Timbral Segregation of Sonar Transients Using Auditory Processing Techniques

Simon Tucker and Guy J. Brown

Department of Computer Science, University of Sheffield, Regent Court,

211 Portobello Street, Sheffield S1 4DP, UK.

{s.tucker,g.brown}@dcs.shef.ac.uk

Abstract

By listening to sonar acoustic signals, operators can effectively distinguish man-made sounds from biological events. It has proved difficult to replicate such classification performance using conventional automated signal analysis techniques. This study investigates the timbre of sonar transients using a psychophysical multidimensional scaling study, and elucidates a three dimensional timbre space in which spatial distance corresponds to timbral similarity. Through analysis of this space, acoustic cues to the timbre of transient sounds are identified, following which an auditory model is constructed to automatically assign novel events to a position in the timbre space. The performance of the model was evaluated on a small data set consisting of mechanical and biological transients. The results are promising, showing a good segregation of biological and mechanical transients in the timbre space.

1 Introduction

Transient classification is typically undertaken by trained operators who perform classification largely through auditory perception. Human listeners are highly skilled in such classification tasks; therefore it seems likely that an auditory motivated approach to sonar signal analysis may prove advantageous. The use of auditory models for sonar signal analysis has largely been neglected in the literature; an exception is Teolis and Shamma's (1) study of the use of auditory models for transient classification. In their work, a wavelet transform was combined with a hair cell model as an auditory front-end, and the resulting representation was used as the input to a neural network classifier. Comparison of the auditory front-end with a conventional Fourier analysis showed that the auditory approach was superior.

Since sonar transients are isolated temporally, judgement is made on the perception of the event in isolation; therefore an approach which could prove successful is to discriminate transients by their timbre. Timbre is defined as being "that attribute of auditory sensation in terms of which a subject can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" (2). Attempts to quantify timbre have included the use of multidimensional scaling techniques (3), which isolate the acoustic properties used by human subjects to discriminate sounds of similar loudness, pitch and duration.

The use of multidimensional scaling (MDS) in timbre studies requires that a group of subjects rate the similarity of pairs of sounds; a space is then constructed in which distances between the sounds correspond to the measures of similarity provided by the subjects. Sounds rated as similar are clustered together in the space. The dimensions of this space are then correlated with acoustic properties of the sounds, in order to identify the properties used to distinguish them. MDS has been used to study the timbre of sustained musical instrument sounds, and indicates that properties such as spectral centre of gravity, onset slope and inharmonicity are used to discriminate timbre.

Since most timbral analysis uses sustained musical signals as stimuli, transient sounds have been largely overlooked. An exception is the study by Lakatos (4), which examined a set of musical and percussive sounds and found that spectral centroid, onset time and an unspecified measure of spectral richness accounted for the timbre of the percussive set. The signals examined here are inherently non-musical, so a goal of this study is to determine whether the timbre of non-musical transient signals can be measured using similar properties to that of musical sounds.

MDS studies have also been used to analyse the timbre of non-musical signals. Most notably, Howard (5) applied multidimensional scaling to ratings of similarities of selected sonar recordings, and identified a two-dimensional timbre space; the acoustic correlates found were spectral envelope shape and the degree of low frequency periodicity. This study will complement Howard’s work by examining short-duration signals, as opposed to the long-duration signals analysed in his study.

In summary, the study described here has three main motivations; firstly to examine the use of auditory models in sonar signal classification, secondly to examine the timbre of transient signals in general, and thirdly to examine the feasibility of analysing the timbre of sonar signals. Of particular interest, at this stage, is the ability to quantify the timbre of transient events and to be able to compare timbre of events. Also of interest is the ability to discriminate biological and mechanical timbres, since human listeners are highly skilled in this task.

Accordingly, a MDS study was undertaken using sonar transients as stimuli. A timbre space was constructed and acoustic correlates were found; these acoustic correlates were then used to recreate the space, and a new space was constructed from signals not part of the original stimuli set. An auditory model was used to process the sonar stimuli, allowing direct measurement of the acoustic properties relating to timbre.

