Analysis of 7395 dog barks
(A BME Technical Report)

Nick Campbell, Géza Németh, Ádám Miklósi

Sept 11th 2010

Abstract

This paper presents the results of a preliminary analysis of the sounds of dog barks in various everyday situations. Seven types of bark were recorded from twenty-seven dogs of breed Mudi. Barks were recorded in the following situations: alone,playing with a ball, fighting with a human, food time, at play with a human, warning of a stranger, and taking a walk. Not all dogs were recorded in every situation, but a few were, and the analysis of these barks forms the core of this technical report. Statistical analysis of the acoustic parameters of each bark was performed, and a principal components analysis determined which were the main contributing factors in each case. A statistical model was then trained (we used both classification trees and state-vector machines) to predict the most likely category for each bark by training either with barks from other dogs opr from barks of the same dog with a subset held back for testing. N-fold cross-validation was tested at n=5, n=10, and n=15. Good categorisation was possible for each dog when trained on similar material from the same dog, but models trained on one or more different dogs were very poor at categorising the barks of another.

1 Introduction

Recordings were made available by Professor Ádám Miklósi and the analysis was carried out by Nick Campbell while working with Géza Németh as a guest of the BME during the summer of 2010. Table 1 shows counts of the barks recorded in each of the seven categories. Table 2 shows the total number of barks recorded for each of the twenty-seven dogs for each category.of bark. It is clear from Table 2 that not all dogs were recorded in all conditions, and that some dogs only produced a small number of barks. Those dogs that were not prolific were clustered into a single category (shown by an 'x' in Table 3) which was used for non-specific training and testing. The most prolific dogs were d12, d23, and d24. Only d24 barked in all categories. These three dogs formed the core of our training and testing material, and numbers of data-points are given in Table 3.

Table 1: Showing the number of barks recordes for 7 categories of bark: a=alone, b=ball, f=fight, fo=food, p=play, s=stranger, and w=walk

a	b	f	fo	p	s	w
853	1105	1236	877	880	1593	851

Table 2: Showing the total number of barks (in all categories) recorded for each dog. Dogs are numbered from d01 to d27 to preserve their privacy, though identities are noted in a separate file kept in a locked drawer.

d01	d02	d03	d04	d05	d06	d07
22	40	15	10	303	20	40

d08	d09	d10	d11	d12	d13	d14
95	728	108	20	1045	8	480

d15	d16	d17	d18	d19	d20	d21
33	354	38	277	20	720	9

d22	d23	d24	d25	d26	d27	-
9	970	1701	91	111	128	-

Table 3: Showing the number of barks in each category for the dogs with most recordings. A separate category 'x' has been made to contain the data of dogs with fewer barks. See Table 1 for explanation of category labels.

	a	b	f	fo	p	s	w
d05	10	-	-	-	158	135	-
d09	-	50	-	163	-	487	28
d12	152	193	345	99	-	194	62
d14	168	158	34	101	-	9	10
d16	-	20	100	3	44	118	69
d18	48	63	-	-	147	10	9
d20	-	177	161	137	146	9	90
d23	140	201	139	187	0	150	153
d24	277	165	235	168	313	208	335
d27	-	23	64	-	-	26	15
x	58	55	158	19	72	247	80

2 Principal Components Analysis

The acoustic features for each bark were calculated using the Snack speech processing library [?], part of the Tcl/Tk programming language. The features we tested were measures of fundamental frequency, speech amplitude, and spectral tilt. For fundamental frequency and power, we calculated the mean, maximum, and minimum values measured across each bark, as well as the position of the maximum in relative percentage values within the waveform. We measured spectral tilt by the difference between the first harmonic and the amplitude of the third formant, after Hansen '90 [?], and included duration of the bark as well as the amount of voicing it contained. The following values were obtained for the set of barks as a whole:

```
    fmean             fmax             fmin             fpct
 Min.   :118.5   Min.   :121.5   Min.   :113.8   Min.   :0.0000
 1st Qu.:403.5   1st Qu.:510.3   1st Qu.:278.9   1st Qu.:0.3300
 Median :529.3   Median :634.5   Median :394.6   Median :0.4600
 Mean   :503.6   Mean   :585.8   Mean   :392.0   Mean   :0.4822
 3rd Qu.:618.5   3rd Qu.:698.2   3rd Qu.:504.0   3rd Qu.:0.6100
 Max.   :774.1   Max.   :824.6   Max.   :767.5   Max.   :0.9800

     pmean            pmax             pmin             ppct
 Min.   :26.43   Min.   :43.86   Min.   : 0.04   Min.   :0.110
 1st Qu.:43.11   1st Qu.:67.09   1st Qu.:15.31   1st Qu.:0.290
 Median :50.01   Median :71.41   Median :24.24   Median :0.340
 Mean   :50.31   Mean   :71.81   Mean   :25.18   Mean   :0.351
 3rd Qu.:57.14   3rd Qu.:77.90   3rd Qu.:33.10   3rd Qu.:0.410
 Max.   :76.32   Max.   :83.82   Max.   :74.21   Max.   :0.790

      h1h2              h1a3              h1                a3
 Min.   :-23.710   Min.   :-17.25   Min.   :-61.69   Min.   :-70.82
 1st Qu.:  3.558   1st Qu.: 15.14   1st Qu.:-45.35   1st Qu.:-63.35
 Median :  8.080   Median : 20.84   Median :-39.87   Median :-61.32
 Mean   :  8.278   Mean   : 21.62   Mean   :-39.00   Mean   :-60.62
 3rd Qu.: 12.550   3rd Qu.: 27.16   3rd Qu.:-33.80   3rd Qu.:-58.60
 Max.   : 35.520   Max.   : 70.95   Max.   :  7.98   Max.   :-38.17

      fvcd              dn
 Min.   :0.0300   Min.   :0.1100
 1st Qu.:0.3400   1st Qu.:0.2600
 Median :0.5600   Median :0.3000
 Mean   :0.5523   Mean   :0.3183
 3rd Qu.:0.7600   3rd Qu.:0.3700
 Max.   :1.0000   Max.   :0.9200
```

