Audio Engineering Society
Convention Paper
Presented at the 114th Convention
2003 March 22-25 Amsterdam, The Netherlands

Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds

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

A psychophysical experiment was undertaken to investigate whether human listeners are able to perceive the material properties of struck plates when they are suspended in air, and also when they are artificially damped by suspension in water. Listener's judgements of the size, shape and material of the plates were found to be less reliable in the damped condition. A computational model was developed which estimates the material properties of an impulsive sound by measuring the decay rate of significant acoustic components at the output of an auditory filterbank. The model provides a good overall match to the pattern of human responses in the psychophysical study. The auditory model has also been used to classify transient sonar sounds with encouraging results.

1. INTRODUCTION

Ecological psychology regards auditory perception as an analysis of the structure placed on an acoustic signal by its source. This approach is highlighted by Gaver's [1] distinction between *everyday* listening (in which the focus is the source of the signal) and *musical* listening (in which the focus is the signal itself). In musical listening, a particular sound might be described as “a fluctuating medium frequency harmonic complex”, but in everyday listening the same sound might be described as “a saxophone solo”. According to the ecological theory, everyday listening is the dominant mode of auditory perception. There is some support for this in the literature; for example, listeners are able to discriminate a bouncing bottle from a breaking one on the basis of a sound recording of the impact [2].

This paper describes an investigation into the perception of source properties from acoustic signals. We begin by reviewing the literature on the auditory perception of source properties, and then describe a psychophysical experiment which examines the auditory perception of the size, shape and material of struck plates. Finally, a computational model of material perception is proposed and evaluated.

One motivation for the work described here is the prospect of applying computational models of material perception to the analysis of passive sonar recordings. Such models might provide a means of discriminating transient sonar sounds of biological origin from those of non-biological origin. Automatic judgements of the material properties of a sound would be very useful in this respect; for example, a ‘metallic’ sound would be of particular interest to sonar operators since it could represent a threat.

1.1. Auditory Perception of Size and Shape

An early theoretical investigation of shape perception was reported by Kac [3], who asked whether the frequency components of a drum uniquely describe its shape. The results of his investigation were inconclusive, although it was later shown [4] that it is possible for two differently shaped drums to yield the same acoustic spectrum. In practice, however, such acoustic homologues are rarely found because of the difficult in constructing two drums with sufficient accuracy.

Lakatos *et al*. [5] asked subjects to identify the cross-section of metal and wooden bars from their impulse responses. Subjects were played a pair of sounds relating to two bars of the same material but of different cross-sections, whilst seeing visual representations of the bars on a screen. The task was to correctly assign each visual representation to each of the sounds in the pair. The results showed that subjects performed better when the difference in cross-sections was more pronounced. Furthermore, subjects concentrated on acoustic cues which were relevant to a comparison task rather than those which could be used in an identification task.

Of particular relevance to the study reported here is the work of Kunkler-Peck and Turvey [6], who examined the perception of shape and material from the sound of struck plates. Experiments took place in free field, with plates hung behind a screen and struck in the centre. Two experiments examined the perception of the shape of rectangular plates, in which listeners were asked to estimate the height-width ratio of a plate by adjusting a frame to match it’s perceived size. The results showed that subjects underestimated the absolute ratios, but the relative estimates were highly correlated with the relative height-width ratios of the actual shapes.

Two more experiments by Kunkler-Peck and Turvey used triangular, circular and square plates of metal, steel and wood. Subjects were asked to identify both the material and shape of the plate after hearing it struck in the centre. Their results showed that material perception was excellent and shape perception was good, especially for metal and wooden plates, and the authors noted a strong correlation between the modal frequencies of the plates and subject’s responses.

1.2. Auditory Perception of Material

Theoretical work on the perception of material was undertaken by Wildes and Richards [7]. They suggest that the decay time of an impulse response (or correspondingly, its bandwidth) are good indicators of material. For example, the impulse responses of steel objects have a large bandwidth and long decay, whereas those of plastic objects have a small bandwidth and a short decay. Although the role of these acoustic features was not verified experimentally, other work was cited [8] which provides support for their theory.

Lutfi and Oh [9] examined the perception of the material of synthesised struck bars. The experiment utilised a physical model of a vibrating bar and, through perturbation of the parameters of the physical model, the salient acoustic cues to material could be identified. They concluded that listeners gave significant weight to pitch in preference to decay and intensity cues, although they acknowledge that the perturbations used could have led to more noticeable differences in pitch than decay or intensity.

Giordano [10] investigated the perception of the material of iron, glass, wood and plastic plates of different areas and height-width ratios. In some conditions, plates were artificially damped. Subjects displayed confusion between iron and glass plates, and between wood and plastic. Furthermore, glass and wood were better identified when the plates were small, and iron and plastic were better recognised for larger plates. Damping increased the degree of confusion and caused glass to be confused with plastic and wood. The author explains this finding in terms of the change in decay rates caused by damping.

