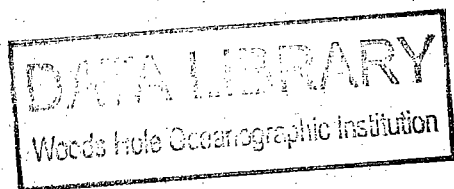**WHOI-94-13**

c. 2

Woods Hole
Oceanographic
Institution

Marine Animal Sound Classification

by

Kurt M. Fristrup and William A. Watkins

October 1993

Technical Report

Marine Animal Sound Classification

by

Kurt M. Fristrup and William A. Watkins

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

October 1993

Technical Report

Approved for public release; distribution unlimited.

Approved for Distribution:

George V. Frisk

George V. Frisk, Chair
Department of Applied Ocean Physics
& Engineering

Table of Contents

Introduction

Marine animals produce a remarkable variety of sounds (Watkins and Wartzok 1985). A primary goal of our bioacoustic program at the Woods Hole Oceanographic Institution (WHOI) has been to parse this variation into biologically significant classes of signals. Marine mammal sounds exhibit distinctive features associated with species (op. cit.), individual identity (Caldwell, Caldwell and Tyack 1990), and certain behaviors . These features have never been examined quantitatively, comparing the sounds of a wide variety of species. Do these sounds remain distinctive as the scope of comparison broadens? Experienced researchers can aurally and visually (via spectrographic analysis) identify acoustic features that appear to be species-specific, and sometimes features unique to individual animals; can we specify numerical algorithms that objectively recognize these distinctions?

The logistic requirements for addressing these questions remain formidable. Many biological and environmental conditions potentially contribute to acoustic variability. To quantify the interspecific and intraspecific variability in marine animal sounds, large numbers of sounds must be accumulated and analyzed for each distinct class of sounds. Several results indicate that correct identification of sounds is significantly improved by utilizing all available biological information during the construction or "training" of the classifier (Fristrup and Watkins 1992). Therefore, numeric features extracted from acoustic data must be conveniently referenced to species, population, group, social context, behavior, activity, individual identity, sex, reproductive situation, age, season, geographic location, water depth, and sound propagation.

The SOUND database system of marine animal sounds (Watkins, Fristrup, and Daher 1991) provided this capability. The databases and associated files contained thousands of digitized sound segments. The database described the time, geographic location, recording conditions, identity of the animal(s) producing the sounds, the behavioral observations associated with sound production, etc. These SOUND databases represented years of work by several people. The ONR Ocean Acoustics Program (Marshall Orr) provided the initial funding, but continued development and expansion were funded by a blend of Navy and private sources. The feature extraction and classification program would not have been feasible without the SOUND resources. In turn, development of feature extraction and sound classification, funded by TRICCSMA, resulted in significant structural improvements in the database systems. New, relational database structures were implemented to permit flexible and convenient integration of statistical results with the biological and environmental information.

The ability to select and to analyze acoustic measurements based on related biological or environmental observations was crucial for these data. This could have been done by segregating data files for different species, activities, locations, etc. and independently processing each batch. However, it would have been increasingly cumbersome and difficult to manage data segregation as the scope and complexity of analyses increased. Maintaining the integrity of the data (correct file assortment by attributes, labeling processed output) would be problematic. A more powerful technique was to process all

4

sound cuts in one batch, and to attach an identifier to the vector of measurements from each sound cut. Automatic feature extraction for all available sound cuts proceeded unattended, with one command. The resulting measurements, with the attached identifier, were imported as a table in a relational database program. The identifier provided a unique link between the vector of quantitative acoustic features describing a sound and the biological observations associated with that sound. Interactive exploration of relationships among statistics and biological or environmental factors followed, exploiting the convenience and flexibility of relational database queries.

The SOUND text databases for the recordings and the digital sound sequences (Watkins, Fristrup, and Daher 1991) could have accommodated new numeric data from the statistical analyses, but the INMAGIC software used to develop this system was unsuitable. It required restructuring the entire database each time the number of numeric fields changed. This was not feasible: the analyses required many iterations and modifications. Therefore, PARADOX software (relational database support, with visual, query-by-example interface) was used. The text information from the SOUND databases remained unmodified as distinct tables, and additional tables were created for the acoustic results. Statistical summaries of subsamples were generated with specific queries. This structuring of information also permits queries using sound characteristics to identify species and locations that have previously exhibited similar sounds.

The acoustic feature extraction program (AcouStat) was called with one command line parameter, the name of the file containing a digital sound cut. AcouStat processed these data, and sent the results to standard output (stdout in the C language). Redirection of this output was used to store the data, or to pipe the acoustic features to another program for additional processing. For the analyses described here, these data were appended to a text file that was later imported into a PARADOX table. PARADOX queries were used to link text and acoustic features, and the results were exported to SPLUS and SYSTAT (data analysis packages) for classification analysis.

There is no scientific precedent for quantification of time-frequency characteristics of animal sounds on this scale. No prior work has dealt with so many species and such a variety of repertoires from individual animals. The WHOI studies of marine animal acoustics, a continuous program initiated by William E. Schevill in the late 1940's, have provided the heuristic basis for selecting features and designing algorithms. Our personnel utilize many different acoustic features to describe sounds and diagnose their identity. As a first step toward the development of an automatic, objective system for identifying animal sounds, we devised statistical measures to resolve familiar acoustic features.

The Feature Extraction Algorithm

There is no *a priori* basis for selecting statistics that will maximize classifier performance. Our approach has been iterative, guided by the following criteria:

- Each statistic was designed to emphasize particular parameters of animal sounds that we recognized as important for distinguishing species.
- Each statistic had to be insensitive to temporal artifacts introduced by ocean propagation (multipath, fading).
- Most statistics had to be relatively insensitive to noise levels (resistant to outliers, possessing a high breakdown point).
- Most statistics had to yield consistent results despite variation in the shape of the ambient noise power spectra.
- Many statistics needed to relate to obvious features in time-frequency displays of these sounds (duration, frequency range, ...). This eased interpretation of success and diagnosis of flawed performance.

The signal processing was relatively simple, using power spectra derived from a Fast Fourier Transform. For most files, FFT size was 256 sample points, but for very short files (low sampling rates) the FFT size was decreased to obtain a minimum of 16 FFT data blocks per file. Adjacent blocks overlapped by 25%. The samples were level-shifted to obtain a block mean of zero, tapered with a Hamming window, and level shifted again to remove the DC bias introduced by tapering. The complex FFT values were multiplied by their complex conjugates (to form the magnitude-squared values), and the energy in the "negative" frequency bins was added to the corresponding "positive" frequency bins. Thus, the sum of the first $N_{ff}/2 + 1$ bins equaled the sum-squared energy. Windowing smoothed the power spectra. Overlapping increased time resolution, and extracted useful information from data that would otherwise have been "lost" in the tails of the window.

More precise time-frequency analyses could be substituted for this procedure (Wigner-Villle, RID), but resolution of signal characteristics on these scales would be sensitive to phase perturbations. These techniques might require explicit source extraction (environmental deconvolution) to provide signals of sufficient quality. Such efforts were not indicated in the course of our analyses, but they remain an attractive option for very brief signals.

Noise Compensation

Our noise compensation technique starts with an estimated "average" noise power spectrum. This was computed as the median of the data blocks comprising the initial and terminal 5% of the sound file. By convention, our transient extraction protocols included a leader and trailer of background sound. Some signal energy was occasionally present in one of these regions, but the median spectrum was not grossly inflated by these signals. Previously, we used a large number of data blocks, taken at fixed intervals throughout the sound cut, and computed a noise spectrum from the quietest sections. The newer routine is comparable and faster.

A multiple of this noise spectrum is subtracted from each data block's spectrum; all negative results are set to zero. Previously, we subtracted a constant multiple -- about 7x the noise spectrum -- from all data blocks. This multiple was subjectively determined by examining a variety of spectrographs. The newer technique was "adaptive" because the level of the noise spectrum was adjusted prior to subtraction from each block's spectrum. This enhancement was prompted by frequent observations of "swelling" noise backgrounds: the shape of the noise spectrum seemed relatively constant, but the noise energy fluctuated widely in some cuts. Initial attempts to model this utilized orthogonal decompositions to identify principal components of noise spectrum variability. Utilization of more than two orthogonal components was found to introduce spectrographic artifacts in the form of frequency banding, due to the partial correlation of some noise vectors with transient sound spectra (biological signals). Also, the most significant improvement was seen to result from allowing the first component (essentially the mean spectrum) to vary. Thus, a simplified algorithm was devised.

