oscillations,
- vibrations and noise superimposed on usually clean signals,
- unusual cyclic modes (data repeated, where it would normally be expected to progress)
- drifts in parameters that would normally be constant,
- non-standard variations in power or level,
- a change in the correlation between two parameters (e.g. signals which are normally synchronized becoming decoupled),
- Discontinuities or 'jumps' in data which is in general smooth or constant.

Identification of such events helps to pinpoint problems in the aircraft, and can provide enough information to launch a further, more focused, investigative procedure.

We wish to determine whether any information from the data series is going to be lost when rendered sonically rather than graphically. So, for the purposes of this experiment we have identified five basic attributes of data which we study both visually and aurally. These are 1) Noise, 2) Repetitive elements, 3) Oscillations at fixed frequencies, 4) Discontinuities, and 5) Signal power level.

If a human analyst perceived the presence of one or more of the first four attributes, (or a change in overall signal strength), in an area of the signal where it would not be expected, this would prompt further investigation. So, our experiment determines whether subjects rate the presence of the first four attributes, and the average level of the signal power, to the same degree using a) visual and b) aural presentation.

2. EXPERIMENTAL AIMS & HYPOTHESIS

The aim of the experiment is to compare how users rank the above five attributes when a series of data sets is presented visually or aurally. We are looking to see whether aural presentation allows the identification of each attribute to the same degree as visual presentation. We are interested in the average response across a large group of subjects, rather than identifying whether an individual subject can use visual or audio presentation equally well.

2.1. Hypothesis

The experimental hypothesis is that for each data series, there will be a strong correlation between the recognition of each of the five data attributes in the visual domain and audio domain. If this hypothesis is proved, then we have a strong basis for trusting the analysis of the data using sound alone.

In this experiment we only try to verify if the sound portrays the data attributes at least as well as the visual display. If there is poor correlation, with this experiment, we cannot infer the reasons. We would need other experiments to discover the reasons for a poor correlation.

2.2. Structure of the data under test

In consultation with the flight handling qualities group at Westland helicopters we have gathered 28 sets of time-synchronized data taken from a half hour test flight. Each data set is taken from a sensor on the aircraft under test. The details of the aircraft and the mapping of each individual sensor are being kept confidential.

Each data set contains 106500 samples which were originally sampled at 50Hz. The helicopter parameters measured are of highly differing natures: from the speed of the rotors, to engine power, etc. Most of the data sets represent physical parameters that change over time. For this experiment, the knowledge of what each channel represents in the helicopter system is not important, only whether the user perceives the presence of noise (etc.) in both the visual and audio displays.

2.3. Overview of the experimental task

The visual display used in this experiment is the spectrogram of each data set. The audio display is the audification of the data.

The subjects were presented with a screen containing thumbnail pictures of the spectrograms of all the data sets. After having had an overview of all the spectrograms, they were asked to examine and score each spectrogram (on an integer scale from 1 to 5) for the following characteristics:

- a) presence of noise;
- b) presence of a repetitive element in time;
- c) oscillations at fixed frequencies;
- d) presence of discontinuities or jumps in amplitude;
- e) signal power.

For the sonic display, the subjects were presented with icons – one for each data set, which played the audification when clicked. Subjects were asked to listen to all the sounds at least once. Then they were asked to listen to each sound as many times as required, then score it using the same categories as for spectrograms.

2.4. The audifications

Kramer describes the audification of data as “a direct translation of a data waveform to the audible domain” [13]. The audifications in this experiment were created by linearly scaling the 28 data arrays between -1 and 1 and by converting each array into a wave file of sampling rate 44100Hz in Matlab. Each audification was therefore around 2.5 seconds long.

2.5. The spectra

The spectrograms, of the same data channels, were created by using the Matlab function ‘specgram’. The sampling frequency specified when computing the spectrograms was ‘fs = 50’, which corresponded to the original sampling frequency of the data (50Hz). The minimum and maximum values of the color scale of the spectrograms were set the same for each spectrogram so that the spectrograms were comparable to each other. All the spectrograms were saved as .jpg files.

