EXPLORATORY SOUND ANALYSIS: SONIFYING DATA ABOUT SOUND

Sam Ferguson and Densil Cabrera

Acoustics Research Laboratory
Faculty of Architecture, Design and Planning
The University of Sydney
sferguson@arch.usyd.edu.au

ABSTRACT

Sound is commonly analysed subjectively by listening to it. However, when we want to analyse a sound objectively, we often switch domains, and change to visual or numerical displays. While it is likely that, generally speaking, the visual sense dominates other senses, when the data being explored are sound the question naturally arises as to whether these data may be statistically represented in that same domain. This paper describes and demonstrates a general scheme for building statistical representations of sound that exist entirely within the auditory domain, and use the original audio data to present descriptive data.

1. INTRODUCTION

This paper presents an approach to sound analysis that is based in auditory display. Rather than mapping analysis data to auditory graph parameters, the approach presented here uses the original sound recording - fragmented and rearranged - to convey the analysis to the listener's ears. This approach is called 'exploratory sound analysis' (ESA), after Tukey's 'exploratory data analysis', and it draws on the approach to concatenative synthesis taken by Schwarz [1] and others [2]. The concept of using the sound itself to represent the sound has great potential in providing a transparent analysis method that straightforwardly reveals relationships between sound events and the analysis parameters, relationships between analysis parameters, and indeed explains the meaning of analysis parameters by way of example. This introduction provides a brief overview of several concepts that underlie exploratory sound analysis, before it is described in section 2.

Previous research has discussed methods for building sonifications that assist users in understanding acoustic and auditory phenomena. Cabrera et al. [3] developed methods for using sonification in a teaching context, to demonstrate certain acoustical phenomena and characteristics of audio systems. These sonifications were mostly interactive, and assisted students to explore ideas about the physical nature of sound and sound recordings. Ferguson et al. [4] discussed a method for extracting information about sound from a set of psychoacoustic models and using abstract signals to represent it, somewhat simplifying comparisons between two signals. It used a large lookup table to pre-calculate the appropriate signal parameters, given that psychoacoustical models are not reversible.

Schwarz [1] proposes a concatenative synthesis system named *Caterpillar*, which uses a partnership between data describing grains of sound and the corresponding audio data. He uses the term *unit* to describe these. Schwarz stores each audio frame, along with appropriate metadata, and several types of descriptor data in a database. The metadata fields (unit type, base file, start time, end time, duration, segmentation confidence and description) are used more often when non-regular segmentation schemes are used. The purpose of the *Caterpillar* system is to build a new concatenated version of a target phrase that has also been described using descriptor data – the *units* available in the database are transformed and concatenated to build a new version of the target phrase.

Associating descriptor data about a piece of sound with that piece of sound is crucial to the framework we will describe here. Schwarz's *units* are very similar to the grains of sound used in granular synthesis, and Roads [5] describes the development of these theories in detail. The fundamental difference between the way that granular synthesis and *Caterpillar* function is that granular synthesis often chooses grains from a particular time range within a sound, to ensure they are somewhat similar, while *Caterpillar* takes a more rigorous approach, employing the use of descriptor analysis algorithms to describe the grains of sound so they may be used in a detailed database selection method.

Tukey's seminal work, *Exploratory Data Analysis* [6], is a masterpiece of statistical thinking. It developed many new methods for approaching batches of data, and described several visual representations, such as the boxplot, that are in wide use today. It describes the statistical purpose of exploratory data analysis as that of the detective, looking for clues, as opposed to the process of confirmatory data analysis. The name 'Exploratory Sound Analysis' (ESA) is used to draw attention to the correspondence between Tukey's purpose and the current framework's. Many of the representations developed are based on transposing Exploratory Data Analysis representations to the auditory domain.

Computational Information Design, a process defined by Ben Fry, brings together the specialised fields of computer science, data mining, statistics, graphic design, and information visualisation [7]. It synthesises these disparate fields into a unified framework, whose purpose is to collect, manage and then understand data. The framework, briefly, contains seven steps: acquire, parse, filter, mine, represent, refine, and interact. A simple data representation example (from Fry) may be described using these steps as: data about the postcode system is: *acquired* from a website, *parsed* into fields representing latitude and longitude, *filtered* for a particular region's postcodes, *mined* to work out how large a representation grid should be, *represented* as points on a longitude/latitude grid, *refined* by altering contrast and colour attributes, and *interacted* with by using zoom and label actions. Whilst Fry's domain of interest is information visualisation, and his framework addresses this particular interest, the purpose of the thesis is to extend the framework to general data representation tasks, such as the current framework.

