J. Mar. Biol. Ass. U.K. (2007), **87**, 27–34
Printed in the United Kingdom

doi: 10.1017/S0025315407049508

A comparison of model and non-model based time–frequency transforms for sperm whale click classification

M. van der Schaar*, E. Delory*, J. van der Weide[†], C. Kamminga[‡], J.C. Goold[ʃ], N. Jaquet[§] and M. André*[¶]

*Laboratori d'Aplicacions Bioacústiques (LAB), Universitat Politècnica de Catalunya, Spain. [†]Department of Statistics, Delft University of Technology, The Netherlands. [‡]Information Theory Group, Delft University of Technology, The Netherlands. [ʃ]Institute of Environmental Science, University of Wales, Bangor, UK. Present address: Bridge Marine Science Group, 10 Bridge Street, Menai Bridge, Anglesey, UK. [§]Texas A&M University, Galveston (TX), USA. Present address: Center for Coastal Studies, 5 Holway Street, Provincetown (MA), USA. [¶]Corresponding author, e-mail: michel.andre@upc.edu

We tried to find discriminating features for sperm whale clicks in order to distinguish between clicks from different whales, or to enable unique identification. We examined two different methods to obtain suitable characteristics. First, a model based on the Gabor function was used to describe the dominant frequencies in a click, and then the model parameters were used as classification features. The Gabor function model was selected because it has been used to model dolphin sonar pulses with great precision. Additionally, it has the interesting property that it has an optimal time–frequency resolution. As such, it can indicate optimal usage of the sonar by sperm whales. Second, the clicks were expressed in a wavelet packet table, from which subsequently a local discriminant basis was created. A wavelet packet basis has the advantage that it offers a highly redundant number of coefficients, which allow signals to be represented in many different ways. From the redundant signal description a representation can be selected that emphasizes the differences between classes. This local discriminant basis is more flexible than the Gabor function, which can make it more suitable for classification, but it is also more complex. Class vectors were created with both models and classification was based on the distance of a click to these vectors. We show that the Gabor function could not model the sperm whale clicks very well, due to the variability of the changing click characteristics. Best performance was reached when three subsequent clicks were averaged to smoothen the variability. Around 70% of the clicks classified correctly in both the training and validation sets. The wavelet packet table adapted better to the changing characteristics, and gave better classification. Here, also using a 3-click moving average, around 95% of the training sets classified correctly and 78% of the validation sets. These numbers lowered by only a few per cent when single clicks, instead of a moving average, were classified. This indicates that, while the features may show too much variability to enable unique identification of individual whales on a click by click basis, the wavelet approach may be capable of distinguishing between a small group of whales.

INTRODUCTION

Recordings of sperm whales (*Physeter macrocephalus*) often contain a mixture of clicks from different whales, which are difficult to separate into individual sequences. This separation is usually done manually, a time-consuming and difficult process. While a true model for the production of a click still does not exist, we can expect this function to depend, at least partly, on the animal's morphology, since the signal results from a complex, and not yet fully understood, reverberation pattern within the spermaceti–junk complex. Although different animals may have a similar directivity pattern regardless of their body size, the resulting temporal and spectral content of the click will inevitably be size- and hence individual-dependent. There may also be a more deterministic or behavioural influence, but the extent to which an animal can control its click production remains unknown. We tried to find characteristics that would enable discrimination between clicks from different whales, at least during a group dive, and reconstruction of their individual click trains. We also explored the possibility of using these features as a biometric in order to uniquely identify a sperm whale. Recent papers suggest that the latter may be difficult due to the directionality of the click (Møhl et al., 2003) and the influence of hydrostatic pressure (i.e. the diving depth) on the whale (Thode et al., 2002), which both affect the frequency content. However, for the purposes of discriminating between clicks from different animals this variability is not necessarily as important, especially when the changes are gradual. Detection algorithms could potentially adjust the classification parameters in real-time, although real-time classification is not developed in this paper. We used two different approaches to find discriminating features in sperm whale clicks.

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

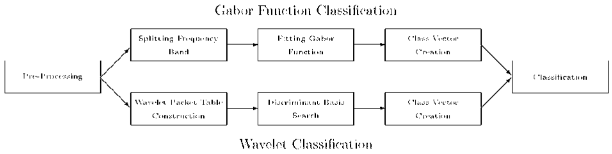

Figure 1. Signal processing steps taken for both the methods in order to create the class vectors and to perform classification. The steps are detailed in the Appendix.

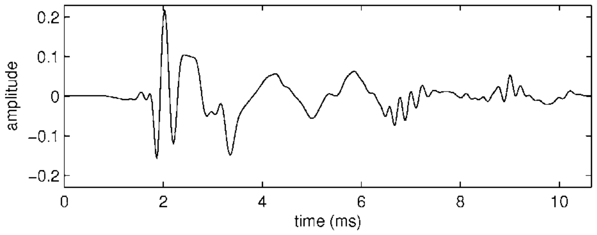

Figure 3. Typical example of a sperm whale click, low pass filtered at 5 kHz.