2.6. The subjects

The subjects for this test were selected according to the following criteria.

It was considered that the end user of such an auditory display would be an experienced analyst, able to interpret spectrograms and able to distinguish various characteristics in a sound’s signal such as noise, repetitions, frequencies, discontinuities and signal level.

Apart from this specific knowledge, the user could be any gender or age or from any cultural background. A between-subjects design, in which there are 2 groups of different subjects (one of which scores the spectra and the other the sounds), would have been ideal for this experiment. This would have required the recruitment of too many subjects, which was not realistic. Instead, we chose a group of 23 subjects and used a mixed within-subjects / between-subjects design, in which mostly the same group of people scored both the spectra and the sounds, but some only did one or the other. This design was due to the fact that some subjects were available only for a short time.

In order to minimize the errors in the results due to the order of presentation of the task, the order in which the spectra and the sounds were presented to each person was randomized between subjects and tasks.

Out of the 23 subjects tested, 21 were men and 2 were women. The average age of the subjects was 33. All the subjects were lecturers, researchers or postgraduate students in media and electronic engineering (with a specialisation in audio and music technology) and one person was a computer music composer. They all had experience in working with sounds and spectrograms. The subjects’ understanding of sounds and spectrograms was considered to be similar to the expected understanding of the ideal end user. Subjects were from different nationalities. All the subjects declared that had no known problems with their hearing and that they had good sight or, if the sight had some defect, it was fully corrected by spectacles.

2.7. Procedure

Firstly, each subject was given a single-page written document which explained the task. Then the subject was asked to fill in a questionnaire to gather the information about occupation, gender, age, nationality, his/her familiarity with spectrograms and sound interpretation, and any known hearing or sight problems.

The audio test was carried out in a silent room (mostly in the recording studio performance area at York). Good quality headphones (DT990 Beyerdynamic) were used with a wide frequency response (5 - 35,000Hz). This minimized the errors that could be due to external sounds. The volume of the sounds was maintained the same for all subjects.

The spectrogram test was also conducted in a generally quiet room which allowed concentration.

Subjects who were able to do both tests in one sitting were asked to take at least a 2 minute rest between the visual and the audio parts of the test. The total test, for each subject, lasted about 45 minutes. Subjects were asked to record on a piece of paper any comments about the test they thought could be valuable.

For the experiment, a program was created in PD (Pure Data [14]) and all the results of the test were automatically recorded in a text file. Before presenting the spectrograms and the sounds to each subject, the order of presentation of each data set on the screen was randomized. This should minimize errors due to the order of presentation.

The test began with an overview of all the spectrograms (see Figure 1). Then by clicking on each thumbnail image a larger version of the spectrogram appeared. A click on the ‘Test’ button brought up a further window, consisting of a series of radio buttons (labeled from 1 to 5) for each parameter being scored (noise, repetition, frequency, discontinuity and signal power).

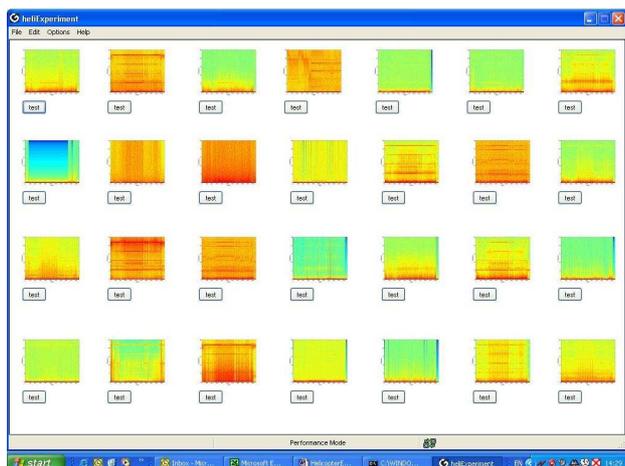


Figure 1: *Thumbnails of Spectrograms*

For the second part of the experiment the subject was presented with a set of buttons, one for each audification. Before starting scoring, the subject was asked to listen to all the sounds at least once.