A principal components analysis showed these parameters to be largely independent, and a combination of the first seven components cumulatively accounted for 83% of the variance in the data:

Importance of components:

	Cmp.1	Cmp.2	Cmp.3	Cmp.4	Cmp.5	Cmp.6
Standard deviation	1.7597	1.5894	1.3210	1.12837	1.0773	1.00134
Proportion of Variance	0.2211	0.1804	0.1246	0.09094	0.0829	0.07162
Cumulative Proportion	0.2211	0.4016	0.5262	0.61724	0.7001	0.77177

	Cmp.7	Cmp.8	Cmp.9	Cmp.10	Cmp.11	Cmp.12
Standard deviation	0.94532	0.89067	0.71722	0.69154	0.62758	0.253054
Proportion of Variance	0.06383	0.05666	0.03674	0.03415	0.02813	0.004574
Cumulative Proportion	0.83560	0.89226	0.92900	0.96316	0.99130	0.995876

	Cmp.13	Cmp.14
Standard deviation	0.240279	3.220e-04
Proportion of Variance	0.004123	7.4104e-09
Cumulative Proportion	0.999999	1.000e+00

Loadings: (shown here only for the first 9 components):

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
fmean	-0.429		0.224	-0.406			-0.194	0.189	
fmax	-0.447	0.120	0.147	-0.375			-0.192	0.252	
fmin	0.149	-0.194	0.442	-0.183	0.318	-0.363		-0.157	-0.209
fpct	-0.178	0.247	-0.407		-0.319	-0.117	0.115	0.209	-0.622
fvcd	-0.248	0.314	-0.315	0.151	0.184	0.121		0.102	0.476
pmean	0.164	0.540	0.271						
pmax		0.319	0.351		-0.490	0.104		-0.111	-0.195
pmin	0.192	0.484	0.207						0.250
ppct		0.134	-0.282	-0.124		-0.769	0.252	-0.325	0.176
h1h2	-0.163	0.201	-0.189	-0.139	0.222	0.279	-0.346	-0.764	-0.215
h1a3	0.443		-0.234	-0.427			-0.104	0.152	
h1	0.452		-0.240	-0.290	0.119		-0.283	0.200	
a3		0.110		0.544	0.121	-0.366	-0.672	0.172	-0.133
dn		-0.265		-0.167	-0.651		-0.409	-0.166	0.364

As can be seen from the loadings, the first principal component is largely related to a combination of (mean & max) fundamental frequency and spectral slope, the second to power of the bark (mean & min), and the third to minimum fundamental frequency and position of the peak. These three axes alone account for little over half of the variance observed in the sound of the barks. Blanks in the table above indicate non-significant relations, held out for ease of viewing.

Figures 1 and 2 plot the acoustic data for two dogs (d12 and d24) to illustrate the difference in the distribution of barks in the two most relevant dimensions. Figures 3 and 4 show the equivalent data for dog D12 plotted according to the PCA, Figure 3 showing the first and second components (rotated features) with vectors representing the similarity of the various features. Figure 4 is an amalgam showing the same data plotted according to combinations of the first four dimensions.