2 Procedure

2.1 Participants

Nine subjects were chosen from postgraduates at the University of Sheffield and one was chosen from outside the university; none reported any hearing problems. All were naïve with regard to sonar signals. Since the signals were generally unpitched, musical knowledge was not considered a confounding factor in this study.

2.2 Stimuli

Classes of underwater transients were selected to represent possible submarine transient emissions. Six classes of transients were chosen, with two exemplars being included for each class. Five of these classes represented submarine emissions, the final class being counter signals. Additionally, four sounds of biological origin were included, so that discrimination of biological and mechanical sounds could be studied. In total, therefore, sixteen stimuli were used. All recordings were single-channel with a sampling frequency of 16384 Hz and a sample size of 16 bits. Since the data was selected from continuous recordings, it was necessary to apply a shaped window to each signal which prevented discontinuities occurring at the signal boundaries. All signals were chosen to be of similar perceived duration, and the intensity of each signal was equalised for loudness by the first author. All signals had a relatively static noise background and this was removed using spectral subtraction (6); listening to the signals revealed that the spectral subtraction had not introduced any acoustic artefacts.

2.3 Apparatus

A MATLAB program was used to control the experiment, present the signals and collect the results. Signals were presented diotically in a quiet laboratory over Sennheiser HD25SP headphones.

2.4 Procedure

The subjects first adjusted the volume of the signals to a comfortable level using a test tone. Stimuli were then presented in random pairs and subjects rated the similarity on a five point scale; the presentations could be repeated as many times as subjects wished before making their decision. The first eight pairs presented consisted of sounds not in the stimuli set, but still of denoised sonar transients, which allowed the subjects to familiarise themselves with the experimental procedure. Each pair was presented once, giving a total of 120 presentations. Each experiment took 15-25 minutes.

2.5 Analysis

Analysis was undertaken using the weighted measure of distance, as used in the INDSCAL approach to MDS (7). The weighting compensates for the differences in saliences that each subject has for different acoustic properties and also allows an examination of the importance of each dimension of the timbral space for the discrimination task. Some timbre studies use latent classes (8) where subjects are partitioned into classes, depending on their dimensional saliences. However this was considered unnecessary for this study since, because of the abstract nature of the sounds and naïvety of the subjects, it was expected that the weightings would be largely similar between listeners. Furthermore, specificities (9), which have been included in modern timbral studies, were not considered useful here. Specificities are a per-sound translation in timbre space, added to deal with unique acoustic factors which are not covered by the acoustic correlates. Since these factors are accounted for by subject ratings they cannot be calculated for sounds which are not rated, and therefore would complicate any subsequent modelling work.

3 Results

The subjects showed good consistency between each other; the mean correlation of the results vectors was 0.7386, with a standard deviation of 0.050; indicating that subjects were making similar difference ratings. An examination of several dimensional spaces indicated that a 3-dimensional space best captured the data and ensured that each dimension was significantly salient to all subjects.

Dimension 1	Dimension 2	Dimension 3
0.2503 (0.0756)	0.5356 (0.0540)	0.4450 (0.0820)

Table 1 : Mean salience (standard deviation in brackets)

The saliences indicate the weightings that the subjects gave to each dimension; thus it can be seen from table 1 that dimension 2 was highly salient for this discrimination task. An examination of individual salience values indicates that there were no significant subject differences. Figure 1 shows the 3 dimensional space projected into two planes.