The first approach was motivated by the results of Goold & Jones (1995), who showed that clicks from a single click train can have fairly constant dominant frequencies in at least two different frequency bands. For a male sperm whale, these frequencies can be found around 500 Hz and 2000 Hz. For female sperm whales, these frequencies are 1200 Hz and 3000 Hz respectively. We tried to find a parametric estimation for the signal in these frequency bands using the so-called elementary signal of Gabor (1947). The Gabor function is suitable for describing harmonic frequency signals without reverberations or echoes and has been used successfully in the past to model clicks from porpoises and dolphins (Kamminga & Cohen Stuart, 1995, 1996). The dominant frequencies will then appear as parameters in the model. With an accurate model it might be possible to get a combination of parameters which uniquely identify a sperm whale.

For the second method, we tried to find characteristic features in a click using a local discriminant basis, constructed with a wavelet packet approach as described in Saito & Coifman (1994). Similar methods have been applied successfully in the classification of signals from other underwater mammals (Delory & Potter, 1998; Huynh et al., 1998; Delory et al., 1999), and a discrete wavelet transform has been used with some success in an attempt to classify sperm whales (Dougherty, 1999). The advantage of using the wavelet-packet approach is that the entire method can be implemented in hardware as a recursive filter, which allows the method to be employed in the field for real-time analysis.

MATERIALS AND METHODS

Data collection and preparation

Sperm whale click recordings were collected from an inflatable boat during four field seasons spanning four to ten weeks each (from 1997 to 1999) at Kaikoura, New Zealand (Jaquet et al., 2001). Recordings were made of solitary diving male sperm whales using an omnidirectional hydrophone (Sonatech 8178; frequency response 100 Hz to 30 kHz ±5 dB) lowered to 20 m. The hydrophone was first connected to a fixed gain amplifier (flat response from 0 to 45 kHz), and then to one channel of a Sony TCD-D10PROII digital audio tape recorder (frequency response 20 Hz to 22 kHz ±1 dB with an anti-alias filter at 22 kHz). The digital audio tape recorder samples at 48 kHz with 16-bit data resolution. The data recordings were subsequently filtered with a band-pass linear phase filter between 100 Hz and 12 kHz. The use of data from solitary sperm whales gave the advantage that we knew with certainty which click belonged to which animal, and this allowed us to train the classifiers with sets that contained no errors. It should also be noted that the recordings were made during different seasons, which may have had an influence on the data.

In the following we give a textual overview of the classification process shown in Figure 1; a more detailed description of the methods and analytical expressions is given in the Appendix.

In order to compare the Gabor and wavelet methods, five click trains were chosen. The clicks in these sequences were automatically detected, and then visually filtered to remove surface echoes and clicks that appeared to be badly affected by noise. Before the clicks were used they were de-noised using a standard soft-thresholding algorithm, available in WaveLab (Donoho et al., 1999). Only the first 10 ms of the clicks and frequencies below 5000 Hz were used in order to minimize the effect of reverberations within the whale head and the effects of click directivity. The remainder of a click was found to be unsuitable for useful feature extraction.

For the Gabor model parameterization we separated the data into two frequency bands with two band pass filters

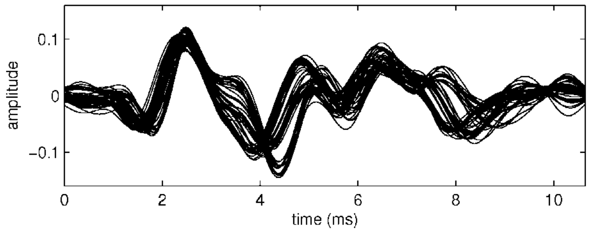

Figure 2. Cumulative plot of the low frequency component of 50 clicks from Set 4. The plot shows a second pulse coming into the signal at different time delays, complicating the synchronisation of the clicks.

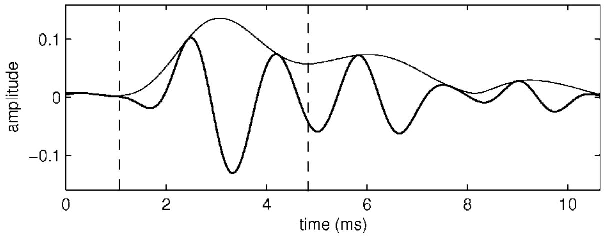

Figure 4. Low frequency component of the click in Figure 3 with its envelope. In this case the envelope allows reliable isolation of the main pulse in the click.

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

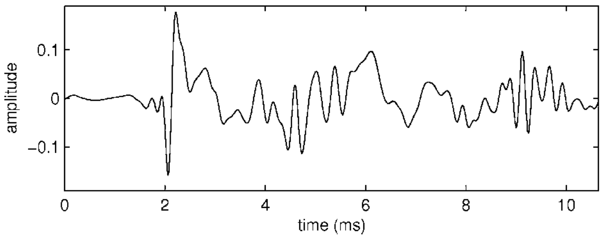

Figure 5. Example of a click, low pass filtered at 5 kHz, with a 'fast' reverberation.