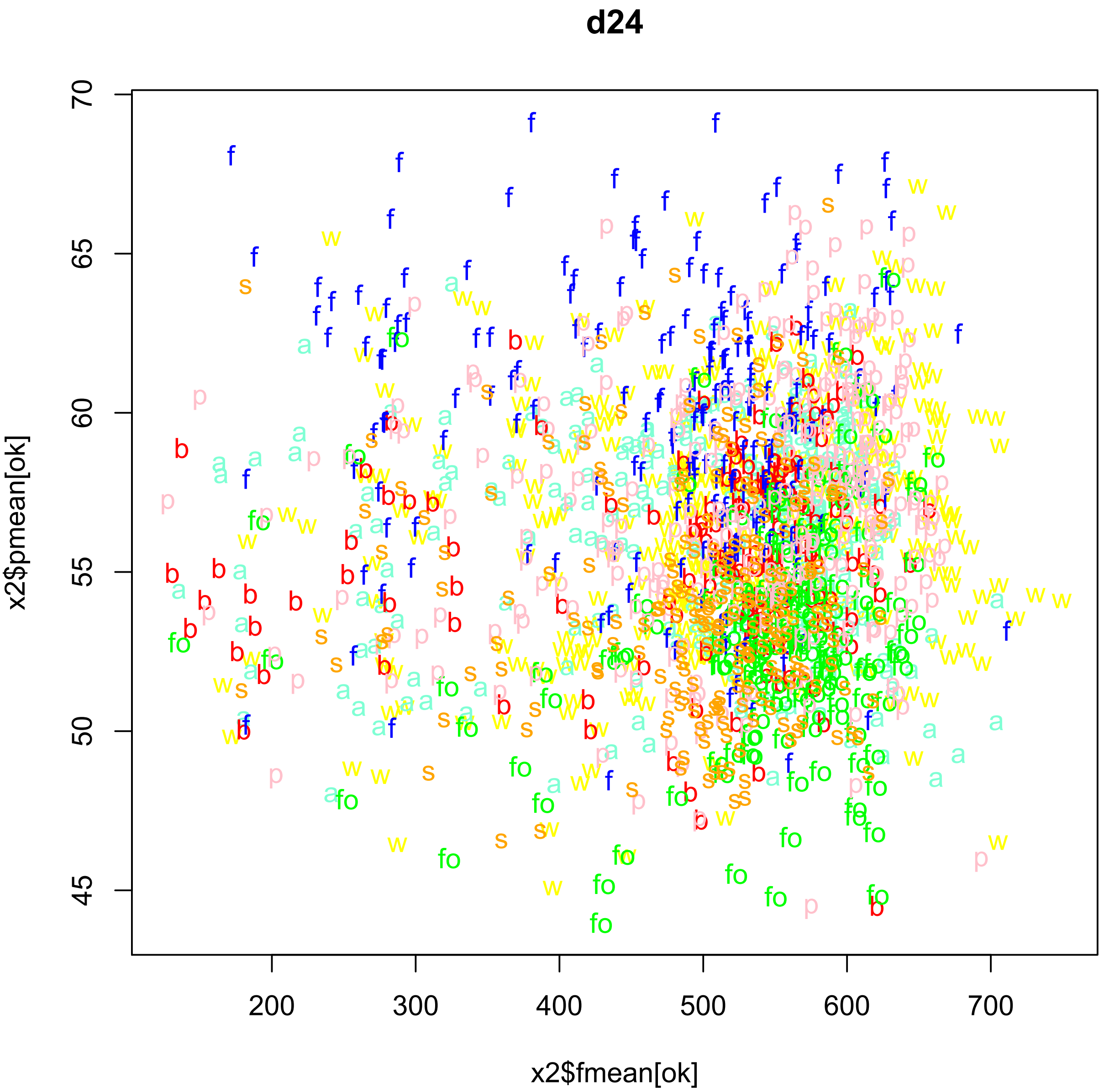

Figure 1: **Plotting the seven types of bark (see Table 1 for explanation) in terms of pitch and power of the acoustics. Note for example that 'f' (fighting) is characterised by relatively high mean power, and that 'b' (playing with a ball) is characterised by relatively low mean power. Most barks are in the region of high pitch (fundamental frequency) and there is considerable spread of bark categories across the whole of this two-dimensional feature space.**

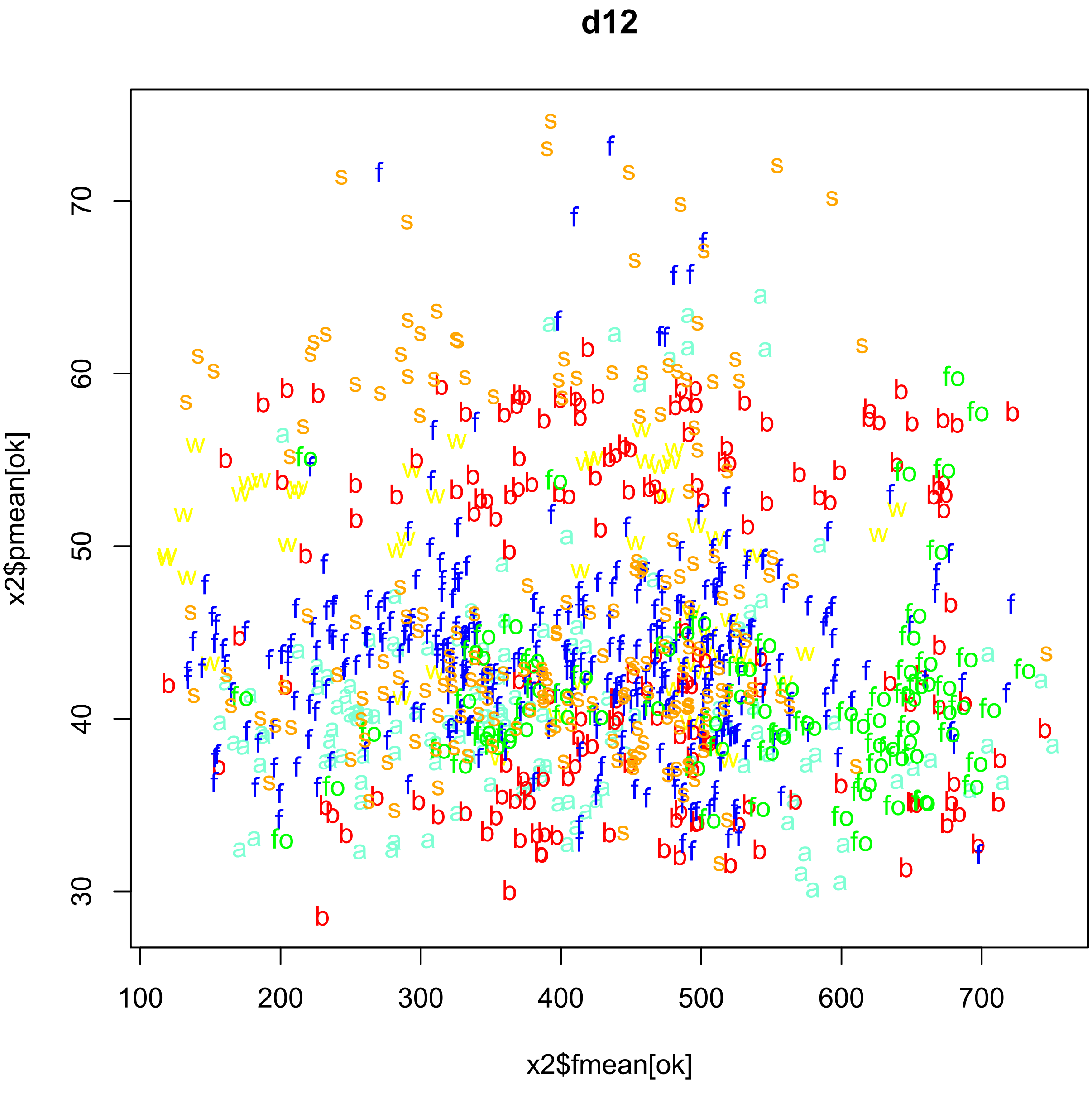

Figure 2: **Dog 12 shows a very different distribution of categories across the same feature space, making better use of the range of fundamental frequency (fmean) Note especially that the 'b' category appears to exhibit two clear variants or subsets; one with high mean power, overlapping with 's' (stranger), and the other with relatively low mean power, overlapping with 'fo' (food). One is tempted to speculate here that the dog is playing two types of game with the same ball perhaps?**

