

INITIAL RESULTS ON SIZE DISCRIMINATION OF SIMILAR UNDERWATER OBJECTS USING A HUMAN HEARING MODEL

P D Fox Institute of Sound and Vibration Research, University of Southampton, UK
S Bleeck Institute of Sound and Vibration Research, University of Southampton, UK
P R White Institute of Sound and Vibration Research, University of Southampton, UK
T G Leighton Institute of Sound and Vibration Research, University of Southampton, UK
V F Humphrey Institute of Sound and Vibration Research, University of Southampton, UK

1 INTRODUCTION

During conversation humans are good at discriminating between the size of speakers by listening to them talking¹. We can also estimate the size of inanimate objects by the sounds they emit when excited, for example when perceiving the length of a falling rod². It has also been shown that humans can be as effective as dolphins in discriminating underwater objects by listening to dolphin-like scattering sounds which have been down sampled to bring them within the range of human hearing³. We report here on preliminary results obtained from a project aimed at investigating the potential of using human hearing models for underwater classification. The project is at an early stage so the intention of this paper is to introduce the general structure of the approach giving some initial results obtained; a number of developments are planned for various aspects of the work and these are the subject of ongoing research. We currently simulate simple scattering models from analytic solutions available in the literature, and then process the scattered signals through an adaptation of the current MATLAB *aim-mat* model developed in recent years for modelling the human auditory process. Initial results are presented, with emphasis on investigating whether any size discrimination of the objects can be obtained from such a model.

2 TARGET SCATTERING – SIMULATIONS

Pulses scattered by infinitely long, water-filled cylindrical shells submerged in water and interrogated by a plane wave normal to the shell are considered. The scattered pulses were generated using a normal mode solution^{4,5} to obtain an analytic frequency domain solution for the scattering of an incident plane wave, from which the corresponding time domain pulses were obtained by inverse Fourier transformation. A set of 10 similar water-filled aluminium cylindrical shells of different outer radius a were considered, in which ratio of the inner to outer radius b/a was fixed at 0.99 for all values of a . For each shell, the insonification pulse was the same 2-cycle cosine weighted 1 kHz pulse of unit amplitude giving a spectrum with the majority of its energy in the 0-2.5 kHz region (Figure 1). The backscattered pulse from each shell was calculated at a fixed range of $r = 200$ m from the cylinder centre and the 10 cylindrical shells varied linearly in radius from $a = 0.2$ to $a = 2.0$ m in steps of 0.2 m, such that the ratio r/a varied from 1000 for the smallest cylinder down to 100 for the largest cylinder. This means that in all cases the backscatter observation point was effectively far from the cylinder ($r/a \gg 1$) and, though an exact normal mode solution for arbitrary range was used to calculate the scattered signal in practice, the scattering characteristics of the cylinder can be approximated by the cylinder's form function as shown in Figure 2. The energy in the incident pulse is concentrated over the range $f = 0$ to 2.5 kHz for all cylinders, which corresponds to wavenumber values of $k = 0$ to 10.5 m^{-1} and hence ka values from 0 to $10.5a$. Therefore the energy of the incident pulse spans a range which is dictated by the cylinder radius a , meaning that as the cylinder radius increases so does the range of ka values over which the incident pulse contains energy. For the range of values $a = 0.2$ to 2.0 m considered in this paper, this corresponds to ranges of ka from $ka = 0$ to 2.09 for $a = 0.2$ m, up to $ka = 0$ to 20.9 for $a = 2$ m. Observing the form function of Figure 2 we note, therefore, that small values of a will result in relatively few peaks in the spectrum of the scattered signal. In contrast, greater values will excite a large number of peaks and as a result, a range of scattered responses are expected. In turn the

frequency spectrum of the scattered signals which will be processed by the human auditory model will also change considerably for the different sizes of cylinders considered; the interest of the paper here is to see whether a reasonable size discrimination of the objects can still be obtained in this case. Note also that the numerical examples here have been based on the frequency spectrum range $f = 0$ to 2.5 kHz, since these are auditory frequencies which the human hearing model has been designed to process. In the more general case however, scattering at higher frequencies can still be analysed by merely rescaling the sampling rate of the scattered signals generated so as to shift their spectrum down into the human hearing range before processing them through the hearing model. For simplicity of discussion here, we consider only frequencies and object ranges which give scattering responses directly in the human hearing range.