By clicking on a button the subject could hear each audification through headphones. Again, a click on the ‘Test’ button brought up the scoring window, identical to that used for the spectrograms (see Figure 2).

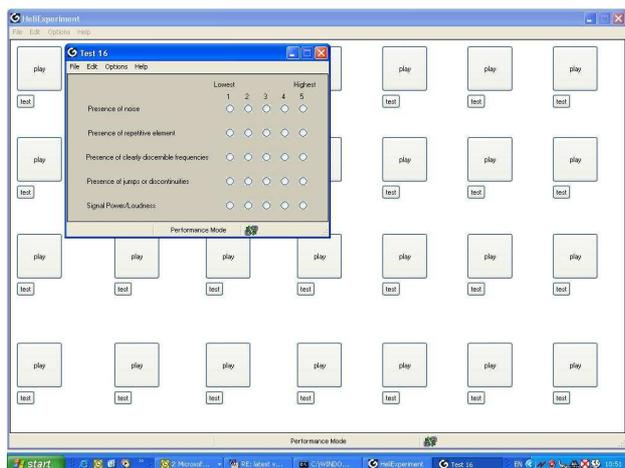


Figure 2: *The scoring window superimposed upon the buttons for each sound*

3. RESULTS

The scores were divided and analyzed by the 5 attributes being tested (noise, repetition, frequency, discontinuity and signal power).

For each of the 28 data sets (i.e. channels of sensor information from the helicopter) two mean scores were calculated across all subjects; one for the sound display and one for the visual display. Therefore for each attribute being tested (noise, repetition, etc) we have 2 arrays of average scores (one for the sound and one for the spectra), with an average score across all subjects for each data set.

If the two displays portray information in exactly the same way, then we might expect the two arrays of scores to be exactly the same. A scatter plot (x axis = spectra scores, y axis = sounds scores) was plotted for each of the five attributes under test.

This helps us to see if a linear relationship exists between the spectra scores and the sound scores. Then the correlation factor (1) was calculated.

Correlation factor

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}} \quad (1)$$

3.1. Presence of noise

In all the following scatter plots, the continuous line represents the line where ideally the dots should sit, while the segmented line represents the regression line calculated from the actual points. Each dot is the average score across all subjects for one (of the 28) data sets.