Information Visualisation has shown many examples of effective representation of large sets of data. For instance, Shneidermann's Treemap visualization [8] has been used in a large number of applications, including the visualization of prices and companies on the stock exchange [9], and of online news [10]. Similar visual display principles have been applied to musical sound by Wattenberg [11] and by Snydal [12]. The principles underlying the field are most briefly summarised by the Information Visualization Mantra, discussed by Shneidermann et al. [13] – 'Overview First, Zoom and Filter, Details on Demand'. This paper is an attempt to work towards applying some of these principles to the sonification of sound.

1.1. Auditory Graphing

Auditory Graphing is the representation of numerical data through auditory means. Visual graphing is commonplace and achieved easily – it is taught from a young age, and its applications are numerous. On the other hand, techniques for auditory graphing commonly rely on computer-based sound synthesis, impractical until comparatively recently, and thus they have not been widely employed.

Auditory graphing maps numerical data to some kind of auditory attribute, commonly fundamental frequency. The data is often presented over time, as the auditory system does not scan a scene in the same way that the visual system can. A generally recognised problem with auditory graphing is that it is not immediately clear which mapping to use for a given type of data [14]. The shape of the resulting profile is easily perceived when particular mappings (such as pitch) are used, but some other forms of mappings (for instance loudness) can be quite problematic [15]. Walker [16] has investigated the usefulness of various mapping as applied to different types of data. Vickers [17] describes some of the problems with some common techniques for auditory graphing, and shows how the multiple abstractions used affect its usability and perceptibility.

2. EXPLORATORY SOUND ANALYSIS OVERVIEW

Typical methods for analysing sound result in numbers, often represented through visual graphs, such as a graph of the spectrum as it changes over time, or a graph of harmonics to noise ratio. These graphs often require significant training to be understood, as the concepts represented are purely auditory in nature. To assess the numerical values represented by the visual graph sometimes it is necessary to attempt to create an mental image of sounds that have similar values. For instance, when interpreting a graph of sound pressure level, it is possible to refer to tables that describes what common sound sources (eg. quiet conversation, traffic noise, jackhammer, jet aeroplane) those levels might be produced by. For other audio descriptor algorithms (eg. Spectrum Standard Deviation, Tonal Dissonance) these tables do not exist, and therefore it is difficult to interpret the visual graph or the numeric values by reference to a common sound source. Auditory phenomena do not transfer directly to the visual domain without a loss of some of the characteristics particular to the auditory domain, and many of these characteristics are difficult to precisely describe– it is easier and more straightforward to experience them.

Also, some of the attributes of visual graphs may not be appropriate for the representation of auditory data. For instance, spectra are commonly represented using a line graph. However, a line graph equally weights peaks and troughs at various frequencies. A strong peak, due to the line graph, may seem as important as sharp trough that is next to this peak, although any sound at this point would be completely masked by the strong peak. So, when using visual graphs only to represent auditory information we lose some of the important characteristics of the auditory phenomena, we also gain some artefacts caused by the representation method.

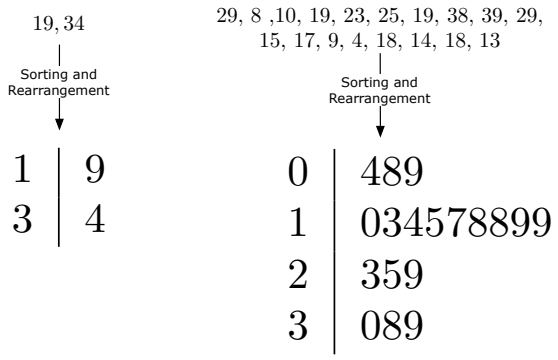

Figure 1: A stem and leaf plot is an effective way of representing the composition of a numerical dataset.

The purpose of Exploratory Sound Analysis is to provide a method for exploring characteristics of recorded sounds through listening to sounds, rather than through comprehending numerical values about sounds. Descriptor algorithms commonly break up the sound into small frames (of around 10-50 ms) and then calculate a value to describe that frame. It is reasonable to expect that if a frame is played back to the user, it will sonify the value that has been assigned to it by the descriptor. This is the fundamental notion that underpins Exploratory Sound Analysis.

ESA is a method for using frames of audio in a manner similar to the way that numerical values are manipulated. Rearranging these frames appropriately can form a representation that provides a kind of audio summary of the sound. A numerical analogy to this rearrangement process is the stem and leaf plot, invented by Tukey, whereby a set of numbers is sorted and tabulated to form a type of histogram of the numbers [6]. Each two digit number is divided at the decimal place, so that all the numbers between, say, 10 and 19 inclusive are represented on a single row, with a 1 at the beginning of the row, and a digit to represent each number following this initial 1 (see Figure 1). Because the plot uses the original data in the representation, the link between the representation and the original data is much stronger.

The rearrangement of audio data can be used in much the same way, to elucidate the structure and importance of elements within the descriptor data. The method of rearrangement can be different, depending on the many different exploratory purposes, outlined in Section 3. The fundamental notion of the framework of Exploratory Sound Analysis is to better explore sound by using audio representations based on the original audio.