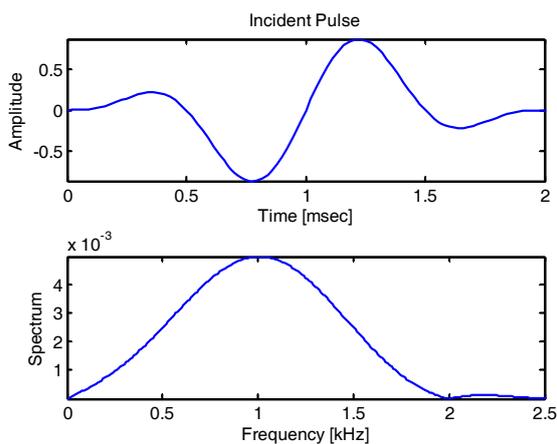


Figure 1: Incident plane wave pulse and spectrum.

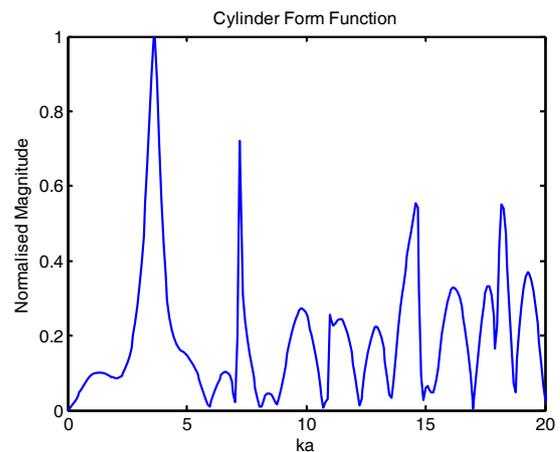


Figure 2: Form function magnitude for water-filled aluminium cylindrical shell (with $b/a = 0.99$).

3 HUMAN HEARING MODEL

3.1 Overview of Human Hearing Model and AIM-MAT

The auditory image model (AIM) is a time-domain model of the signal processing in the hearing system, where each step can be associated with a specific stage of the ascending auditory pathway. The ‘auditory image’ was first described by Patterson et al⁶; it is intended to simulate the neural representation which underlies our first conscious awareness of sound. A computational implementation of AIM in MATLAB code (*aim-mat*), which is frequently used by the auditory community to simulate the construction of auditory images⁷, was used here. The principle functions of *aim-mat* are to simulate:

1. the basilar membrane motion (BMM) produced in the cochlea,
2. the neural activity pattern (NAP) observed in the auditory nerve and cochlear nucleus,
3. the identification of the neural peak times (strobe points) used to construct the auditory image (STROBES),
4. the stabilised auditory image (SAI) that forms the basis of auditory perception and
5. the size and shape transformation (SIZE and SHAPE) to get information about the objects physical properties.

The auditory image model is optimised for detecting temporal regularity and temporal fine structure in periodic sounds. This is achieved by averaging information from repeated periods of the sound. In the problem addressed in this paper only a single non-periodic pulse was obtained for each incident pulse, so that in order to make the stimulus appropriate for processing with the auditory image model, the scattered pulse was repeated to make a periodically repeating scattered click. This manipulation only serves the purpose of making the sounds analysable by *aim-mat* and does not add further information to the sound. The processing steps are described in more detail in the following sections.

BMM: basilar membrane motion

The BMM module simulates the spectral analysis performed by the auditory system with a linear, Gammatone auditory filterbank⁹. Here 100 channels were used with centre frequencies from 100 Hz to 6 kHz. All of the signals were processed as wav files with 16 kHz sampling rate, after normalisation to unit maximum magnitude for each pulse. (Note therefore that the relative amplitudes of the different scattered pulses are not conveyed to the human model's processing in this case). The impulse response of the Gammatone filter is

$$g_t(t) = a t^{n-1} \exp(-2\pi b \text{ERB}(f_c) t) \cos(2\pi f_c t + \varphi)$$