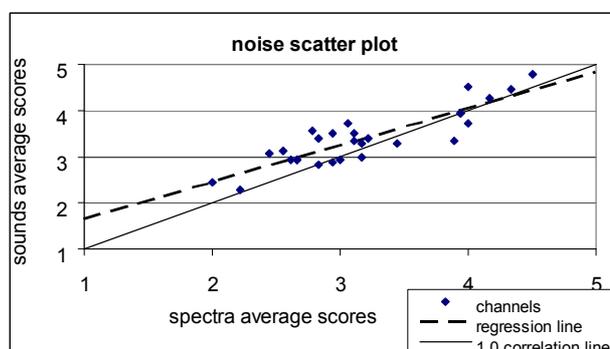


Figure 3: *Scatter plot for the attribute ‘Presence of Noise’*

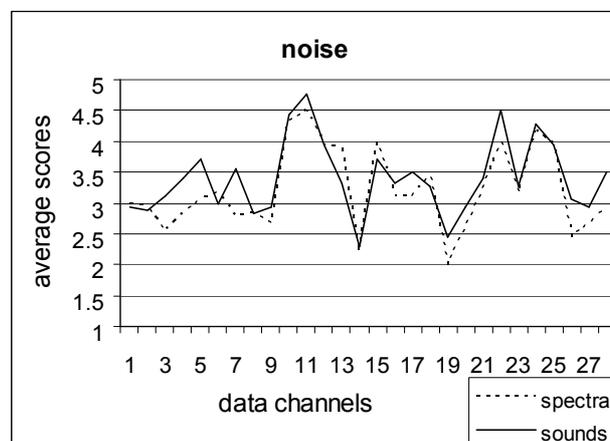


Figure 4: *Average Noise scores for each data set*

For each category a second plot was made (e.g. see Figure 4) in which along the x axis are the individual data sets (the ‘channels’) and along the y axis are the average scores across all subjects. The solid line represents the results for the sound display and the segmented line represents the spectra results.

The correlation (**r: 0.88**) between the auditory display scores and the visual display scores is very high. Thus the average scores for presence of noise is very similar whether people are presented with a spectrogram or an audification of the data sets. Another way of looking at this is as follows. Let

us round the average scores to the nearest integer (remembering that people were asked to score with a 5-step integer scale) and we calculate the absolute value of the difference between the rounded spectra scores and rounded audio scores for each channel (see Table 1).

Rounded scores spectra	Rounded scores sounds	Abs(difference)
3	3	0
3	3	0
3	3	0
3	3	0
3	4	1
3	3	0
3	4	1
3	3	0
3	3	0
4	4	0
5	5	0
4	4	0
4	3	1
2	2	0
4	4	0
3	3	0
3	4	1
3	3	0
2	2	0
3	3	0
3	3	0
4	5	1
3	3	0
4	4	0
4	4	0
2	3	1
3	3	0
3	4	1

Table 1: Difference in rounded scores

We can see that only 7 data sets out of 28 are scored differently in the visual display than in the audio display (for the degree of noise present) and the difference is only 1 point.

We now present the data in the same formats (scatter-plot and average channel scores) for each of the remaining attributes.

3.2. Presence of a repetitive element

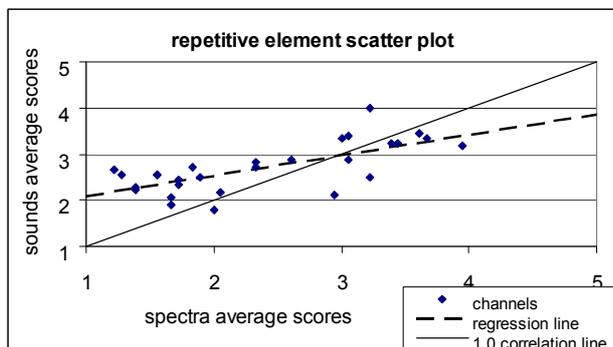


Figure 5: Repetitive element scatter plot

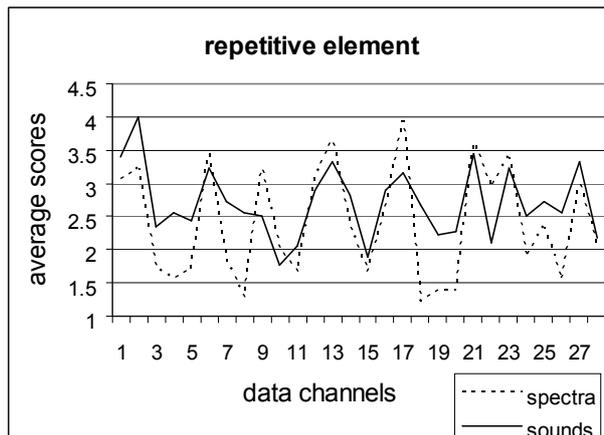


Figure 6: Repetitive element scores

The correlation ($r: 0.70$) is still quite high but less than in the noise case. 15 out of 28 rounded average scores are different between the visual and the audio display. In 13 the difference is by 1 point and in 2 by 2 points.