2.1. Units

Schwarz uses the term *units* to describe the chunks of sound and their associated meta and descriptor data. In Schwarz's system each unit is added to a large database, from which appropriate units are selected based on the target phrase to be synthesised. In the system described in this paper, each audio file is separate, and the purpose

of producing each set of units is to represent that file through the selection, rearrangement and concatenation of the units in the set.

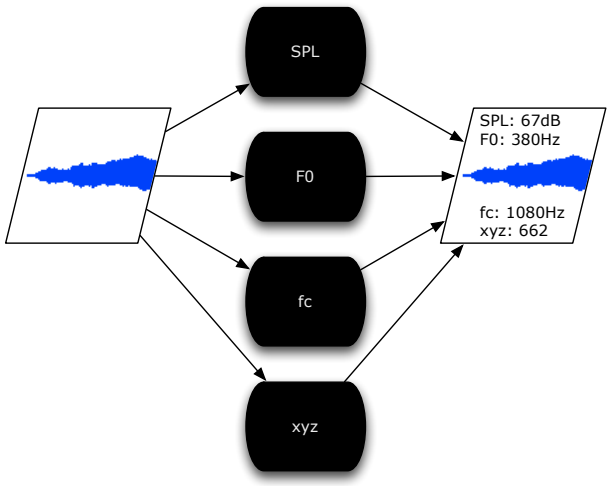

Figure 2: We can treat most descriptor algorithms as a black box, if they provide a single numerical value to be associated with each frame of sound. In this way, each timeseries descriptor value is associated with the frame it represents, and a set of units are built that can now be used to sonify data about the sound file.

Each audio frame is passed through the descriptor algorithm and the result stored in a database. Other important metadata is also stored, such as the frame's position in the original audio file, and the time step information.

2.2. Segmentation

The audio waveform that is to be segmented can be broken up in a number of ways, based on what the purpose of the segmentation is. Two purposes are immediately apparent: reorganising the audio waveform (or a large proportion of it) into a new, altered version of the original sound, or taking one particular segment and expanding it (through repetition) into a longer sound of some kind.

The most straightforward way to segment the waveform is the common method based on windowing an audio sample. Frames of the audio waveform are extracted and saved in a matrix sequentially, with the frames slightly overlapping each other so they can be interleaved without discontinuities. This results in a sequential set of frames of audio that are all exactly the same length. While this is a parsimonious approach, it totally ignores the content of the waveform and the descriptor data associated. Other methods of delineating segments in a sound, such as onset detection, chord change information, thresholding, spectral flux, descriptor flux information, may form another basis for segmenting a file, but these are outside the scope of this paper.

3. BUILDING A REPRESENTATION

Using these grains of sound we are able to build any number of different representations that fulfil different exploratory purposes. Three are demonstrated and explained below:

- Measures of location, such as the median of a sample.

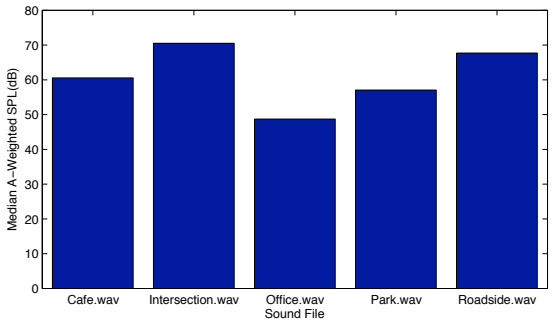

Figure 3: Five background noise recordings are represented in terms of their median A-weighted SPL. A graphical representation of A-weighted level is not as direct as listening to the median level. Audio example.

- The cumulative distribution describes the distribution of descriptor data across the time duration of the sample.
- The boxplot demonstrates this distribution in a different manner.
- Difference analysis can label and extract peaks and troughs in fluctuating timeseries descriptor data.

3.1. Representing Numbers about Sound

We have already chosen to use concatenative synthesis based on the input audio waveform as the output representation, but many possibilities remain for designing a representation. The particular frame that represents the number to be represented, such as the median, is often far too short to be properly understood when played back on its own. A solution that may be considered is to repeat the frame multiple times, so that it forms a waveform of a longer duration. However, when a frame is concatenated regularly in such a manner sidebands are produced, based (irrelevantly) on the repetition of the waveform. The resulting sidebands are related to the choice of window size, which is an arbitrary decision unrelated to the data being represented. This is likely to confuse the listener, and so an alternative is proposed.