1.3. Interim Summary

The literature concerning object property perception has shown that listeners possess the ability to identify the size, shape and material of struck objects. However, listeners generally underestimate absolute sizes, and shape perception is affected by the material of the object. Furthermore, materials form perceptual macro-categories, such that many confusions occur within such macro-categories but few confusions occur between categories.

Recall that our study is motivated by the need for an automatic classifier of transient sonar sounds. Human sonar operators are able to reliably discriminate transient sounds of biological origin from those of non-biological origin. It is interesting to consider to what extent this ability might be underlain by mechanisms of material, size and shape perception. Furthermore, since sonar recordings are derived from the vibration of objects in water rather than in air, the effect of damping on the perception of source properties is of interest here.

A psychophysical experiment and modelling study are now described. The former investigates the auditory perception of the size, shape and material of plates suspended in air and water. The in-air and underwater conditions are compared in order to assess the effect of damping on listeners' performance. In the modelling study, we propose a model of the auditory perception of material, and show that this provides a good match to listeners' data. Furthermore, the model has been evaluated on a range of sonar recordings, and shows promise as a pre-processor for automatic classification of sonar transients.

2. DATA COLLECTION

Recordings were made of plates being struck in two different environments. This section describes both the plates and the recording environments.

2.1. Materials

Fifteen plates were used for the recordings. Square, circular and triangular plates were used to investigate the perception of shape. Each plate had the same surface area, and the triangle was equilateral. To investigate the perception of size, three square plates were used whose surface area progressively quadrupled. The smallest square was common to both sets of plates. Table 1 shows the dimensions of each plate.

Plate	Dimensions (cm)	Area (cm^2)
Large Square	Width = 100	10000
Medium Square	Width = 50	2500
Small Square	Width = 25	625
Circle	Radius = 14	625
Triangle	Edge Length = 38	625

Table 1: Dimensions and areas of the plates.

Furthermore, to investigate the perception of material each of the plates in Table 1 was constructed from aluminium, plastic and wood. All plates had a thickness of 5 mm, with the exception of the wooden plates which had a thickness of 4 mm. Weights were hung on the wooden plates to prevent them from floating. Weight was added incrementally until the plates sunk in the water so that the vibration of the plate was not affected.

In both experiments, the plates were struck by a metal spike, which was 169 mm long, 17 mm in diameter and had a tip diameter of 6 mm. To ensure that the spike struck the underwater plates with sufficient energy it was necessary to attach a 1 kg weight to the spike during these recordings. This weight was attached to the rear of the spike, so that only the tip was in contact with the plate during the strike.

2.2. Underwater Recordings

The underwater recordings were made in a large tank, depicted in Figure 1. The plates and the hydrophones were suspended from a walkway over the tank. Two hydrophones were used, one placed perpendicular to the centre of the plate (H_0) and one placed at the side of the plate (H_1). Recordings of the strikes were made to DAT at a sampling rate of 48 kHz. These DAT recordings were transferred digitally to a PC for editing and playback. An attenuator was used, as required, to prevent signal clipping. The side walls and floor of the tank were treated with foam pyramids to reduce echoes.

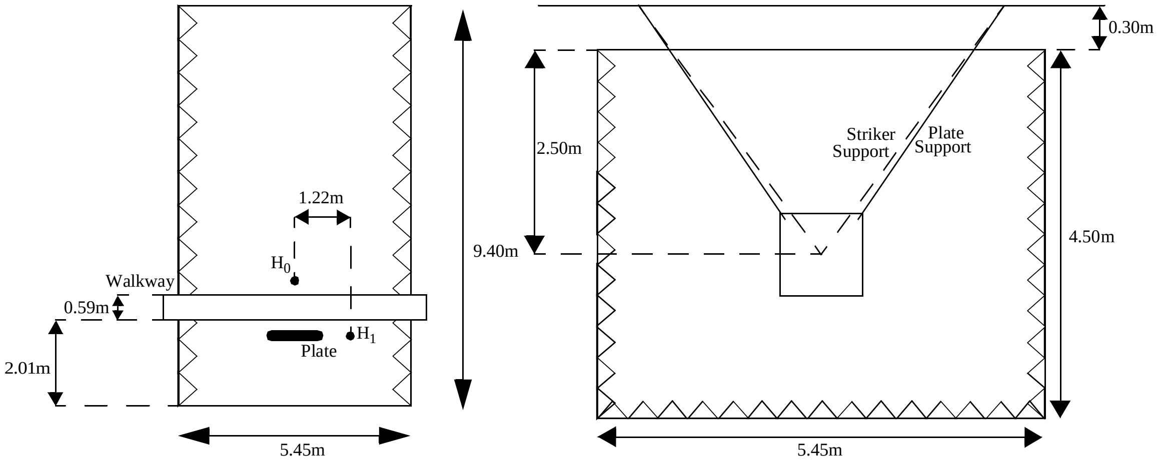

Figure 1: Schematic of the underwater recording setup. Triangles indicate acoustically treated walls.

