DEVELOPING NEW METRICS FOR THE INVESTIGATION OF ANIMAL VOCALIZATIONS

JOHN PLACER
Department of Computer Science and Engineering
Box 15600
Northern Arizona University
Flagstaff, AZ

AND

C. N. SLOBODCHIKOFF
Biological Sciences
Box 5640
Northern Arizona University
Flagstaff, AZ

ABSTRACT—An increasing body of evidence shows that non-human animal languages may be far more complex than previously assumed. Semantic content in animal alarm calls has been found in the vocalizations of vervet monkeys, some types of ground squirrels, dwarf mongooses, chickens, and Gunnison’s prairie dogs. This paper presents a classification system that provides important evidential support to earlier work on prairie dog communications. It does so using an entirely different system of analysis and a more fully automated experimental procedure. Furthermore, the application of fuzzy logic led to the development of a new system of metrics that can be used to investigate the difficult issues of how information is encoded in animal vocalizations and the level of linguistic complexity of those vocalizations.

Key Words: acoustic; alarm calls; fuzzy logic; neural nets; prairie dogs

1. INTRODUCTION

Many scientists today believe that animal vocalizations are merely affective by nature expressing anger or fear but not containing structured information. However, an increasing body of evidence shows that the alarm calls of some animals contain meaningful information of a semantic nature. Semantic content in predator-specific alarm calls has been found in the vocalizations of some ground squirrels and in vervet monkeys. Owings and Virginia found that the California ground squirrel has a different call for aerial predators than it does for terrestrial predators [1]. Seyfarth et al. discovered that vervet monkeys have different alarm calls for several species of predators, such as the leopard, martial eagle, and python [2]. Semantic information has also been found in the alarm vocalizations of dwarf mongooses [3], in the alarm calls of chickens [4], and in the alarm calls of red squirrels [5].

The techniques of analysis used to establish these semantic differences have fallen into two main categories. One category involves visual inspection of sonograms of vocalizations, and the separation of sonograms on the basis of their qualitative differences, i.e., one sonogram looks different from another. A sonogram is a plot of time versus frequency where the plotted data points are often represented with colors that denote different values along a third dimension of power. Thus, in a visual analysis of sonograms, if vocalizations elicited in the presence of different predator species produced qualitatively different sonograms, this is assumed to be evidence that the animals are producing different vocalizations for different species of predator [2]. Another category of analysis involves some form of quantitative
measurement of the structure of a sonogram, followed by statistical analysis of the measurements [6-8]. A few papers have used neural networks to classify animal vocalizations [9,10], but these have not been in the context of addressing the semantic nature of the vocalizations, but simply to classify different sounds of unknown function.

This paper describes how an attempt to corroborate the finding that Gunnison’s prairie dogs have different alarm calls for different species of predators led to the creation of an effective classification system that uses both fuzzy logic [11,12] and artificial neural networks [13,14]. The paper shows how an automated system using an entirely different set of techniques provides further support for the semantic nature of Gunnison’s prairie dog vocalizations, without the constraints imposed by previous forms of analysis, which employed statistical measurements. Section 2 discusses relevant earlier work demonstrating that Gunnison’s prairie dogs have different alarm calls for different species of predators. It also discusses the limitations of the techniques used in the earlier studies and explains why the work reported on in this paper was needed in order to lend evidential support to the claim that Gunnison’s prairie dogs have different alarm calls for different species of predators. Section 3 explains the experimental procedures utilized in two separate phases of the investigation being reported. In the first phase, the same types of parameters that had been used effectively in earlier work by Slobodchikoff et al. [15] were provided as input to a neural network classification system. In the second phase, a new set of metrics based on fuzzy logic was used as input to the same type of neural classifier used in phase one. Section 4 considers the results of both of these efforts. Section 5 offers conclusions concerning the experimental results obtained. It also discusses potential future work in which the new metrics that were developed promise to shed light on how information is encoded in animal vocalizations as well as on the level of linguistic complexity of those vocalizations.