Rather than choosing a single frame to represent a number, a small selection of them (approximately 10) are chosen. Random selections from this set are concatenated until a reasonable length of sound is built. This means there will be regular grain repetition in any perceivable sequence, whether long or short. This produces a sound whose quality is based on the original part of the sound the grains have been taken from, but whose repetition does not form amplitude modulation. The large number of grains produced when an audio file longer than, say, 10 seconds is broken into 100 frames for every second, means that there is usually a very large number of data points. If 1000 data points exist and 10 are chosen to represent a number, then 1% of the data is being used to represent a single number. However, if there were only 100 data points, the use of 10 of them becomes problematic because they may have values ranging across 10% of the data range and the representation is then fairly inaccurate compared with the range of the data. The descriptor being represented also affects the necessity for accuracy within a representation – for instance, a representation of pitch needs a lot more accuracy than one of sound pressure level, due

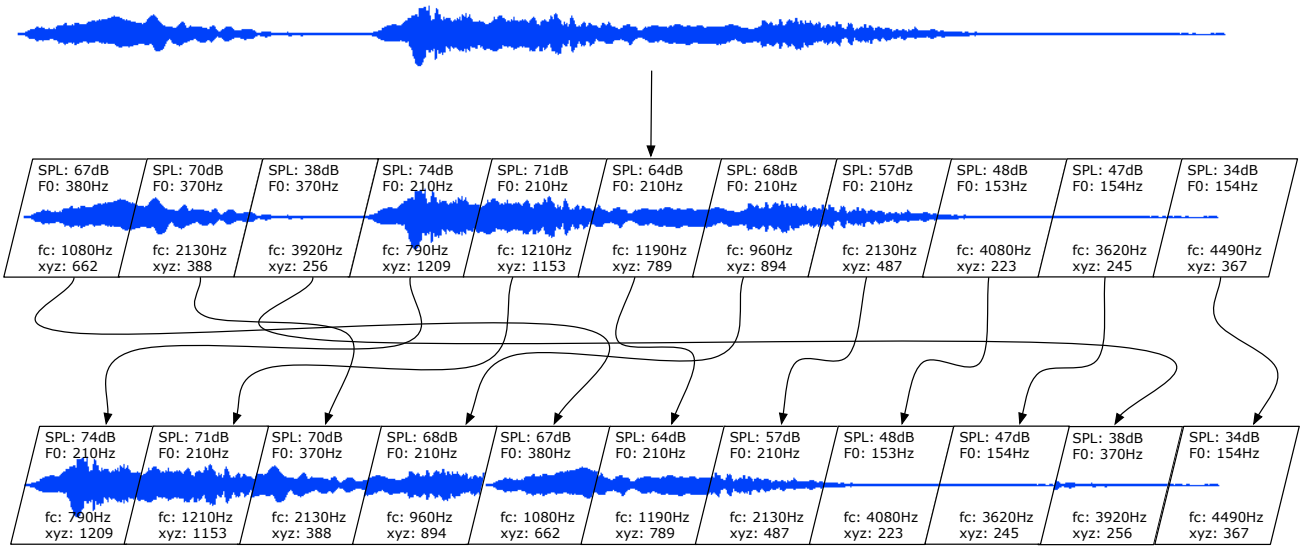

Figure 4: By breaking up a sound into a set of small frames, and rearranging those frames in order of the value of acoustic descriptors, we can create a representation of the cumulative distribution. In the implementation there would usually be a much larger number of frames than the 11 shown in this simplified figure.

to the great difference between auditory acuity in regards to these two auditory attributes. Short sound files, large window sizes or small overlaps all decrease the number of frames, and therefore the number of data points, attainable from an audio file. Attention to any of these three elements may relieve a problem with either of the other two.

Five background noise recordings are represented in this manner as a demonstration. Figure 3 shows a typical bar chart of the median A-weighted SPL for each recording, but with an ESA based auditory representation of the median level the differences are directly audible.

3.2. Cumulative Distribution Function

The cumulative distribution function describes the probability of a single auditory parameter x having a particular value. The cumulative distribution is the probability that a random choice of a number in the set x is less than a reference value. For instance, if a sound sample exhibits a sound pressure level greater than 60 dB for 160 s of a 200 s recording we may say the cumulative probability of 60 dB is 0.8. By representing all the values of x taken from each recorded time window in monotonically increasing order, we produce a contour that represents the cumulative probability as a function of x.

This is often represented visually, but it can also be sonified through reorganisation and concatenation. As demonstrated in Figure 4, once the descriptor data has been calculated, the rearrangement order can be based on this data. The frames are then interleaved using a short overlap and concatenated to form the sonification.

As an example, a sample of shakuhachi performance has been analysed and represented. This is a short sample of 15 s length, chosen for the combination of a large range within intensity, breathiness and pitch . In Figure 5 we will take the Cumulative Distribution of the F_0 of the note (as determined by the SWIPEP algorithm [18]).