Each bin in a data block's spectrum was divided by the corresponding bin in the noise spectrum, yielding a vector of possible multipliers. "Multipliers" indicates that if the noise spectrum is multiplied by one of these values, and subtracted from the data block's spectrum, the corresponding bin in the data spectrum will be exactly cancelled. These values were sorted, and the value corresponding to the 6th percentile order statistic (8th of 128) was used. This multiplier always underestimates the proper scaling for the noise spectrum, but it also is very unlikely to be inflated by signal energy: it is a **consistent** underestimate. This order statistic must be magnified by a constant value for best performance. The magnitude of that adjustment was determined by analyzing noise compensation performance with a variety of parameter settings.

To measure noise compensation performance, we needed to measure the relative amounts of noise and signal energy that were removed. All of our sound cuts represented single channel recordings containing noise and signal. Thus, synthetic signals resembling marine mammal sounds were generated, as were randomly generated noise sequences resembling the backgrounds in our cuts. The noise compensation algorithm was applied to pure signal and noise sequences respectively, and the residual energy after compensation was measured. Values were sought that preserved as much signal energy and removed as much noise energy as possible.

Before discussing the results, our index of performance merits explanation. Residual signal-to-noise ratio did not prove to be a useful measure, because this metric resulted in excessively high levels of noise subtraction. The "optimal" values reached with this metric resulted in spectrographs that retained only the very loudest portion of the signal. Critically important components of the signal (for classification) were subtracted out. A more useful metric proved to be the percent signal energy remaining minus the percent noise energy remaining. A simple interpretation provides heuristic justification for this criterion. One estimate of expected residual signal energy is the residual noise energy multiplied by the original signal-to-noise ratio. This would be accurate if both signal and noise energy were reduced in proportion by the noise compensation technique. Our

7

metric is proportional to the "signal excess", the actual residual signal energy minus the expected residual signal energy. Optimal parameter values derived with this metric agreed with subjective judgments of spectrograph quality by experienced observers.

The most problematic sounds were relatively broad band, because more of the possible multipliers could be inflated by signal energy. This suggested that the best performance would be realized with low order statistics. Figure 1 presents the results of a simulation that used a broad-band signal and noise generated by forcing a sixth order autoregressive model with normally distributed white noise. Higher levels of signal excess represent better performance. Each vertical line represents performance at varying multiplier values, holding the order statistic constant. The leftmost vertical line starts at the bottom with a multiplier of 8, and ends with a multiplier of 160. The diagonal segment to the next vertical line denotes the performance value with the largest multiplier value on the left, and the performance of the next order statistic with its smallest multiplier on the right. Successive vertical lines represent different ranges of multipliers, ending with a range of 1->20 for the 50% order statistic. The graph illustrates the falling levels of performance with increasing order statistic number, and our success in bracketing the best multiplier values for each order statistic. On the basis of these and other tests, we chose the 6th percentile order statistic and a multiplier value of 75.

Figures 2 and 3 illustrate the effect of noise compensation, and compare the fixed noise compensation technique used previously with the adaptive technique. Both of these signals have poor signal-to-noise ratios, much worse than our typical sound cut. Note the improved retention of signal energy: fewer dropouts in the *Lagenodelphis* whistles, clearer representation of the soft, introductory moan in the right whale signal. The marked speckling in the *Lagenodelphis* adaptive spectrograph also represents preserved signal energy: echolocation clicks.

This noise compensation algorithm, and the methods we used to develop and test it, represented significant improvements over our previous work, but we do not represent this as the optimal or state-of-the-art technique. It allows us to achieve impressive classification performance. In our judgment, further improvements in this area are desirable, but not essential. The software has been designed to facilitate replacement of this module if we become aware of a better alternative.

After noise compensation, seven measurements were extracted from each data spectrum and stored. The first was **amplitude**, computed as the sum of the residual spectrum energy. This exploited Parseval's relation (Oppenheim and Schafer 1989, p. 574) to measure loudness after noise compensation. The remaining measurements described spectral characteristics. The frequency that bisected the energy in the power spectrum was saved as the **median**. The frequency corresponding to the largest energy value in the spectrum was saved as the **mode**.

Three estimates of "bandwidth" were saved. The minimum number of spectral bins needed to accumulate half of the total spectral energy was computed (including a fraction

derived from linear interpolation); we designated this the **concentration**. The highest and lowest frequencies encountered in this integration were saved as the **upper** and **lower** frequencies; the difference between these provided a broader estimate of bandwidth, designated as **spread**. The ratio of total energy to the energy in the modal spectral bin was saved as the **modewidth,** the most compact bandwidth estimate of the three. We rescaled these three bandwidth estimators by dividing them by the sample interval represented by a single FFT block, so the resulting units were Hertz/s (otherwise, the same signal would have yielded different values when sampled at different rates or processed with different FFT sizes).

An analog of skewness, designated as **asymmetry,** was computed as (**upper-median)/(upper-lower). Asymmetry** varied between 0.0 (**median** equal to **upper**) and 1.0 (**median** equal to **lower**). **Spectral asymmetry** of 0.5 indicated a symmetrical density; so we later may shift these values by subtracting 0.5 from them to render the results more intuitive (this would not affect classifier performance).

The lists of short-term signal measurements were sorted to extract the upper quartile (75th percentile), **median** (50th percentile), and lower quartile (25th percentile) values. When the computed index for one of the quartiles had a fractional component, the nearest values were used to linearly interpolate the desired value. The **mode** was estimated by finding the most tightly grouped set of five consecutive values, and selecting the middle of these. The quartiles were used to compute **spread** (upper quartile-lower quartile) and **asymmetry** (upper quartile-median)/(upper quartile-lower quartile). These statistics were analogous to the standard deviation and skewness, but they performed better. **Amplitude** was treated differently from the other short-term measurements. Its magnitude was arbitrary, so we divided **mode** and **spread** by the **median** to render them dimensionless. A total of 27 statistics resulted from these calculations:
- **Amplitude: mode/median, spread/median, asymmetry**
- **Frequency Mode: mode, median, spread, asymmetry**
- **Frequency Median: mode, median, spread, asymmetry**
- **Spectral Spread: mode, median, spread, asymmetry**
- **Spectral Concentration: mode, median, spread, asymmetry**
- **Spectral Modewidth: mode, median, spread, asymmetry**
- **Spectral Asymmetry: mode, median, spread, asymmetry**

Nonparametric correlations were computed among the short-term measurements, to quantify relationships among time, amplitude and frequency. We employed the Spearman Rank-Order Correlation (Press, W. H. et al. (1989), Numerical Recipes in C, Cambridge Univ. Press, pp. 507-509), and utilized the deviation of the sum-squared difference of ranks from its expected value, scaled in standard deviations. A large negative value indicated strong positive correlation, a large positive value indicated strong negative correlation (a sign change might be introduced later to ease interpretation). The 15 statistics resulting from these calculations were:
- **Time-Amplitude Deviance**
- **Time-Frequency Mode Deviance**
- **Time-Frequency Median Deviance**

- **Time-Spectral Spread Deviance**
- **Time-Spectral Concentration Deviance**
- **Time-Spectral Modewidth Deviance**
- **Time-Spectral Asymmetry Deviance**
- **Amplitude-Frequency Mode Deviance**
- **Amplitude-Frequency Median Deviance**
- **Amplitude-Spectral Spread Deviance**
- **Amplitude-Spectral Concentration Deviance**
- **Amplitude-Spectral Modewidth Deviance**
- **Amplitude-Spectral Asymmetry Deviance**
- **Frequency Median-Spectral Spread Deviance**
- **Frequency Median-Spectral Asymmetry Deviance**

To measure "flat" frequency contours, which were often important in distinguishing among odontocete whistles, we timed the longest section in the signal exhibiting minimal change in **frequency mode (maxflat)**. We computed the fraction of neighboring signal blocks in which the latter had more energy than the former **(attack fraction)**, and in which the latter had a higher **frequency median** than the former **(upsweep fraction)**. We also computed the average of all changes in **frequency median (upsweep mean)**, and the average absolute value **(sweep mean)** of such changes.