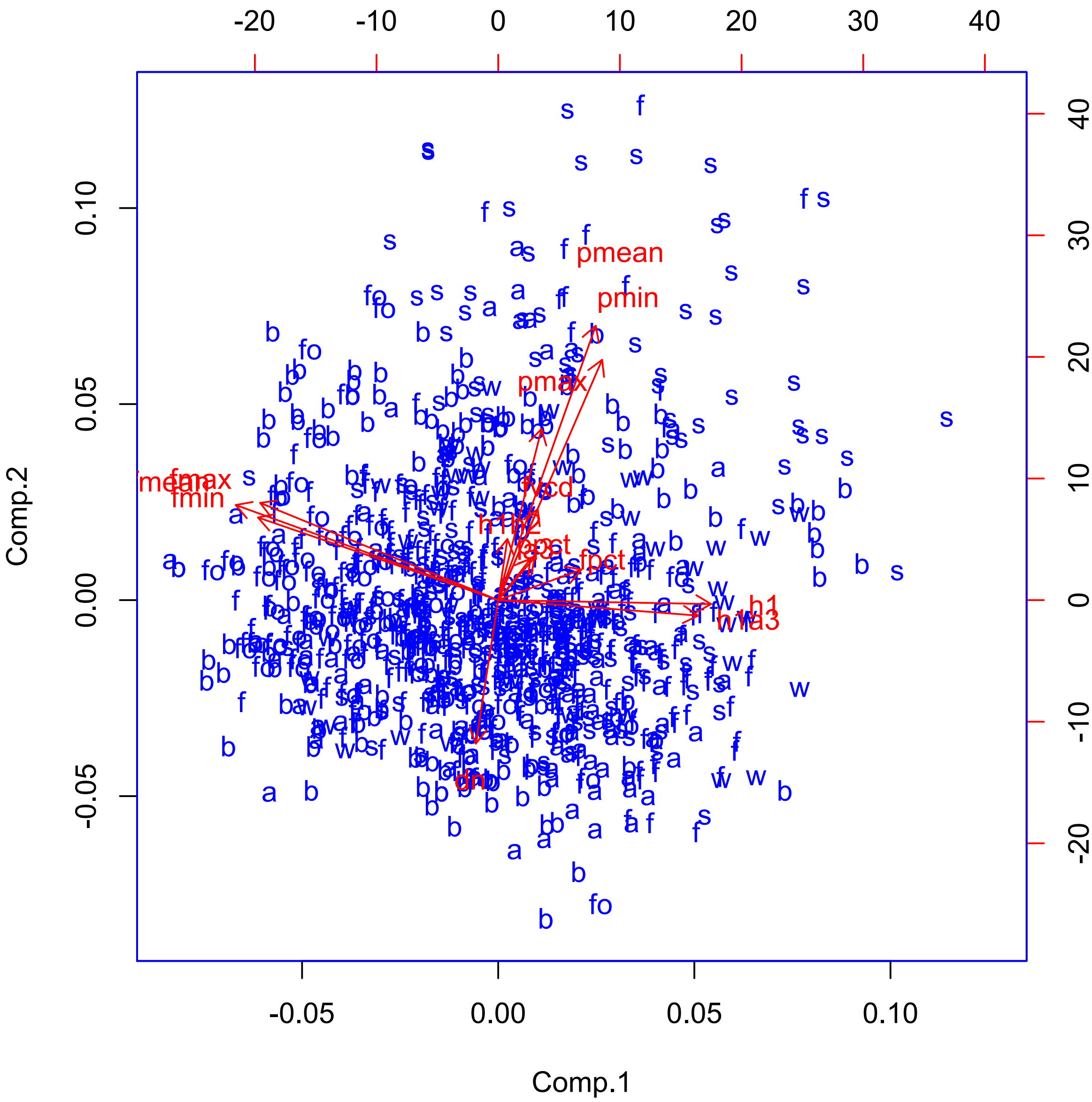

Figure 3: **Biplots of the first two principal components of dog D12's acoustic features and associated categories.**