3.3. Boxplot

Contour based representations are useful in describing the fine detail of a single distribution, but are not so useful for comparing multiple distributions. Each contour provides so much information that like by like comparison is likely to overwhelm. The 'boxplot' was based around the five-number summary which Tukey invented as a way of numerically summarising a distribution [6]. These numbers were (in order of graphical placement): the median; the 25th and 75th percentile; the maximum and minimum. The boxplot is a representation of these numbers, with a line in the middle for the median, a box around the 25th to 75th percentile range (the inter-quartile range), and whiskers extending to the upper and lower points that are 1.5 times the interquartile range from the box edge. Using the method described in Section 3.1 to create a sound for each of the five numbers, and then concatenating the resulting five sounds, it is possible to produce a sound that represents the five numbers; an auditory boxplot. This is a relatively short sound, and thus many of them can be listened to quickly, allowing comparisons to be made.

As a demonstration, the shakuhachi performance discussed above is reanalysed, and an auditory boxplot demonstrates the loudness distribution in the sample (Figure 6).

3.4. Peak Analysis

All varying signals contain peaks or troughs of some kind. In many cases these peaks or troughs are the main points of interest an investigator uses to understand a signal. In other cases the peaks and troughs form a periodic signal that is understood as a whole, as

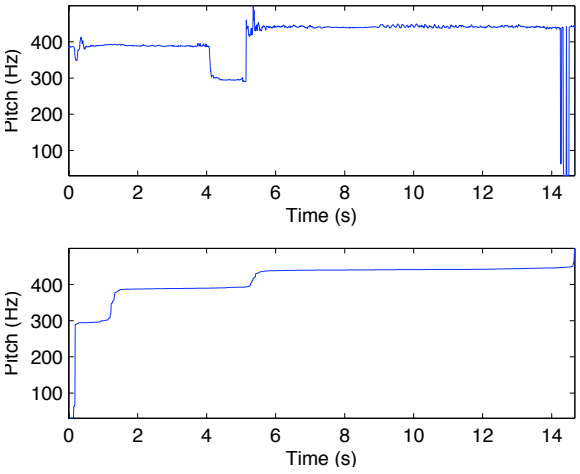

Figure 5: The cumulative distribution of the pitch of a shakuhachi recording.
Audio example.

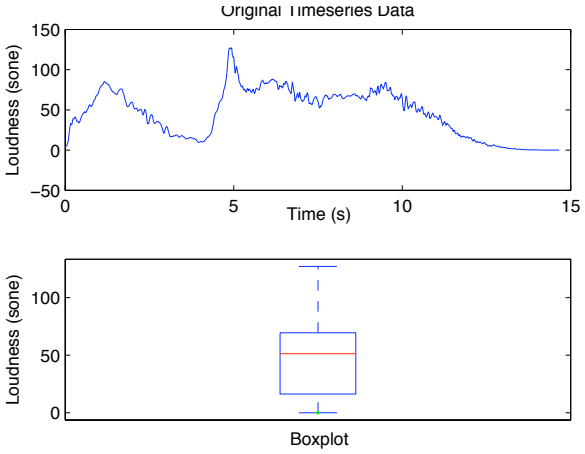

Figure 6: A visual boxplot of the loudness of a shakuhachi recording. The values that form each of the horizontal lines in the plot are sonified directly in the ESA representation.
Audio example.

for instance, the vibrato of a singer is understood. In both cases we can use timeseries analysis to find these points, and then follow up with further analysis.

A simple algorithm to find the points that are considered to be peaks is to look for changes in the direction of the rate of change. Firstly, the signal is differenced by subtracting each point from its neighbour. This results in a signal 1 sample shorter than the original signal that represents the amount of change between neighbouring points. The sign of the resulting difference signal represents whether the original signal was increasing or decreasing at each point, and change in the sign represents either a peak or a trough in the signal. Peaks result in changes from positive slope to negative slope, and troughs go from negative to positive.

The time information for each of these peaks and troughs can be used to find the time between peaks, or between peaks and troughs, which can describe periodicity. The values of peaks and troughs can be subtracted to find the range, which is useful for assessing whether a signal is continuous or varying.

In Figure 7 an example of vibrato analysis (of a single sung note) shows a fairly constant frequency difference between peak and trough. We can choose to represent this information by zeroing out each of the frames in between the peaks or troughs, or by extracting the frames at wither the peaks or troughs and concatenating them. There will be an almost identical number of peaks and troughs, so each set may be placed in two ears for simultaneous comparison.

4. DISCUSSION

Some of the strengths of the proposed technique are worthwhile demonstrating.

4.1. Domain Correspondence

The correspondence between the data and the domain assists the understanding of the data. A benefit already discussed is that this display is entirely within the auditory domain, and as such it can be used by blind or visually impaired users. A more direct method of audio analysis is likely to be of benefit to these users.

