Strategic Hyperspectral Detection through Sonic Transformation

M. Bernhardt, J.P. Heather and P.F. Edmundson
Waterfall Solutions Limited
Parklands, Guildford, Surrey, GU2 9JX, UK

Abstract

There are applications in which it is necessary for people to interact with hyperspectral imagery data. In particular these include strategic and other reconnaissance tasks. Hyperspectral data is high-dimensional and therefore intrinsically hard for people to properly comprehend (in the way they can easily understand a single-band image). Therefore additional tools to assist analysts are required. This paper describes work that has made use of sound as an additional means of encoding the properties of hyperspectral data. A transformation has been devised that renders spectra into audible tones in such a way that background clutter makes 'consonant' sounds whereas anomalies sound dissonant. We show the results of trials that indicate that using this method to screen RX detections can reduce the false alarm rate.

Keywords: Sonic transformation, HCI, hyperspectral data

Introduction

Hyperspectral reconnaissance imagery has a number of uses at one end of the scale of uses might be its tactical use on the battlefield for the real-time identification of targets of opportunity of confirmation of pre-designated targets. This type of application requires fast (real-time) processing algorithms. At the other end of the scale is its use in intelligence gathering and monitoring. In these circumstances the data is analysed by a hybrid approach involving both people and automated processing. An analyst would typically make use of tools such as ENVI with specific libraries of processing algorithms that could be employed under her direction. Even in these circumstances it is crucial for the analysis timeline i.e. the time from receiving the data to producing the analysis product (specific intelligence information) be as short as possible.

In strategic intelligence gathering from hyperspectral imagery the data is analysed by a hybrid approach involving both people and automated processing. People and in particular image analysts have a very highly developed ability to recognise things spatially. Comprehending high-dimensional data sets though is something that people, even with experience, are rather poor at. Because of this any tools that can genuinely assist in the comprehension of hyperspectral data will assist an analyst in her task.

We describe a novel tool to assist human operators in their interpretation of hyperspectral data via real time sonic transformations based upon the notions of consonance (where simultaneous sounds are pleasant to the ear) and dissonance (where simultaneous sounds are harsh and unpleasant). These notions can be quantified to an extent by drawing upon psycho-acoustical research outlined in this paper.

The human ear is a very complex and (at least from a neurological point of view) not fully understood acoustic sensor [1]. One important aspect of the ear is that it has about 20 000 tiny hair-like nerve cells which (after a complex mechanical process) vibrate in sympathy with acoustic waves in the liquid of the inner ear. The important feature of these cells is that by their physical make-up they are each ‘tuned’ to resonate at a different frequency. Hence the response from this collection of nerves could be regarded as a kind of continuous ‘analogue Fourier transform’. This effectively makes the ear in some sense inherently a spectral sensor – albeit in the acoustic domain.

Although we are not aware of anyone having applied the idea of transforming complex data into sounds in order to assist comprehension to remote sensed hyperspectral data the use of this approach for other types of data has been investigated before. However we feel that it might be particularly pertinent to this area of investigation.

Theory of Consonance and Dissonance.

Musicians have known for centuries which frequencies (pitches) have a pleasing or consonant effect when played simultaneously, so for example the so-called perfect fifth (e.g. playing C and the G above it together on a piano) has a very consonant quality. The frequency ratio of the higher to the lower pitch in this example is $\frac{3}{2}$ (in the ‘just’ intonation tuning system – actually in the modern ‘equal temperament’ tuning system (devised by Bach to enable) this is $2^{7/12} = 1.4983$ which differs from the ‘ideal’ value by about 0.1%). Other frequency ratios are particularly un-pleasant or dissonant. For example the minor second (corresponding to playing C and C# together on a piano) with a frequency ratio of about 1.06 is particularly unpleasant! Generally speaking the better approximated a frequency ratio by a low denominator rational number the more pleasing it sounds. The reasons for this are not fully understood, however it appears that there are both physical and psychological reasons. There has been quite a lot of research done in this area which we will exploit.

Studies in dissonance [2] and more recently [3], based on a curve fit to a large amount of experimental data on volunteers. Experiments were conducted in which different frequency sine waves were simultaneously played to people who rated their ‘degree of dissonance’. This data was collated, normalised and analysed to produce the following equation:

$$C(f_1, f_2, v_1, v_2) = v_1 v_2 \left(\exp(-as\left|f_1 - f_2\right|) - \exp(bs\left|f_1 - f_2\right|) \right)$$

where f_1 and f_2 are the frequencies of the two sine waves, v_1 and v_2 are their respective amplitudes, s is given by

$$s = \frac{d^*}{s_1 \min(f_1, f_2) + s_2}$$

and $d^*=0.24$, $a=3.5$, $b=5.75$, $s_1=0.021$, $s_2=19$ are empirically determined constants.