where a , b , n , f_c , and φ are filter parameters. The term $\text{ERB}(f_c)$ is the rectangular bandwidth of the filter, represented⁹ as $\text{ERB}(f_c) = 24.7 + 0.108 f_c$ for moderate hearing levels (60-80dB). The filter gets its name from the fact that the envelope formed by the function and the exponential is a gamma distribution function, and the cosine carrier is a tone when it is in the auditory range. The amplitude spectrum of the Gammatone filter is essentially symmetric on a linear frequency scale. The Gammatone has been used to characterise spectral analysis in humans at moderate levels where the amplitude characteristic of the auditory filter is nearly symmetric on a linear frequency scale. In a later stage of the project we intend to substitute the Gammatone filterbank with the more powerful and physiological justified, nonlinear and wavelet-like "Gammachirp" filterbank¹⁰.

NAP: neural activity pattern

A crude simulation of the transformation of acoustic information into neural activity is used. The BMM output is converted into *aim-mat*'s simulation of the neural activity pattern (NAP) in the auditory nerve using three sequential operations: half-wave rectification, compression and low-pass filtering. The half-wave rectification makes the response to the BMM uni-polar like the response of the hair cell. The compression simulates cochlear compression and reduces the slope of the input/output function. A logarithmic compression was used as per normal convention in the auditory image literature⁶. The low-pass filtering simulates the progressive loss of phase-locking as frequency increases. The filter has a cutoff frequency of 1200 Hz and decays 3 dB per octave.

STROBES: strobe finding

Perceptual research on pitch and timbre indicates that at least some of the fine-grain time-interval information in the NAP is preserved in the auditory image¹¹. This means that auditory temporal integration cannot, in general, be simulated by a running temporal average process, since averaging over time destroys the temporal fine structure within the averaging window⁶. It is the fine-structure of periodic sounds that is preserved rather than the fine-structure of noises, and it was shown⁷ that this information could be preserved by a) finding peaks in the neural activity as it flows from the cochlea, b) measuring time intervals from these strobe points to smaller peaks, and c) forming a histogram of the time-intervals, one for each channel of the filterbank. This two-stage temporal integration process is referred to as 'strobed' temporal integration (STI). It stabilises and aligns the repeating neural patterns of periodic sounds such as vowels and musical notes. The complete array of interval histograms is *aim-mat*'s simulation of our auditory image of the sound. The auditory image preserves all of the fine-structure of a periodic NAP if the mechanism strobes once per cycle on the largest peak. Provided the image decays exponentially with a half life of approximately 30 ms, then it builds up and dies away with the changing characteristics of the sound in a reasonable way⁷. In principle the strobe finding mechanism is inspired by "onset" neurons in the auditory pathway. These neurons mark "important" times in each frequency channel, usually points with high amplitude.

SAI: stabilised auditory image

Once the strobe points have been found, the NAP is converted into an auditory image. Strobed temporal integration converts the time dimension of the neural activity pattern into the time-interval dimension of the stabilised auditory image (SAI) image, and it preserves the time-interval patterns of repeating sounds⁶. In *aim-mat* the mechanism operates as follows: When a strobe occurs it initiates a temporal integration process during which NAP values succeeding the strobe point are scaled and added into the corresponding channel of the SAI as they are generated; the time interval between the strobe and a given NAP value determines the position where the NAP value is entered in the SAI. In the absence of any succeeding strobings, the process continues unabated for 35 ms

and then terminates. If more strobes appear within 35 ms then each strobe initiates a temporal integration process, but the weights on the processes are constantly adjusted so that the level of the auditory image is normalised to that of the NAP. The SAI decays exponentially with a half life of 30 ms to avoid buffer overrun. The process is described in more detail in reference⁷. Figures 4, 6 and 8 show examples of SAIs for the scattered pulses from three different sized cylinders with radii of 0.2, 1.6 and 2 m, along with their scattered pulse classical time and frequency characteristics in Figures 3, 5 and 7 respectively. The dominant structures in the SAI are diagonal stripes of activity. These are a consequence of the ringing basilar membrane filter.