Each short-term spectrum also contributed to two cumulative power spectra. One averaged all of the short-term spectra; this produced the marginal spectral density of the spectrographic representation of signal, the **total spectrum**. The second accumulated energy from the loudest element of each residual spectrum, the **modal spectrum**. Figure 4 exhibits the relationship of the **total spectrum** (frequency marginal energy density) and **amplitude** envelope (time marginal energy density) to a noise compensated signal. The dark regions represent the portions of these densities that concentrate 75% of the total signal energy. These cumulative spectra were summarized with the same spectral statistics as the short-term spectra. This produced 6 **total spectrum** and 6 **modal spectrum** statistics: **medians, modes, spreads, concentrations, modewidths, and asymmetries**. In total, 91 fields were produced by the feature extraction program for each sound.

Classification Performance

Two techniques were used to quantify the usefulness of these acoustic features for distinguishing among species. The first was a classical linear classifier (Morrison 1976, ch. 6), which would be optimal if the species differed in their group means, but shared a common multivariate normal dispersion (common covariance matrix). This was applied to a subset of the data consisting of isolated sound elements; it produced 73% correct classification (208 errors for 784 sounds). The distribution of mistakes is illustrated by the bubble graph in Figure 5. Most of the errors were located in a square at the lower left corner of the plot, which indicated confusion of one baleen whale sound for another. A weaker tendency was incorrect identification of some baleen whale sounds as seals.

Linear classification analysis of all sounds revealed poorer performance: only 50% correct (1037 errors for 2104 sounds). Figure 6 indicates the distribution of errors. Confusion

among baleen whale sounds was again an important feature, but the horizontal banding in the plot indicated that a few species are responsible for most of the confusion. This structure is partly an artifact of sample sizes: more heavily sampled species will of course produce more incorrect classifications (these bubbles are not scaled to account for species sample size). This did not appear to be a complete explanation, and this phenomenon will be studied at greater length in the future.

An alternative technique for classifying these sounds utilized tree-based models (Clark and Pregibon 1992, ch. 9). This technique recursively partitioned the data, using a single variable at each binary split. At each point in the tree (called a node), a measure of diversity called "deviance" can be computed. It is defined as:

$$\text{deviance} = -2 \sum_{k=1}^{N} y_{ik}\log(p_{ik}), \quad y_{ik} = 1 \text{ if the } k^{th} \text{ individual is of class } i, \text{ 0 otherwise;}$$

p_{ik} = the probability that the kth individual is of class i,
estimated as the fraction of individuals in the node of class i.

This is equivalent to minus twice the log-likelihood function. Each interior node (including initial node containing all sounds) is split such that the residual deviance of the resulting pair of nodes is maximally reduced. Thus, the process of splitting results in successively "purer" nodes, with the process terminating when a node is sufficiently pure or there are insufficient individuals in the node to support another split. This process provides both a simple technique for classifying unknown sounds (a series of true/false questions) and clues to the important variables for diagnosis. It also accommodates diversity within a class: if a species produces two or more distinct types of sounds, a tree-based analysis will not be compromised (unlike a linear or quadratic classifier).

Figure 7 exhibits the tree-based classifier for the isolated sounds. The vertical distance associated with each split graphically depicts the reduction in deviance achieved by that split. The initial split was based on the **median** short-term spectral **concentration**; the next two splits were based on the **median** frequency of the **total spectrum** and the spectral **concentration** of the frequency modulation spectrum. This analysis permits a species' sounds to be split into more than one "leaf," depending upon their relationships to the other sounds in the sample. Correct classification was 85%; this is nearly a 50% reduction in misclassification relative to the linear classifier. A tree-based classification analysis of all sounds yielded 66% correct classification (figure 8). Tyack, Fristrup and McIntosh (submitted) have shown that similar analyses of signature whistles in young bottlenose dolphins correctly identified the individual for 90% of the sounds tested.

The linear classifier had one advantage over the tree-based classifier: it provided a measure of similarity to help judge the correctness of the identification. The tree-based technique must be augmented to provide this capability, using some distance metric generated from the terminal groupings. A straightforward adaptation would be to calculate a sample covariance matrix for each terminal grouping, and use Mahalanobis distance to measure the similarity of an unknown to that group. This adaptation, and tests of alternative classification schemes (quadratic classifier, kNN voting, hybrid designs), will be pursued further.

Literature Cited

Caldwell, M. C., D. K. Caldwell and P. L. Tyack 1990. *The Bottlenose Dolphin: recent progress in research*. Academic Press, San Diego.

Clark, L. A., and D. Pregibon 1992. *Statistical Models in S*. J. M. Chambers and T. J. Hastie (eds.). Wadworth and Brooks/Cole, Pacific Grove, CA.

Morrison, D. F. 1976. *Multivariate Statistical Analysis*, 2nd ed.. McGraw-Hill Inc., New York.

Tyack, P. L., K. Fristrup and J. S. McIntosh submitted. Signature whistle development in three captive bottlenose dolphins *Tursiops truncatus*. Animal Behavior.

Watkins, William A., Kurt Fristrup, and Mary Ann Daher 1991. Marine animal SOUND database. Technical Report WHOI-91-21, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, 51 pp.

Watkins, William A., and Douglas Wartzok 1985. Sensory biophysics of marine mammals. Marine Mammal Science 1:219-260.

Figure 1. Noise compensation performance: higher levels of "signal excess" represent better performance. Each vertical line represents the range of performance achieved with a fixed multiplier order statistic and varying magnification. The leftmost vertical line starts at its minimum with a magnifier of 8, and terminates at the intersection with the diagonal line with a multiplier of 160. The initial and terminal values of subsequent settings of multiplier order statistic are indicated by the intersections with diagonal lines on the left and right. The range of magnification decreases with increasing order statistic, but the maximum performance is clearly bracketed in each case.

Noise Compensation Performance
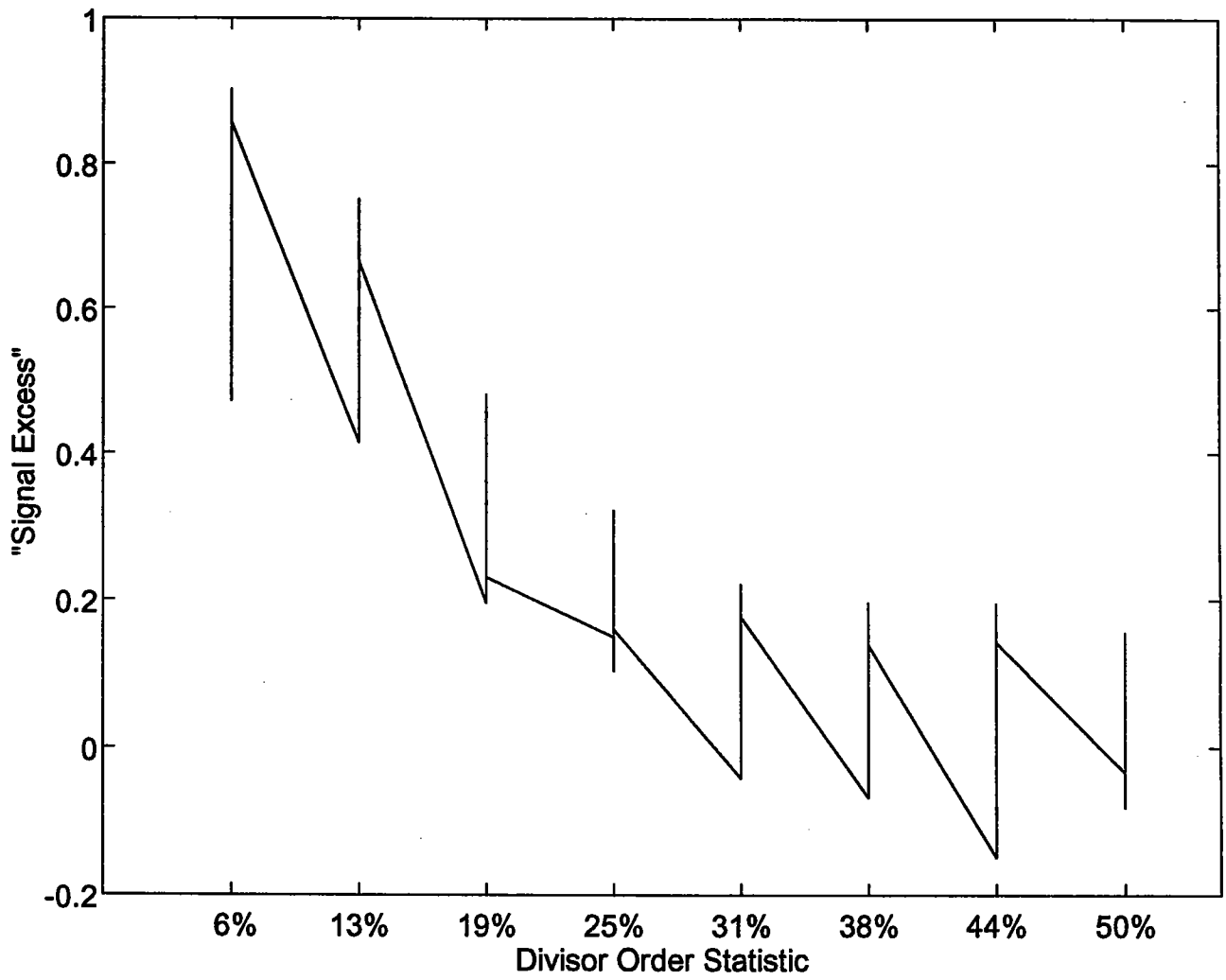
Figure 2. Noise compensation performance: this sequence of spectrographs illustrates a recording of *Lagenodelphis hosei,* the Fraser's dolphin. The first panel is the unmodified signal. The second illustrates the same signal processed using the older, fixed compensation algorithm. The third the illustrates the effect of processing with the newer, adaptive compensation algorithm. The pronounced speckling in all spectrographs represents echolocation clicks by the dolphins.

