

SAR Detection Through Sonic Transformation

W. Oxford
Waterfall Solutions, Parklands,
Guildford, Surrey GU2 9JX.

Abstract

In surveillance applications there is generally an operator in the loop. The active nature of Synthetic Aperture Radar (SAR) image formation represents a particularly challenging problem to all but experienced image analysts. Also, the imagery available for exploitation may not represent all of the useful information contained within the SAR data. This project investigated the utility of using sound (sonification) for the discrimination of targets in SAR imagery. Several sonic mappings were developed to translate image features into sounds. An assessment of the sonic mappings showed that users were able to discriminate between targets and false alarms as successfully as an automatic classification process. Together with previous work this implies that sonification is best applied to problems described by high-dimensional features.

Keywords: Synthetic Aperture Radar, Sonification, Sonic Mappings, Classification

Introduction

It is tempting to treat a SAR image as if it were collected by an electro-optic (EO) sensor. However, it is formed very differently and its interpretation presents unique challenges. This project was motivated by previous research conducted by Waterfall Solutions Ltd [1][2] which demonstrated the benefit to analysts in 'listening' to images containing targets that were difficult to see by eye, and was particularly applicable to high-dimensional (hyperspectral) data.

SAR data bears some resemblance to the high-dimensional data used previously in that it may contain useful information that is difficult to visualise (e.g. high dynamic range and phase map). This makes SAR data a candidate for sonification, which is defined as the 'use of non-speech audio to convey information'.

Sonification is enabled by means of an acoustic mapping that encodes the

information contained in the data as changes in sound timbre (also known as sound 'quality' or sound 'colour'). Timbre is the quality of sound that distinguishes it from other sounds and includes spectrum as well as envelope.

The aim of this project was to develop methods of turning SAR data into sounds to assist human operators. The study began by examining image exploitation tasks in concert with the aspects of SAR phenomenology that lend themselves most naturally to the development of sonic mappings. Alongside the nature of the available test data this directed the study to examine the task of tactical target discrimination.

A framework capable of assessing this task was developed. The framework included the insertion of simple targets into the test data, feature extraction, sonification, and assessment.

Several acoustic mappings were employed by the framework. The first was the most successful mapping developed under the previous study. New mappings were also developed using an implementation of a digital synthesiser. This was capable of generating musically recognisable sounds more pleasing to experience than those generated in the previous study.

Finally, the effectiveness of the sonic mappings was evaluated via user trials. This provided quantitative performance results for the comparison of the different mappings with the performance attained by automatic algorithms.

This paper now discusses the selection of the exploitation task, introduces elements of the framework, and presents an analysis of the results from the user assessment.

Exploitation Tasks

Previous work [1][2] concluded that the tasks which benefit most from the addition of sound are those in which the visual display cannot convey all of the available information to the user, such as sub-pixel target discrimination in hyperspectral imagery. In relation to the exploitation of SAR data, this includes the tasks that follow.

Reduced resolutions: It is often convenient to visually scan an overview of the imagery. This means that the image will be re-sampled and spatial detail will be lost. Also, SAR imagery typically has a very high dynamic range that far exceeds the capability of the human eye, and mapping the data to the visual display is lossy. A sonic mapping may be capable of representing this 'missing' information, or of informing the image analyst of regions that are not well reproduced to indicate that closer inspection is required.

Time series: A SAR image may be produced at a reduced spatial resolution by using only a fraction of the available synthetic aperture. A number of these images constitute a time series in which dynamic events (e.g. glint) may be observed that would otherwise be overlooked. By representing the temporal information via a sonic mapping, the analyst would then have the benefit of the spatial resolution afforded by the entire synthetic aperture.

Multi-polarisation: Some SAR systems are capable of transmitting and receiving energy of different polarisation states. The resulting multi-band images formed contain information useful for target detection and discrimination. It may be possible to construct sonic mappings that are indicative of the polarisation signature of targets of interest.

Target discrimination: Targets can appear very different in SAR imagery as to how they would in EO imagery. In a tactical scenario (e.g. a fast jet) the pilot or operator may not have the time nor experience to visually identify potential targets. Sonic mappings may be useful in assisting an overloaded operator in the task of discriminating targets in clutter.

The data available to the project comprised complex imagery. The tasks suited to this data were the reduced resolution and the target discrimination tasks. It was decided that an investigation of the latter task would provide better prospects for exploitation, and so this formed the focus of subsequent research.