SIZE and SHAPE: measuring physical properties

The process that was described so far was historically optimised to measure properties of pitch and fine structure frequency information-like formants in human speech. Here we consider how the information in the auditory image might also be used to measure size and could be used in future to gather information about the shape/properties of the scattering object. The information in the auditory image was used recently to extract information about the size of a human speaker^{12,1}. The idea pursued in these investigations was that physical object properties influence the activity pattern in the SAI deterministically. Two physical trends play a role: smaller objects have resonances at higher frequencies than larger objects of the same material and shape; and the decay time of these frequencies are shorter for smaller objects. When comparing two objects of the same shape and material, smaller ones have higher frequency resonances, but these decay faster. When the SAI is plotted on log-log axis, these physical properties become interpretable: the location of activations in the SAI shift *linearly* from low frequency, long time intervals to high frequency, short decay time. The log-log representation is provided in our model in the frequency dimension by logarithmic spacing of the basilar membrane centre frequencies, and on the time interval axis by plotting logarithmic time intervals.

In reference¹² the "Size-Shape image" is calculated from the SAI by multiplying each frequency channel by its centre frequency, thus aligning the peaks in the SAI and correcting for the longer time delay in the lower frequency filters. This representation was used¹² for differentiating and subsequent normalising the signals from different sized human speakers. In this paper we apply a different method which yields essentially the same result by projecting along tilted, but parallel lines through the SAI. The approach is easier to visualise, because it is the same as looking at the SAI under a 45° angle. The result of both methods – normalised for the different length of the projections – are vertical stripes instead of diagonals. The above mentioned physical properties however, make the transformed axis interpretable: the vertical "size" axis indicates the size of an object: both effects of decreasing the size of an object (increasing frequency and decreasing decay time) lead to a shift to higher "sizes". The horizontal "shape" axis indicates all other aspects of the sounding object. Since the units on the axis depend on the cutoff frequencies of time and frequency, they are defined arbitrarily. At the current state of the model it is only meaningful to compare objects of the same material relative to each other.

3.2 Application to Scattering Simulations

The processing described in Section 3.1 was applied to the scattered pulses from the 10 different cylindrical shells. Figures 3, 7 and 5 show examples of the responses for three different sizes: the smallest ($a = 0.2$ m), the largest ($a = 2.0$ m) and one in between ($a = 1.6$ m). The top graph of each figure shows the time history of the scattered pulse and the lower graph the frequency profile. Note the increase in the number of peaks in the spectrum of the scattered signals with increasing object size as discussed in Section 2. Figures 4, 8 and 6 show the corresponding stabilised auditory images for the three pulses. The smallest shell (Figures 3 and 4) has a relatively uniform spectrum in the frequency domain and this shows up as a smooth pattern in the SAI, whereas the larger objects (Figures 5-8) have spectra with more peaks and resonances, and these show up as localised points in the SAI. The scale and shape parameter are measured by projecting the SAI at 45° angles on to axes parallel to, and perpendicular to, the activity ridges from bottom left to top right.

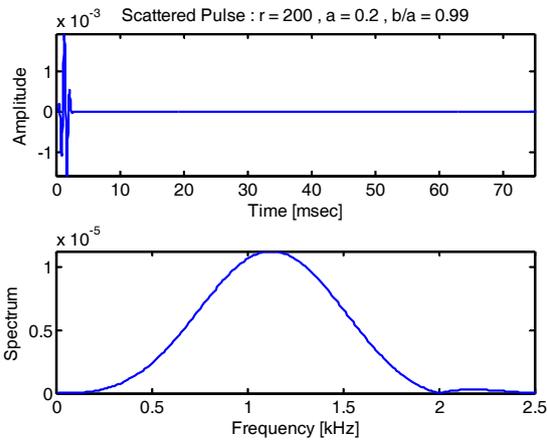


Figure 3: Scattered pulse for cylindrical shell with radius $a = 0.2$ m. Top panel: time history of pulse; Bottom panel: frequency spectrum.

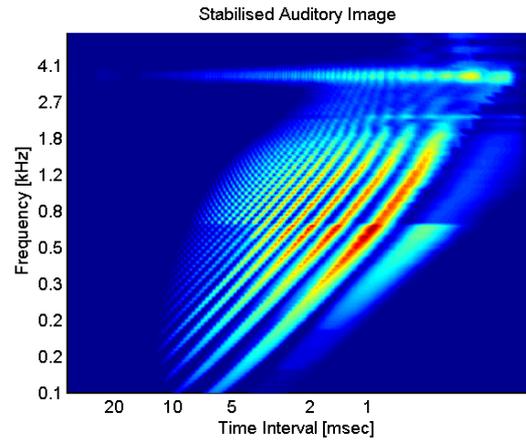


Figure 4: Stabilised Auditory Image for $a = 0.2$ m. x-axis: time interval since last strobe; y-axis: centre frequency of basilar membrane filter.

