# Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types

Michael R. W. Dawson, Isabelle Charrier, and Christopher B. Sturdy[a)]
*Department of Psychology, Centre for Neuroscience, University of Alberta, Edmonton, Alberta T6G 2E9, Canada*

(Received 9 May 2005; revised 17 October 2005; accepted 2 March 2006)

The "chick-a-dee" call of the black-capped chickadee (*Poecile atricapillus*) contains four note types, A, B, C, and D that have important functional roles. This provides strong motivation for studying the classification of acoustic components of the call into different note types. In this paper, the spectrograms from a sample of A, B, and C notes (370 in total) were each described as a set of 9 summary features. An artificial neural network was trained to identify note type on the basis of these features, and was able to obtain better than 98% accuracy. An internal analysis of this network revealed a distributed code in which different hidden units generated high activities to different subsets of notes. By combining these different sensitivities, the network could discriminate all three types of notes. The performance of this network was compared to a discriminant analysis of the same data. This analysis also achieved a high level of performance (95%). A comparison between the two approaches revealed some striking similarities, but also some intriguing differences. These results are discussed in terms of two related issues: developing a research tool for note classification, and developing a theory of how birds themselves might classify notes. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2189028]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Jz [JAS] Pages: 3161–3172

## I. INTRODUCTION

Songbirds are a popular model for the study of vertebrate acoustic communication (e.g., Slater, 2003). One reason for this is that the circumstances under which songbirds learn song from a model can be viewed as being analogous to human language learning (Doupe and Kuhl, 1999; Goldstein *et al.*, 2003). That said, song is not the only class of vocalizations that songbirds learn from a model. For example, songbirds also learn *calls* that are used in the context of specific behavioral interactions. As a result, more general studies of acoustic signaling and communication in songbirds are broadening their scope to include examinations of learned calls and other vocalizations in addition to learned songs (e.g., Hailman and Ficken, 1996; Hughes *et al.*, 1998; Vicario *et al.*, 2002). The purpose of this paper is to report the results of one such examination of the notes of a learned call of a small songbird native to North America, the black-capped chickadee (*Poecile atricapillus*). We show that when the notes of this call are represented as small sets of acoustic features, a nonlinear statistical method (artificial neural network) and a linear statistical method (linear discriminant analysis) are both capable of using these features to correctly classify notes into different note types.

The black-capped chickadee is one songbird that has provided a great deal of information about learned calls. One of the learned calls for this species is the "chick-a-dee" call, for which the species was named (Ficken *et al.*, 1978). Chick-a-dee calls contain four note types, labeled A, B, C, and D. Chick-a-dee calls are interesting because while their component notes are generated in a fixed order (A→B→C→D), in a particular call note types can be either repeated or omitted. As a result, a seemingly infinite variety of different chick-a-dee calls can be produced (e.g., ACCCCD, ABDDD; see Figs. 1 and 2 for examples of calls and the constituent note types, respectively). Indeed, the combinatorial nature of the chick-a-dee call has been used to draw interesting analogies between it and human speech (Hailman, 1985).

The different note types from which the chick-a-dee call is composed are important for a number of reasons. First, chickadees appear to vary the note-type composition of their chick-a-dee calls depending on the particular context (e.g., Ficken *et al.*, 1994; Gaddis, 1985; Smith, 1972), suggesting that different distributions of note types within calls convey different information. Second, chickadees fail to respond in a species-typical fashion to the playback of chick-a-dee calls that have an atypical syntax. Reduced responsiveness can be caused by gross syntactical violations, such as complete reversal of call syntax (DCBA rather than ABCD); (Charrier and Sturdy, 2005), or from more subtle syntactical violations (CACACA instead of AAACCC; Clucas *et al.*, 2004). Third, even more subtle alterations of notes within chickadee calls can cause birds to treat calls as atypical. Charrier and Sturdy (2005) found that chickadee calls with the correct syntax (ABCD) but in which each individual note was time-reversed, thus altering both amplitude and frequency modulation patterns of the notes within the calls, led to significantly reduced responding compared to normal calls. Fourth, black-capped chickadees are themselves able to discriminate and categorize the different note types of the chick-a-dee call, although they sometimes confuse adjacent note types (Sturdy *et al.*, 2000). Finally, different chick-a-dee call note