Raw Spectrograph: Lagenodelphis Whistles

Fixed Compensation
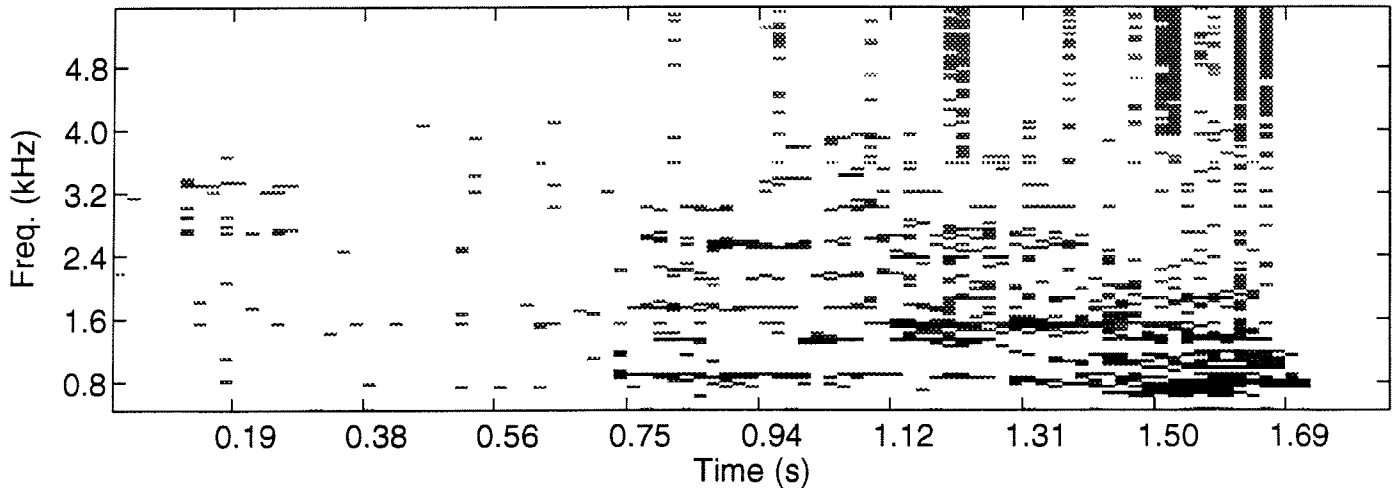
Adaptive Compensation

17
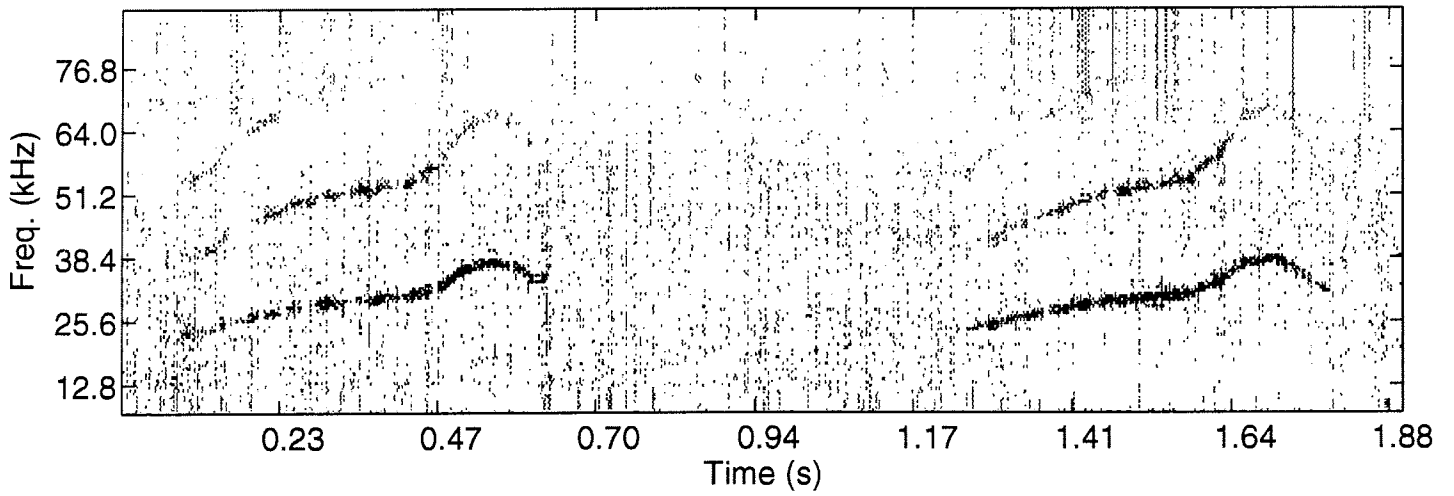
Figure 3. Noise compensation performance: this sequence of spectrographs illustrates a recording of *Eubalaena glacialis*, the northern right whale. The first panel is the unmodified signal. The second illustrates the same signal processed using the older, fixed compensation algorithm. The third the illustrates the effect of processing with the newer, adaptive compensation algorithm. Note the preservation of a faint, introductory moan in the third panel, near the left edge.