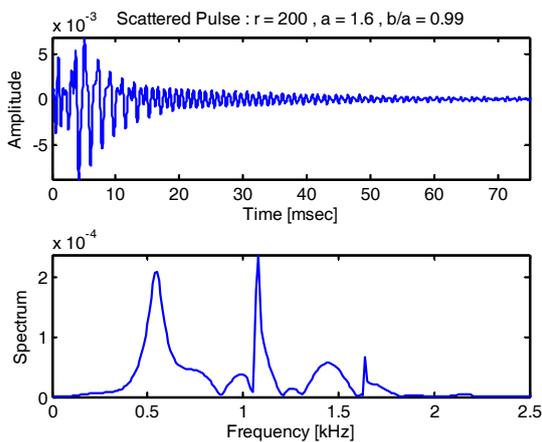


Figure 5: Scattered pulse for cylindrical shell with radius $a = 1.6$ m. Top panel: time history of pulse; bottom panel: frequency spectrum.

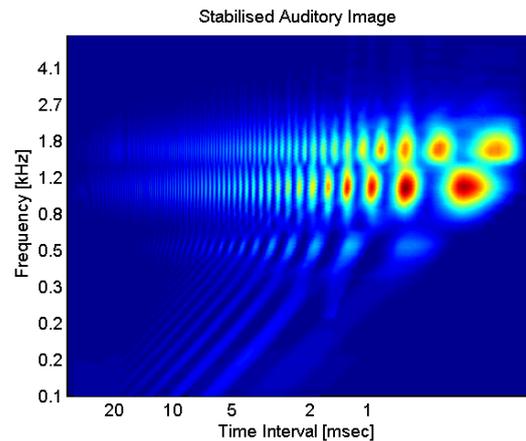


Figure 6: Stabilised Auditory Image for $a = 1.6$ m. x-axis: time interval since last strobe; y-axis: centre frequency of basilar membrane filter.

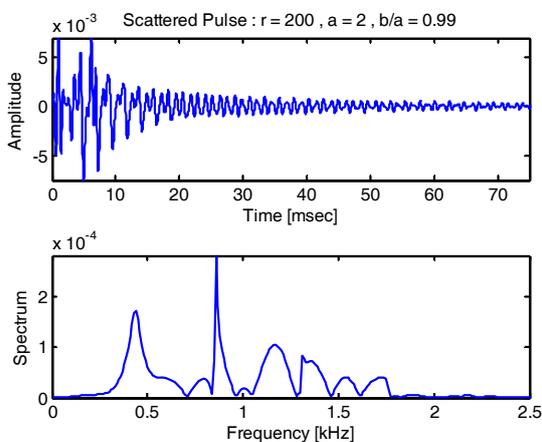


Figure 7: Scattered pulse for cylindrical shell with radius $a = 2.0$ m. Top panel: time history of pulse; bottom panel: frequency spectrum.

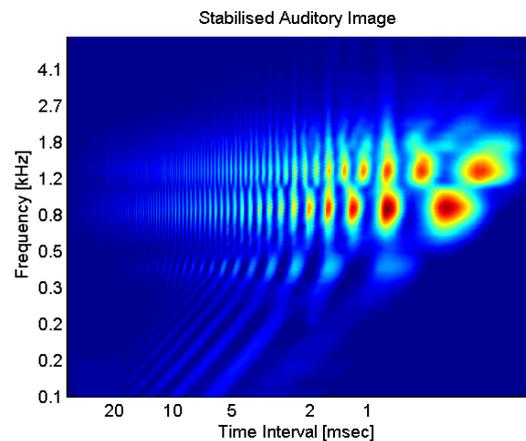


Figure 8: Stabilised Auditory Image for $a = 2.0$ m. x-axis: time interval since last strobe; y-axis: centre frequency of basilar membrane filter.