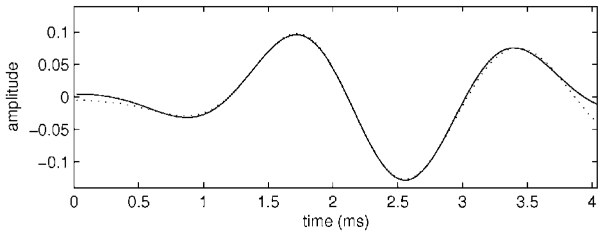

Figure 7. The Gabor model fitted on a low frequency fit part of a click, the original signal is dotted, the model is solid.

on the original signal. One frequency band was selected between 100 and 1000 Hz and the second band between 1000 and 5000 Hz. The bands were intended to focus on the dominant frequencies of male sperm whale clicks. The data used for the wavelet method were band filtered between 100 and 5000 Hz. All clicks were time-aligned using a cross-correlation technique. This was important especially for the wavelet approach, as the position of the wavelet coefficients are a function of time. A shift in the wavelet coefficients may result in using the wrong coefficient as classifier. As can be seen in Figure 1, which illustrates the phase error for the low frequency components of the clicks in set four, there was still some drift in the phase, and it was necessary to allow a phase error of about 30 sample points. After phase alignment the clicks were normalized in energy. The training sets consisted of the first 50 clicks in the 'cleaned' sequences; the remaining clicks in each train were used as validation sets.

Description of the Gabor algorithm

The Gabor function, given by equation (1), can only be used for signals that contain a single central frequency and pulse. Therefore, we are limited to only using those frequency bands of the signal that contain the dominant frequency of the sperm whale click, and we can only apply it to the first part of a click, before the first reverberations or echoes enter the signal. For low dominant frequencies, occurring around 500 Hz, part of the first reverberation will always overlap the end of the first pulse which makes it more difficult to fit the model. The higher frequencies suffer less from this problem, but are more easily affected by the direction and distance of the animal, and may not give constant parameters.

$$G(t) = He^{-\alpha^2(t-t_0)^2}\cos\left(2\pi f_0(t-t_0) = \varphi\right) \qquad (1)$$

where H=amplitude, α=sharpness, f_0=frequency, φ=phase and t_0=mid epoch.

We will describe the extraction of the main pulse with the example click shown in Figure 3. First, the click is filtered in the two frequency bands around the dominant frequencies. Then, its envelope, shown for the low frequency band in Figure 4, is used to isolate the pulse by taking the area of the click between two local minima of the envelope. In this case the area between the two dashed vertical lines includes roughly 3.5 ms and two cycles of the pulse. This should give enough points to reliably fit the Gabor model on the pulse. However, when the click is followed by a reverberation, or an echo within a very short time interval (less than 4 ms), this method may not give enough points for an accurate estimate of the Gabor parameters. This is illustrated in Figure 5, where a click is almost immediately followed by an echo or reverberation. Comparing with Figure 3, it can be seen that the click contains many more high frequency pulses around 4 ms, which makes it less likely that a sufficient number of uncorrupted points of the first pulse can be found. Its low frequency component and envelope are shown in Figure 6. In comparison with Figure 4, the echo made the second half of the pulse unusable, not even allowing a single period of the main pulse for modelling. From the figure itself, it can be seen that the frequency estimate will be too high, as the first pulse is shortened and corrupted on the right side by the reverberation. Even though the model might fit perfectly on the reverberating pulse, it will not reflect the real frequency of the emitted signal. The effect of the changing echo time-delay of arrival on a click can also be seen in Figure 2, where the clicks could not be synchronised correctly.

After the first pulse was localized, standard line fitting algorithms were used to fit the model on the click. The results of fitting the model on the click of Figure 3 are shown in Figures 7 and 8, for both the low and high frequency bands. Despite a slight error at the signal boundaries, the centre of the signal seems to be closely described by its model.

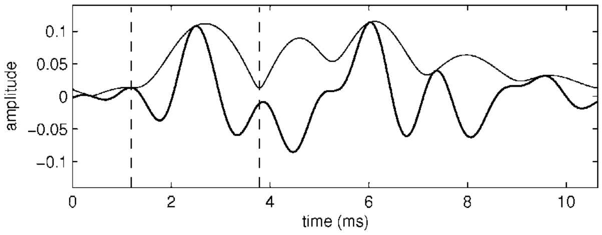

Figure 6. Low frequency component of the click in Figure 5 with its envelope. Comparing to Figure 4, there is almost one millisecond less information available.

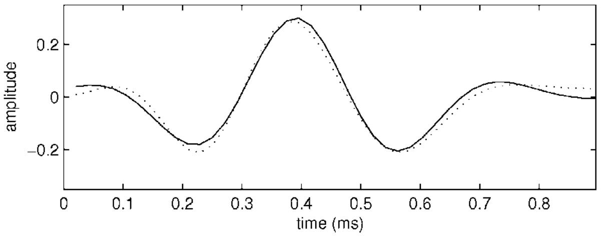

Figure 8. The Gabor model fitted on a high frequency part of a click, the original signal is dotted, the model is solid.