Raw Spectrograph: Eubalaena moans
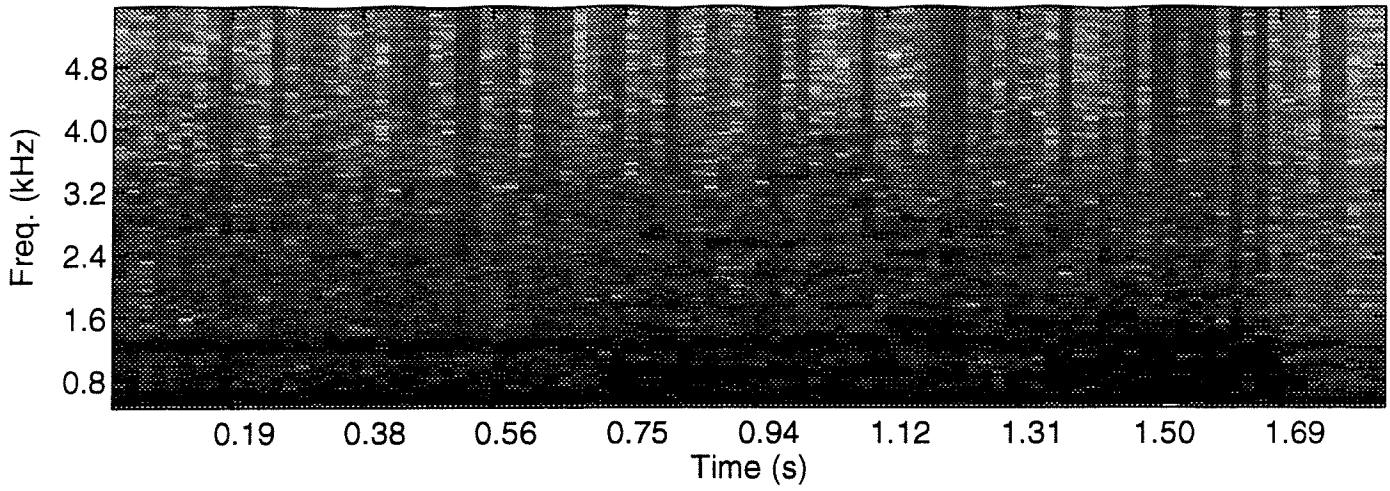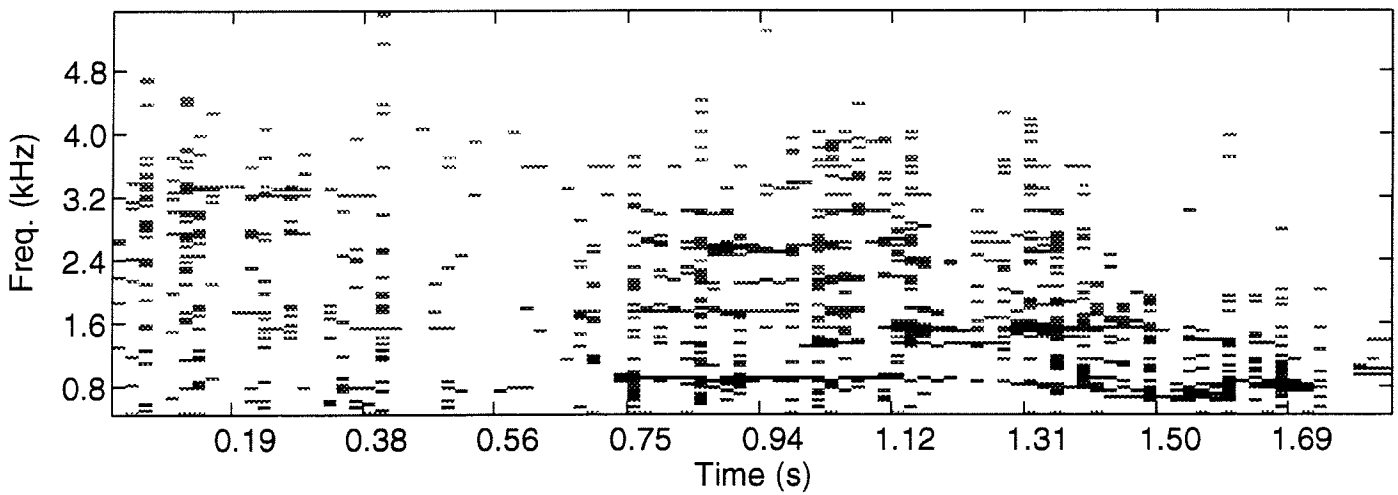
Fixed Compensation

Adaptive Compensation

Figure 4. The time and frequency marginal energy densities, in relation to the spectrograph that produced them. The dark areas of the energy densities indicate the portions included in the calculation of **concentration, upper**, and **lower** values. The original spectrograph was of very poor quality, not useable for classifier training. Note the leakage of low frequency noise energy, and the appearance of this energy in the shaded portions of the marginal distributions.

time marginal energy density

frequency marginal
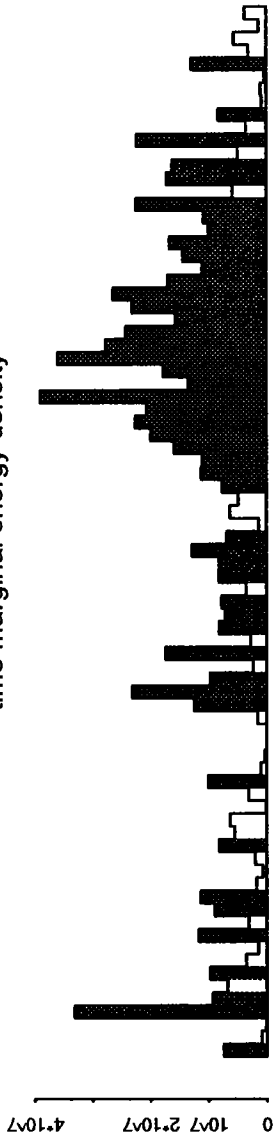
Peponocephala burst pulsed sound
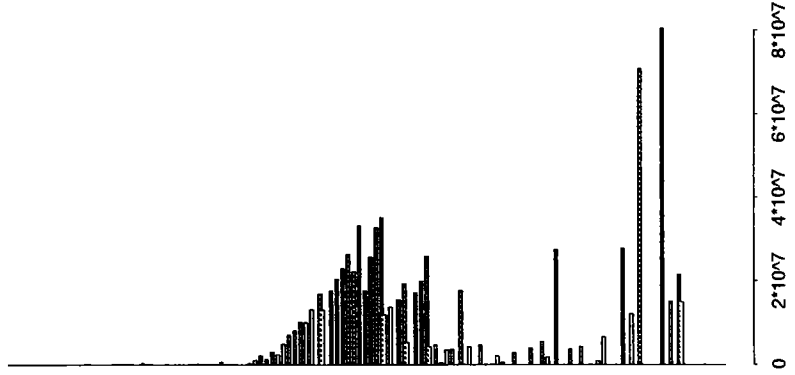
frequency in Hz

time in seconds

21
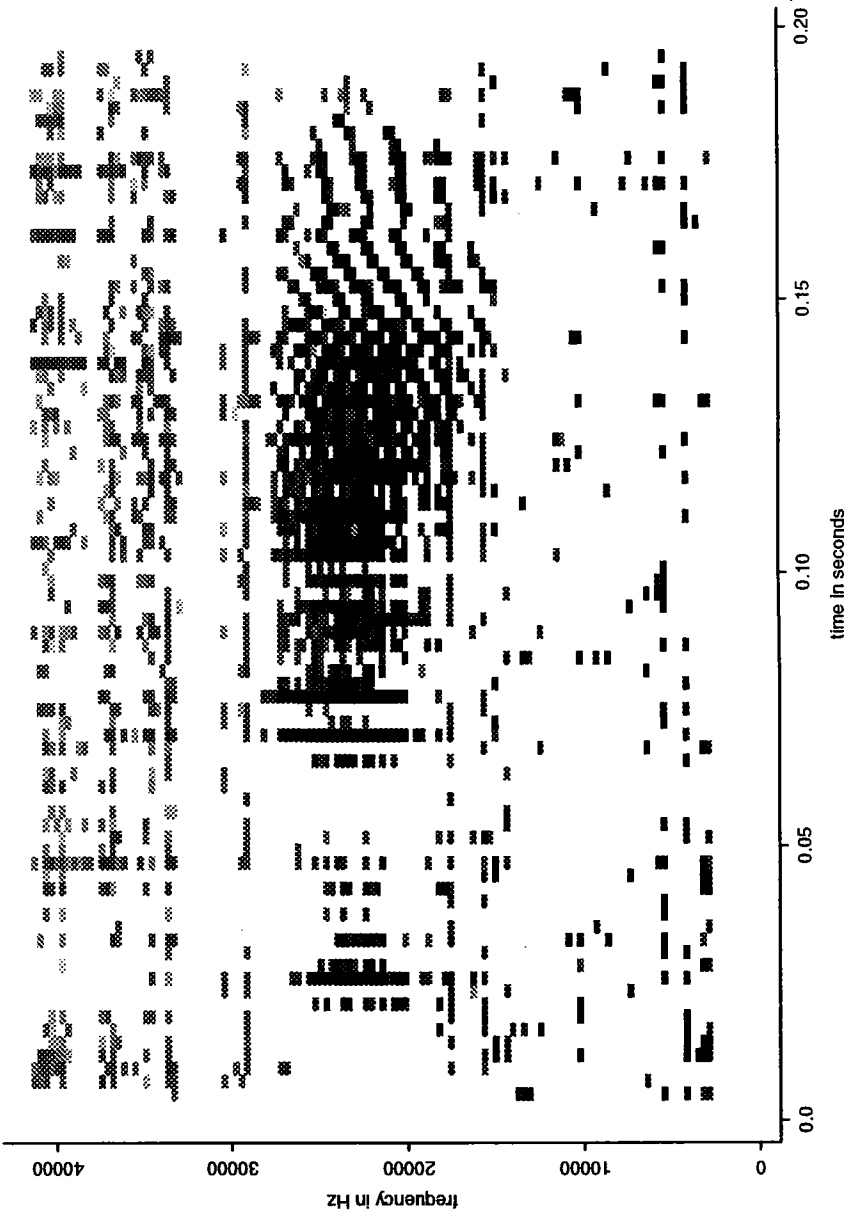
Figure 5. Linear classifier performance, isolated sounds: the x and y axes represent the species tested, the area of the circle represents the number of mistakes. The numeric ordering closely follows systematic ordering (historical relatedness). Numbers 1-8 are baleen whale species; numbers 9-22 are toothed whale species, 23-31 are seals, and number 32 is a manatee. Specifically,