Data Generation

A part of the assessment aims, full control over the signature or strength of the targets of interest was required. Random sub-images of the test image were extracted, into which small targets (10x10 pixels) of a

specified strength were inserted. A spatial matched filter was then used to find an equal number of false alarms.

Sonic Mappings

The first sonic mapping, the spiral mapping, was developed under the previous work [1], and was selected since it was the most successful in discriminating unresolved targets from false alarms in hyperspectral imagery. The spiral mapping generated sounds directly from the imagery using the pixel values in the vicinity of the potential targets.

The remaining mappings were based on the development of a digital synthesiser. The motivation for this development was to generate sounds that were not unpleasant to experience, a fair criticism of previous sonic mappings. The timbres of the sounds generated by the digital synthesiser were modified via the features extracted from the test imagery in the vicinity of the potential targets.

Feature Extraction

The features required for the Spiral Mapping were simply the pixel values in the vicinity of the potential targets. The mapping is so-called as the sampling locations, relative to the potential target, moved in a spiral pattern. This idea was taken further in this project by representing the image at various scales.

The image was firstly down-sampled three times to create an image pyramid at different scales (i.e. 1, 4 and 16). The sampling locations for the spiral were determined by a maximum radius and the speed with which the locus moved outward/inward in the spatial dimension of the image (spatial speed) and also back and forth through the image pyramid (spectral

speed). Figure 2 shows an example of the sampling locations for a maximum radius of 10 pixels.

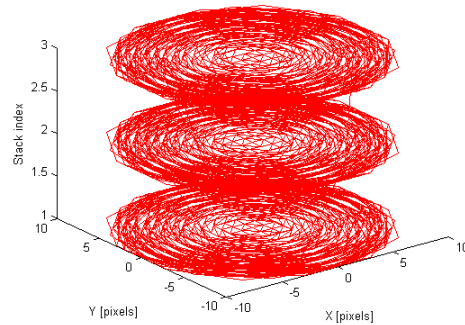


Figure 2 – Typical spiral sampling pattern

The timbre of sounds generated by the digital synthesiser could potentially be manipulated using a large number of parameters. However, it was found that the manipulation of three key parameters gave a wide range of sounds whilst retaining their musical character. Therefore, three discriminatory image features were extracted to map to the parameters of the digital synthesiser. These features leveraged the structured shape and size of the synthetic targets.

Sound Generation

The spiral sampling locations were generated using a maximum radius, spatial speed and spectral speed selected based on the size of the targets and a brief assessment of the discrimination afforded between targets and false alarms. The pixel values extracted from the image pyramid around each cue location were treated as the output audio signal.

The remaining mappings were based on sounds generated by a digital synthesiser. The synthesiser was constructed in a manner inspired by early electronic instruments, as shown in Figure 3.

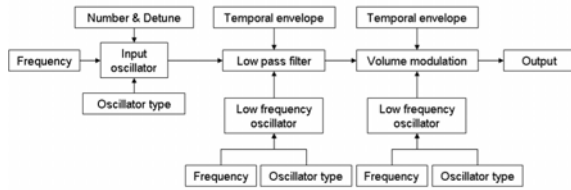


Figure 3 – Digital synthesiser

The input oscillator was defined by the waveform and frequency. The mapping of image features to oscillator frequency (as described later in this section) was performed linearly on the Mel Scale [3] to correct for the mismatch between actual and perceived pitch. More than one input oscillator could be defined (e.g. for a string instrument) in which case they would each be detuned slightly from the fundamental frequency.

The low pass filter (LPF) was a discrete-time replica of a voltage-control filter invented by Dr. Robert Moog [4]. This filter is well known for its characteristic sound and was a breakthrough development in electronic music at its inception in 1969.

The continuous time transfer function of this filter is given by:

$$H_a(s) = \frac{1}{s^4 + 4\omega_0 s^3 + 6\omega_0^2 s^2 + 4\omega_0^3 s + \omega_0^4 (1 + feedback)} \quad [1]$$

where s is the complex frequency, ω_0 (Hz) is related to the resonant frequency and feedback controls the amount of resonance, in the range ~ 0.4 to 4. Figure 4 shows the typical frequency response of the filter for different levels of resonance.