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

Table 1. *Average Gabor fit statistics. The energy residues are the averages of the percentages of energy not explained by the model. Correlation low and high give the averages of the correlation between the clicks and their Gabor models.*

	Click Train				
	1	2	3	4	5
Energy residue % low	4.9	6.5	3.8	9.0	6.4
Energy residue % high	5.4	1.2	3.7	4.3	3.2
Correlation low	0.976	0.969	0.984	0.957	0.963
Correlation high	0.973	0.995	0.978	0.979	0.988

A summary table of the results of fitting the Gabor model is shown in Table 1. The first two rows give the (mean) percentage of energy not explained by the model, i.e. *E*(*signal–model*)/*E*(*signal*). The last two rows give the (mean) correlation between the model and signal. The high frequencies seem to be modelled more accurately. This can partly be explained by the reduced influence of noise, and by the fact that the number of points that have to be described by the model is smaller than at lower frequencies. However, this can motivate the use of a higher weight when the parameters are used for identification. Only the sharpness and frequency parameters of both the low and high frequency bands were used for identification, as the amplitude and phase were found to be too variable. The class vector was constructed by taking the mean of the four features in the training set, and classification was based on the distance of a click's characteristics to these vectors.

Description of the wavelet algorithm

A complete description of the relationship between wavelets and filters can be found in Strang & Nguyen (1997). From a signal processing point of view, a wavelet transform will split a signal in two frequency bands using a high- and low pass filter, keeping the high frequency wavelet coefficients and re-splitting the low frequency scaling coefficients. A schematic of the discrete wavelet transform (DWT) is shown in Figure 10. The signal $a_{0,k}$ enters the filters, which produce the new coefficients at a lower scale $a_{1,k}$ (low pass) and wavelet coefficients $b_{1,k}$ (high pass). The subscript k is the index of the coefficient, which in our case ran from 0 to 511. The outputs of these filters contain some redundancy as there are now 2×512 samples. To remove this redundancy, the filter outputs are down sampled by two resulting in 2×256 samples.

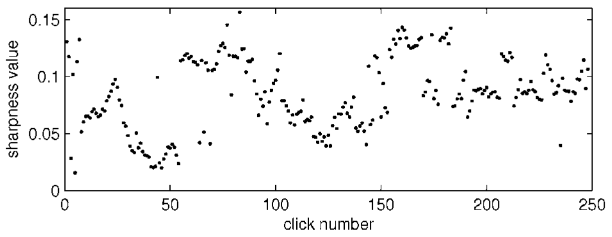

Figure 9. Sharpness parameter from the parameterization of the high frequency component of the clicks in Set 1.

$$\begin{tikzcd}
a_{0,k} \arrow[r, "L"] \arrow[dr, "H"'] & a_{1,k} \arrow[r, "L"] \arrow[dr, "H"'] & a_{2,k} \arrow[r, phantom, "\cdots\cdots"] & {} \arrow[r, "L"] \arrow[dr, "H"'] & a_{N,k} \\
& b_{1,k} & b_{2,k} & & b_{N,k}
\end{tikzcd}$$

Figure 10. Discrete wavelet transform through a low pass and high pass filter. The signal $a_{0,k}$, where k is the coefficient index, enters at the left. The filter creates the down sampled scale ($a_{1,k}$) and wavelet ($b_{1,k}$) coefficients. The scale coefficients are then run through the filter again until the lowest scale has been reached. The decomposition in dyadic frequency bands is shown in Figure 11.

The DWT continues with filtering the scale coefficients, until a lowest scale is reached. The signal is then represented by the wavelet coefficients $\left\{b_{i,k}\right\}_{i=1}^{N}$ and the lowest scale coefficients $a_{N,k}$. A wavelet packet algorithm will continue entering both outputs $a_{j,k}$ and $b_{j,k}$ back into the filters, storing all (down sampled) coefficients at every level. Figure 11 shows how frequency bands are dyadically decomposed every time the coefficients are run through the filter. A standard discrete wavelet transform would only contain the 0–3 ($a_{3,k}$), 3–6 ($b_{3,k}$), 6–12 ($b_{2,k}$) and 12–24 kHz ($b_{1,k}$) intervals (bins). The entire wavelet packet table gives a redundant representation of the signal, and many different bases can be selected from it to rebuild the original. For example, from Figure 11, a possibility would be the 0–6, 6–9, 9–12 and 12–24 kHz intervals, another choice could be 0–3, 3–6, 6–12, 12–18 and 18–24 kHz.

For identification we were interested in finding a basis that will emphasize the differences between classes. The details of finding a discriminating basis from a wavelet packet table can be found in Saito & Coifman (1994), a description of the algorithms and methods we used is given in the Appendix. Although creating complete wavelet packet tables might be time consuming, once the basis is chosen, then only a small part of the tree has to be rebuilt for the classification process. We used WaveLab (Donoho et al., 1999) to perform the wavelet operations.