Comparing Figures 6 and 8 illustrates the change in the SAI when the target size changes. The activity maxima of the smaller object (Figure 6) glides along the diagonal downwards for the bigger object. Figure 9 and 10 show the sum of activity of the two “projected” 45° SAIs in direction of the diagonals, for all 10 cylinders. Figure 9 indicates the “Scale Profile” corresponding to the 45° diagonal axis and Figure 10 the “Shape Profile” corresponding to the axis perpendicular to it. From these it is possible to apply various methods to compare the size of two different objects from these profiles. In Figure 11 the first moment of the scale profile has been plotted for all 10 cylinders. One observes a general shift along the horizontal axis of the scale profile curves for different cylinders, and the first moments of area give an indicator of this shift. We note that the moments generally increase with increasing cylinder size, although there appears a lack of ability to discriminate well between the 4 smallest cylinders. Figure 12 then shows the same results for water-filled steel cylindrical shells of the same outer radii values a , and same ratios $b/a = 0.99$ for all cylinders. Again a general increase in the function value is observed with poor discrimination of the 4 smallest cylinders.

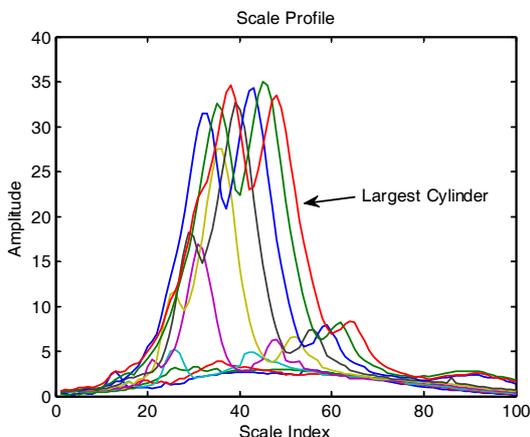


Figure 9: Scale Profile curves for 10 aluminium cylindrical shells ($b/a = 0.99$) with $a = 0.2$ m to $a = 2.0$ m. Axis units are arbitrary.

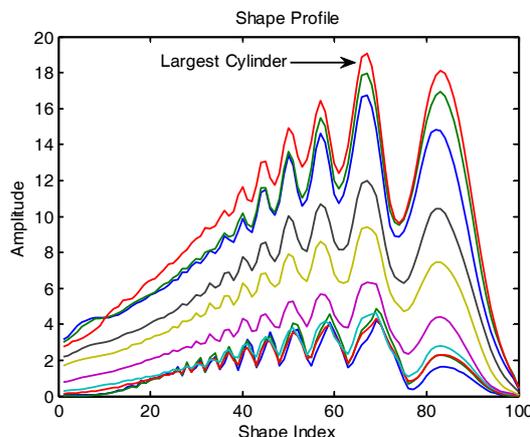


Figure 10: Shape Profile curves for 10 aluminium cylindrical shells ($b/a = 0.99$) with $a = 0.2$ m to $a = 2.0$ m. Axis units are arbitrary.

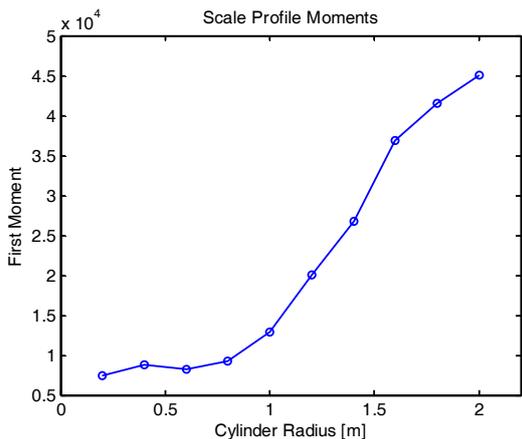


Figure 11: First area moments of the scale profile functions for all aluminium cylindrical shells: $b/a = 0.99$.

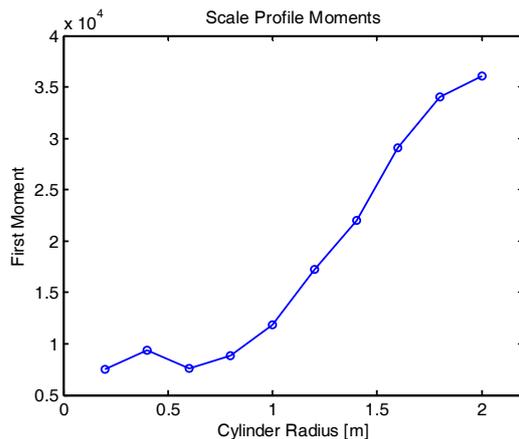


Figure 12: First area moments of the scale profile functions for all steel cylindrical shells: $b/a = 0.99$.