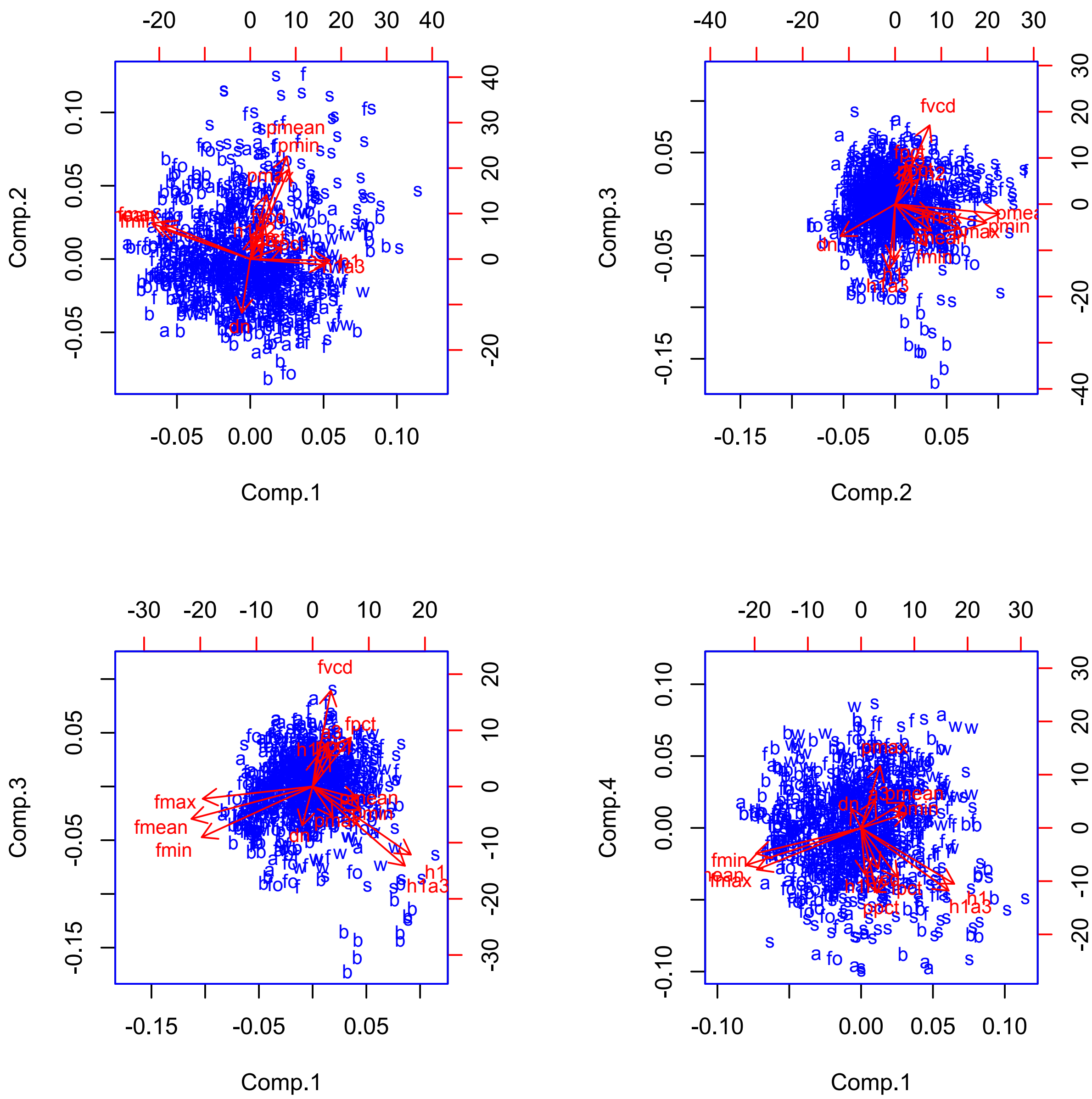

Figure 4: **Biplots of the first 4 principal components of dog D12's acoustic features and their categories, showing the overall shape of the data space.**

3 Prediction using Classification Trees

We used the 'R' statistical package [?] to perform a statistical modelling and prediction of the above acoustic material, learning the correlation of each acoustic category to the category of barks and predicting from the characteristics of unseen samples to classify each sound into one of the seven types of bark.

Both Classification Trees and SVMs were tested; the former being relatively weak at prediction but very useful for examining the contribution of the factors, and the latter being perhaps the strongest classifier available for general use.

The classification tree was trained with the 'R' command 'Tree', taking d2$ctxt (the object containing the contexts or categories of the barks, and learning the factors of x2 (the object containing the acoustic variables) that best predict those categories. Here we examine the subset containing the barks of dog D24 first. The '>' symbol indicates an 'R' command line:

```
> ttt=tree(d2$ctxt ~ ., x2,subset=d2$id=="d24")
> summary(ttt)
Classification tree:
tree(formula = d2$ctxt ~ ., data = x2, subset = d2$id == "d24")
Variables actually used in tree construction:
[1] "dn"    "pmax"  "a3"    "ppct"  "pmean" "fvcd"
Number of terminal nodes:  10
Residual mean deviance:  2.889 = 4886 / 1691
Misclassification error rate: 0.575 = 978 / 1701
```

Next we use the tree to predict a classification for the same data. Since the tree has been pruned so strongly, it only predicts less than half of the tokens correctly, but the confusion matrix allows us to see the typical confusability in the data (see also Figure 5):

```
> ttp=predict(ttt,type="class")
> table(ttp,ctxt[d2$id=="d24"])

ttp    a   b   f  fo   p   s   w
  a   43  13  36   2   0   1  17
  b   22  77  14  59   4  13  19
  f   23  42 140  19   5  46  25
  fo  29   6   0  40  39   0  18
  p   70   0   1  16 128  14  56
  s   48   5   4  19  11 117  22
  w   42  22  40  13 126  17 178
```

Play (p) and Stranger (s) appear to be well predicted, and we can assume that they have characteristic or distinctive acoustical features that distinguish them, but Ball (b) and Fight (f) show some confusion, indicating that their acoustic features are not so clearly distinguished at this simple level of statistical decision making. Not surprisingly perhaps, Walk (w) and Play (p) appear to be highly confused (similar?). Stranger (s) is similarly confused with Fight (f). For completeness, we look next at the performance through detailed counts (note the use of '==' here):

```
> table(ttp==d2$ctxt[d2$id=="d24"])

FALSE  TRUE
  978   723
```