Finally, in order to classify clicks, class vectors were created from the most discriminatory coefficients in the local basis. The classification was then based on the distance of a click's characteristics to these vectors.

RESULTS

The results of classifying both the training set and the remainder of the data for all five click trains using the Gabor

split level	Frequency Bands							
0	0-24 kHz							
1	0-12 kHz				12-24 kHz			
2	0-6 kHz		6-12 kHz		12-18 kHz		18-24 kHz	
3	0-3 kHz	3-6 kHz	6-9 kHz	9-12 kHz	12-15 kHz	15-18 kHz	18-21 kHz	21-24 kHz

Figure 11. Wavelet packet table, showing the decomposition of the frequency bands every time the signal is sent through the filter. For example, bin (3,1,·), containing the 3–6 kHz band, corresponds to the wavelet coefficients $b_{3,k}$ in Figure 10. However, bin (3,2,·), covering the 6–9 kHz band, corresponds to the scale coefficients obtained after re-filtering the wavelet coefficients $b_{2,k}$ and are not normally available in the discrete wavelet transform.

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

Table 2. *Click classification with Gabor parameterization (moving average), values are correctly classified percentages.*

	Training set					Validation set				
Classified as	1	2	3	4	5	1	2	3	4	5
Set 1	73	21	23	19	17	43	6	3	15	0
Set 2	0	79	0	8	6	30	86	18	15	0
Set 3	8	0	71	0	17	1	4	79	0	22
Set 4	0	0	0	73	0	15	4	0	58	0
Set 5	19	0	6	0	60	11	0	0	12	78

Table 3. *Click classification with a local discriminant basis (moving average), values are correctly classified percentages. Both the training and validation sets classify significantly better than when the Gabor model is used.*

	Training set					Validation set				
Classified as	1	2	3	4	5	1	2	3	4	5
Set 1	100	0	4	0	0	70	0	0	0	0
Set 2	0	100	0	0	0	14	71	34	0	17
Set 3	0	0	90	0	2	0	0	64	0	0
Set 4	0	0	0	100	0	16	0	0	100	0
Set 5	0	0	6	0	98	0	29	1	0	83

function are shown in Table 2. To create these results a moving average of three clicks was used to strengthen the characteristics. Also, the classifiers from the high frequency band were given a weight four times higher than those from the lower band. The training data already classified quite poor, with correct classification percentages between 60% for class five and 79% for class two. The validation set did not do much better with percentages between 43% for class one and 78% for class five. It is notable that the validation set mixed up different classes than the training set. For example, validation set three shows 27 clicks in class two and only four clicks in class one, while training set three attributed 11 clicks to class one and none to class two. This can indicate a significant variability of the characteristics throughout the click trains. Taking a closer look at the parameter variation, Figure 9 shows the sharpness parameter for clicks from the first data set in the high frequency band. Both local trends and large jumps can be seen. This kind of variation can be caused by the variation of echo time delay of arrivals. At small variations of the parameter the algorithm could still separate the first pulse and the echo, while at a jump the algorithm may no longer have been able to make this distinction and included a part of the echo resulting in significant parameter variation. This kind of behaviour is difficult to correct and is a weakness of the Gabor model.

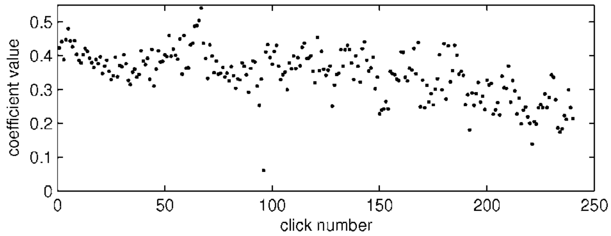

Figure 12. Values of the strongest discriminatory wavelet coefficient for all clicks in Set 1.

Table 4. *Click classification with a local discriminant basis, values are correctly classified percentages. The classification performs only slightly worse than when using a 3-click moving average.*

	Training set					Validation set				
Classified as	1	2	3	4	5	1	2	3	4	5
Set 1	98	0	2	0	0	69	0	0	0	0
Set 2	2	94	2	0	0	14	60	25	0	22
Set 3	0	0	82	0	4	1	3	66	0	2
Set 4	0	0	2	100	0	16	4	2	100	0
Set 5	0	6	12	0	96	0	33	7	0	76

Table 3 shows the results when the same dataset of five click trains was classified with a local discriminant basis, also using a three click moving average. It is clear that wavelets are doing a much better job. In the training sets between 90% and 100% of the clicks were now classified correctly, while the validation set had percentages between 64% for class three and 100% for class four. Comparing with the Gabor classification, validation set one shows similar errors in both tables, indicating a close similarity between sets one, two and four. It also seems that the wavelet coefficient discriminators were focused more on separating sets one and four, and there were no strong differentiators found for sets two and three. Plotting the most discriminating coefficient for the first dataset (Figure 12) shows a strong trend. Other coefficients showed trends as well and this may indicate that the method cannot be used for larger groups of whales. In Table 4 the results are shown when a moving average was not used; these results are similar to the ones with a moving average, with slightly lower percentages. The clicks in the training sets classified correctly for around 94%, and the validation sets classified correctly between 60% for set two and 100% for set four. Apparently the values of the coefficients change slowly enough so that they do not significantly change over three clicks. This can be important when the method is used for something other than unique identification.