3.3. Presence of oscillations at fixed frequencies

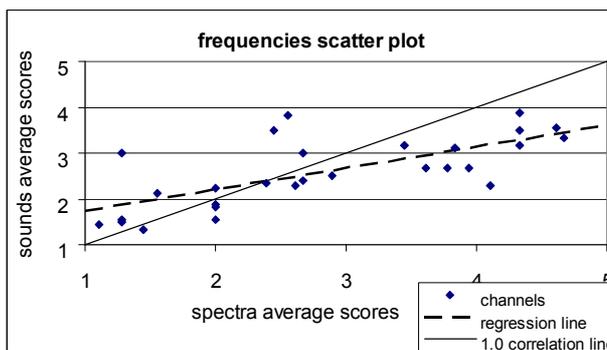


Figure 7: Frequencies scatter plot

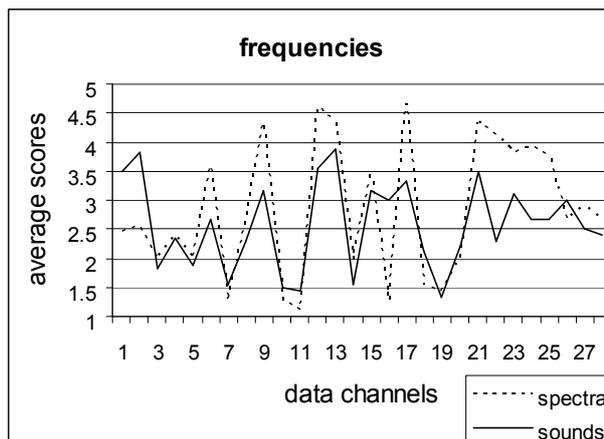


Figure 8: Frequencies scores

The correlation ($r: 0.71$) is close to the correlation calculated for the repetitive element. 15 out of 28 rounded average scores are different between the two displays: 11 have a difference of 1 point, and 4 by 2 points.

3.4. Presence of discontinuities

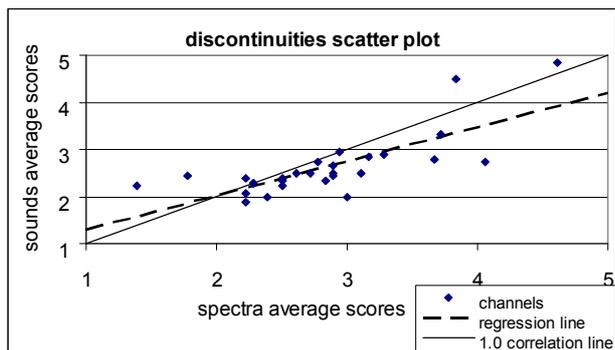


Figure 9: Discontinuity scatter plot

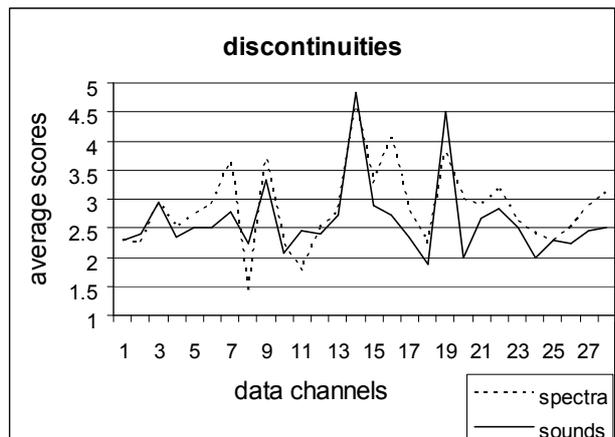


Figure 10: Discontinuity scores

The correlation ($r: 0.76$) is quite high. 11 out of 28 rounded average scores are different between the 2 displays: in all cases the difference is by 1 point.