The metal spike was attached to a rope which allowed the striking device to be positioned and released from the near wall of the tank. The energy of each strike was kept constant by releasing the striking device from the same point for each recording. To record a plate the support ropes were marked with tags, indicating the appropriate height from the water's surface. A recording was made for each plate at three different locations: the centre, the edge and the midpoint of these two. For each configuration a one minute signal was recorded. Strikes were separated by ten seconds to allow the plate to finish resonating, and to ensure that the plate was in the same location for each strike.

2.3. In-Air Recordings

The in-air recordings took place in an acoustically isolated booth. The configuration of the equipment in the booth matched the underwater configuration with two exceptions. First, due to the recording equipment available, signals were recorded direct to disk at a sample rate of 50 kHz. Secondly, the striking device travelled a significantly shorter distance in the in-air configuration. This meant that the strike energies were different in the two environments; however, the effect of this difference was reduced by equalising the loudness of the recordings (see below). The in-air recording configuration is depicted in Figure 2. As with the underwater recordings, plates were hung such that their centres were aligned with the centre of the microphone. Since the plates resonated for longer in air, it was necessary to make 2 minute recordings and allow 20 seconds for the plate to finish resonating between successive strikes. The striking device, support ropes, and plates were the same in both environments.

2.4. Post-Processing of Recordings

Following the recording phase one centre strike was selected for each plate and environment. The signal to noise ratio for the wooden plates recorded underwater was found to be significantly lower than that for the recordings of the plastic and metal plates. To prevent noise being used as a cue to material, the noise background was normalised for the underwater recordings by adding noise from non-event portions of the recordings. This approach was preferred to removal of the noise floor by spectral subtraction [11], since this can introduce acoustic artefacts. Further to this, all the signals were equalised for loudness by the first author.

3. EXPERIMENTAL PROCEDURE

Using the data collected from the two environments a psychophysical experiment investigated the perception of size, shape and material. To simplify the experimental procedure, material and shape perception were examined together. Therefore two listening tests were run on two data sets, giving a total of four experiments.

3.1. Invariants

The same twelve subjects participated in each experiment. The subjects were split into four groups of three, each group attempting the experiments in a different order. In cases where the preceding experiment examined the perception of a different property, at least two days separated each experiment. In cases where the preceding experiment examined the perception of the same

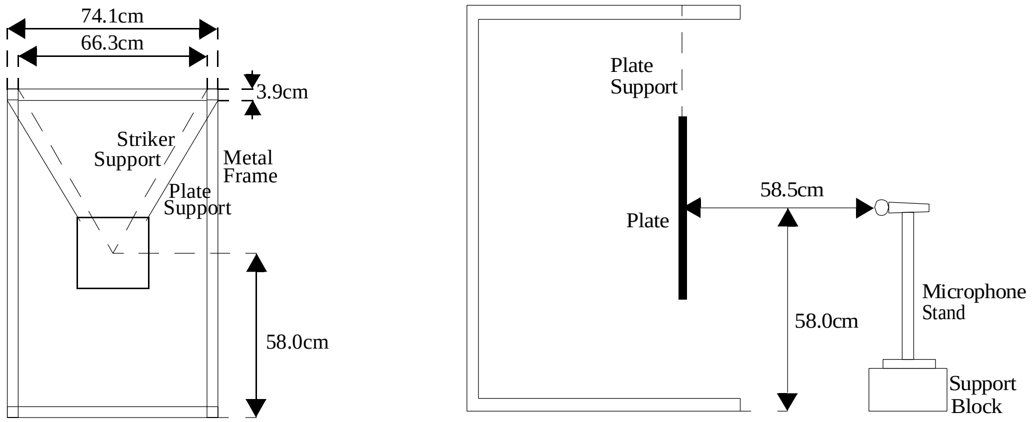

Figure 2: Schematic of the in-air recording configuration.

property but with a different data set, at least a week separated the experiments. All subjects were chosen from postgraduates at the University of Sheffield; none reported any hearing problems.

All experiments took place in an acoustically isolated booth. Signals were presented diotically to listeners over headphones. Subjects were informed of the nature of the sounds but they did not hear any of the signals before beginning the experiment. Following each experiment, subjects were encouraged to make any comments they might have regarding their performance.

3.2. Auditory Perception of Size

In the size perception experiment, subjects heard successive pairs of sounds of the same material. The subjects viewed an on-screen representation of two squares, along with a slider. Moving the slider to the left made the left square bigger than the right, and vice versa. The slider represented the full range of the plate sizes subjects would hear. Subjects heard pairs in a random ordering with the constraint that they would not hear successive pairs of the same material. Since each pair was played in both orderings there was a total of 18 comparisons in each experiment - although subjects were not informed that they would hear pairs in both orderings.

3.3. Auditory Perception of Shape

Prior to the start of the experiment the subjects were shown the shapes of the plates they would be hearing. Subjects were played single sounds in a random order with the constraint that successive sounds had to be of a different material. The subjects were required to select the shape and material from a list of options displayed on screen. Only the possible shapes and materials were given as options. Subjects heard each set of sounds three times in distinct blocks, giving a total of 27 shape and material estimates for each experiment.