DISCUSSION

We have shown that identification of individual sperm whales from a large group of whales based on single clicks using a linear classifier may not be feasible. In order to be successful, the number of animals that the system can recognize has to be kept small in order to cope with the variance in the features. Also, at this moment, there are no complete studies concerning click variability of a single animal over different seasons and under different environments, and we do not know how these might alter a classifier's performance. One problem is that, in the field, it is generally not possible to keep track of the position and

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

orientation of an animal at depth. Changes in orientation may cause alterations in click characteristics throughout a click train, especially with respect to a stationary listener, and it is difficult to find constant information that will allow unique classification. Identification based on an entire click train is more feasible, and can be done when a significant percentage of the clicks falls within one class. In that case, when classification is based on click trains or perhaps on entire dives, coefficients could be taken not just from the first 50 clicks of a train, but randomly from a train or dive. This should improve performance, as was shown in Dougherty (1999). However, we did not have enough data to investigate this in detail. Therefore, in this paper we focused on the more practical case where an expert manually classifies the first few clicks in a recording, and then an automatic classifier continues with the remainder. The classification process could also be improved by limiting it to those clicks that satisfy predefined conditions, for example a certain time delay before the first reverberation, a specific depth at which it was emitted or only using on-axis clicks.

The worse performance of the Gabor function approach can be explained by the fact that it is based on a fixed model that searches for a global feature, by fitting a single frequency on the duration of the first pulse of the signal. The function contains a high bias, as it can only approximate a Gabor-type pulse. In the case of sperm whale clicks, the reverberation made this fitting procedure difficult. Wavelets, on the contrary, allow searching for local features; especially a wavelet packet table contains coefficients for many different time–frequency resolutions, and offers an extensive choice for possible identifying characteristics. The wavelet approach is more flexible and does not try to fit a specific model.

Finally, considering that the click characteristics change slowly, using the wavelet method it might be possible to separate mixed click trains within a recording, as often encountered during group dives. In this case the first few clicks will have to be processed manually in order for the system to learn each whale's characteristics.

This study was funded by the Rolex Awards for Enterprise.

REFERENCES

Baraniuk, R., 1994. Wavelet soft-thresholding of time-frequency representations. In *IEEE International Conference on Image Processing* (Austin, TX), **1**, 71–74.

Beitsma, R., 1989. *Reverberaties in Dolfijnsignalen*. Msc thesis, Delft University of Technology, Delft, The Netherlands.

Coleman, T. & Li, Y., 1996. An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, **6**, 418–445.

Delory, E. & Potter, J., 1998. Signal processing aspect of signal detection masking and noise suppression. In *Proceedings of Acoustics and Vibration Asia'98, Singapore*. Published on CD-ROM (arl.nus.edu.sg).

Delory, E., Potter, J., Miller, C. & Chiu, C.-S., 1999. Detection of blue whales A and B calls in the northeast Pacific Ocean using a multi-scale discriminant operator. In *13th Biennial Conference on the Biology of Marine Mammals, Maui, Hawaii*. Published on CD-ROM (arl.nus.edu.sg).

Donoho, D., 1995. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, **41**, 613–627.

Donoho, D., Duncan, M., Huo, X. & Levi, O., visited 1999. Version 8.02, http://wwwstat.stanford.edu/~wavelab/

Dougherty, A., 1999. *Acoustic identification of individual sperm whales (*Physeter macrocephalus*)*. Master's thesis, University of Washington, Seattle, USA.

Gabor, D., 1947. Acoustical quanta and the theory of hearing. *Nature, London*, **159**, 591–595.

Goold, J. & Jones, S., 1995. Time and frequency domain characteristics of sperm whale clicks. *Journal of the Acoustical Society of America*, **98**, 1279–1291.

Huynh, Q., Cooper, L., Intrator, N. & Shouval, H., 1998. Classification of underwater mammals using feature extraction based on time-frequency analysis and BCM theory. *IEEE Transactions on Signal Processing*, **46**, 1202–1207.

Jaquet, N., Dawson, S. & Douglas, L., 2001. Vocal behavior of male sperm whales: why do they click? *Journal of the Acoustical Society of America*, **109**, 2254–2259.

Kamminga, C. & Cohen Stuart, A., 1995. Wave shape estimation of delphinid sonar signals, a parametric model approach. *Acoustics Letters*, **19**, 70–76.

Kamminga, C. & Cohen Stuart, A., 1996. Parametric modelling of polycyclic dolphin sonar wave shapes. *Acoustics Letters*, **19**, 237–244.

Møhl, B., Wahlberg, M., Madsen, P., Heerfordt, A. & Lund, A., 2003. The monopulsed nature of sperm whale clicks. *Journal of the Acoustical Society of America*, **114**, 1143–1154.