A plot of this function is shown in figure 1

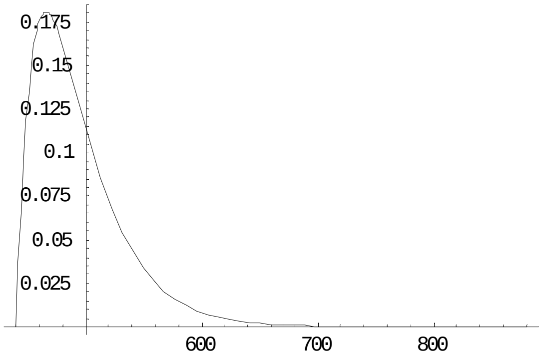

Figure 1 *Dissonance between sine wave at 440Hz and sine wave whose frequency is given on horizontal axis*

This analysis extends to more complicated waveforms by Fourier analysis through an assumption of linearity where the overall dissonance is just the sum of dissonances over all pairs of Fourier components. For example if a saw-tooth waveform is chosen then the dissonance curve looks like figure 2 where the horizontal axis is now calibrated as a multiple of the base frequency (between 1 and 2). Note that many of the minima of this curve occur at the frequency ratios of the notes of the western Major scale [3].

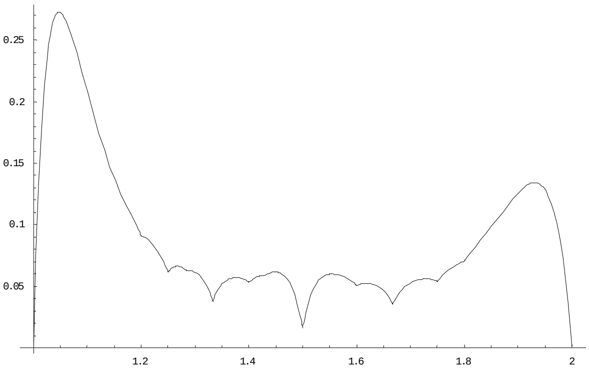

Figure 2 *Dissonance curve for saw-tooth waveforms of frequency 440Hz and 440m Hz. Horizontal axis is multiple of (m) of base frequency.*

Our aim is to make use of these established psycho-acoustical properties in determining good mappings from the spectral domain into the sonic domain. We have devised a mapping such that the background clutter has a very consonant sound whereas the any anomalies stand out by virtue of their dissonance in relation to the background.

A transformation was devised based upon mapping spectral bands (or linearly projected spectral bands) directly to sets of partials whose frequencies were determined by optimising over the dissonance curve. In addition to the transformation based upon the dissonance theory described above we also devised a baseline transformation based upon a simple logarithmic scaling of frequencies between a minimum and maximum frequency chosen by the user.

Description of the Tool

We have developed a tool in Java (for portability). The tool allows a user to read in a hyperspectral cube in any of the main formats and any general format specified by the dimensions of a cube and the orientation of band data. This cube is then displayed in the window, allowing visual inspection of any chosen band. There are controls that allow the user to choose the particular mapping under investigation and determine the basic waveform to be used with that mapping. If the mouse is moved over the image display then the pixels over which it moves are rendered into sound in real time according to the selected mapping.

For initial tests, an option to read in ground truth data associated with a cube together with algorithm detection data was implemented. This data was represented as a set of red dots on the screen at the locations of targets and false alarms. The user is then invited to listen to the data and click on any red dots that they feel are actually genuine anomalies. When they have finished the program generated their score in the form of a ‘confusion matrix’ showing the true class (anomaly or false alarm) in the rows versus the chosen class in the columns, allowing initial trials to be conducted.

Initial Experiments

The aim of the initial experiments was twofold: firstly to assess the potential benefit of sound processing or “sonification” – i.e. does it carry any useful information that can be used to assist in understanding hyperspectral data. Secondly to see whether the ‘dissonance’ approach described earlier had more utility than a baseline approach which did not make use of the notion of dissonance at all.

The experiments were designed around a simple concept of operation for a sonification tool (many other concepts of operations are possible). This concept supposes that it is unlikely that sonification would ever be used in isolation, that is it is very unlikely that an operator would search a whole cube for targets purely using a sonic representation of each pixel. A much more likely scenario is to use sonification in conjunction with existing algorithms. The concept upon which the trials were based is that another algorithm (for example RX) would be used initially to screen a hyperspectral data cube to select a number of pixels or areas of interest. These would then be interrogated by the user using sonification in order to determine a priority order or remove false alarms.