1. *Balaena mysticetus*
2. *Eubalaena glacialis*
3. *Eubalaena australis*
4. *Eschrichtius robustus*
5. *Balaenoptera acutorostrata*
6. *Balaenoptera borealis*
7. *Balaenoptera physalus*
8. *Megaptera novaeangliae*
9. *Physeter catodon*
10. *Delphinapterus leucas*
11. *Monodon monoceros*
12. *Peponocephala electra*
13. *Steno bredanensis*
14. *Delphinus bairdii*
15. *Delphinus delphis*
16. *Grampus griseus*
17. *Lagenorhynchus acutus*
18. *Globicephala macrorhynchus*
19. *Globicephala melaena*
20. *Orcinus orca*
21. *Pseudorca crassidens*
22. *Phocoena phocoena*
23. *Arctocephalus forsteri*
24. *Eumetopias jubatus*
25. *Odobenus rosmarus*
26. *Phoca fasciata*
27. *Phoca largha*
28. *Ommatophoca rossi*
29. *Erignathus barbatus*
30. *Halichoerus grypus*
31. *Leptonychotes weddellii*
32. *Trichechus manatus*

Linear classifier errors (208/784): isolated sounds
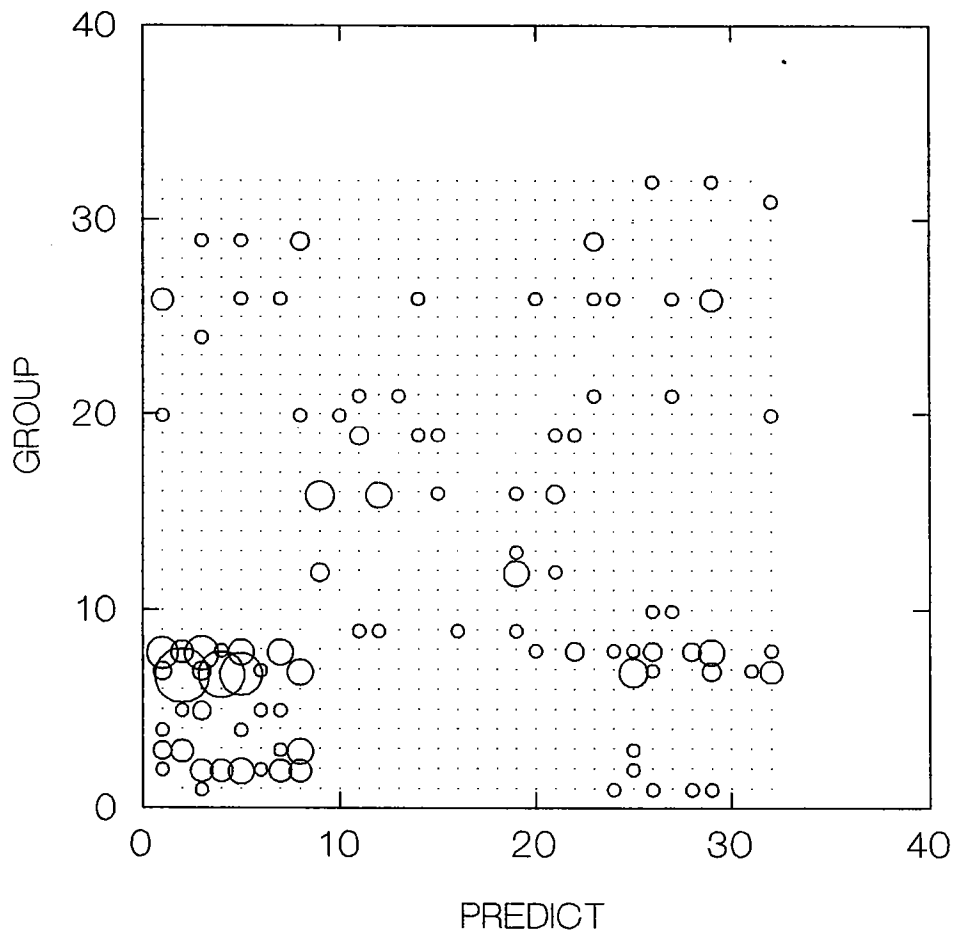area of circles is proportional to number of errors

Figure 6. Linear classifier performance, all sounds: the x and y axes represent the species tested, the area of the circle represents the number of mistakes. The numeric ordering closely follows systematic ordering (historical relatedness). Numbers 1-9 are baleen whale species; numbers 9-39 are toothed whale species, 40-53 are seals, and number 54 is a manatee. Specifically,

1. *Balaena mysticetus*
2. *Caperea marginata*
3. *Eubalaena glacialis*
4. *Eubalaena australis*
5. *Eschrichtius robustus*
6. *Balaenoptera acutorostrata*
7. *Balaenoptera borealis*
8. *Balaenoptera physalus*
9. *Megaptera novaeangliae*
10. *Physeter catodon*
11. *Delphinapterus leucas*
12. *Monodon monoceros*
13. *Peponocephala electra*
14. *Sotalia*
15. *Sousa*
16. *Stenella attenuata*
17. *Stenella clymene*
18. *Stenella coeruleoalba*
19. *Stenella longirostris*
20. *Steno bredanensis*
21. *Tursiops catalania*
22. *Tursiops truncatus*
23. *Cephalorhynchus commersonii*
24. *Cephalorhynchus heavisidii*
25. *Delphinus bairdii*
26. *Delphinus delphis*
27. *Grampus griseus*
28. *Lagenodelphis hosei*
29. *Lagenorhynchus acutus*
30. *Lagenorhynchus albirostris*
31. *Globicephala sp.*
32. *Globicephala macrorhynchus*
33. *Globicephala melaena*
34. *Globicephala scammoni*
35. *Orcinus orca*
36. *Pseudorca crassidens*
37. *Phocoena phocoena*
38. *Neophocaena phocaenoides*
39. *Inia geoffrensis*
40. *Arctocephalus forsteri*
41. *Eumetopias jubatus*
42. *Odobenus rosmarus*
43. *Phoca fasciata*
44. *Phoca groenlandica*
45. *Phoca hispida*
46. *Phoca largha*
47. *Ommatophoca rossi*
48. *Cystophora cristata*
49. *Erignathus barbatus*
50. *Halichoerus grypus*
51. *Leptonychotes weddellii*
52. *Enhydra lutris*
53. *Trichechus manatus*