4.2. Abstraction Limitation

Another difference with this method is that the process of representation is not as abstract as numerical/visual methods. When a pitch (for example) for a particular audio sample is calculated it becomes a single representative number – all the other features within that audio sample are ignored. If that pitch is then represented visually it is devoid of its context, and the users ability to aurally correlate the existence of that pitch with other features that may coexist at the same time in the sample is diminished. Through the proposed technique we preserve this user ability, and extend it by reorganising the sound so that this ability may be easily employed. If the sound is reorganised based upon pitch, we will obviously hear increasing pitch in the reorganised sample, but we may also hear a correlation rising sound pressure level if a correlation between pitch and sound pressure level exists in the sample. This is useful for finding these correlations.

4.3. Grouping through similarity

Another benefit is that groups of sounds will pop out. It is possible to see in a visual graph of the SPL of a sound (for instance) that there may be various groups of sounds. However, it is still quite difficult to listen to those sounds, or to associate that group with some characteristic of the sounds except by inference. As an example, we may infer that the loudest sounds in the sample of singing are likely to be the transient consonants produced at the start of notes, and that the quieter parts of the sound are the ends of phrases. However, when we listen to a reorganised version of the sound we can hear a number of transient like sounds at the early part of the sound, amongst the rest of the low SPL sounds. We then infer that a lot of the transient sounds are being assigned low SPL by the descriptor algorithm. The transients have a high SPL, but preceding them there is often silence, and thus an average of the two results in a much lower SPL being assigned them. This effect would not be represented in a visual graph, but in this auditory graph it sticks out.

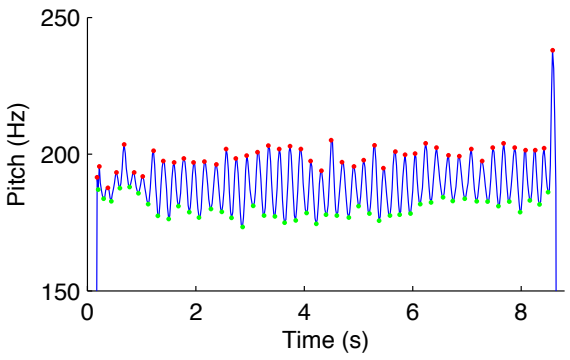

Figure 7: Peaks (in red) and troughs (in green) of timeseries pitch data. Playing only the frames associated with either the peaks or troughs lets us listen to them in isolation.

4.4. Implicit Feature Description

This method of representation also explains the feature being represented as it applies to the sample. Some acoustic features are quite complex to describe, as they often involve multiple signal processing stages, possibly followed by statistical summaries being applied to the output of these signal processing stages. Especially those effects within the spectral domain are extremely complex to describe. By reorganising the sound we create a monotonically increasing example of change within this descriptor, as long as there is substantial change within this feature. Not only this, but an example is created of how this feature applies to this audio sample, rather than to some artificial case. Through reorganisation we solve the problem of how to explain the graph's subject at the same time as presenting it.

Reorganisation is also useful as it can be used to intersperse reference points within the sound, that can be rearranged along with the sound itself. As long as the reference point is sufficiently recognisable and precisely stable, it will be relatively easy to pick it out in the reorganised version, as it will all be reorganised to be one group together. This could be as easy as recording a calibration tone at the beginning of the sound, and then reorganising the sound based on Sound Pressure Level. This would mean all levels below the calibration level would be organised to precede the calibration level, and all higher would follow it. A set of calibration tones at various levels (or one calibration tone altered with systematic level shifts) could be concatenated with the original sound before the sound is reorganised. The effect of this would be that the tones would be interspersed with the audio data forming a rhythm. The time between successive tones will obviously be shorter depending on how much audio data exists between the tones. Alterations to this technique can be used for the other analysis algorithms.

4.5. Auditory Relevancy

Some acoustic descriptors are strongly correlated with various perceptible changes in the sound source, while others are often quite irrelevant. This is sometimes difficult to see in a visual graph – while there may be many changes in the contour of the graph, as it is difficult to work out which exact part of the sound is related to the data. ESA allows quick auditory comparison between sounds reorganised based on perceptually relevant or irrelevant descriptors. The reorganised sounds should yield a contour that can be perceived by the user, if the descriptor is perceptually important. If the descriptor is irrelevant however, the reorganised sound will be chaotic and will not form a contour that can be heard by the user.

Using a cumulative distribution function ESA representation, it is possible to compare descriptors for particular sound files, to see whether the variation of a descriptor corresponds with a perceptible change in the cumulative distribution function.