We ran the well known RX [4,5] algorithm on AVIRIS [6] data which consisted of vegetation backgrounds. Into this vegetation some anomalies had been embedded which were pixels from an urban area of the same cubes. These pixels were embedded at a range of mixing fractions with their local background – i.e. all anomalies were sub-pixel. The location of these anomalies was chosen randomly by the algorithm so that the trial would be blind. The RX algorithm was run on the cube with the threshold level set so that it detected all of the embedded anomalies. When run at this threshold the RX algorithm also detects a number of false alarms some of these were retained for use as false alarms in the experiments.

Shown in tables 1 and 2 are confusion matrices produced from blind trials where anomalies are embedded with 50% embedding strength. Tables 3 and 4 show confusion matrices for 30% embedding strength.

		Declared as	
		RX False Alarm	Anomaly
Truth	RX False Alarm	90%	10%
	Anomaly	10%	90%

Table1 *Confusion matrix for dissonance based sonification algorithm*

		Declared as	
		RX False Alarm	Anomaly
Truth	RX False Alarm	80%	20%
	Anomaly	20%	80%

Table2 *Confusion matrix for baseline based sonification algorithm*

		Declared as	
		RX False Alarm	Anomaly
Truth	RX False Alarm	84%	16%
	Anomaly	14%	86%

Table3 *Confusion matrix for dissonance based sonification algorithm*

		Declared as	
		RX False Alarm	Anomaly
Truth	RX False Alarm	74%	26%
	Anomaly	28%	72%

Table4 *Confusion matrix for baseline based sonification algorithm*

Tables 5 and 6 show confusion matrices for a 20% embedding strength.

		Declared as	
		RX False Alarm	Anomaly
Truth	RX False Alarm	61%	39%
	Anomaly	41%	59%

Table 5 *Confusion matrix for dissonance based sonification algorithm*

		Declared as	
		RX False Alarm	Anomaly
Truth	RX False Alarm	56%	44%
	Anomaly	46%	54%

Table 6 *Confusion matrix for baseline based sonification algorithm*

Note that even for the first experiment where the embedding strength is relatively high that RX has a considerable number of false alarms in order to detect all the targets.

It is clear from these results that sound mapping contains substantial information that can be used to allow users to discriminate between objects that RX cannot. It is also clear that the dissonance based approach is better than the baseline algorithm – although even the baseline algorithm manages quite good performance. At embedding strengths much above 50% the human performance was virtually perfect (whereas RX was still generating large numbers of false alarms). This performance drops off steadily until an embedding strength of just below 20% is reached when there appears to no longer be enough information in the sounds for users to perform any better than guessing.

Although the initial experiments were not over a large data sample, it is large enough to justify the overall trend that we have observed.

Summary

It is also clear from these early experiments that sonification can be a very useful tool to assist the analysis of hyperspectral data. The experiments showed that it is possible to outperform RX in terms of differentiating between genuine anomalies and false alarms. This in itself should not be too surprising since the human brain is very sophisticated compared to RX. However this indicates that there was much useful information coded in the sounds produced by the tool. It is also interesting to note that the dissonance based approach which made use of psycho-physical theoretical results did outperform the baseline algorithm – screening out about twice as many false alarms. It is also worth noting that some testers noted that the dissonance based approach was also a little ‘easier on the ear’. It appears that using our current mapping, 20% embedding strength is about the threshold where it becomes quite hard to differentiate between RX false alarms and genuine anomalies.

We are now working to devise a number of new spectral to acoustic mappings which will improve the ease of interaction and increase the number of tasks that this approach can be used for.

References

1. F.H. Bess. “Audiology: the fundamentals”, Baltimore: Williams and Wilkins. 1995 ISBN 0683006207
2. R. Plomp and W.J.M. Levelt, “Tonal consonance and the relationship between timbre and scale”, J. Acoust. Soc. Am. 38, 548-560. (1965).

3. W.A. Sethares, “Local consonance and the relationship between timbre and scale”, J. Acoust. Soc. Am. 94, 1218-1228, (1993)
4. I.S. Reed, X. Yu, “Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution”, IEEE Trans. Acoustics, speech and signal processing”, vol. 38, no. 10, Oct. 1990
5. X. Yu, L.E. Hoff, I.S. Reed, A.M. Chen, L.B. Stotts, “Automatic target detection and recognition in multiband imagery: a unified ML detection and estimation approach”, IEEE Trans. Image Proc. vol. 6, pp143-156, Jan. 1997
6. http://aviris.jpl.nasa.gov

Acknowledgements

The work reported in this paper was funded by the Electro-Magnetic Remote Sensing (EMRS) Defence Technology Centre, established by the UK Ministry of Defence and run by a consortium SELEX Sensors and Airborne Systems, Thales Defence, Roke Manor Research and Filtronic.