Oppenheim, A. & Schafer, R., 1989. *Discrete-time signal processing*, pp. 311–312. Englewood Cliffs, New Jersey: Prentice-Hall.

Saito, N. & Coifman, R., 1994. Local discriminant bases, in mathematical imaging: wavelet applications in signal and image processing II, (ed. A. Laine and M. Unser). *Proceedings of SPIE—the International Society for Optical Engineering*, **2303**, 2–14.

Strang, G. & Nguyen, T., 1997. *Wavelets and filter banks*. Wellesley MA: Wellesley-Cambridge Press.

Thode, A., Mellinger, D., Stienessen, S., Martinez, A. & Mullin, K., 2002. Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico. *Journal of the Acoustical Society of America*, **112**, 308–321.

Submitted 20 December 2004. Accepted 28 November 2006.

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

APPENDIX

This appendix will briefly describe the entire signal process pictured in Figure 1.

Pre-processing

The clicks were automatically detected using a threshold on the signal amplitude, in combination with a required number of zero crossings (to find a high frequency). Next they were 'visually filtered' to remove surface echoes and clicks that appeared to be badly affected by noise. The 'visual filtering' was done by an experienced operator inspecting the individual click waveforms and making a judgement on whether to use them or not. This was done to ensure that the training sets consisted of clicks that were as noise-free as possible. Using a Butterworth filter the data were filtered between 100–5000 Hz. This filter has the property that it has a 'maximally flat magnitude' in its pass band, which leads to minimum phase distortion. To obtain a zero phase delay the data were filtered first forward and then backward (Oppenheim & Schafer, 1989).

The clicks were synchronized using cross-correlation with a typical example click and de-noised with a soft-thresholding algorithm (Baraniuk, 1994; Donoho, 1995). This algorithm first calculates the wavelet coefficients of the signal for a specific wavelet (we used the Symmlet wavelet for reasons explained below), and sets those coefficients smaller than a threshold to 0. The magnitude of the remaining coefficients is linearly reduced and the coefficients are then used to reconstruct the de-noised signal. The thresholds were calculated based on a noise measurement in every individual recording.

For every click, only the first 10 ms (512 samples) were used for classification, as the remainder of a click was found to be unsuitable for feature extraction. The clicks were normalised in energy, and the first 50 clicks of every data set were used to find characteristic class vectors.

Gabor function classification

Splitting the frequency band

The frequency band was split in two separate bands, 100–1000 Hz and 1000–5000 Hz. This was done in order to isolate two dominant frequencies (Goold & Jones, 1995). The separation of the data in a low- and high frequency band followed the same protocol as the filtering in the pre-processing of the data.

Fitting the Gabor function

The first pulse was isolated from the click using the click's envelope (see text). The Gabor function (1) was then fitted on this pulse minimizing the squared error function,

$$\sum_i \left(G\left(t_i; H, \alpha, t_0, f_0, \varphi\right) - y_i\right)^2$$

where y_i are the points of the first pulse. The minimization was performed by a standard algorithm provided by Matlab, which is based on the interior-reflective Newton method (Coleman & Li, 1996). An initial estimate of the parameters was made in order to assure fast convergence. The amplitude H and frequency f_0 are easily estimated by the amplitude and number of zero-crossings of the first pulse. The other parameters can be estimated using linear regression (Beitsma, 1989). Writing the analytical representation of the Gabor signal and taking the natural logarithm gives,

$$G^*(t) = e^{-\alpha^2(t-t_0)^2} e^{j\left(f_0(t-t_0)+\varphi\right)}$$

$$\ln\left(G^*(t)\right) = -\alpha^2\left(t-t_0\right)^2 + j\left(f_0\left(t-t_0\right)+\varphi\right).$$

Then, after taking the real and imaginary parts,

$$\Re\left(\ln\left(G(t)\right)\right) = -\alpha^2\left(t-t_0\right)^2 = -\alpha^2 t^2 + 2\alpha^2 t_0 t - \alpha^2 t_0^2 =$$

$$= \beta_2 t^2 + \beta_1 t + \beta_0$$

$$\Im\left(\ln\left(G(t)\right)\right) = f_0 t + \varphi \qquad (\mathrm{mod}(2k\pi))$$

$$= \gamma_1 t + \gamma_{0.}$$

Using the points from the first pulse and linear regression, the above equations allow estimates of the other parameters, $\alpha = \sqrt{-\beta_2}$, $t_0 = \beta_1/(2\alpha^2)$, and $\varphi = \gamma_0$.

Class vector creation and classification

As described in the text, only the sharpness and frequency parameters were stable enough to be used for classification. Combining the low frequency and high frequency parameters, this resulted in four features. The class vectors were then created by taking the mean of each of the parameters over the training set. Subsequent classification was based on the Euclidean distance between these four characteristics of a click and the class vectors.