2. BACKGROUND

A diversity of information of a semantic nature has been associated with Gunnison’s prairie dog alarm calls. This includes 1) different alarm calls for different species of predators [16,17] 2) information about the size, shape, and color of different individuals within a predator species [15] 3) information related to the direction and speed of approach of a predator [17] 4) and dialects in alarm calls between colonies [18,19].

![Figure 1. Sonogram showing variables used in Slobodchikoff et al.'s. [15] analysis.](image-url)
The work presented in this paper began with an attempt to corroborate the finding by Slobodchikoff et al. [15] that Gunnison’s prairie dogs have different alarm calls for different species of predators. Slobodchikoff’s work began by generating audio waveforms (sonograms) made from prairie dog alarm calls. Visually, a sonogram of a prairie dog alarm call bark consists of a number of chevrons stacked one on top of the other. Some small distance that represents a frequency difference separates each chevron from the one above it. Generally there are no alarm barks found below frequencies of 500 Hz or above frequencies of 8000 Hz. See Figure 1. Using these sonograms Slobodchikoff manually measured the following parameters: 1) the highest frequency contained in the most energetic chevron 2) the highest frequency contained in the chevron just above the most energetic chevron 3) the difference between these two frequencies 4) the slopes of these two chevrons 5) the time durations of these two chevrons related to the up-slopes 6) the down-slopes and 7) the entire chevrons 8) the time duration between alarm call barks.

Several problems can be identified with this form of analysis. One major problem is that of subjectivity of measurement. Measurements performed by one person may not be exactly the same as measurements performed by another person, introducing an artificial source of variation into the data set. Another major problem is in the form of the statistical analysis. Slobodchikoff used multivariate discriminate analysis, which is a form of multivariate analysis of variance that measures the within-group variance to the total variance found in the data set [20]. Based on such ratios of variances, a set of discriminant functions is generated, which then provides the statistical tools for establishing a classification table. This table assigns individual data points into categories. In the case of prairie dogs, the data points are the alarm barks from an individual prairie dog in response to a specific predator and the categories relate to barks elicited by a specific category of predator, such as a coyote. This statistical technique is sensitive to low sample sizes. Typically, it is not possible to have large sample sizes when studying prairie dogs because the size of each colony of animals is relatively small. The technique also might distort relationships as a result of the variance created by individual differences in the voices of the prairie dogs (e.g., male prairie dogs tend to have lower frequencies in their vocalizations than female prairie dogs, simply as a function of larger body size).

To address the above problems, the research described in this paper was required to satisfy the following constraints and goals.

- The data analysis must be accomplished with a more automated system that relies on fewer manual procedures for the measurement of alarm call parameters.
- The new system must utilize a powerful and flexible analysis technique.
- The new system must match or exceed the 80% to 85% classification accuracy achieved in the work completed by Slobodchikoff et al. [15].

Requiring a fully automated system avoided the problem of subjectivity mentioned earlier. Requiring a powerful and flexible method of analysis addressed the need to move closer to the ultimate goal of the research. Since final proof of the actual level of complexity of animal communications must rely on some exposition of the manner in which these communications encode information, a method of analysis is needed that allows in-depth examination of the structure of the alarm calls. Although the analysis technique based on multivariate statistics used by Slobodchikoff is capable of identifying important alarm call parameters and, to some degree, their relative importance, multivariate statistics cannot easily be used to ascertain the manner in which these factors encode information in animal vocalizations. The automated classification system described in this paper addresses all of the goals and constraints just discussed. Using a sophisticated numeric-processing environment, the new system of analysis is implemented as a fully automated software package where all measurements are made under software control. In addition to this, the power of fuzzy logic and artificial neural networks is used to analyze and classify the prairie dog alarm calls with high accuracy into predator-specific classes. Furthermore, the new system promises to support more sophisticated future research into the manner in which information is encoded in alarm calls.