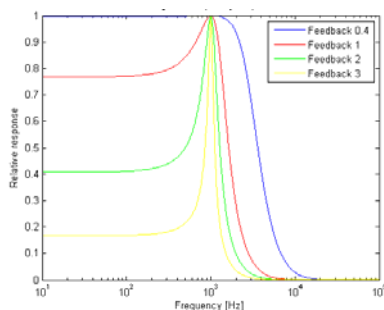


Figure 4 – Moog filter frequency response

The figure shows that at low values of feedback, the Moog filter behaves

somewhat similarly to a low pass filter. As feedback is increased, the resonant peak becomes more pronounced until, at a value of 4, the filter begins to self-resonate. For the generation of sounds, a feedback of 3 was used throughout.

The discrete-time version of the Moog filter was implemented by taking the Z-transform of Equation 1 and taking into account the phenomenon of frequency warping.

Both the LPF and volume modulation elements used temporal envelopes. Figure 5 shows a highly stylised temporal envelope defined in terms of its attack, decay, sustain and release (ADSR).

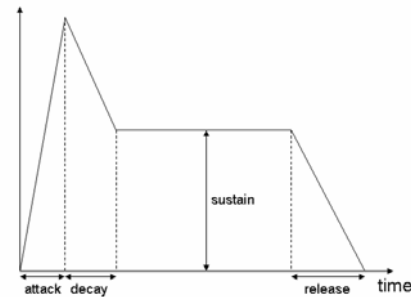


Figure 5 - Temporal envelope

For volume modulation, the envelope was used to directly multiply the input signal. For the LPF, the envelope was scaled by a gain factor, G , and used to dynamically alter the resonant frequency of the Moog filter, ω_0 . A low value of G resulted in a muddy sound whereas a higher value of G resulted in a brighter, cleaner sound.

The low frequency oscillators (LFO) were used to modulate the temporal envelopes used by the LPF and volume modulation elements. In both cases, the low frequency modulation was a sine wave and was simply added to the temporal envelope (in the case of the LPF, this was defined in units of Hz).

With an appropriate selection of parameters, the digital synthesiser was capable of generating sounds that resembled musical instruments. The key

parameters controlling the characteristics of the sound were found to be the LPF gain (G), LPF and volume envelopes, and the LPF LFO. Figure 6 shows the temporal envelopes of the instruments used in the assessment.

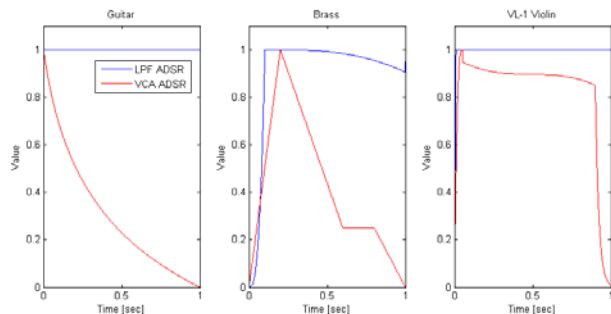


Figure 6 – Musical instrument envelopes

Feature Mapping

Three sonic mappings were devised using the digital synthesiser: the first used one instrument; the second used two instruments played simultaneously; and the third used one instrument which consisted of a combination of two instruments.

The key parameters that affect the timbre of a sound were judged to be frequency (pitch), LPF gain (G) and the LPF LFO amplitude. The manner in which the image features were mapped onto these parameters is an aspect of sonification that is highly empirical; the actual choice was made after a great deal of experimentation.

In summary, the one-instrument mapping bestowed targets with a high, pure tone and false alarms a low, poor quality tone; for the two-instrument mapping targets sounded like high frequency string and low frequency brass whereas false alarms were the reverse; and for the merged-instrument mapping targets sounded like a high frequency guitar whilst false alarms sounded like a low frequency brass.

User Assessment

The framework was executed for ten iterations at five target strengths. Each time the framework was executed sounds were generated for all four sonic mappings of 1sec duration sampled at 22kHz. A total of 104 assessments were made by four users.

In each assessment (defined by the sonic mapping and target strength) the users were first permitted to listen to training sounds and then asked to listen to and choose the class of test sounds. At any point the user could replay the test sound and compare it to a randomly selected target and false alarm sound from the training set. Once the user had completed the test, the results were recorded in terms of a confusion matrix.