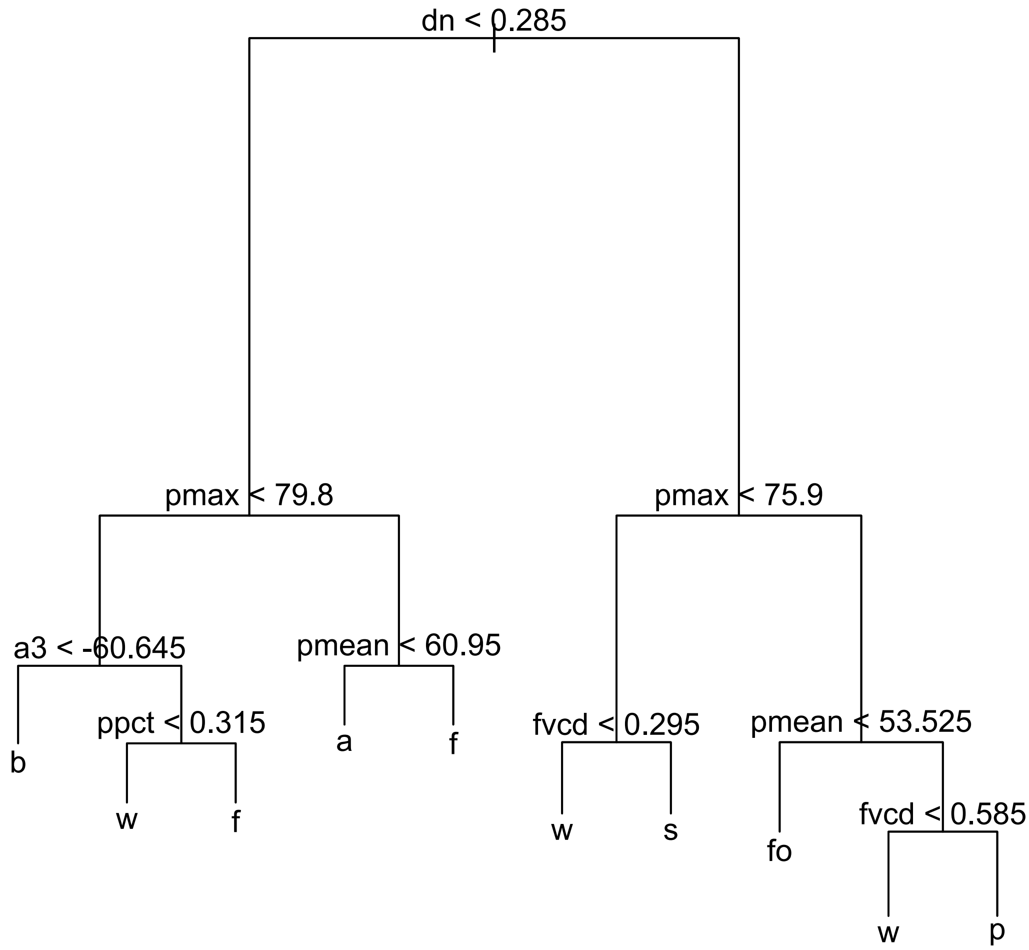

Figure 5: **The Classification Tree for predicting barks from D24 - with 10 leaves, showing duration to be the first determining factor, followed by pmax, a3, and f0-voicing respectively. This simple tree predicts less than half of the barks correctly, but confirms that most of the features are being used in making the classification.**

We turn next to dog D12.

```
> ttt=tree(d2$ctxt ~ ., x2,subset=d2$id=="d12")
> summary(ttt)

Classification tree:
tree(formula = d2$ctxt ~ ., data = x2, subset = d2$id == "d12")
Variables actually used in tree construction:
[1] "pmin"  "pmax"  "dn"    "pmean" "fmean" "fvcd"  "fmin"
Number of terminal nodes:  15
Residual mean deviance:  2.087 = 2150 / 1030
Misclassification error rate: 0.4134 = 432 / 1045
> table(ttp,d2$ctxt[d2$id=="d24"])

ttp    a   b   f  fo   p   s   w
  a   43  13  36   2   0   1  17
  b   22  77  14  59   4  13  19
  f   23  42 140  19   5  46  25
  fo  29   6   0  40  39   0  18
  p   70   0   1  16 128  14  56
  s   48   5   4  19  11 117  22
  w   42  22  40  13 126  17 178

> table(ttp==d2$ctxt[d2$id=="d24"])

FALSE  TRUE
  978   723
```

This time the tree predicts almost 60% of the tokens correctly, but as the plot (Figure 6) shows, a completely different set of features is used. Finally we test the entire dataset. The tree only develops 5 terminal nodes, and fails to predict more than 69% of the tokens, with no predictions of Walk, Food, or Alone barks Interestingly, Fight and Stranger appear to be well predicted but internally confusable. However, it is clear that there is considerably individuality in the remaining categories of barks for these dogs.

```
Classification tree:
tree(formula = d2$ctxt ~ ., data = x2)
Variables actually used in tree construction:
[1] "pmin" "dn"   "a3"   "pmax"
Number of terminal nodes:  5
Residual mean deviance:  3.581 = 26470 / 7390
Misclassification error rate: 0.6944 = 5135 / 7395
>  ttp=predict(ttt,type="class")
> table(ttp,d2$ctxt)

ttp     a    b    f   fo    p    s    w
  a     0    0    0    0    0    0    0
  b    72  398  134  370   67  182  134
  f   215  117  549   76  167  308  234
  fo    0    0    0    0    0    0    0
  p    66   93   24   38  259   49   93
  s   500  497  529  393  387 1054  390
  w     0    0    0    0    0    0    0
```