3. THE EXPERIMENTAL PROCEDURE

Initially a database of individual alarm call barks was created using tape recordings of alarm calls taken from one hundred Gunnison’s prairie dogs obtained over a period of ten years at two separate prairie dog colonies. Copies of these data recordings were digitized using a generic sound card and software package on an IBM compatible PC with an Intel Pentium processor. The sound card had 16 bit resolution and
sampled data at 44,100 hertz. Following the digitization process a standard "cut and paste" sound editor was used to manually extract individual prairie dog "barks" and save them in separate files. These files of single alarm call barks formed a library of files that was used for all of the classification tests described later in this paper.

All subsequent tests and data manipulations were performed using a system of computer programs created by one of the authors. These programs were implemented using the high-performance numeric computation software called MATLAB and the Neural Network Toolbox associated with MATLAB. Experimental protocol and software design were tested using classification tests run on all combinations of the different predator species associated with the alarm calls. In other words, all six combinations of two different species were tested, all four combinations of three different predator species were tested, and data associated with all four species were tested.

The automated classification system first preprocesses each of the prairie dog alarm calls and then trains and tests a neural network. In the preprocessing stage, the frequency ratios contained in each alarm call bark are determined and then vectors of fuzzy values are created that characterize each alarm call bark. These vectors are used as input for the training and testing of the neural network. A block diagram of the classification system is given in Figure 2.

**Figure 2. Diagram of the classification system, showing the steps involved in processing and classifying alarm calls.**
When an experiment was to be run that attempted to classify a targeted combination of predator-specific alarm calls, each file used in that experiment was converted from a simple time domain recording of an alarm call bark into a matrix of sound frequencies contained within that bark. This conversion process was implemented in the following way. Each digitized data file representing an alarm call bark was divided into partitions of 256 data points. Since alarm calls were recorded at a rate of 44100 samples per second, each bark was effectively divided into a number of time periods of 0.0058 seconds each. The Fast Fourier Transform algorithm was then used to transform each partition of data points (i.e. each time period) into a normalized power spectrum of the sound frequencies that were expressed during that time period. Thus each alarm call bark was transformed into a matrix of values representing sound frequencies where each column in the matrix represented a different time period in the bark and recorded all of the frequencies that were expressed within that time period. This matrix represented the raw data associated with a sonogram.

The work described in this paper consisted of two different phases, where each phase investigated a different set of input parameters. However, both phases used the same general feedforward backpropagation neural network classifier. The network contained an input neuron for each separate input measurement, one hidden layer of fifty neurons, and an output layer containing as many neurons as there were predators to distinguish in a given classification experiment. Figure 3 illustrates a typical set-up of the neural classifier used when separating alarm calls associated with dogs from those associated with hawks. Each output node represented a particular species; the output node with the largest value for a given input vector named the species associated with that input. By running experiments with varying numbers of hidden nodes it was determined that 50 hidden nodes gave a good combination of low training times and high test accuracies. The hidden layer neurons utilized a sigmoid transfer function and the output neurons used a linear transfer function.

![Figure 3. A typical set-up for distinguishing dog from hawk associated alarm calls.](image)

The alarm call barks selected for a given classification test were divided into two roughly equal-sized sets. One set was used to train the neural network how to identify which predator was associated with each alarm call bark. The other data set was used to test the trained network for its ability to correctly guess which predator was associated with each alarm call bark. The two sets were disjoint which means that none of the vectors used to test the accuracy of the neural network were used to train the network. Training of the neural nets continued until either no errors occurred when classifying input vectors or until the training algorithm no longer improved the classification accuracy of the network. In none of the classification experiments did the network reach 100% accuracy. Therefore, in each test conducted the iterative training regime stopped itself when no further improvement was seen in three consecutive training sessions. The network weights that existed just prior to these three failed training sessions were then used as the final weights for the trained network.

3.1 The Input Parameters for Phase One.

The experimental design for phase one and phase two differed only in what the inputs to the network represented. In phase one, the inputs to the neural network consisted of measurements of all the parameters utilized in Slobodchikoff's [15] earlier work as discussed in Section 2. The only exception to this was the time duration between alarm call barks which had to be deleted because the database of digitized alarm call barks did not retain this information. Recall that Slobodchikoff manually measured attributes of visual
representations (graphs) of sonograms. In these experiments the measurements of all the parameters utilized were made under software control.