Wavelet classification

Wavelet packet table construction

An extensive description about wavelets and filter banks can be found in Strang & Nguyen (1997). The low pass and high pass filters shown in Figure 10 are described by the equations

$$a_{j+1,k} = \sum_l c\left(l-2k\right) a_{j,l} \quad \text{and} \quad b_{j+1,k} = \sum_l d\left(l-2k\right) a_{j,l}$$

where $c(n)$ are the low pass and $d(n)$ the high pass filter coefficients; $a_{j,l}$ is coefficient l of the signal at filter step j. Note the recursive nature of these filters, as the output $a_{j+1,k}$ is the input of both filters in the next filter step. The down sampling-by-two operation is expressed in the $2k$ term. The connection between these filters and wavelets is given by

$$\varphi(t) = \sum \sqrt{2} c(n) \varphi(2t-n) \quad \text{and} \quad w(\mathrm{t}) = \sum \sqrt{2} d(n) \varphi(2t-n).$$

The equation on the left is called the dilation equation, and the one on the right the wavelet equation. It can be seen that, in the case of a quadrature mirror filter bank, the low pass filter coefficients $c(n)$ directly define both the dilation equation and its corresponding wavelet. The discrete wavelet transform is calculated recursively by the filter equations, the scaling coefficients $a_{j+1,k}$ by the low pass filter, and the wavelet coefficients $b_{j+1,k}$ by the high pass filter. In the case of a wavelet packet table (WPT), both scaling and wavelet coefficients are used as inputs in the filter bank, leading to the redundant decomposition of a signal in multiple frequency bands as shown in Figure 11. The filter

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press

coefficients $c(n)$ we used for this paper correspond to the Symmlet wavelet, which has the property that it is almost symmetrical. This almost symmetry gives a minimal phase distortion to the filtered signal. We used a wavelet with eight vanishing moments, which, considering the 512-point signals, gave a maximum number of five frequency splits (number of times the signal was run through the filter).

Discriminating basis search

The discriminating basis construction is described in detail in Saito & Coifman (1994). After a WPT has been generated for every click in the training set, a representation of the signal is searched that maximizes the distance between the different classes. One way to measure the difference between classes is to measure the difference in energy in a specific time–frequency bin. In order to do this, the WPTs of the clicks from one class have to be combined. This is achieved with a time–frequency energy map, defined as

$$\Gamma_c(j,k,m) = \sum_i^{N_c} \left(\hat{x}_i^c(j,k,m)\right)^2 / \sum_i^{N_c} \left\|\mathbf{x}_i^c\right\|^2$$

where (j,k,m) denotes the position in the WPT, at splitting level j, frequency band k and coefficient m within the bin (for example, in Figure 11, the position (3,4,5) corresponds to the sixth coefficient in the 12–15 kHz frequency band); $\hat{x}_i^c$ denotes the wavelet coefficient of click sample i and class c at position (j,k,m); $\mathbf{x}_i^c$ the original click sample i of class c; Nc the number of training samples in class c.

After all the tables have been collapsed into one energy map per class, the discriminating power of a specific bin $(j,k,\cdot)$ can be measured by summing the differences of the coefficients in the bin between every pair of classes,

$$\mathcal{D}\left(\{\Gamma_c(j,k,\cdot)\}_{c=1}^{C}\right) = \sum_{p=1}^{C-1} \sum_{q=p+1}^{C} \mathcal{D}\left(\Gamma_p(j,k,\cdot), \Gamma_q(j,k,\cdot)\right).$$

In the above expression $\mathcal{D}$ is an additive discriminant measure, for which we used the squared l^2-norm; C is the total number of classes. A high value for $\mathcal{D}$ indicates that coefficients between at least two classes lie far apart, and that the bin may be able to differentiate between at least two classes.

The discriminating basis can now be constructed with the following rule, if the discriminant measure over a bin, $(j,k,\cdot)$, is higher than the sum of the measures taken over the two bins it splits into, $(j+1,2k,\cdot)$ and $(j+1,2k+1,\cdot)$, then the 'parent' bin $(j,k,\cdot)$ is selected, otherwise it is split up. For example, using Figure 11, if $\mathcal{D}$('0–12 kHz')>$\mathcal{D}$('0–6 kHz')+D('6–12 kHz'), then the frequency band 0–12 kHz is selected, otherwise the two half-bands are used.

Class vector creation and classification

The discriminant basis gives a representation for the signals that maximizes the distances between the different classes, under selected measures. To create a class vector, the strongest coefficients from this basis are chosen. In this report, these coefficients were selected with Fisher's class separability (Saito & Coifman, 1994), which calculates the discriminating power of a wavelet coefficient at position (j,k,m) as follows:

$$\frac{\sum_c \left(\text{mean}_i(\hat{x}_i^c(j,k,m)) - \text{mean}_c(\text{mean}_i(\hat{x}_i^c(j,k,m)))\right)^2}{\sum_\text{c} \text{var}_\text{i}(\hat{x}_i^c(j,k,m))},$$

mean_i and var_i take the mean and variance of the coefficient over all samples in the class and mean_c takes the mean over all classes. In our case, choosing the 15 strongest coefficients gave optimal results. Classification of a click was done by calculating the Euclidean distance of the click to the class vectors.

https://doi.org/10.1017/S0025315407049508 Published online by Cambridge University Press