4. RESULTS

This section outlines the results of the experiments, beginning with the listening tests for material perception.

4.1. Material Perception Results

Table 2 shows the overall confusions for perception of material (left panel) and the same confusions categorised by the shape of the plate (right panel), for the in-air condition.

	Actual Material		
Perceived Material	M	P	W
M	108	0	5
P	0	61	46
W	0	47	59

	Actual Shape		
	C	S	T
M	37	37	37
P	37	33	37
W	34	38	34

Table 2: Listeners' confusions for the perception of the material of plates struck in air (M = Metal, P = Plastic, W = Wood, C = Circle, S = Square, T = Triangle).

The left panel shows that the perception of material is good. In particular, discrimination of metal from plastic and wood is excellent, indicating that the responses of listeners can be organised into macro-categories. If such categories are assumed to be present, then chance rating of wood and plastic would be 54.

The right panel of Table 2 shows that there appears to be no significant effect of shape on the perception of material in this environment ($G^2(4) = 0.60, p > 0.2$).

Corresponding results for material perception of the underwater recordings are shown in Table 3.

	Actual Material		
Perceived Material	M	P	W
M	108	3	0
P	0	69	52
W	0	36	56

	Shape		
	C	S	T
M	39	36	36
P	43	36	42
W	26	36	30

Table 3: Listeners' confusions for the perception of the material of plates struck underwater (M = Metal, P = Plastic, W = Wood, C = Circle, S = Square, T = Triangle).

The results suggest that there is a similar level of confusion to that seen for the in-air recordings. The pattern of confusions occurring in both environments is, however, different. With the in-air recordings subjects tended to be of two types: those that consistently identified plastic and wood plates, and those that consistently confused them. With the underwater recordings, subjects intermittently confused plastic and wood. A measure of subject consistency can be calculated by examining the average highest rating for plastic and wood shapes for each subject. An average closer to 9 indicates that subjects are consistent in their judgements, even if those judgements are wrong. The averages for the in-air and underwater data are 8.1 and 7 respectively. The average for the in-air data is significantly higher ($t = 2.17, p < 0.025$), indicating that subjects were less consistent in their judgements with the underwater recordings.

A more detailed analysis of the results outlined in the right panel of Table 3 indicates that subjects had difficulty identifying the material of square plates. There was no overall effect of shape on the perception of material ($G^2(4) = 2.53, p > 0.2$).

4.2. Shape Perception Results

Table 4 shows the overall confusions for the perception of shape (left panel), and the same confusions categorised by the material of the plate (right panel), for the in-air recordings.

Listeners' performance on shape perception (left panel) is significantly worse than that for material perception. The results here show little correspondence between the actual shape and the perceived shape ($G^2(4) = 1.33, p > 0.2$). It is also possible to examine the perception of shape by material. The right portion of Table 4 reveals that subjects favoured plastic squares and wooden circles ($G^2(4) = 26.01, p < 0.01$).

	Actual Shape		
Perceived Shape	C	S	T
C	46	38	42
S	36	39	37
T	26	31	29

	Actual Material		
	M	P	W
C	40	30	56
S	27	53	32
T	41	25	20

Table 4: Listeners' confusions for the perception of the shape of plates struck in air. (M = Metal, P = Plastic, W = Wood, C = Circle, S = Square, T = Triangle).

	Actual Shape		
Perceived Shape	C	S	T
C	37	39	32
S	49	42	45
T	22	27	31

	Actual Material		
	M	P	W
C	38	28	42
S	30	53	53
T	40	27	13

Table 5: Listeners' confusions for the perception of the shape of plates struck underwater. (M = Metal, P = Plastic, W = Wood, C = Circle, S = Square, T = Triangle).

Corresponding results for the perception of shape with the underwater recordings are shown in Table 5. Again, it can be seen that shape perception is poor and the results appear to be largely random ($G^2(4) = 2.82, p > 0.2$). The apparent tendency for subjects to choose squares is not significant and is probably due to subjects favouring this shape when guessing.

The results categorised by material (right panel of Table 5), show that listeners favoured squares for both plastic and wood shapes. This could be caused by a mismatch between the sound of a plate and listener's expectations of it. For example, subjects may expect triangular plates to sound markedly inharmonic, whilst the acoustic signal, in this environment, only displays a small degree of inharmonicity. This could cause listeners to ignore triangles when making their selections.

The theory that shape judgement in this condition was related to harmonicity is supported by the results for the in-air recordings. Since the metal plates generate sounds with a complex acoustic spectrum, listeners are able to differentiate plates which are highly harmonic from those which are relatively inharmonic. The results for plastic and wood can be ascribed to the former's lack of either noticeably harmonic or inharmonic sounds, and the latter's tendency to constantly sound harmonic. The selection of squares for the underwater recordings could, therefore, be attributed to damping, which preventing subjects from making accurate judgements of harmonicity.