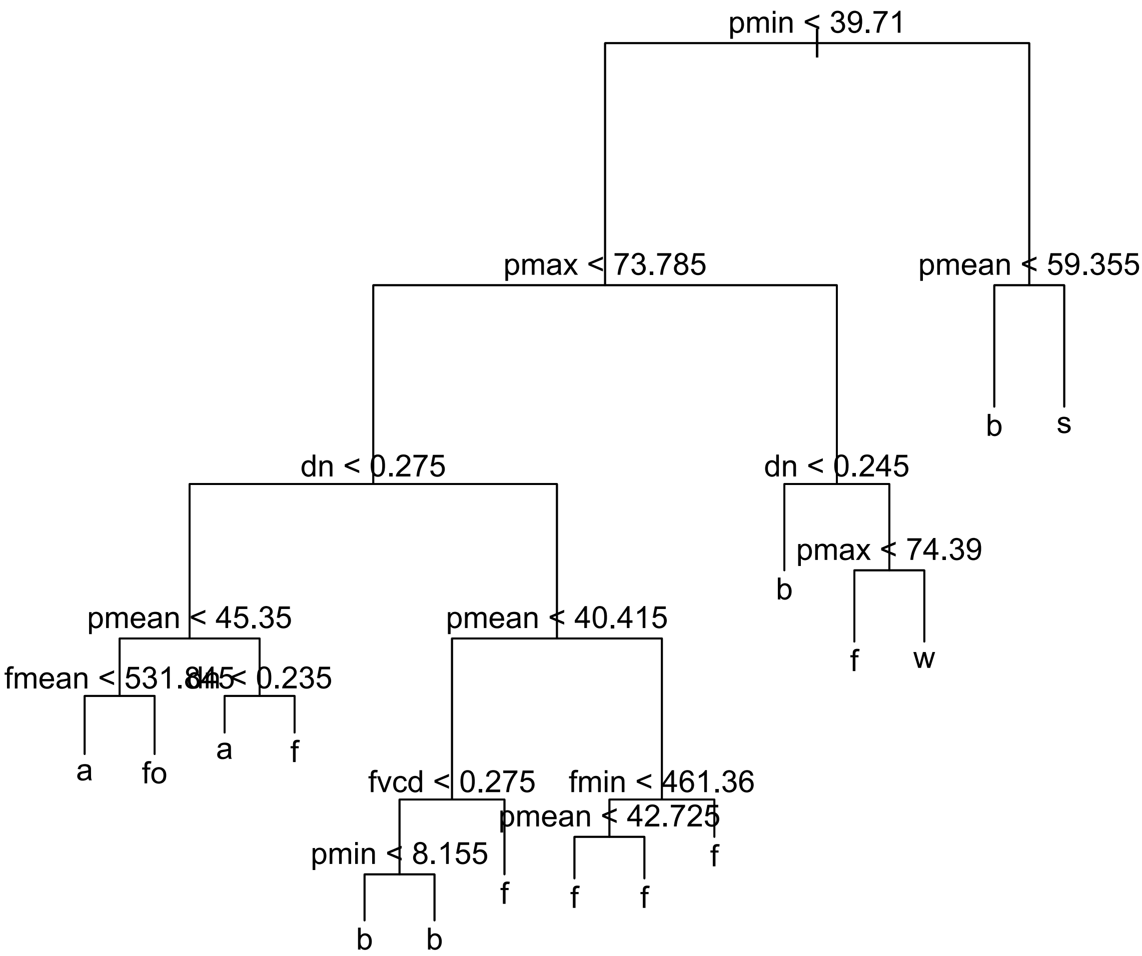

Figure 6: **The Classification Tree for predicting barks from D12 - with 15 leaves, using a different set of parameters and parameter ordering from that determined for D24, starting from min-pitch, then looking at mean and max of the same. This tree predicts about 60% of the barks correctly,**

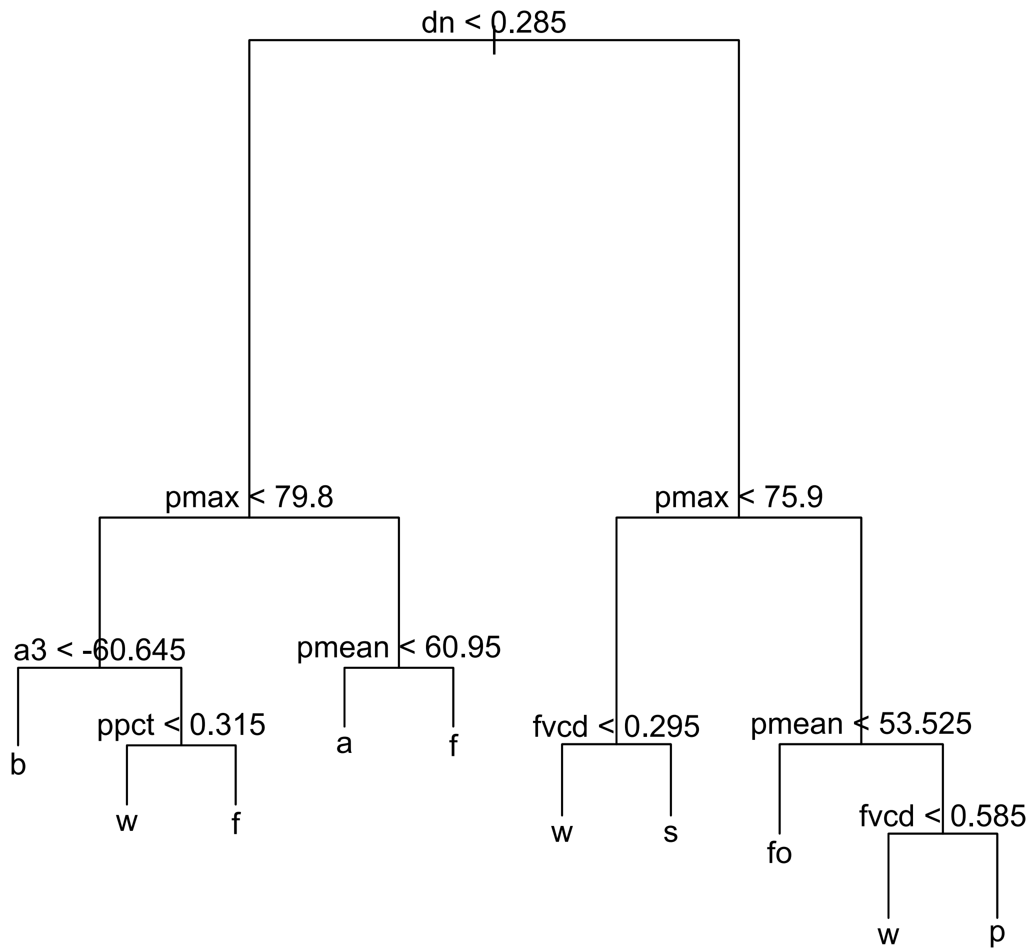

Figure 7: **The Classification Tree for predicting barks from D24 - with 10 leaves, showing duration to be the first determining factor, followed by pmax, a3, and f0-voicing respectively. This simple tree predicts less than half of the barks correctly, but confirms that most of the features are being used in making the classification.**

4 Prediction by SVM

Support Vector Machines are included in the library **e1071** in 'R'. Training an SVM on the same data gives a much more successful result, since it employs a total of 6744 Support Vectors to predict the data, rather than the 10 to 15 terminal nodes determined by the Classification Tree. Since it is not feasible to examine those vectors individually (as it was with the branches of the tree) we use n-fold cross validation to hold out n% of the data, train with the rest and then test with the held-out samples. Here we set n=15 (it is usually about 10, so this is a stricter test perhaps) and find that on average the SVM can successfully predict less than half of the samples.

```
> sss=svm(d2$ctxt ~ ., x2,cross=15)
> summary(sss)

Call:
svm(formula = d2$ctxt ~ ., data = x2, cross = 15)

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.07142857

Number of Support Vectors:  6744

 ( 1070 845 1339 1019 839 828 804 )

Number of Classes:  7

Levels:
 a b f fo p s w

15-fold cross-validation on training data:

Total Accuracy: 45.94997
Single Accuracies:
 38.94523 50.91278 46.65314 48.27586 44.62475 47.66734  43.20487
 46.4503 48.88438 47.66734 47.87018 46.04462 45.23327 46.24746 40.56795
```

However, when we limit the prediction to a subset of the barks, predicting in this case those of dog D12, we achieve an accuracy of greater than 60% in the general case:

```
Call:
svm(formula = d2$ctxt ~ ., data = x2, cross = 15, subset = d2$id == "d12")

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.07142857

Number of Support Vectors:  923
```

```
 ( 149 157 280 89 188 60 )

Number of Classes:  6

Levels:
 a b f fo p s w

15-fold cross-validation on training data:

Total Accuracy: 60.86124
Single Accuracies:
 31.88406 54.28571 62.85714 62.31884 75.71429 77.14286  75.36232
 64.28571 65.71429 69.56522 65.71429 62.85714 57.97101 51.42857 35.71429
```