Linear classifier errors (1037/2104): all sounds
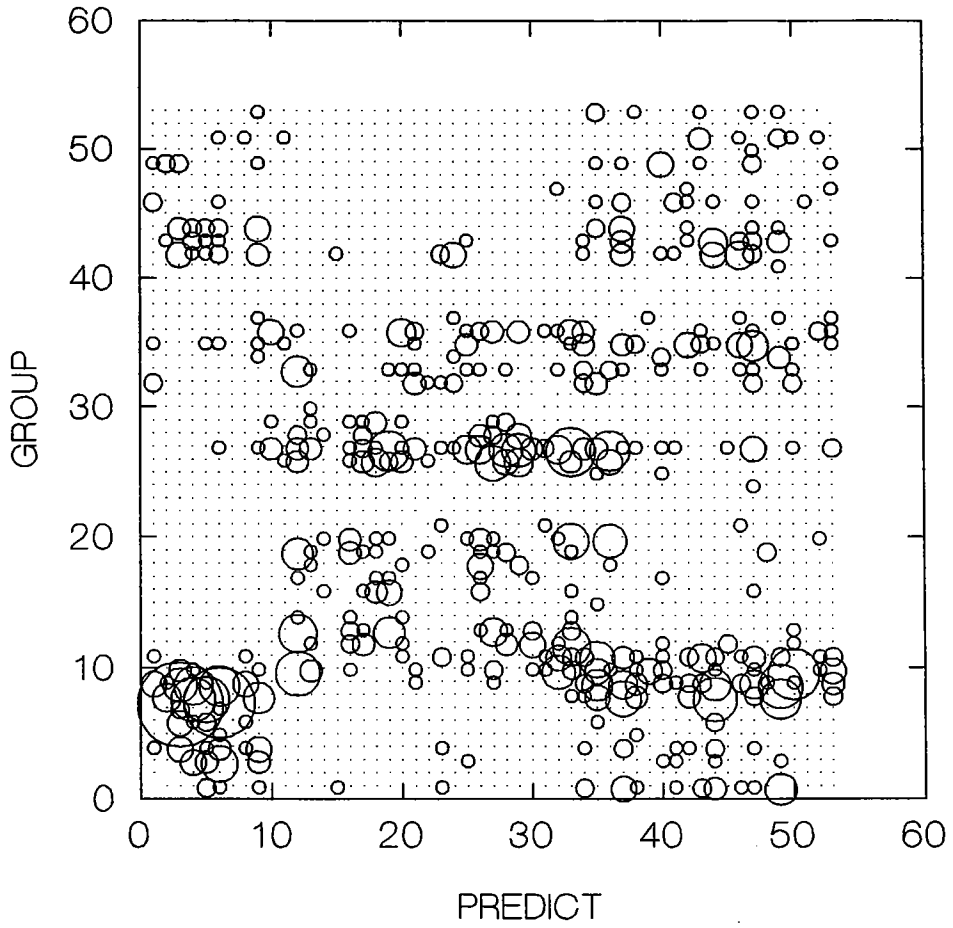area of circles is proportional to number of errors

Figure 7. Tree-based classification, isolated sounds: the minuscule labels at each interior node describe the criterion that was used to split the sounds at that node into two subnodes. The terminal nodes ("leaves") of the tree, located at the bottom of the figure, are labeled with a code describing the identities of the dominant fractions of the sounds in those nodes. The code reflects the systematic hierarchy. The translations are:

```
CONCmed:median st concentration
TSmed:total spectrum median frequency
MSmed:modal spectrum median frequency
ERGmxmd:maximum/median amplitude
FMEDSPRDr:median freq. X spread corr.
EGDmodw:amplitude modewidth
MODWmod:mode of st modewidth
FMSmod:FM spectrum mode
TSmod:total spectrum mode
MSmodw:modal spectrum mode
ATAKfrac:attack fraction
FMSmed:FM spectrum median
SPRDsprd:spread of st spread
AMSmod:AM spectrum mode
UPSWfrac:upsweep fraction
TSmodw:total spectrum modewidth
ASYMmod:modal st asymmetry
AMSupp:AM spectrum upper frequency
MODWmed:median st modewidth
TSasym:total spectrum asymmetry
FMSconc:FM spectrum concentration
SPRDasym:asymmetry of st spread
AASYMr:amplitude X st asymm. corr.
FMSsprd:FM spectrum spread
AFMODWr:amplitude X st modewidth corr.
TSupp:total spectrum upper frequency
MSupp:modal spectrum upper frequency

st == short term
```

AA1A: *Balaena mysticetus*
AA3A: *Eubalaena glacialis*
AA3B: *Eubalaena australis*
AB1A: *Eschrichtius robustus*
AC1A: *Balaenoptera acutorostrata*
AC1B: *Balaenoptera borealis*
AC1F: *Balaenoptera physalus*
AC2A: *Megaptera novaeangliae*
BA2A: *Physeter catodon*
BB1A:. *Delphinapterus leucas*
BB2A:. *Monodon monoceros*
BD10A: *Peponocephala electra*
BD17A:. *Steno bredanensis*
BD3A: *Delphinus bairdii*
BD3B: *Delphinus delphis*
BD4A:. *Grampus griseus*
BD6A:. *Lagenorhynchus acutus*
BE3B: *Globicephala macrorhynchus*
BE3C: *Globicephala melaena*
BE7A: *Orcinus orca*
BE9A: *Pseudorca crassidens*
BF2A: *Phocoena phocoena*
CA1F: *Arctocephalus forsteri*
CA3B: *Eumetopias jubatus*
CB1A: *Odobenus rosmarus*
CC12F: *Phoca fasciata*
CC12L: *Phoca largha*
CC14A: *Ommatophoca rossi*
CC2A: *Erignathus barbatus*
CC3A: *Halichoerus grypus*
CC5A: *Leptonychotes weddellii*
DB1B: *Trichechus manatus*
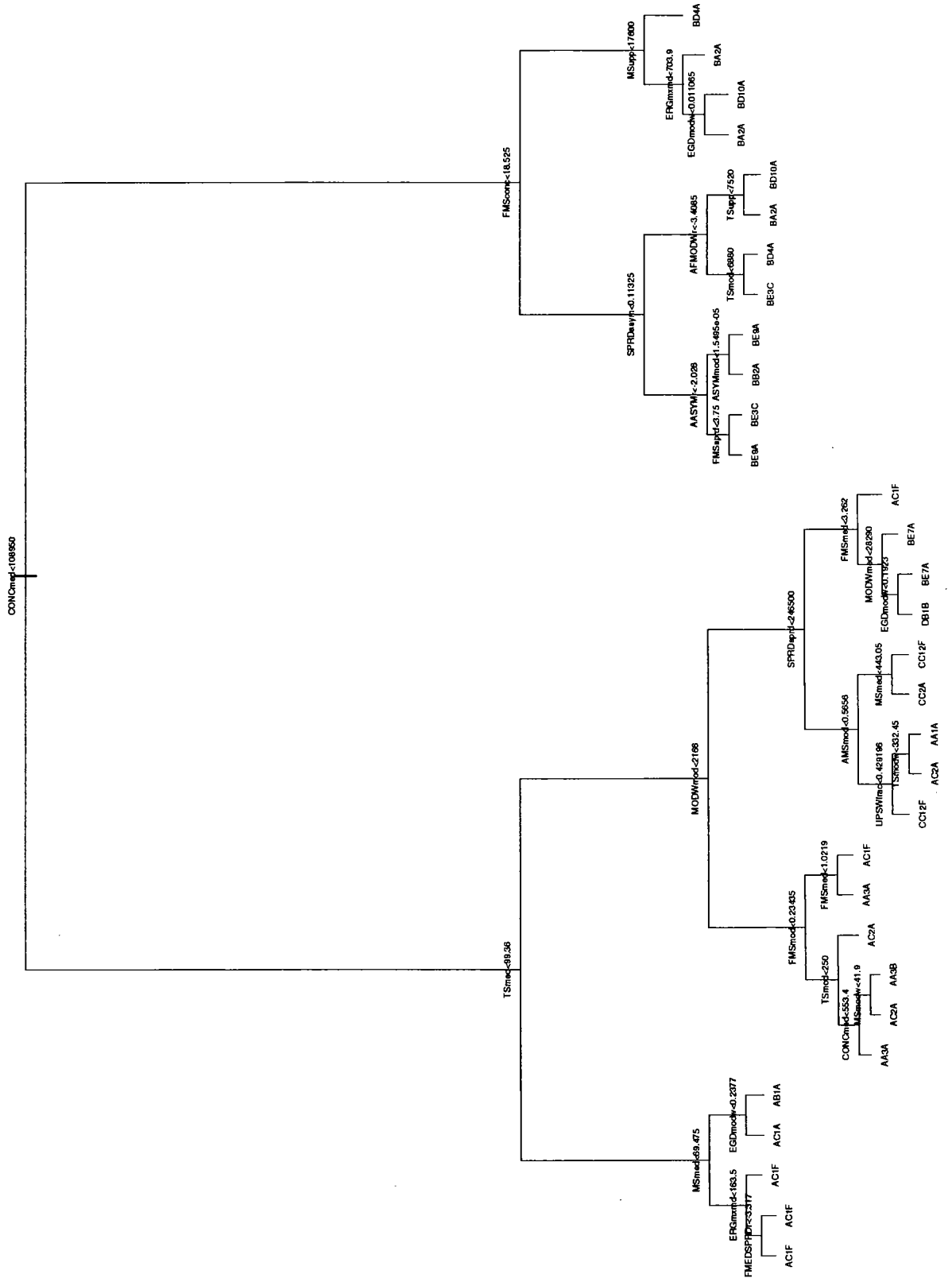
Isolated Sounds: 84% correct