4.3. Size Perception Results

When listeners estimate the size ratio of two plates, there are two errors which could be made. Listeners could reverse the order of the size ratio, so that after hearing a small plate followed by a larger plate the listener rates the first plate as being larger than the second. Listeners could also identify the correct order of the ratio but make an inaccurate estimate of its magnitude. Here, we refer to the former as *gross* errors and the latter will be termed *accuracy* errors.

Gross Errors

The gross errors made in judging the size of plates in both environments are given in Table 6.

	Metal	Plastic	Wood
In-air	14	15	21
Underwater	19	33	19

Table 6: Listeners' gross size errors in both environments, for each material.

Table 6 shows that the number of gross errors is greater for the underwater recordings. It is interesting to note that the in-air gross errors were largely confusions between medium and large plates, although the cause of this effect is unclear. The poor performance on recordings of plastic plates could be explained by the relatively short decay times for this material, meaning that the amount of information available for judging plate size is reduced.

Accuracy Errors

Figure 3 shows the accuracy of subjects in estimating the relative size of the second plate in the pair that was heard. The size of the centre box indicates the estimated size of the second plate (the actual size of the first plate is given on the abscissa), and its position represents the mean size judgement. The extent of the lines indicate the range of the given estimates. For example, the left area of the figure corresponds to a condition in which subjects heard a small metal plate followed by a large metal plate, for which they rated the size of the large plate as being the size of the medium plate.

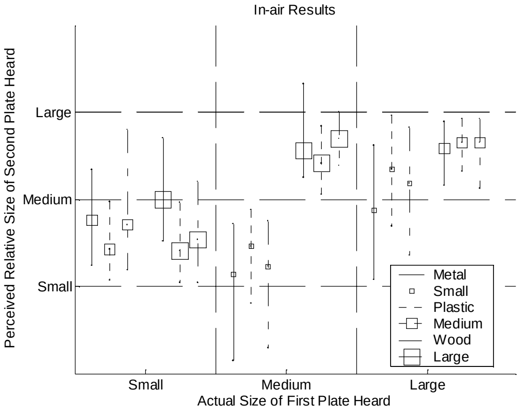

Figure 3: Listeners' perception of size for the in-air recordings. The abscissa shows the actual size of the plate in the pair. The ordinate shows the relative perceived size of the second plate. Line extrema show the range of responses, and markers indicate the average response over all subjects.

The accuracy of listeners in judging plate size is poor; listeners consistently underestimated the size ratio of the plates. The relative performance is better, however, with listeners perceiving that larger plates were larger. For example, when the large plates were heard first (right hand area of Figure 3) subjects were not accurate, but medium plates were estimated to be larger than smaller plates.

It is also of interest to note that, like the results seen for material perception, subjects performed better on the metallic plates. Subjects performed worst with the plastic plates, indicating that the decay time could be a useful cue to plate size. Figure 4 shows the size judgements with the underwater recordings. Again, this shows that subjects performed better on recordings of metallic plates than on plastic or wood. It should also be noted that performance is worse than that seen for the in-air recordings - the average estimate is worse, and the range of responses is reduced. This indicates that subjects perceived the plates to be smaller when listening to the underwater recordings. It should be noted that subjects are still able to correctly identify the relative ratios of the plates.

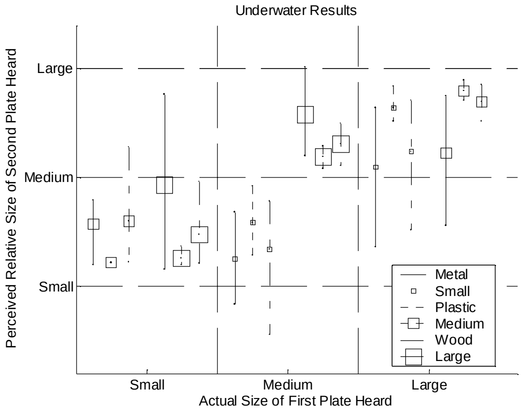

Figure 4: Listeners' perception of size for the underwater recordings. The abscissa shows the actual size of the plate in the pair. The ordinate shows the relative perceived size of the second plate. Line extrema show the range of responses, and markers indicate the average response over all subjects.

4.4. Conclusions from the Perceptual Study

Several conclusions can be derived from the experiments presented above. Firstly, there is a degradation of performance with the underwater recordings relative to the in-air recordings; specifically, there were increased gross size confusions. Material perception was largely good, with materials forming appropriate macro-categories.

Shape perception was poor for both sets of recordings. Certain shapes exhibited exceptions to this and were better identified (for example, metal circles were better recognised by listeners). Subjects also tended to underestimate size ratios although performance in recognising relative size differences was good.

It is apparent from the results that the decay of the impulse response plays a significant role in the perception of both size and material. The following section exploits this finding in a computational model of material perception.

5. MODELLING

Following from the psychophysical experiment an attempt was made to simulate listener's perception of material with a computer model. The modelling stage had two primary motivations. The first was to investigate representations and models which could best reflect the experimental results seen in Section 4. Secondly, we wanted to assess the suitability of the model for the classification of biological and non-biological sonar transients.