3.5. Rating of signal power

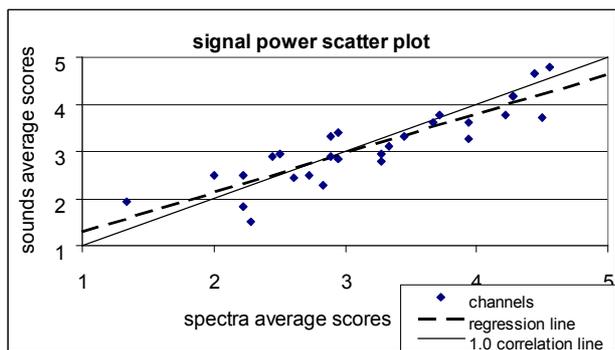


Figure 11: signal power scatter plot

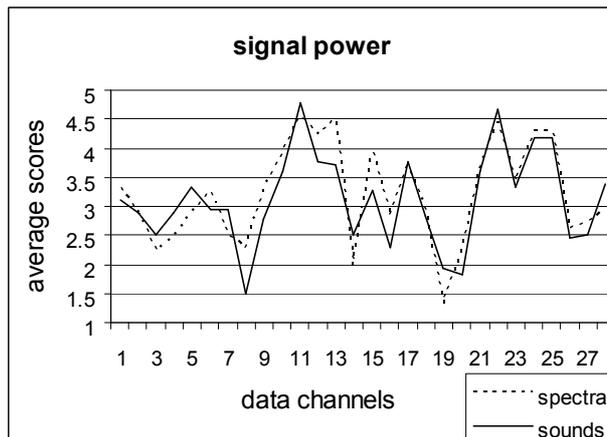


Figure 12: Signal power scores

The correlation ($r: 0.88$) is very high. Only 9 out of 28 average rounded scores are different between the displays and these are by only 1 point.

4. DISCUSSION

For each of the five attributes the average scores for the spectra show high correlation with the average scores for the sounds. This means that the two displays do indeed allow users to gather some basic information about the structure of the data to a similar degree.

It is reasonable to think that the degree of similarity of the two displays could be improved by considering the following:

- The audio display could be improved by choosing a different data scaling informed by sound perception principles.
- The subjects were presented with a complex task. They had to score 28 data sets for each of the 5 attributes, both in the visual mode and in the audio mode. It is possible that an easier task (e.g. score 10 channels for one category only at the time) could show an even higher similarity between the data.
- The subjects had to score very complex sounds containing (to varying degrees) noise, clicks, the presence of many frequency components, and often a complex evolution of the sound over time. Again with easier sounds, i.e. simpler data structures, the similarity in the scores could be higher.
- The test questions were often ambiguous. For example, subjects often wondered if the noise of the clicks, produced when there is a discontinuity in amplitude, should count for 'presence of noise' since it was already accounted for under 'presence of discontinuity'. These ambiguities have surely contributed to the increase in variance in the results. Less ambiguous questions could yield better results.

The correlation between average scores for the visual display and the auditory display in the Noise and Signal Power attributes is higher than that for the other three. The reason for this difference can probably be found in the nature of the data displayed and the way the displays were built. For instance, the perception of frequency influences the perception of loudness, e.g. to perceive a 100Hz and a 1000Hz sound with the same loudness, the level of the 100Hz sound needs to be higher than that of the 1000Hz sound [15]. It is possible, therefore, that

frequencies that can be seen in the spectrogram are not easily perceivable in the audification. The difference could also be due to the different characteristics of the visual and the auditory sense: one could be better at picking up certain elements than the other. For instance, the ear could be better at perceiving repetitive elements in time, since we are used to recognizing rhythmic structures in sound, while repetitions could be harder to spot in a spectrogram. A more precise analysis of the results for each particular channel, focusing in particular on the differences in scoring between audio and visual display, will be done in the near future. From the results of such deeper analysis new hypothesis could be made regarding the degree of similarity and difference of these two displays, which will then need to be tested with new experiments.