Initial user feedback influenced modifications to the assessment. A reference tone was added before each sound in order to assist judgments concerning relative pitch and the two-instrument mapping was discarded, as users simply focused on the most noticeable instrument that did not itself carry information regarding all three of the image features.

Results

The results, in terms of correct classifications, are presented in Figure 7 (strength 1 is the highest and 5 the lowest).

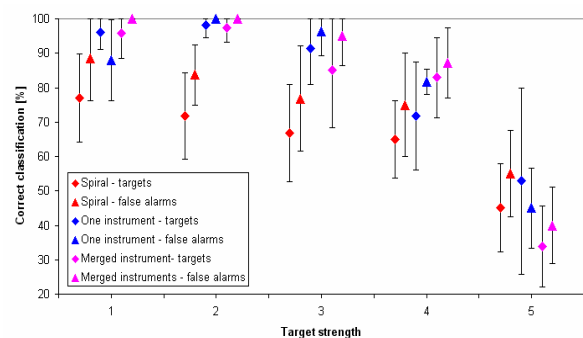


Figure 7 – User assessment results

Figure 7 shows that all three mappings were successful to varying degrees and the one- and merged-instrument mappings were significantly more successful than the spiral mapping. Overall, the merged-instrument mapping was the best performing.

At the weakest target strength, the merged-instrument mapping indicates that the users were able to discriminate targets but with the opposite effect to that intended. This is consistent with feedback which indicated that the discriminators, learned only using the highest target strengths (possibly in haste) swapped roles at the lowest target strength.

The features used for the one- and merged-instrument mapping were used in a Mahalanobis distance and nearest neighbour classifiers, repeatedly trained using one set of features and tested on the remaining sets. Figure 8 shows a comparison with the merged-instrument results.

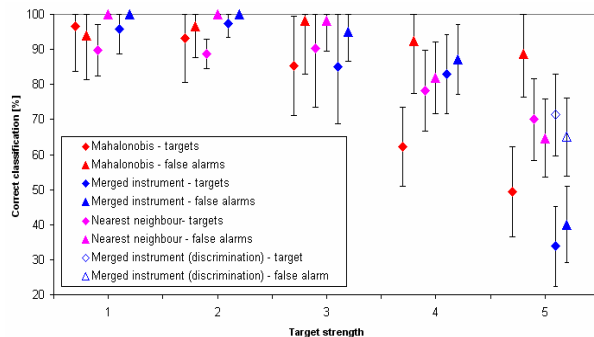


Figure 8 – Automatic classification and sonification results

Figure 8 includes results for the merged-instrument mapping using discrimination rather than correct classification as the performance metric. The figure shows that sonification performed very similarly to the automatic classifiers for all but the weakest target strength and, if the discriminators at this target strength had been learnt, for all target strengths.

Discussion

Overall, this study has shown that sonification provides performance on a par with automatic processes and so is not a technique that will add benefit for the task of SAR target discrimination. This result, valuable in its own right, contrasts with the outcome of the previous study that concluded that sonification had significant benefit when applied to the task of sub-pixel target discrimination in hyperspectral imagery. This seemingly contradictory result can be explained by the fundamental differences in the data used by the two studies and implies that the benefit of sonification lies in cases where the problem can be described by high dimensional features. The data available for this study did not support a high-dimensional analysis but the new sound generation techniques certainly do since the timbre of the resulting sounds can be potentially driven by a large number of parameters.

Several useful lessons were learnt during the development of the sonic mappings and the execution of the user trials. These include using short assessments, audio references, quick decision making, and feedback on the construction of future sonic mappings. These lessons will be very valuable for future work in this area in which we have gained a deeper understanding of the problems that might benefit from sonification and developed the tools needed to devise suitable new mappings.

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 of SELEX Galileo, Thales UK, Roke Manor Research and Filtronic.

References

1. Waterfall Solutions Ltd., Strategic Hyperspectral Detection Through Sonic Transformation, WS/06/EMRS/SON/ TR004, 27th October 2006.
2. Waterfall Solutions Ltd., Strategic Hyperspectral Detection Through Sonic Transformation: Human Factors Study, WS/07/EMRS/SON/TR005, 30th March 2007.
3. http://en.wikipedia.org/wiki/Mel_scale
4. www.till.com/articles/moog/patents.html#US03475623