Initially, modelling experiments were made with simple spectral and correlogram [12] representations, which were used as inputs to both single layer perceptrons and Gaussian mixture models [13]. To assess the suitability of this approach, the models were trained on the in-air recordings and then tested on the underwater recordings.

The results showed that whilst the models worked well on the training data, the performance on the test data did not match the experimental findings. Hence, a different approach was then taken which drew on the work on material perception discussed in Section 1.2, and is motivated more strongly by the notion of ecological perception.

5.1. Decay Modelling

Wildes and Richards [7] suggest that the decay time of the impulse response is a significant cue to object material. They show that decay time is related to a shape invariant constant, the coefficient of internal friction, as follows:

$$\tan\varphi = \frac{1}{\pi f t_e} \qquad (1)$$

Here, $\tan\varphi$ is the coefficient of internal friction, f is frequency and t_e is the time taken for the signal to decay to $1/e$ of its initial energy. From (1) it can be seen that the frequency is inversely related to the decay time. A computer model which exploits this relationship is described below.

Auditory Model

Figure 5 outlines the processing stages of the auditory model. First the signal is split into long frames (200 ms) with a short overlap (10 ms). Each frame is then passed through a bank of 128 Gammatone filters [14], with centre frequencies distributed between 50 Hz and 12 kHz on the equivalent rectangular bandwidth (ERB) scale. The Hilbert envelope [15] is extracted from each filter channel. Filterbank channels which are responding to significant acoustic components are selected by examining the zero-crossings of the channel energy, convolved with the second order derivative of a Gaussian. For each of these selected channels, the exponential decay rate is calculated using the Prony method [16]. It was found that the Prony method required a precise estimate of the onset of

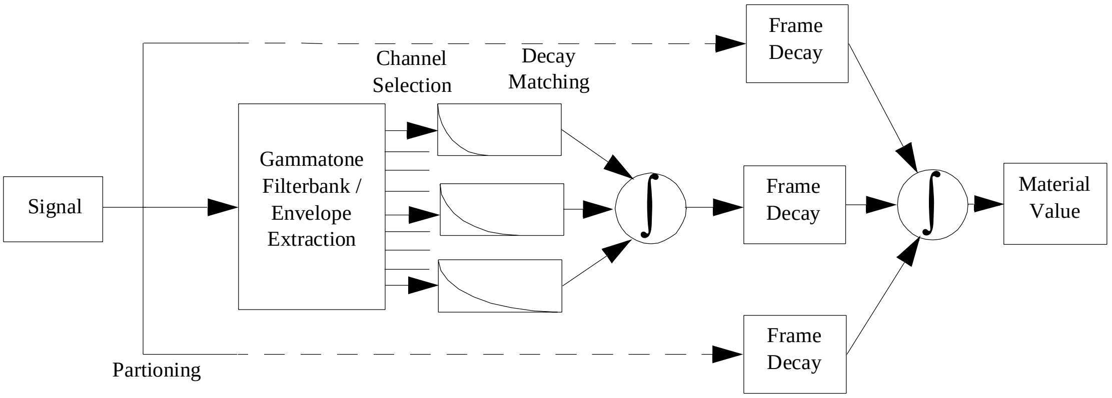

Figure 5: Outline of the perceptual model.

the impulse response, and so the maximum energy in the first half of the time window is used as a starting point for the decay rate calculations.

The decay rates are then normalised by the corresponding channel centre frequencies, giving a set of normalised decay rates for each frame. A single measure for each frame is then calculated by taking the median of this set of decay rates. Finally, to calculate the overall material value for the signal, the log of the average of the measures for each frame is taken, weighted by the frame energy.

Model Performance on Plate Recordings

A free parameter in the calculations is the width of the Gaussian derivative used to identify significant frequency components. The results for a range of Gaussian widths were calculated and the optimal result was defined as that which led to the highest correlation with the experimental findings.

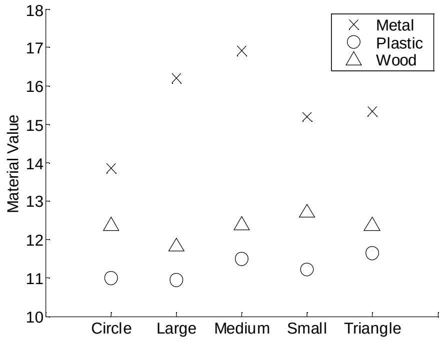

Figure 6: Model classification of material for the in-air recordings. Optimal Gaussian width was 4.6 ERB

Figure 6 shows the results for the in-air recordings. The model results demonstrate a good match to the experimental data ($r^2 = 0.69$). The model clearly distinguishes metal from plastic and wood, thus reflecting the organisation of materials into macro-categories that was seen in the experimental results.

The results for the underwater recordings are shown in Figure 7. Again the recordings of metal plates are clearly distinguished from those of wood and plastic plates. The model also displays a degree of confusion between plastic and wood recordings. Overall the correlation between the model results and the experimental findings in this condition is good ($r^2 = 0.62$).