3.2 The Input Parameters for Phase Two.

While the input parameters used in phase one represented several different types of features in a prairie dog alarm call bark, only the single feature of frequency ratios was investigated in phase two. Now, after an alarm call bark was transformed into a matrix of frequencies as described earlier, a complete collection of frequency ratios was computed in the following way. For each time period (matrix column) the set of frequency ratios was computed by dividing each frequency expressed in that time period into all other smaller frequencies expressed within that period. Thus, every time period was represented by a set of frequency ratios, all of which lay within the range of values 0 to 1. The frequency ratios computed in this manner for each time period were then combined into one total set of values that represented the entire alarm call bark that was being analyzed. Figure 4 gives a simplified illustration of this process.

![Diagram](image)

Figure 4. Steps involved in transforming an alarm call into a collection of frequency ratios. In the time interval shown in the figure, there are three frequencies contained within the interval. To obtain the frequency ratios, each frequency within that time interval is divided into every other lesser frequency.
In order to characterize this collection of frequency ratios in a meaningful way, each collection was divided into twenty-one sub-ranges of ratios. Once this division was made, a vector of twenty-one fuzzy membership values was computed. Each fuzzy value in the vector represented the degree to which a corresponding sub-range of frequency ratios approximated a particular fuzzy frequency ratio.

To make this more explicit, consider that the frequency ratios computed for a given bark lie in the range of values 0 to 1. These ratios were divided into twenty-one sub-ranges as indicated in Figure 5. If the entire spectrum of sub-ranges was shown in Figure 5, there would be nineteen of the larger overlapping triangles (B to T) and two of the smaller right triangles, one on each end of the diagram (A and U). Each of these triangles encompasses a particular sub-range of frequency ratios and represents a membership function for a fuzzy set. For example, triangle C in Figure 5 represents the fuzzy membership function “approximately 0.10”. Similarly, each of the other membership functions represents a fuzzy frequency ratio. The fuzzy value corresponding to fuzzy frequency ratio \( i \) where \( i = \{ \text{"approximately 0.05"}, \text{"approximately 0.10"}, \text{etc.} \} \) was computed using the following formula:

\[
\mu_i = g_i \ast \bigvee \left( \mu_{i1}, \mu_{i2}, ..., \mu_{iN} \right)
\]

where

- \( \mu_i \) is the fuzzy membership value for fuzzy frequency ratio \( i \).
- \( \bigvee \left( \mu_{i1}, \mu_{i2}, ..., \mu_{iN} \right) \) is the standard fuzzy union (i.e. the maximum) of all the membership values that lie in the range of ratios associated with fuzzy frequency ratio \( i \). For example, if \( i = \text{"approximately 0.10"}, \) then the expression
  \[ \bigvee \left( \mu_{i1}, \mu_{i2}, ..., \mu_{iN} \right) \]
  would represent the maximum membership value of all the frequency ratios in the range 0.05 to 0.15.

\( g_i \) is the number of fuzzy frequencies associated with fuzzy frequency ratio \( i \) by 10

For example, if \( i = \text{"approximately 0.10"}, \) then the value of the numerator in the expression above would be the number of frequency ratios computed that lie in the range 0.05 to 0.15. The entity \( g_i \) is a simple weighting factor.

A number of different formulations of the equation given above were investigated. However, the equation given produced the best results.
Thus a distribution of twenty-one fuzzy frequency ratios was used to represent each time period examined in the individual barks in prairie dog alarm calls; these were the twenty-one values that were provided as input to the neural network. Each fuzzy frequency ratio distribution formed a particular signature that represented the presence of a specific combination of fuzzy frequency ratios.

4. RESULTS

The complete set of results for all classification tests in phase one is given in Table I and the results for all tests in phase two are given in Table II. Each row of each table represents a classification experiment for a specific combination of target predators where every classification experiment was run twenty separate times. Each time an experiment was run the neural network used was initialized with a new set of random weights. The first column in Tables I and II gives the mean of the testing accuracy achieved for each series of experiments. The parenthesized values in the first column are the standard error values of the means that were obtained. In columns three through six that specify predator type, the number of alarm call barks used in each neural network training session is given as a numerator and the number of barks used in each testing session is given as a denominator. The mean of the total number of sessions needed to train the neural network for each combination of predator species over a series of twenty experiments is given in the second column of each table. The parenthesized values in the second column are the standard error values of the means that were obtained.