4.6. Decimation for Timebase Contraction

The most obvious difficulty with representing a sound with a reorganised version of the same sound is that you have to listen to the whole sound again, albeit reorganised. This is not a problem if the sound is short, like it is for a speech phrase, but it becomes impossible if the sound is much longer. The representation of background noise is an obvious example where a measurement time would often be in the order of many hours or even days. Putting to one side the obvious impracticality of this situation, there is a more serious problem with the perception of long audio representations. Auditory adaptation rates are relatively fast, and the reorganisation process is likely to place windows of sound next to each other that have similar values for the descriptor concerned – very likely to within the difference limen for that descriptor.

The use of a shorter sample to represent the same is likely to ameliorate some of the confusion caused through adaptation over the playback period. As we see each of the windows within the sound as separate data points, unconnected with the rest of the sound, we are free to reorganise them as we see fit. While this reorganisation must still represent the data accurately, there seems no requirement to use all of the windows to achieve this. As they have been organised in this case to form a contour, based on an increasing or decreasing parameter, selecting a smaller subset of these data points will not affect the representation of the contour substantially. This technique is often used for simplifying the rendering speed and memory requirements of computer based visual representations of sound. However, in the case of timebase decimation it also has an effect on the usefulness of the display – and is not just for performance reasons.

5. IMPLEMENTATION

The above concepts are implemented in the software package *PsySound3*, which has been discussed previously by [19]. Many of the representation techniques discussed above can be applied to any audio descriptor data that PsySound3 can produce. As the design of PsySound3 is modular, any numerical descriptor data that can be produced by an audio analysis algorithm may be used to build an audio analyser in the software. The framework of ESA is generic and may be applied to any descriptor algorithm that outputs one numerical datapoint per frame. This means that the various ESA representations may be applied to any timeseries descriptor output that can be produced by PsySound3. The various representations may be applied to timeseries data outputs from PsySound3's output interface, and depending on the number of outputs selected various options are highlighted for use (see Figure 8). Decimation and other options are also available for selection when building these auditory representations. During the oral presentation, the software will be