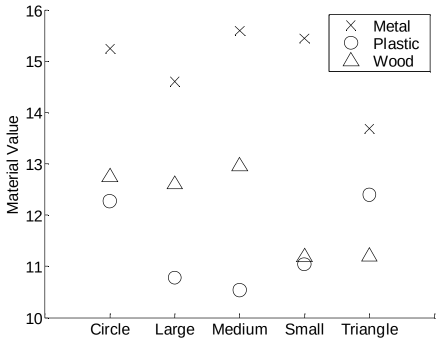

Figure 7: Model classification of material for the underwater recordings. Optimal Gaussian width was 8.2 ERB

Model Performance on Sonar Transients

In order to investigate the performance of the auditory model on real sonar signals, 40 transients were selected from a set of sonar recordings. The source of each transient varied, but each was categorised as being of either biological or non-biological origin. Each signal was isolated by hand before being processed by the perceptual model.

When determining class membership, two different classifications could occur. A signal that does not belong to a class could be classified as being so - this is termed a false-positive. A correct classification is termed a true-positive. The Receiver Operating Characteristic (ROC) curve plots the true-positive rate against the false-positive rate. The closer the curve is to the top left corner the better the performance of the classifier. Figure 8 shows a ROC curve plotted by calculating the convex hull over the true-positive and false-positive rates for a number of model configurations, formed by varying the width of the Gaussian derivative.

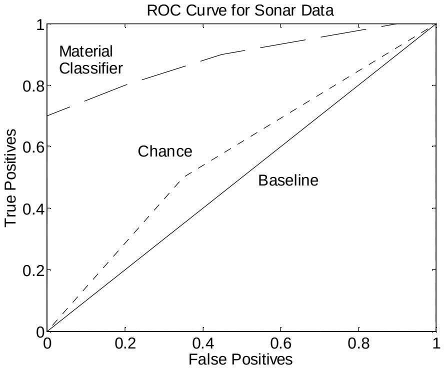

Figure 8: ROC curve for model classification of transient sonar sounds as biological or non-biological in origin.

As can be seen, the material classifier performs reasonably well on the real sonar recordings. The ROC curve is well above chance and shows a true-positive rate of 70% with no false positives. Although the test set is relatively small the results are promising, and suggest that the material classifier could form the basis for discriminating sonar transients of biological and non-biological origin.

6. CONCLUSION

The perception of the size, shape and material of struck plates in two environments was investigated. Recordings were made of plates of different sizes, shapes and materials both freely vibrating in air and damped underwater. These recordings were used in two experiments: one investigated the perception of shape and material, the other examined the perception of size. Following these experiments, a model of the perception of material was constructed which matched the experimental data well. The model also displayed some ability to discriminate real sonar transients of biological and non-biological origin.

The results of the experiment raised several salient points. It is clear that performance was better on the in-air data than on the underwater recordings. In the underwater condition, the plates were highly damped, and therefore the duration of their impulse responses was relatively short. If it is assumed that listeners have some notional model of how such plates would sound in air when struck, then two causes of the performance degradation are suggested. Firstly, our subjects were not experienced in listening to underwater signals and so the application of the in-air models to the underwater recordings could lead to judgement errors. Secondly, the short duration of the signals means that there is less information available to derive the relevant perceptual features.

The experiment provides further support for the notion of material macro-categories. Subjects clearly distinguished plastic and wood from metal in both recording conditions. Considering this, together with other work in the literature [17][18], it could be argued that macro-categories are formed from non-overlapping decay ranges. The decay times of metal and plastic are significantly different, whilst those of plastic and wood are similar. The increased confusion described in Section 4.1 could be an effect of a reduction in the range of the decay times, which in turn would increase the overlap between decay times of different materials.

The results here showed that listeners had difficulty estimating the shape of the plates; the informal interviews suggested that subjects had little confidence in their judgements of shape. The results did reveal an effect of material on shape judgement, and the pattern of responses in both environments was similar ($r^2 = 0.65$). This evidence suggests that the poor performance of listeners could be attributed to a mismatch between subjects expectations of the acoustic cues to shape and the actual values of those cues.

Listeners’ perception of size was better with the in-air recordings. The informal interviews suggested that subjects were using a combination of frequency and decay time cues to estimate the size of the plates. The cause of the consistent underestimation of size seen here and elsewhere in the literature is unclear, although a possible cause is the lack of experience in performing the task.

The results of the modelling phase of the experiment are at odds with the work of Krotkov *et* *al*. [19]. They examined the acoustic response of two bars of different lengths with the aim of assessing whether the decay time could be used to identify the object material. They found that the relationship between the decay time and frequency was not linear, as predicted by Wiles and Richards [7], but that a quadratic relationship provided a better fit to the data. There are, however, several differences between their work and the experiments outlined here. We found that a linear fit to log energy, as described in [19], was less reliable than directly fitting an exponential using the Prony method. Furthermore it was found that a more accurate measure of signal decay could be found by measuring the decay across long windows of the signal, rather than analysing the entire signal as a whole.