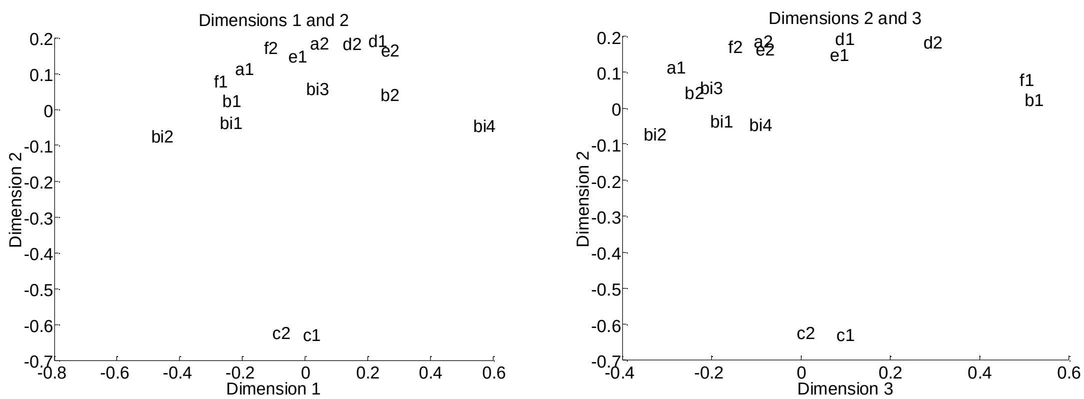

Figure 1 :Projections of the timbre space (bi = biological, a = bow movements, b = unspecified metallic clanks, c = counter signals, d = mast movements, f = pump sounds).

It should be noted that the counter sounds used in the study were tonal transients which had a distinct pitch; both were clearly distinguishable from the rest of the stimulus set, explaining the clear discrimination seen in dimension 2 of figure 1. It should also be noted that the second and third dimensions showed good discrimination of the biological signals from the mechanical signals.

To find dimensional correlates acoustic measures were made of the stimuli set, using measures previously found in the literature as a guide. Each measurement was correlated with each dimension, the selected acoustic factor being the one with the best correlation. Dimension one correlated with a measure of the change in the spectrum over time (r = -0.74778, t = -4.2141). Dimension two correlated highly with a measure of the lowest frequency component present in the signal (r = -0.90168, t = -7.8022). Dimension three correlated with a measure of the log of the spread of the spectral energy present in the signal (r = -0.75464, t = -4.3034).

The results seen here can be related to other timbre studies seen in the literature. Spectral change over time was identified as a correlate of timbre by McAdams (10) and Krumhansl (11), who denote it 'spectral flux'. Grey (12) identified the spectral distribution as being a discriminatory factor, and this may relate to the log of spectral spread found in our study. Our finding that lowest frequency component is a correlate of timbre appears to be unique. This factor may not have been identified in previous studies because they typically use harmonic sounds which are normalised to be of equal pitch. In our study, the majority of stimuli were inharmonic but some (notably the counters) gave a weak pitch percept.

Howard's (5) study of the timbre of sonar signals found that measures of the modality of the signal spectra and the degree of low-frequency periodicity accounted for the dimensions of the timbre space. Neither of these measures were found by this study; however the types of sonar signals used in Howard's study differed from the ones used here. Spectral centroid, rise time and an unspecified measure of timbral richness were identified by Lakatos' (4) study of percussive timbres. Lakatos also reported poor correlation with a measure of spectral flux, although the measure used was significantly different to the one applied here. It could be argued that timbral richness is related to log spectral spread, although no attempt was made to quantify this factor.

4 Modelling

The motivation for the modelling stage was to examine the feasibility of using the correlated measures to quantify the timbre of transients which were not part of the stimulus set. The analysis involved two stages; firstly the acoustic properties were measured, and secondly they were mapped into the timbre space described in section 3. Previous studies have shown that translation of objects in a timbral space is viable (10); hence signals that are not in the stimulus set can also be placed into the space. Figure 2 shows a schematic diagram of the model.