Finally, during the test, the subjects were free to write down any comments about the spectra or the sounds or the test procedure. 13 out of 23 subjects chose to comment and here is a summary of the most common observations:

- it is difficult to score in particular noise, discontinuities and repetitions (7 comments);
- I can hear more detail in the sounds than in the spectra (2 comments);
- I feel that I get better at scoring as I go along (4 comments);
- Some data sets actually *sound like* a helicopter (2 comments).

5. CONCLUSIONS

This paper has described an experiment which compares a visual display and an auditory display in their abilities to portray basic information about complex time-series data sets. The subjects of the experiments were asked to score the spectra and the audifications of the data sets on an integer scale from 1 to 5 for the following attributes: presence of noise, presence of a repetitive element, presence of discernible frequencies, presence of amplitude discontinuities and overall signal power. It was found that the scores for each data set, averaged over all subjects, showed high correlation between the visual and auditory displays for all five attributes. This means that these two displays portray similarly well some basic information about this data set.

6. ACKNOWLEDGEMENTS

The data used in this paper was gathered as part of the project 'Improved data mining through an interactive sonic approach'. The project was launched in April 2003 and is funded by the EPSRC (Engineering and Physical Sciences Research Council). The research team consists of academics at the Universities of York (User Interfacing and Digital Sound) and Teesside (Physiotherapy) led by Prof. Tracey Howe, and engineers at Westland Helicopters led by Prof. Paul Taylor. Many thanks are due to all the people of KTH (Sweden), the Electronics, Music and Computer Science Departments of York University who participated to this experiment.

7. REFERENCES

[1] Nesbitt K. V. and Barrass S., *Finding Trading Patterns in Stock Market Data*, IEEE Computer Graphics and Applications, September/October 2004, pp. 45-55

- [2] Brown L. M. and Brewster S. A., *Drawing by ear: Interpreting Sonified Line Graphs*, Proc. International Conference on Auditory Display (ICAD), 2003
- [3] Peres S. C. and Lane D. M., *Sonification of Statistical Graphs*, Proc. ICAD, 2003
- [4] Valenzuela M. L., Sansalone M. J., Streett W. B., Krumhansl C. L., *Use of Sound for the Interpretation of Impact-Echo Signals*, Proc. ICAD, 1997.
- [5] Tecumseh Fitch W. and Kramer G., *Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System*, In Kramer G. (ed) *Auditory Display: Sonification, Audification, and Auditory Interface*, Addison-Wesley, Reading, MA, 1994
- [6] Hayward C., *Listening to the Earth sing*, In Kramer G. (ed) *Auditory Display: Sonification, Audification, and Auditory Interface*, Addison-Wesley, Reading, MA, 1994
- [7] Dombois F., *Using Audification in planetary seismology*, Proc. ICAD, 2002
- [8] Dombois F., *Auditory Seismology on free Oscillations, Focal Mechanisms, Explosions and Synthetic Seismograms*, Proc. ICAD, 2002
- [9] Krishnan S., Rangayyan R. M., Douglas Bell G., Frank C. B., *Auditory display of knee-joint vibration signals*, Journal of the Acoustical Society of America, 110(6), December 2001, pp. 3292-3304.
- [10] Pauletto, S., & Hunt, A.D., *A Toolkit for interactive sonification*, Proc. ICAD, 2004
- [11] Hunt, A. & Hermann, T., *The importance of interaction in sonification*, Proc. ICAD, 2004.
- [12] Hunt, A.D., & Hermann, T. (eds), *Special Issue on Interactive Sonification*, IEEE Multimedia, Apr-Jun 2005.
- [13] Kramer, G., 1994, *Some organizing principles for representing data with sound*. In Kramer G. (ed) *Auditory Display: Sonification, Audification, and Auditory Interface*, Addison-Wesley, Reading, MA
- [14] Pure Data: <http://pd.iem.at/>
- [15] Howard, D. and Angus, J., 1996, *Acoustics and Psychoacoustics*, Music Technology Series, Focal Press, Oxford.