Figure 8. Tree-based classification, all sounds: the minuscule labels at each interior node describe the criterion that was used to split the sounds at that node into two subnodes. The terminal nodes ("leaves") of the tree, located at the bottom of the figure, are labeled with a code describing the identity of the dominant fractions of the sounds in those nodes. The code reflects the systematic hierarchy. The translations are:

```
MODWmed:median st modewidth
MSmod:modal spectrum mode
ERGmxmd:maximum/median amplitude
SPRDsprd:spread of st spread
FMSmod:FM spectrum mode
MSmed:modal spectrum median
AMSmod:AM spectrum mode freq.
MODWsprd:spread of st modewidth
MSmodw:modal spectrum modewidth
SWPabs:mean absol. delta freq.
CONCasym:asymm. of st concent.
FMEDmed:median st median freq.
Maxflat:see text description
SPRDmed:median st spread
ATAKfrac:attack fraction
FMSsprd:FM spectrum spread
SPRDmod:mode of st spread
TSsprd:total spectrum spread
EGDconc:amplitude concentration
FMSconc:FM spectrum concent.
FMEDASYMr:median freq. X asymmetry corr.
TSupp:total spectrum upp. freq.
FMSupp:FM spectrum upper freq.
EGDsprd:amplitude spread
AFMODWr:amplitude X modewidth correlation
FMEDSPRDr:med.freq. X spread correlation
TSmed:total spect. median freq.
MODWmod:mode of st modewidth
ASYMasym:asymmetry of st asymm.
FMODmed:median of st mode freq.
CONCmod:mode of st concent.
FMEDmod:mode of st median freq.
AMSconc:concent. of st asymm.
ASYMmed:median st asymmetry
UPSWfrac:upsweep fraction
ERGcv:amplitude coeff. of var.
st == short term
```

AA1A: *Balaena mysticetus*
AA2A: *Caperea marginata*
AA3A: *Eubalaena glacialis*
AA3B: *Eubalaena australis*
AB1A: *Eschrichtius robustus*
AC1A: *Balaenoptera acutorostrata*
AC1B: *Balaenoptera borealis*
AC1F: *Balaenoptera physalus*
AC2A: *Megaptera novaeangliae*
BA2A: *Physeter catodon*
BB1A: *Delphinapterus leucas*
BB2A: *Monodon monoceros*
BD10A: *Peponocephala electra*
BD12: *Sotalia*
BD13: *Sousa*
BD15A: *Stenella attenuata*
BD15B: *Stenella clymene*
BD15C: *Stenella coeruleoalba*
BD15L: *Stenella longirostris*
BD17A: *Steno bredanensis*
BD19B: *Tursiops catalania*
BD19D: *Tursiops truncatus*
BD1A: *Cephalorhynchus commersonii*
BD1C: *Cephalorhynchus heavisidii*
BD3A: *Delphinus bairdii*
BD3B: *Delphinus delphis*
BD4A: *Grampus griseus*
BD5A: *Lagenodelphis hosei*
BD6A: *Lagenorhynchus acutus*
BD6B: *Lagenorhynchus albirostris*
BE3: *Globicephala sp.*
BE3B: *Globicephala macrorhynchus*
BD3C: *Globicephala melaena*
BE3D: *Globicephala scammoni*
BE7A: *Orcinus orca*
BE9A: *Pseudorca crassidens*
BF2A: *Phocoena phocoena*
BF6A *Neophocaena phocaenoides*
BG2A: *Inia geoffrensis*
CA1F: *Arctocephalus forsteri*
CA3B: *Eumetopias jubatus*
CB1A: *Odobenus rosmarus*
CC12F: *Phoca fasciata*
CC12G: *Phoca groenlandica*
CC12H: *Phoca hispida*
CC12L: *Phoca largha*
CC14A: *Ommatophoca rossi*
CC1A: *Cystophora cristata*
CC2A: *Erignathus barbatus*
CC3A: *Halichoerus grypus*
CC5A: *Leptonychotes weddellii*
CD1A: *Enhydra lutris*
DB1B: *Trichechus manatus*

28

All Sounds: 62% correct

University of California, San Diego
SIO Library 0175C (TRC)
9500 Gilman Drive
La Jolla, CA 92093-0175

Hancock Library of Biology & Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA 90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS, B2Y 4A2, CANADA

Commander
International Ice Patrol
1082 Shennecossett Road
Groton, CT 06340-6095

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL 33149

Library
Skidaway Institute of Oceanography
10 Ocean Science Circle
Savannah, GA 31411

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI 96822

Marine Resources Information Center
Building E38-320
MIT
Cambridge, MA 02139

Library
Lamont-Doherty Geological Observatory
Columbia University
Palisades, NY 10964

Library
Serials Department
Oregon State University
Corvallis, OR 97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI 02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX 77843

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA 98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

Maury Oceanographic Library
Naval Oceanographic Office
Building 1003 South
1002 Balch Blvd.
Stennis Space Center, MS 39522-5001

Library
Institute of Ocean Sciences
P.O. Box 6000
Sidney, B.C. V8L 4B2
CANADA

Library
Institute of Oceanographic Sciences
Deacon Laboratory
Wormley, Godalming
Surrey GU8 5UB
UNITED KINGDOM

The Librarian
CSIRO Marine Laboratories
G.P.O. Box 1538
Hobart, Tasmania
AUSTRALIA 7001

Library
Proudman Oceanographic Laboratory
Bidston Observatory
Birkenhead
Merseyside L43 7 RA
UNITED KINGDOM

IFREMER
Centre de Brest
Service Documentation - Publications
BP 70 29280 PLOUZANE
FRANCE

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. WHOI-94-13	2.	3. Recipient's Accession No.
4. Title and Subtitle Marine Animal Sound Classification			5. Report Date October 1993
			6.
7. Author(s) Kurt M. Fristrup and William A. Watkins			8. Performing Organization Rept. No. WHOI-94-13
9. Performing Organization Name and Address Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543			10. Project/Task/Work Unit No.
			11. Contract(C) or Grant(G) No. (C) N-00140-90-D-1979 (G)
12. Sponsoring Organization Name and Address Office of Naval Research through the Naval Undersea Warfare Center Division			13. Type of Report & Period Covered Technical Report
			14.

15. Supplementary Notes

This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept., WHOI-94-13.

16. Abstract (Limit: 200 words)

Software was developed to measure characteristics of marine animal sounds (AcouStat). These measurements proved effective for classifying sounds in several contexts: identifying species, quantifying the repertoire of a single species, and identifying individuals. The sound measures included statistics for aggregate bandwidth, intensity, duration, amplitude modulation, frequency modulation, center frequency, and interactions among these variables. Classification analysis based on these measures suggests they adequately characterize the variability of bioacoustic signals for many problems. Correct classification to species was as high as 85%, and correct classification of dolphin whistles to individual was 90%.

17. Document Analysis   a. Descriptors

marine
sound
classification

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement Approved for public release; distribution unlimited.	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 33
	20. Security Class (This Page)	22. Price

(See ANSI-Z39.18)   See Instructions on Reverse   OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce