# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188



1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
thru 12/31/96

4. TITLE AND SUBTITLE
AUTOMATIC CLASSIFICATION OF BIOLOGICAL SOUNDS IN THE ARCTIC

5. FUNDING NUMBERS
N00014-95-1-0363

6. AUTHOR(S)
Dr. Kurt M. Fristrup (Dr. Joel C. Goldman)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
WOODS HOLE OCEANOGRAPHIC INSTITUTION
WOODS HOLE, MA 02543

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Scientific Officer Code: 322
Thomas B. Curtin
ONR, Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Ambient underwater recordings in the Arctic are generated by a complex mixture of physical processes and biological events. Even for experts, it is difficult and time-consuming to detect and identify biological transients. During this project, improved methods for reviewing multichannel acoustic data and promising techniques for automatic classification of biological sounds were developed.

Two analytical methods demonstrated the promise of automatic recognition for these sounds. The first technique was a Classification Tree. This method produced a classifier consisting of a sequence of simple rules based on individual features.

A classification tree was computed that divided the collection of sounds into 23 categories; these 22 rules were sufficient to correctly identify 591 of 699 sounds to species, or about 85% correct classification.

In addition to the classification tree, a principal component analysis was also conducted on these data. Principal component scores were extracted from the rescaled data, to obtain new features that were mutually orthogonal, and identify which axes expressed the preponderance of the overall variation. The dominant principal component scores were then subjected to a discriminant function analysis, to obtain a set of two-dimensional projections that provide a useful perspective on the distinctiveness of the species' sounds.

14. SUBJECT TERMS



15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

## Final Report: Automatic Classification of Biological Sounds in the Arctic

**ONR**

Kurt M. Fristrup
Bioacoustics Research Program
Cornell Laboratory of Ornithology

Ambient underwater recordings in the Arctic are generated by a complex mixture of physical processes and biological events. There are relatively few experts who are familiar with all of the biological sounds that can be encountered in the Arctic. Even for these experts, it is difficult and time-consuming to detect and identify biological transients. During this project, improved methods for reviewing multichannel acoustic data and promising techniques for automatic classification of biological sounds were developed.

The Cornell Bioacoustics Research Program developed an Acoustic Location System that proved effective during censuses of bowhead whale populations (Clark et al 1996). Technicians were able to review multichannel, real-time spectrograms to look for transients, and interactively select those transients (with time and frequency bounds) for subsequent processing to determine the location of the sound source. This system was based on TEAC RD135 8 channel instrumentation recorders, which were directly interfaced to a Macintosh computers using a Cornell-designed interface. The system's performance was accelerated by a Cornell-designed coprocessor board.

When the principal investigator moved to Cornell from WHOI, it was decided to modify the Cornell ALS system such that it could process previously digitized data. This would allow the power of the ALS system to be applied to data from sources other than the TEAC recorders, including all of the Arctic data collected previously. The further advantage of this approach was that most of these data could be reviewed faster than real-time, with attendant savings in technician effort. These modifications proved more demanding than anticipated, but the new software was completed in early 1997. This system provides an unprecedented opportunity to interactively inspect multichannel acoustic data for acoustic transients, and locate the position of the source responsible for the sounds.

The data collected during the 1994 TAP experiment have not been processed with this system yet. A small file conversion utility is needed to extract the data from the digital tape archives and reformat the multitrack audio into standard AIFF files. Spot inspections of the TAP data have revealed electrical artifacts in the recordings that will complicate processing somewhat, but there are several days of 4 channel data that merit close analysis. No animal acoustic transients have been positively identified at this time. The parallel analysis of MIZEX data provided by Baggeroer et al. has not been initiated.

Preparation for automatic recognition of Arctic biological transients proceeded independently of the software development effort. 699 sound transients encompassing eight species of marine mammals were extracted from library recordings at WHOI (Watkins et al. 1991, 1992). This set included isolated sounds, series of sounds from a single individual, choruses of several individuals. Some cuts that included other species' sounds in the background. This inclusive set was selected because automatic detection routines presently cannot be relied upon to identify clear, uncontaminated sounds. The feature extraction program was based on earlier work (Fristrup and Watkins 1992, 1994), but rewritten to improve its performance with noisy recordings and add features that seemed relevant to Arctic biological sounds. Features extracted from these transients were processed to determine their ability to reveal distinctions among these Arctic species.

Two analytical methods demonstrated the promise of automatic recognition for these sounds. The first technique was a Classification Tree (Chambers and Hastie 1991), which is similar to C4.5 Machine Learning system (Quinlan 1993). This method produced a classifier consisting of a sequence of simple rules that were based on individual features. In addition to simplicity, this method had the advantage that it was insensitive to some idiosyncrasies of the feature set: disparities in the units (scales) of the features, correlations among features, and size of the feature set. CART also helped to identify which features were more important for discriminating among the species' sounds. Finally, it could produce a meaningful classifier when each species' sounds are polymodal in feature space: such structure would be expected if species possess a repertoire of distinct sound "types."

A classification tree was computed that divided the collection of sounds into 23 categories; these 22 rules were sufficient to correctly identify 591 of 699 sounds to species, or about 85% correct classification. The following table presents the "confusion matrix," which quantifies the kinds of mistakes that this classifier made. Large, off-diagonal cells indicated pairs of species that should be examined more closely to identify potential improvements in the feature extraction system.

Table 1: Classification Tree Confusion Matrix
Predicted Identity

| Known Identity | AA1A | BB1A | BB2A | CB1A | CC12G | CC12H | CC1A | CC2A |
|---|---|---|---|---|---|---|---|---|
| AA1A | 47 | 0 | 0 | 3 | 2 | 0 | 0 | 0 |
| BB1A | 2 | 128 | 3 | 11 | 1 | 1 | 0 | 3 |
| BB2A | 0 | 11 | 68 | 3 | 0 | 4 | 0 | 0 |
| CB1A | 4 | 6 | 0 | 239 | 7 | 0 | 0 | 4 |
| CC12G | 0 | 0 | 0 | 5 | 40 | 0 | 0 | 1 |
| CC12H | 0 | 11 | 4 | 2 | 0 | 24 | 0 | 0 |
| CC1A | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| CC2A | 3 | 0 | 4 | 11 | 0 | 0 | 0 | 45 |

Table 2: Species Codes Used in Tables and Figures

| | | | |
|---|---|---|---|
| Species 1 | AA1A | *Balaena mysticetus* | bowhead whale |
| Species 2 | BB1A | *Delphinapterus leucas* | beluga whale |
| Species 3 | BB2A | *Monodon monoceros* | narwhal |
| Species 4 | CB1A | *Odobenus rosmarus* | walrus |
| Species 5 | CC12G | *Phoca groenlandica* | harp seal |
| Species 6 | CC12H | *Phoca hispida* | ringed seal |
| Species 7 | CC1A | *Cystophora cristata* | hooded seal |
| Species 8 | CC2A | *Erignathus barbatus* | bearded seal |

To help interpret the confusion matrix, consider one species: the ringed seal. Looking across the 6th row, we see that 17 of 41 sounds known to be produced by this species were incorrectly attributed to other species. Looking down the 6th column, however, we see that 24 of the 29 sounds attributed to that species were correct. Thus, the classifier failed to recognize almost half of the ringed seal sounds, although it was fairly good at discriminating some kinds of ringed seal sounds from all others. The existing ringed seal classification categories were fairly good, but additional categories likely went unrecognized. The need for 23 categories to identify the sounds of 8 species reinforced this indication allowance for the complexity of some species' repertoires. Note that the classifier did not allocate a category to identify the hooded seal sounds, because there were only two of them in the sample. Figure 1 illustrates the classification tree.

The tree is displayed from the "root," at the top of the diagram, to the "tips" at the bottom. The tips represented the terminal categories, each of which was labeled by the species that comprised the majority of the sounds in that category. At each fork in the tree, an abbreviated name for a feature was juxtaposed with a value in an inequality. This indicated that the sounds were sorted into the left and right branches beneath the node on the basis of the named feature, with samples on the left branch having values less than the displayed values. The length of the vertical segments of each branch provided an indication of the fraction of the overall diversity in sound identities that was resolved by that rule. Thus, branches with long vertical segments helped to identify large numbers of sounds, while branches with short segments were less effective, in the context of the samples analyzed. This indication of the importance of each rule was dependent on the number of sound samples available for each species.

Although the overall pattern was somewhat complex, note that the right-hand fork of the first branch separated a large fraction of beluga whale sounds from the main group (with some narwhal and ringed seal sounds). This distinction was based on a feature that measures the range of frequency modulation in the sounds: beluga whale sounds tended to be more highly modulated. The appendix provides qualitative descriptions of the new features that appear in the classification tree.

The tree-based analysis did not express the multivariate structure in the complete feature set. To provide a balanced view, the acoustic feature data were rescaled such that the mean and variance of each feature were zero and one. Principal component scores were extracted from the rescaled data, to obtain new features that were mutually orthogonal, and identify which axes expressed the preponderance of the overall variation. The dominant principal component scores were then subjected to a discriminant function analysis, to obtain a set of two-dimensional projections that provide a useful perspective on the distinctiveness of the species' sounds. These three steps eliminated artifacts of scaling, reduced the effects of redundant measurements and high correlations among some features, and reduced the dimensionality of the discriminant problem for improved reliability.

Figure 2 presented a series of four plots that displayed the distribution of the sounds with respect to seven discriminant function axes. The numbering of the sounds was in accordance with the listing in Table 2 above. In the first plot (axes 1 and 2), the sounds of belugas (2), narwhals (3), ringed seals (6) , and to a lesser extent, bearded seals (8), were broken out from the mass of other sounds. In the second plot (axes 3 and 4), beluga and ringed seal sounds were distinguished, and walrus sounds (4) were somewhat distinctive. The third and fourth plots illustrated the dramatic distinction between the two hooded seal sounds and all the other sounds (a fact that was lost in the tree-based classifier), and bowhead whale sounds (1) began to emerge. These analyses illustrated the potential for constructing parametric classifiers. The advantage of such methods is the ability to augment identifications with a measure of likelihood or confidence.

References

Clark, C. W., Charif, R., Mitchell, S., and Colby, J. 1996. Distribution and behavior of the bowhead whale, *Balaena mysticetus,* based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska.

Chambers, J. M. and Hastie, T. J. 1991. Statistical Models in S. Wadsworth and Brooks, Pacific Grove, CA.

Fristrup, K. M. and Watkins, W. A. 1992a. Characterizing acoustic features of marine animal sounds. WHOI Tech. Rept. 92-04, 74pp.

Fristrup, K., M. A. Daher, T. Howald and W. A. Watkins 1992b, "Software tools for acoustic database management." W.H.O.I Technical Report 92-11, 86 pp.

Fristrup, K. M. and Watkins W. A. 1994. Marine animal sound classification. WHOI Tech. Rept. 94-13, 33pp.

Quinlan, J. R. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

Watkins, W. A., K. Fristrup and M. A. Daher 1991, "Marine animal sound database." W.H.O.I. Technical Report No. 91-21, 52 pp.

Watkins, W. A., K. Fristrup, M. A. Daher and T. Howald 1992, "SOUND database of marine animal vocalizations." W.H.O.I. Technical Report 92-31, 52 pp.

FMEDsprd(FMODsprd) == the range of center frequency values in the Short-Time Fourier Transforms computed from the file, where center frequency was represented as the MEDian (or MODal) value in the power spectrum. A signal with strong FM modulation or frequency-hopping had a large value for this measurement.

FSPRDmed(FSPRDmod) == the median (modal) STFT bandwidth (bandwidth computed as the range of frequencies contributing the loudest 50% of the signal). A consistently broad-band signal had a large value for this measurement.

AM7upp(AM5upp) == the upper frequency bound encountered in the amplitude modulation spectrum while accumulating the strongest 75%(50%) of the spectral energy. A signal with rapid amplitude modulation had a large value for this measurement.

ENVconc7(ENVconc5) == measures of the duration of the signal, which capture 75% or 50% of the signal energy.

AM7asym == the asymmetry of the amplitude modulation spectrum. A value of 0.5 indicated a symmetric spectrum; lower values indicated spectra whose medians were shifted toward lower frequencies in the range of the spectrum.

SWPabsmag == the absolute value of center frequency differences between adjacent STFT power spectra (expressed in Hz/s). Signals with abrupt, dramatic FM modulation had a large value for this measurement.

AM7mode(equals AM5mode) == the modal value in the amplitude modulation spectrum. A large value for this measurement indicated of rapid amplitude modulation.

FSPRDsprd(CONCsprd) == measurements of the spread in the short-term bandwidth measurements made from STFT power spectra. Large values for these measurements indicated that the signal had both narrowband and wideband components.

UPSmean == the average increase in center frequency values in adjacent STFT power spectra. A positive value indicated the signal's tendency to increase in frequency; a negative value indicated a tendency to decrease in frequency.

FMEDmed == the median of the median frequency values computed from STFT power spectra. Signals that maintained a high pitch would produce large values for this measurement.

MaxFlat == the longest interval in the signal for which the center frequency remained relatively constant.

ERGmxmd == the ratio of the loudest element in the signal to the median amplitude of the signal. Signals with strong, isolated impulses will generate large values for this measurement.

Figure 1:

This Figure shows the classification tree output. Beginning at the top of the tree ("trunk"), each fork in the tree lists the feature used as a discriminator, along with the value used for that decision point. A sound sample would be sorted into the left or right branch depending on the sample's value. The length of the vertical segments represents the proportion of calls that were sorted along that branch path. The terminal portion of the branching structures ("tips") shows the abbreviated species name.

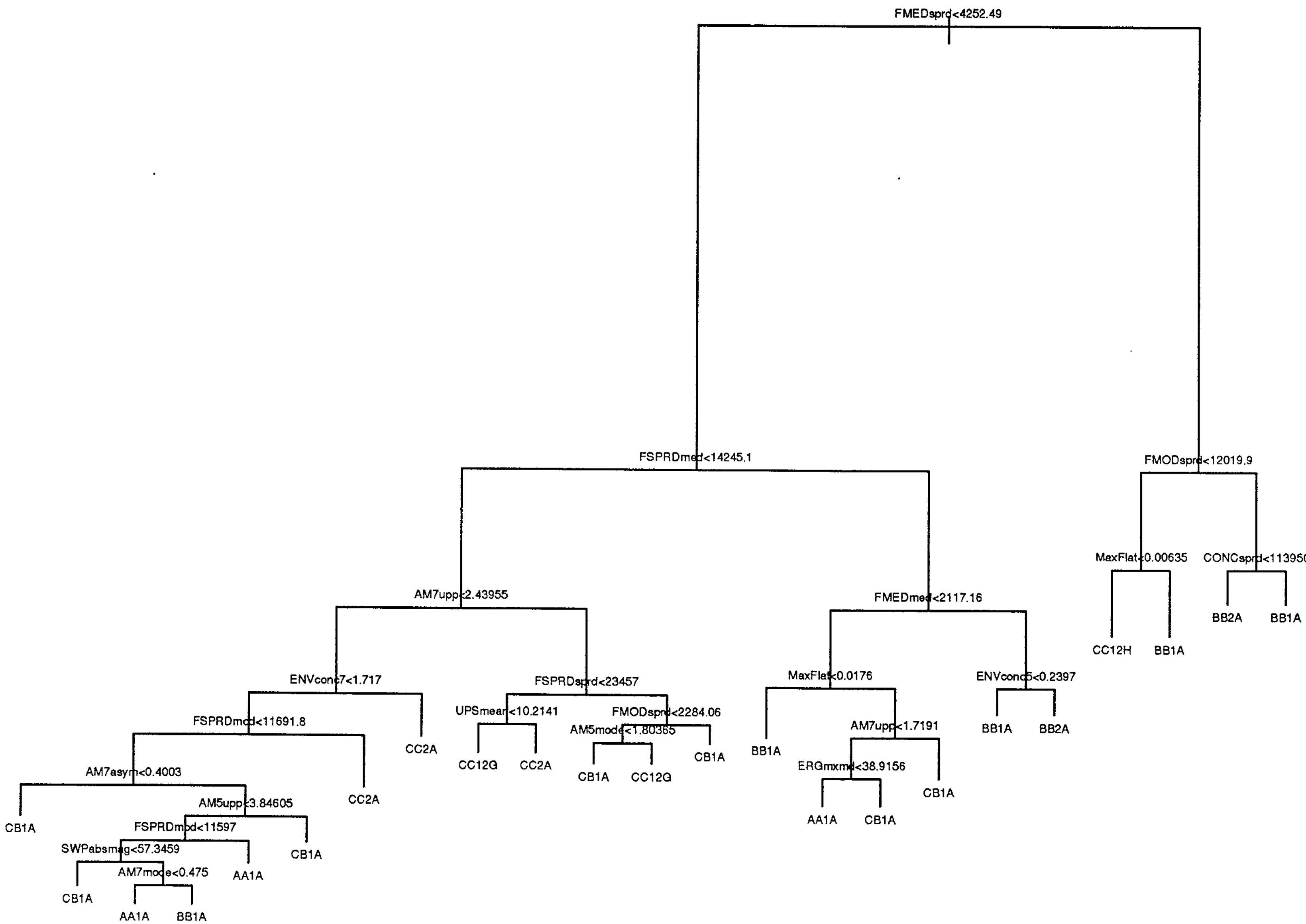
FMEDsprd<4252.49
FSPRDmed<14245.1
FMODsprd<12019.9
MaxFlat<0.00635
CONCsprd<113950
CC12H
BB1A
BB2A
BB1A
AM7upp<2.43955
FMEDmed<2117.16
ENVconc7<1.717
FSPRDsprd<23457
MaxFlat<0.0176
ENVconc5<0.2397
FSPRDmed<11691.8
CC2A
UPSmean<10.2141
FMODsprd<2284.06
AM5mode<1.80365
BB1A
AM7upp<1.7191
BB1A
BB2A
AM7asym<0.4003
CC2A
CC12G
CC2A
CB1A
CC12G
CB1A
ERGmxmd<38.9156
CB1A
AM5upp<3.84605
CB1A
FSPRDmed<11597
AA1A
CB1A
SWPabsmag<57.3459
CB1A
AA1A
AM7mode<0.475
CB1A
AA1A
BB1A

Figure 1

Figure 2A-2D:

This Figure shows four plots of the distribution of calls plotted on eight different discriminant function axes (1-8). Species are represented by number. Bowhead whales (1), belugas (2), narwhals (3), walrus (4), harpd seal (5), ringed seals (6), hooded seals (7) and bearded seals (8) are plotted. Figure 2A shows the discrimination of belugas, narwhals, ringed seals, and bearded seals. Figure 2B beluga and ringed seals were distinguished, and walrus sounds were somewhat distinctive. Figures 2C and 2D show the dramatic distinction between hooded seal sounds and all others, a distinction not made with the tree-based classifier shown in Figure 1. Bowhead calls also begin to appear in Figures 2C and 2D.