When we then examine the individual prediction accuracy for each category by running the SVM in 'predict' mode and comparing with the original categories, we obtain the following confusion matrix:

```
> ssp=predict(sss)
> table(ssp,ctxt[d2$id=="d12"])

ssp    a   b   f  fo   p   s   w
  a   93   2   2  11   0  15   1
  b   11 142   8   4   0   6   1
  f   33  35 320   9   0  68   7
  fo  11  11   7  72   0   6   1
  p    0   0   0   0   0   0   0
  s    3   3   7   1   0  99   1
  w    1   0   1   2   0   0  51
```

This can be compared with a similar exercise for dog D24, achieving similar accuracy (given the larger number of classes) and clear discrimination, as confirmed by the two clear confusion matrices.

```
> sss=svm(d2$ctxt ~ ., x2,subset=d2$id=="d24",cross=15)
 . . . . .
Total Accuracy: 56.61376
Single Accuracies:
 46.90265 56.63717 55.26316 57.52212 64.91228 50.44248 58.40708
 50.87719 66.37168 52.63158 63.71681 67.25664 56.14035 61.9469 40.35088

> ssp=predict(sss)
> table(ssp,ctxt[d2$id=="d24"])

ssp    a   b   f  fo   p   s   w
  a  187   9   6  12  16  18  29
  b   10 109   5  28   1   7  15
  f   12  13 188   1   1  11  17
  fo  15  20   4 103   2   3   8
  p   21   1   7  19 237   2  47
  s   21   8   8   0   5 154  10
  w   11   5  17   5  51  13 209
> ================================
```

5 Cross Prediction (across dogs)

The final exercise in this report concerns the training of a SVM on one dog and then testing the results on another. To do this, we first make a subset of the data containing the material related to each dog. Here we use for example dogs D12, D23, and D24:

```
> d12=subset(d2,id=="d12")
> d23=subset(d2,id=="d23")
> d24=subset(d2,id=="d24")
```

then make a training subset of each, separating the predictor variables from the independent categories:

```
> x12=subset(d12, select=c(-ctxt,-n,-id,-wav))
> y12=d12$ctxt
> x23=subset(d23, select=c(-ctxt,-n,-id,-wav))
> y23=d23$ctxt
> x24=subset(d24, select=c(-ctxt,-n,-id,-wav))
> y24=d24$ctxt
```

and then perform the model training to build the SVM:

```
> m12=svm(x12,y12)
> m23=svm(x23,y23)
> m24=svm(x24,y24)
```

and test the results on the closed data, where perfect prediction can be expected (but considerable confusion remains in certain categories of bark):

```
> table(y12,predict(m12))

y12    a   b   f  fo   p   s   w
  a   93  11  33  11   0   3   1
  b    2 142  35  11   0   3   0
  f    2   8 320   7   0   7   1
  fo  11   4   9  72   0   1   2
  p    0   0   0   0   0   0   0
  s   15   6  68   6   0  99   0
  w    1   1   7   1   0   1  51
```

```
> table(y23,predict(m23))

y23    a   b   f  fo   p   s   w
  a  124   0  11   2   0   1   2
  b    3 155   2  33   0   1   7
  f   16   0 123   0   0   0   0
  fo   3  40   0 133   0   1  10
  p    0   0   0   0   0   0   0
  s    4  27   0   9   0  96  14
  w    0   8   0  20   0  10 115
```

and then between dogs, in the open data case, training first on one and then testing on the other:

```
> table(y12,predict(m23,x12))

y12     a   b   f  fo   p   s   w
  a    20  11  59  29   0   8  25
  b     3  32  58  17   0  41  42
  f    30  71 149  26   0  20  49
  fo   12  20  24  12   0   8  23
  p     0   0   0   0   0   0   0
  s    24  24 109  10   0  18   9
  w     3   5  23   0   0  25   6

> table(y23,predict(m12,x23))

y23     a   b   f  fo   p   s   w
  a    20   0  59  10   0  51   0
  b     0   2 190   0   0   5   4
  f    19   0  55   1   0  64   0
  fo   10   5 161   0   0  10   1
  p     0   0   0   0   0   0   0
  s     6   0  91   0   0  40  13
  w     5   5 128   0   0   9   6
```

or expressing the above in simpler numerical terms:

```
> table(y23==predict(m12,x23))

FALSE  TRUE
  847   123

> table(y12==predict(m23,x12))

FALSE  TRUE
  808   237
```

we see from these very poor cross-prediction results that the previously effective SVM cannot be trained on one dog to predict the bark categories of another. Results are similar for tests with dog D24, but for reasons of space have not been included here.

6 Discussion

From the above results, we conclude that barks are generally consistent in their acoustic realisation and that simple statistical models can be trained to predict the category of bark from the acoustics of that bark type for several different dogs. Within dog, we can obtain good prediction accuracies, of around 60% for a seven-category task, but *across* dogs the prediction accuracy drops considerably. It might be concluded from this that dogs learn to bark in a given situation individually, rather as a species, and that they bark to communicate with their owner, rather than with each other. This is speculation by the first author, a non-expert, but such an interpretation would be supported by the above data.

Within dogs, barks appear to be clearly distinguished and can be recognised to a large extent from their acoustics, but there is also considerable variety that prevents a more successful classification. This remains as work in progress. We will look next at the sequence of barks, to see if there is an effect for position in the bark sequence that might explain the apparent confusion in the data.

7 Acknowledgement

This work reported was carried out at BME in Budapest while the author was a summer guest of the BME (Budapest University of Technology and Economics) and the Hungarian Academy of Sciences as part of the ETOCOM project (TÁMOP-4.2.2-08/1/KMR-2008-0007) through the Hungarian National Development Agency in the framework of Social Renewal Operative Programme supported by the EU and co-financed by the European Social Fund. The first author also wishes to acknowledge the Science Foundation Ireland SFI Stokes Professorship Award 07/SK/I1218).