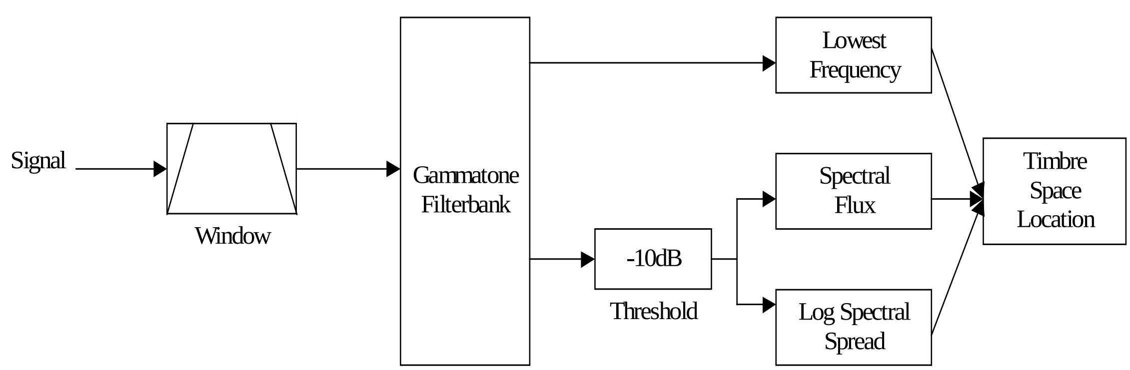

Figure 2 :Schematic of the auditory model

Stimuli were processed using a gammatone filterbank, and the Hilbert envelope was estimated for each frequency channel (13). The filterbank contained 128 channels whose centre frequencies were equally spaced between 20 Hz and 8192 Hz on an equivalent rectangular bandwidth (ERB) scale.

An extension of McAdams' (10) measure of spectral flux was used. This is defined as the mean of the correlation between spectra of successive overlapping time frames:

$$SF = \frac{1}{M}\sum_{p=1}^{M}\left|r_{p,p-1}\right| \qquad (1)$$

Here, M is the number of frames and $r_{p,p-1}$ is the Pearson product moment correlation coefficient between frames at times p and $p-1$. Each time frame was 16 ms long with an overlap of 8 ms. To calculate spectral spread and lowest frequency component, a spectrum was obtained and significant peaks isolated by applying a threshold at a level 10dB below the maximum spectral intensity. Spectral spread was then measured as the difference between the lowest and highest peak boundaries; lowest frequency component was calculated as the frequency value of the lowest peak. The raw values were then scaled according to the linear regression curve of the dimensional location and the corresponding acoustic property. The resulting timbre space is shown in figure 3.

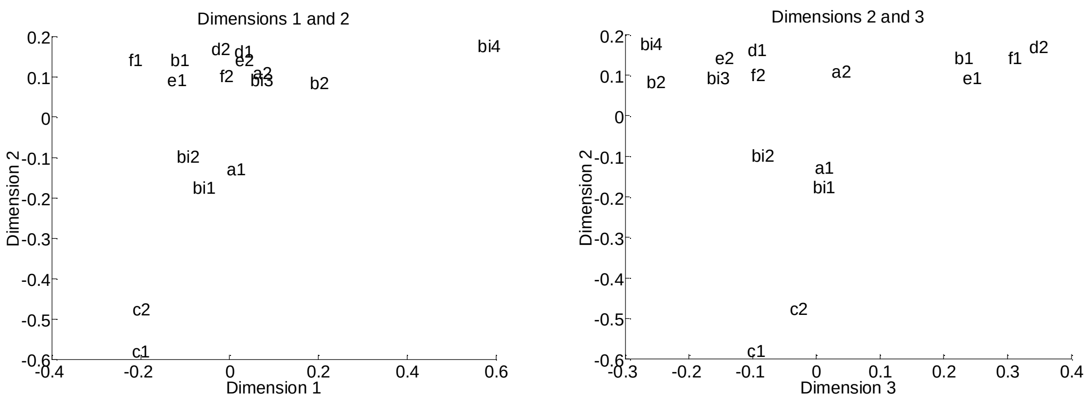

Figure 3 :Re-created timbre space (The cluster in the left figure contains a2 and bi3).

It is important to note that since perfect correlation was not found between the acoustic properties and their spatial locations, the re-created timbral space is different to the original. However it can be seen that the clustering seen in the original is maintained in this recreated space; counter, biologicals and mechanical transients are largely segregated. The major differences between this space and the original is that the clustering is now more distinct, a1 is part of the biological cluster and the sounds are more widely distributed.

Furthermore it is encouraging to note that the re-created space maintains the timbral relationships between the stimuli. A good example of such a timbral relationship is seen in signals b2 and bi4, which are strongly clustered in dimensions 2 and 3 but can be clearly differentiated in dimension 1. Perceptually these sounds are similar except for a characteristic 'grating' present in bi4; since this could be interpreted as a difference in spectral flux, the segregation of these two signals along dimension 1 is consistent with their timbral relationship.

It is important to note that in the timbre space, there is no guarantee that sounds will cluster into semantic classes. However, as figure 3 shows, there is some degree of clustering of biologicals and counter signals, suggesting that members of the classes examined here had timbral similarities. To test the ability of this approach to generalise, twelve sounds (six biological and six mechanical) were also processed that were not part of either the current stimulus set or the practice signals used in the study. The timbral features were calculated for the new data set and normalised according to the values of the current data set; the new signals were then plotted in the timbre space. The space containing the new signals is shown in figure 4.

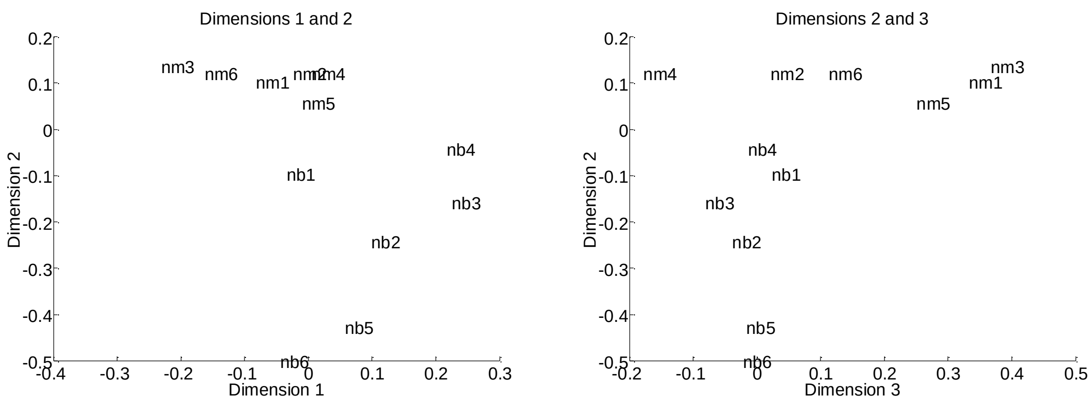

Figure 4 :Timbre space of new signals (nb = new biological, nm = new mechanical).

Similar clustering of sounds can be seen in figure 4 to that seen in figure 3. The new biological sounds form two clusters and the new mechanical sounds cluster in the negative regions of dimensions 2 and 1. Sounds nb6 and nb5 were more tonal, explaining why they clustered in a similar location to the counter signals. Overall, the similarity between the placement of the sounds here and in figure 3 suggests that acoustic events of mechanical and biological origin have general timbral characteristics that allow them to be distinguished.

5 Conclusion

The timbre of underwater acoustic transients was investigated using a multidimensional scaling study. Pairs of denoised sonar signals were played to human subjects who rated the similarity of each sound in the pair; the ratings were used to create a three dimensional timbre space. Acoustic properties were then correlated with the dimensions of the space; it was found that spectral flux, lowest frequency component and log of spectral spread were used by listeners to discriminate the timbre of the stimuli set. Furthermore, the timbral space was recreated using measures of the acoustic correlates, and sounds from outside the stimulus set were placed into the re-created space.