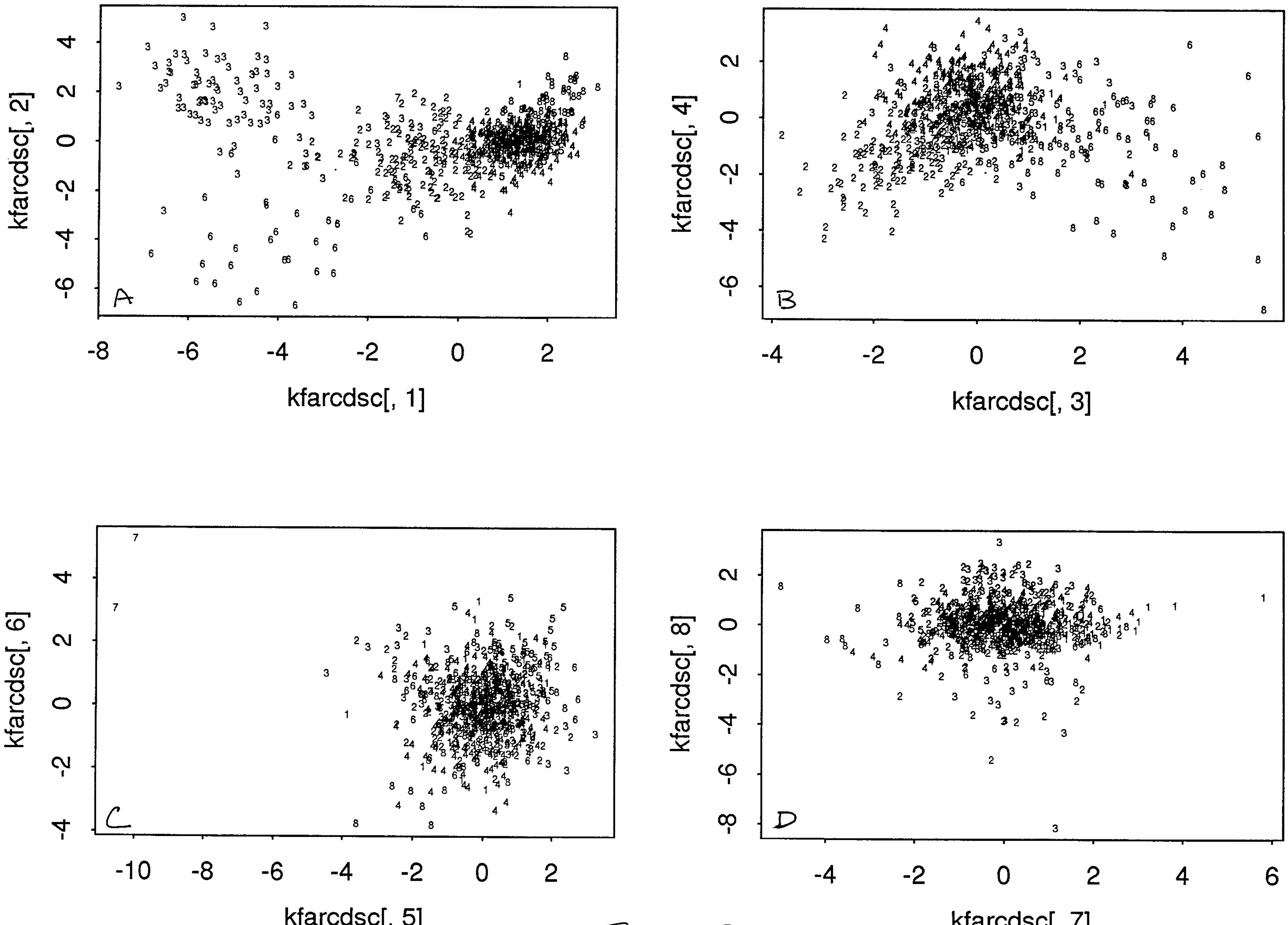
A
kfarcdsc[, 1]
kfarcdsc[, 2]
B
kfarcdsc[, 3]
kfarcdsc[, 4]
C
kfarcdsc[, 5]
kfarcdsc[, 6]
D
kfarcdsc[, 7]
kfarcdsc[, 8]

Figure 2