Whilst a higher order relationship between frequency and decay rate may provide a better fit to the data it could be argued that, given the frequency range examined and type of signals investigated, the linear relationship used here is suitable for a discriminative model. Indeed, our results suggest that a linear fit is suitable for this task. It is possible, however, that a higher order relationship between channel frequency and decay time would improve the results of the classifier. Future work will assess this conjecture by examining the performance of the model using higher order relationships between decay time and frequency.

Our results are also at odds with the findings of Kunkler-Peck and Turvey [6], who found that subjects were good at recognising the shape of struck plates. A key difference between their work and the experiments carried out here is that their listening tests took place in free field, whereas our recordings were presented over headphones. As Lutfi [20] notes, a correlation may exist between their experimental results and spatial acoustic cues. Therefore when recordings are presented in free field there is more acoustic information available to listeners which could lead to the difference in results seen here.

Informal interviews with experimental subjects indicated that, whilst they expected larger plates to have a lower pitch and longer resonance, they were unable to quantify these parameters in order to support their judgements of plate size. Most frequently, subjects would describe their judgements in terms of everyday listening, for example “it sounds like metal” or “it sounds much larger”. Hence, our experiments appear to validate the notion from ecological perception [21] that subjects perceive the source of an acoustic signal, and that they tend to listen in an ‘everyday’ mode rather than an analytical one.

7. ACKNOWLEDGEMENTS

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

8. REFERENCES

[1] Gaver, W.W., “What in the world do we hear?: an ecological approach to auditory event perception”, *Ecological Psychology*, vol. 5(1), pp 1-29, (1993).

[2] Warren, W.H., Verbrugge, R.R., “Auditory perception of breaking and bouncing events: a case study in ecological acoustics”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10 (5), pp 704-712, (1984).

[3] Kac, M., “Can one hear the shape of a drum?”, *American Mathematical Monthly*, vol. 73, pp. 1-23. (1966).

[4] Gordon, C., Webb, D., “You can’t hear the shape of a drum”, *American Scientist*, vol. 84, pp 46-55, (1996).

[5] Lakatos, S., McAdams, S., Causse, R., “The representation of auditory source characteristics: walking sounds”, *Perception & Psychophysics*, vol. 59 (8), pp 1180-1190, (1997).

[6] Kunkler-Peck, A.J., Turvey, M.T., “Hearing shape”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26(1), pp 279-294, (2000).

[7] Wildes, R.P., Richards, W.A., “Recovering material properties from sounds”, in Richards, W.A. (ed), *Natural Computation*, pp 356-363, MIT Press, (1988).

[8] Gemant, A., Jackson, W., “The measurement of internal friction in some solid materials”, *Philosophical Magazine*, vol. 157, pp 960-983, (1937).

[9] Lutfi, R.A., Oh, E.L., “Auditory discrimination of material changes in a struck-clamped bar”, *Journal of the Acoustical Society of America*, vol. 110 (2), pp 1010-1019, (2001).

[10] Giordano, B., “Material recovery from real impact sounds”, in *Physically-based sound spaces and psychophysical scales*, Sounding Objects Group Report, pp 25-30, (2002).

[11] Boll, S.F., “A spectral subtraction algorithm for suppression of acoustic noise in speech”, Proceedings of IEEE ICASSP, pp 200-203, (1979).

[12] Rabiner, L.R., “Digital processing of speech signals”, Prentice Hall, (1978).

[13] Reynolds, D.A., Rose, R.C., “Robust text-independent speaker identification using gaussian mixture models”, *IEEE Transactions of Speech and Audio Processing*, vol. 3(1), pp 72-83, (1995).

[14] Patterson, R.D., Nimmo-Smith, I., Holdsworth, J., Rice, P., “An efficient auditory filterbank based on the gammatone function”, *APU Report 2341*, Applied Psychology Unit, University of Cambridge, UK., (1988).

[15] Cooke, M., “Modelling auditory processing and organisation”, Cambridge University Press, (1993).

[16] Marple, S.L., “Digital spectral analysis: with applications”, Prentice Hall, (1987).

[17] Avanzini, F., Rocchesso, D., “Controlling material properties in physical models of sounding objects”, *Proceedings of the International Computer Music Conference*, Cuba, (2001).

[18] Rocchesso, D., “Acoustic cues for 3-D shape information”, *Proceedings of the 2001 International Conference on Auditory Display*, Finland, (July 29, 2001).

[19] Krotkov, E., Klatzky, R., Zumel, N., “Robotic perception of material: experiments with shape-invariant acoustic measures of material type”, in Khatib, O., Salisbury K., (eds), *Experimental Robotics IV*, Springer-Verlag, (1996).

[20] Lutfi, R.A., “Auditory detection of hollowness”, *Journal of the Acoustical Society of America*, vol. 110 (6), pp 3647-3656, (2001).

[21] Gibson, J.J., “The senses considered as perceptual systems”, Allen and Unwin, (1968).