<table>
<thead>
<tr>
<th>Accuracy Mean (%)</th>
<th>Training Sessions Mean</th>
<th>Humans (#train/#test)</th>
<th>Dogs (#train/#test)</th>
<th>Coyotes (#train/#test)</th>
<th>Hawks (#train/#test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75.2 (.3)</td>
<td>1925 (367)</td>
<td>181/178</td>
<td>86/85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67.0 (.6)</td>
<td>3550 (423)</td>
<td>181/178</td>
<td>138/137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85.7 (.3)</td>
<td>3125 (377)</td>
<td>181/178</td>
<td>107/106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74.1 (.3)</td>
<td>1975 (194)</td>
<td>88/87</td>
<td>138/137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85.0 (.3)</td>
<td>2100 (560)</td>
<td>88/87</td>
<td>107/106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81.1 (.5)</td>
<td>3550 (456)</td>
<td>138/137</td>
<td>107/106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57.0 (.4)</td>
<td>3800 (378)</td>
<td>181/178</td>
<td>88/87</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>69.7 (.4)</td>
<td>3175 (347)</td>
<td>181/178</td>
<td>107/106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62.2 (.6)</td>
<td>5050 (753)</td>
<td>181/178</td>
<td>128/137</td>
<td>107/106</td>
<td></td>
</tr>
<tr>
<td>68.3 (.4)</td>
<td>4175 (468)</td>
<td>88/87</td>
<td>138/137</td>
<td>107/106</td>
<td></td>
</tr>
<tr>
<td>54.2 (.3)</td>
<td>3825 (469)</td>
<td>181/178</td>
<td>86/85</td>
<td>138/137</td>
<td>107/106</td>
</tr>
</tbody>
</table>

In Table I, when alarm calls associated with only two different predator species were being classified, accuracy levels reached their highest peaking at 85.7% for humans and hawks and reached their lowest value of 67.0% for humans and coyotes. However, classification accuracy for tests associated with more than two predator species never went above 68.3% and sank to a low of 54.2% when all four species were being classified. In an attempt to improve the classification results, various combinations of the input measurements were used in the experiments. However, no specific combination of measurements was found that improved the classification results.

The classification accuracy in Table II shows an improvement over the results given in Table I for every type of test conducted. The tests associated with dogs and hawks yielded the highest classification accuracy with results that yielded an average of 96.3% accuracy. In Table 2, even the classification accuracy of the data associated with all four predator species yielded an accuracy of over 78%.

5. CONCLUSIONS AND FUTURE WORK

The classification system described in this paper corroborates earlier work showing that Gunnison's prairie dogs have different alarm calls for different species of predators. The corroboration is strong because it comes through the use of an entirely different analysis technique than that used in the original
Table II. Phase two results using the new fuzzy metrics.

<table>
<thead>
<tr>
<th>Accuracy Mean (%)</th>
<th>Training Sessions Mean</th>
<th>Humans (#train/#test)</th>
<th>Dogs (#train/#test)</th>
<th>Coyotes (#train/#test)</th>
<th>Hawks (#train/#test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>87.0 (.2)</td>
<td>2255 (225)</td>
<td>181/178</td>
<td>86/85</td>
<td>138/137</td>
<td>107/106</td>
</tr>
<tr>
<td>89.1 (.1)</td>
<td>1725 (213)</td>
<td>181/178</td>
<td></td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>95.2 (.1)</td>
<td>1525 (190)</td>
<td>181/178</td>
<td></td>
<td>138/137</td>
<td>107/106</td>
</tr>
<tr>
<td>87.7 (.2)</td>
<td>2000 (310)</td>
<td></td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>96.3 (.2)</td>
<td>2700 (298)</td>
<td></td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>90.3 (.2)</td>
<td>2850 (262)</td>
<td></td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>79.3 (.2)</td>
<td>3025 (300)</td>
<td>181/178</td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>85.6 (.2)</td>
<td>3575 (306)</td>
<td>181/178</td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>85.5 (.2)</td>
<td>2650 (319)</td>
<td>181/178</td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>85.5 (.3)</td>
<td>3175 (284)</td>
<td></td>
<td>88/77</td>
<td>138/137</td>
<td></td>
</tr>
<tr>
<td>78.6 (.2)</td>
<td>4700 (405)</td>
<td>181/178</td>
<td>86/85</td>
<td>138/137</td>
<td>107/106</td>
</tr>
</tbody>
</table>