Regarding the timbre of transient sounds in general, it is interesting to note that, discounting lowest frequency component, the acoustic factors identified by this study have also been identified by timbral studies of musical sounds described in the literature. However, the specific set of properties found here has not been previously reported. Given the range of identified acoustic correlates described in the literature, it seems that subjects use a variety of cues to discriminate timbre and the saliency of the cues depends upon the stimuli being differentiated. The study undertaken here agrees with Lakatos' (4) conclusion that similar acoustic properties can be used to discriminate transient sounds that are used to discriminate sustained sounds. However, an extra acoustic property was found here that has not been reported in the timbre literature, that of lowest frequency component. It can be argued that this finding supports the above hypothesis; namely, that discrimination is achieved with whatever cues prove salient to the task. When dealing with transient signals, the saliences of certain cues diminishes (e.g. offset slopes, rise time) thus requiring that subjects change their approach, introducing new acoustic discriminants.

The model described here could form part of a classification system in which recent sonar events are compared to previously recorded signals, and timbral similarities are identified. It should also be noted that whilst semantic segregation need not occur, biological and mechanical sounds tend to cluster separately in the timbre space, indicating that these two classes of sound have general acoustic properties that allow them to be discriminated. Such separation of mechanical and biological transients also occurred in a timbre space constructed by an auditory model, for sounds that were not rated by listeners. This result suggests that automatic classification of transient sounds on the basis of timbral characteristics is possible.

One limitation of the current study is that it emphasizes static representations of timbre; a measurement of spectral flux was used, but this only captures the gross spectral dynamics. Future work will investigate representations of timbre that capture spectral changes in greater detail, with a view to producing computer models that match the classification performance of human listeners more closely.

Acknowledgements

S. Tucker is supported by an EPSRC CASE studentship with the Marine Acoustics Department, QinetiQ, Winfrith.

References

(1) Teolis, A. and Shamma, S, Classification of transient signals via auditory representations, (1991), Technical Research Report TR 91-99, University Of Maryland.

(2) ANSI, *Psychoacoustical terminology*. S3.20, (1973), American National Standards Institute.

(3) Borg, I. and Groenen, P., *Modern multidimensional scaling: theory and applications*, (1997).

(4) Lakatos, S., A common perceptual space for harmonic and percussive timbres, (2000), *Perception and Psychophysics* 62(7), 1426-1439.

(5) Howard, J.H., Psychophysical structure of eight complex underwater sounds, (1977), *Journal of the Acoustical Society of America*, 62(1), 149-156.

(6) Boll, S.F., A spectral subtraction algorithm for suppression of acoustic noise in speech, (1979), *Proceedings of IEEE ICASSP*, 200-203.

(7) Carroll, J.D. and Chang, J., Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition, (1970), *Psychometrika* 35(3), 283-319.

(8) Winsberg, S. and De Soete, G., A latent class approach to fitting the weighted euclidean model, CLASCAL, (1993), *Psychometrika* 58(2), 315-330.

(9) Winsberg, S. and Carroll, J.D., A quasi-nonmetric method for multidimensional scaling via an extended euclidean model, (1989), *Psychometrika* 54(2), 217-229.

(10) McAdams, S., Perspectives on the contribution of timbre to musical structure, (1999), *Computer Music Journal* 23(3), 85-102.

(11) Krumhansl, C.L., Why is musical timbre so hard to understand?, (1989), in Nielzen, S. and Olsson, O. (eds), *Structure and perception of electroacoustic sound and music*, Excerpta Medica, 43-54.

(12) Grey, J.M., Multidimensional perceptual scaling of musical timbres, (1977), *Journal of the Acoustical Society of America*, 61(5), 1269-1277.

(13) Cooke, M., *Modelling auditory processing and organisation*, (1993).