4 SUMMARY, CONCLUSIONS, FURTHER WORK

We present in this paper a model that discriminates the scattered pulses of underwater objects by object size. The model is based on simulating human auditory processing. The basic principle is

that scattered sound of objects of different size vary in two distinct ways: in the frequency of the response and in the frequency-temporal properties of their decay. Both properties affect the described transformation. The model is in an experimental stage but preliminary results are interesting. Further work includes simplification of the model, a sound description of its mathematical basis and applying it to various objects of different sizes and shapes. Specifically we want to substitute the Gammatone filterbank with the more powerful compressive Gammachirp filterbank that is designed to perform a size-optimised Mellin transformation^{10,2}. The nonlinear dynamics of this approach has a further advantage: In the linear filterbank that is used here, the specular reflection of the scattered pulse carries the highest energy. Since this early reflection only carries information about the original pulse, and not about the scattered object, the analysis may improve when this first high-energy reflection is attenuated compared to the latter information in the pulses. The compressive Gammachirp filter tackles this problem by instantaneous compression, a mechanism that is physiologically inspired by the feedback mechanism provided by outer hair-cells in the mammalian cochlea. Initial indications are that some form of size discrimination can be obtained over a limited range of object sizes. Different object shapes need to be tested to see whether this is just a special-case result for cylinders and more physical insight into reasons for when we should and should not expect effective discrimination would be of value.

5 ACKNOWLEDGMENTS

This work was carried out as part of the Weapon and Platform Effectors Domain of the MoD Research Programme.

6 REFERENCES

1. D.R.R. Smith, R.D. Patterson, H. Kawahara and T. Irino, 'The processing and perception of size information in speech sounds'. *J.Acoust.Soc.Am.* 117, 305-318. (2005).
2. C. Carello, K.L. Anderson, et al., 'Perception of object length by sound'. *Psych.Sc.* 9, 211-214. (1998).
3. C.M. DeLong, W.W. Au and S.A. Stamper, 'Echo features used by human listeners to discriminate among objects that vary in material or wall thickness: implications for echolocating dolphins'. *J.Acoust.Soc.Am.* 121, 605-617. (2007).
4. R.D. Doolittle and H. Uberall, 'Sound scattering by elastic cylindrical shells'. *J.Acoust.Soc.Am* 39, 272-275. (1966).
5. P. Ugingcius and H. Uberall, 'Creeping wave analysis of the Acoustic Scattering by Elastic Cylindrical Shells', *J.Acoust.Soc.Am.*, 43, 1025-1035. (1968).
6. R.D. Patterson, M.H. Allerhand and C. Giguere, 'Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform'. *J.Acoust.Soc.Am.* 98, 1890-1894 (1995).
7. S. Bleack, T. Ives and R.D. Patterson, 'Aim-mat: the auditory image in MATLAB'. *Acta Acustica*, 90, 781-788. (2004).
8. T. Irino and M. Unoki, 'An Analysis / Synthesis auditory filterbank based on an IIR implementation of the Gammachirp'. *J.Acoust.Soc.Jap.* Vol 20(5), 397-406. (1999).
9. B.R. Glasberg, and B.C. Moore, 'Derivation of auditory filter shapes from notched-noise data'. *Hear.Res.* 47, 103-138. (1990).
10. M. Unoki, T. Irino, B. Glasberg, B.C. Moore and R.D. Patterson, 'Comparison of the roex and gammachirp filters as representations of the auditory filter'. *J.Acoust.Soc.Am* 120, 1474-1492. (2006).
11. K. Krumbholz, R.D. Patterson, A. Nobbe and H. Fastl, 'Microsecond temporal resolution in monaural hearing without spectral cues?'. *J.Acoust.Soc.Am.* 113, 2790-2800. (2003).
12. T. Irino and R.D. Patterson, 'Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform'. *Speech Comm.* 36, 181-203. (2002).