research by Slobodchikoff et al. [16] or in subsequent work done by Slobodchikoff et al. [15] and Ackers and Slobodchikoff [21]. The work described here also meets the requirement that the system of analysis be automated. This automation allowed a large volume of field data to be processed where all measurements of relevant parameters were performed through software control. Previous work processed a smaller data set and utilized manual measurement techniques.

It is surprising that the final experimental results presented in this paper demonstrated such high accuracy. The original alarm call field recordings were made over a period of ten years in outside locations at two different prairie dog colonies. There were various sources of noise (including wind, trains, etc.) encountered during the taping sessions which resulted in recordings of varying quality. Additional degradation in sound quality came from the fact that the digitized alarm calls were made from copies of the original field recordings and these copies were played into a PC using a low-quality tape recorder. Thus, no extraordinary measures were taken to assure high fidelity sound reproduction. In addition to using noisy field data digitized with low quality audio equipment, it should be pointed out that no data filtering was done. The automated classification system utilized all the individual alarm call barks that were selected for a given classification test; no data was thrown out. Not only can one assume that some of the alarm call barks were of poor quality due to noise, but it is entirely possible that not all alarm call barks were intended solely to specify species-specific information. For example, it was mentioned earlier that prairie dog alarm calls have been shown to encode other types of information such as size, shape, and color of individual predators as well as direction and approach of a predator.

The use of a large number of alarm call barks recorded over a period of several years in varying conditions at two separate prairie dog colonies helped to guard against inadvertently training the neural network to recognize some specific recurring environmental condition. For example, if the same external sounds or noise were present each time alarm call barks were recorded in the presence of dogs, the neural network might be recognizing that external noise and not be identifying predator-specific information related to dogs. Another possible source of non-species specific information could be characteristics of the vocalizations of specific prairie dog individuals. This capability of neural networks was shown by Reby et al. [9] who trained neural networks to recognize the individual vocalizations of four fallow deer. Given that the data set used for the work described in this paper included vocalizations from about one hundred different prairie dogs, it is unlikely that recognition of individual prairie dog vocalizations could explain the high accuracy of the classification results.

The work presented in this paper strongly suggests that frequency ratios are related in some way to the manner in which information is encoded in Gunnison's prairie dog alarm calls. Establishing the importance of this parameter is one of the major outcomes of the work described here. However, it is perhaps of greater significance that the fuzzy paradigm has supported a new way of thinking about the study of animal vocalizations. Transforming animal alarm calls into fuzzy frequency ratio distributions provides a new metric by which vocalizations can be compared and manipulated. It also suggests a new framework for future work on animal vocalizations. In the work described in this paper all time dependent information
related to frequency ratios is lost when the collections of ratios across all partitions in an alarm call are combined. In future work, the frequency ratio distributions of individual time periods in an alarm call will be examined for patterns that evolve over time. If fuzzy frequency ratio distributions actually represent distinguishable units or phonemes in prairie dog alarm calls, then structured information in these calls should manifest as patterns in the fuzzy frequency ratio distributions that evolve over time. The study of such patterns would provide a means by which to investigate how and at what level of complexity information is encoded in Gunnison's prairie dog alarm calls. Clearly future work is needed to investigate these possibilities.

6. ACKNOWLEDGEMENTS

We thank Anne Epstein for her efforts in digitizing a large library of Gunnison's prairie dog alarm calls.

REFERENCES