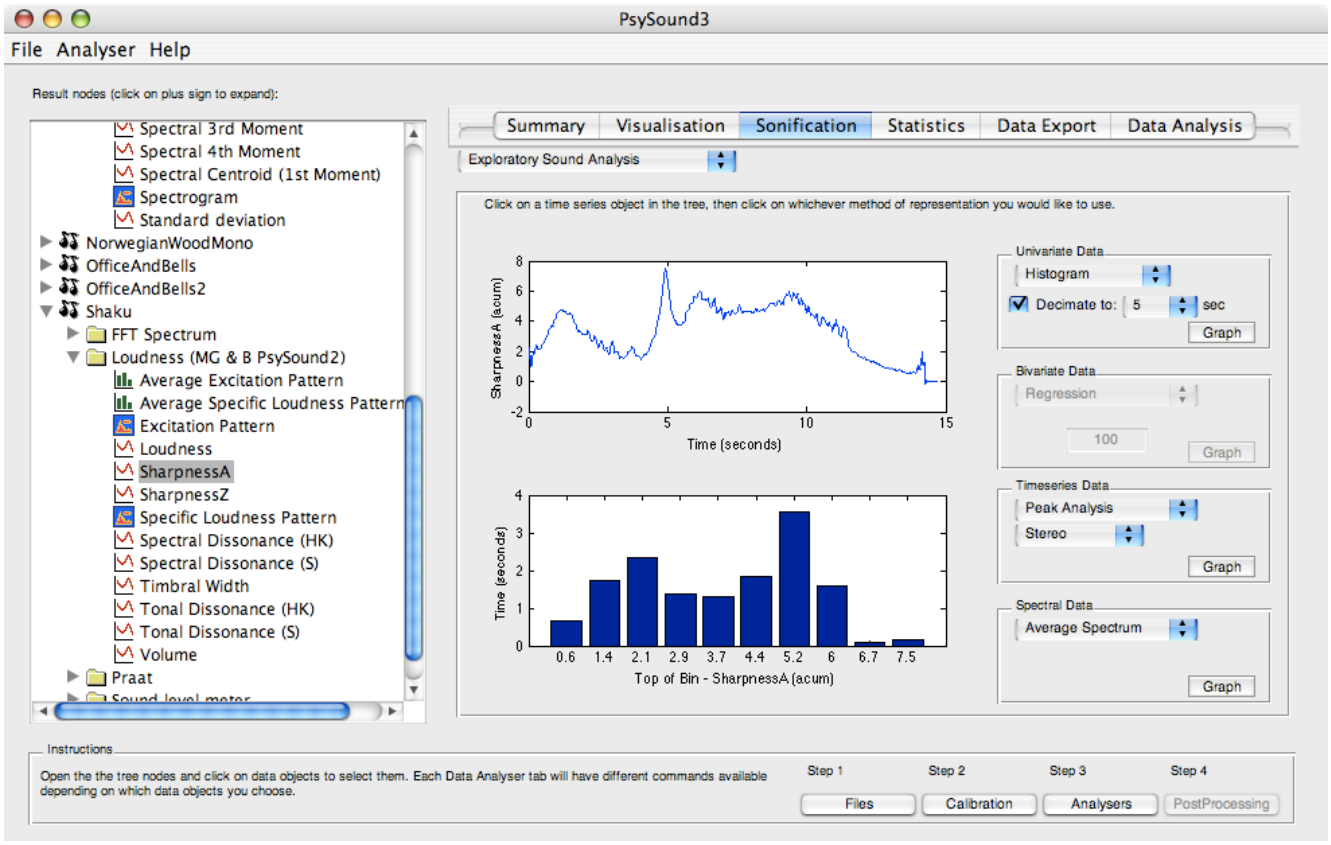

Figure 8: Exploratory Sound Analysis is implemented as an analysis method within PsySound3.

applied to various sound files and the results sonified. PsySound3 is publicly available at http://www.psysound.org.

6. APPLICATIONS

This framework may be applied to many fields within audio and acoustics. As it is a statistical representation technique that can be applied to any set of audio data, it can be applied wherever other statistical techniques are used. The current set of examples are a sample of possible representation techniques, but many other representations that deal with various statistical problems may be built if the sound is decomposed into frames.

7. CONCLUSION

We have described why there is a need for an alternative auditory representation of features within sound. We have presented a technique for building statistical representations by breaking sound up into parts and using those parts, rearranged, to sonify data about the original sound. There are a number of benefits to this technique, including its possible usefulness for blind and visually impaired people, its lack of abstraction, its ability to explain the parameters being represented, the grouping effect it has for similar sounds, and the ease with which it can be decimated to provide a summary of the data.

8. REFERENCES

[1] D. Schwarz, *Data-driven Concatenative Sound Synthesis*, Ph.D. thesis, University of Paris 6 Pierre et Marie Curie, 2004.

[2] D. Schwarz, "Concatenative sound synthesis: The early years," *Journal of New Music Research*, vol. 35, no. 1, pp. 3–22, 2006.

[3] D. Cabrera and S. Ferguson, "Sonification of sound: Tools for teaching acoustics and audio," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, Canada, 2007.

[4] S. Ferguson, D. Cabrera, H-j. Song, and K. Beilharz, "Using psychoacoustical models for information sonification," in *Proceedings of the 12th International Conference on Auditory Display*, London, UK, 2006.

[5] C. Roads, *Microsound*, MIT Press, Cambridge, 2001.

[6] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Reading, Mass., 1977.

[7] B. Fry, *Computational Information Design*, Ph.D. thesis, Massachusetts Institute of Technology, 2004.

[8] B. Shneiderman, "Tree visualization with tree-maps: 2-d space-filling approach," *ACM Transactions on Graphics (TOG)*, vol. 11, no. 1, pp. 92–99, 1992.

[9] M. Wattenberg, "Visualizing the stock market," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI99)*, Pittsburgh, Pennsylvania, 1999.

[10] T. Ong, H. Chen, W. Sung, and B. Zhu, "Newsmap: a knowledge map for online news," *Decision Support Systems*, vol. 39, no. 4, pp. 583–597, 2004.

[11] M. Wattenberg, "Arc diagrams: Visualizing structure in strings," in *IEEE Symposium on Information Visualization 2002*, Boston, Massachussets, 2002, pp. 110–116.

[12] J. Snydal and M. Hearst, "Improviz: Visual explorations of jazz improvisations," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI2005)*, Portland, Oregon, 2005, pp. 1805–1808.

[13] B. Shneiderman and C. Plaisant, "Chapter 14-5: Information visualization," in *Designing the User Interface*, Pearson, Boston, pp. 580–603, 2005.

[14] G. Kramer, B. N. Walker, T. Bonebright, P. R. Cook, J. H. Flowers, Nadine Miner, and John G. Neuhoff, "Sonification report: Status of the field and research agenda," Tech. Rep., National Science Foundation, 1997.

[15] J. H. Flowers, "Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions," in *Proceedings of the 11th International Conference on Auditory Display*, Limerick, Ireland, 2005.

[16] B. N. Walker, "Magnitude estimation of conceptual data dimensions for use in sonification," *Journal of Experimental Psychology: Applied*, vol. 8, no. 4, pp. 211–221, 2002.

[17] P. Vickers, "Whither and wherefore the auditory graph? abstractions and aesthetics in auditory and sonified graphs," in *Proceedings of the 11th International Conference on Auditory Display*, Limerick, Ireland, 2005.

[18] A. Camacho, *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*, Ph.D. thesis, University of Florida, 2007.

[19] D. Cabrera, S. Ferguson, and E. Schubert, "Psysound3: Software for acoustical and psychoacoustical analysis of sound recordings," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, Canada, 2007.