---

[a)]Electronic mail: csturdy@ualberta.ca

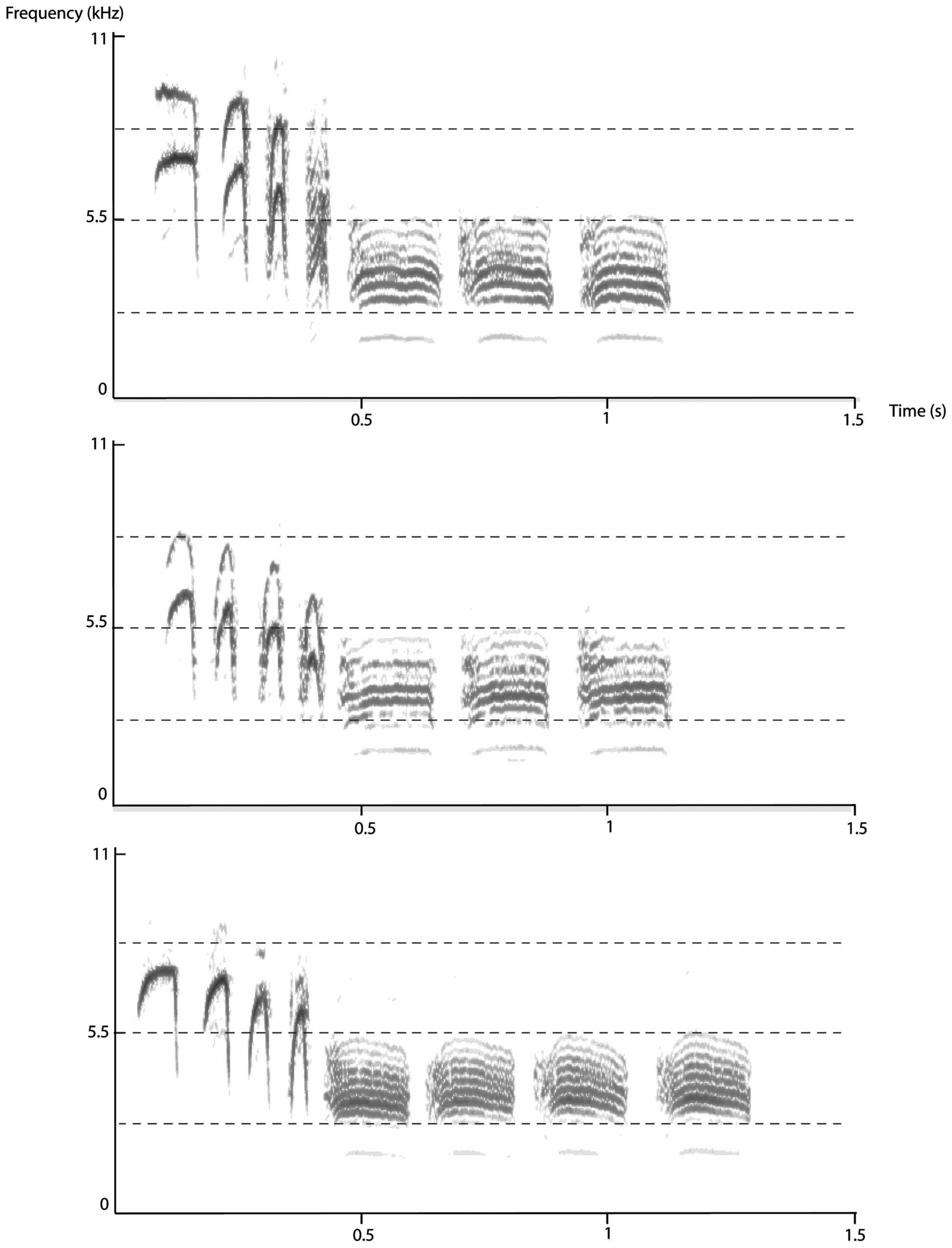

FIG. 1. Sound spectrograms of black-capped chickadee chick-a-dee calls from 3 different birds. The note composition of these calls is AABCDDD (top) AABBDDD (middle), and AAABDDDD (bottom), respectively. (Hamming window size=1024 points, frequency precision=21 Hz).

types have been associated with conveying specific information. For instance, the D note appears to be critical for birds being able to distinguish between chick-a-dee calls from members of their flock from the same kind of calls produced by members of a foreign flock (Mammen and Nowicki, 1981; Nowicki, 1983). The C note appears to be used to indicate the location or availability of food sources (Freeberg and Lucas, 2002). Taken together, it seems clear that the

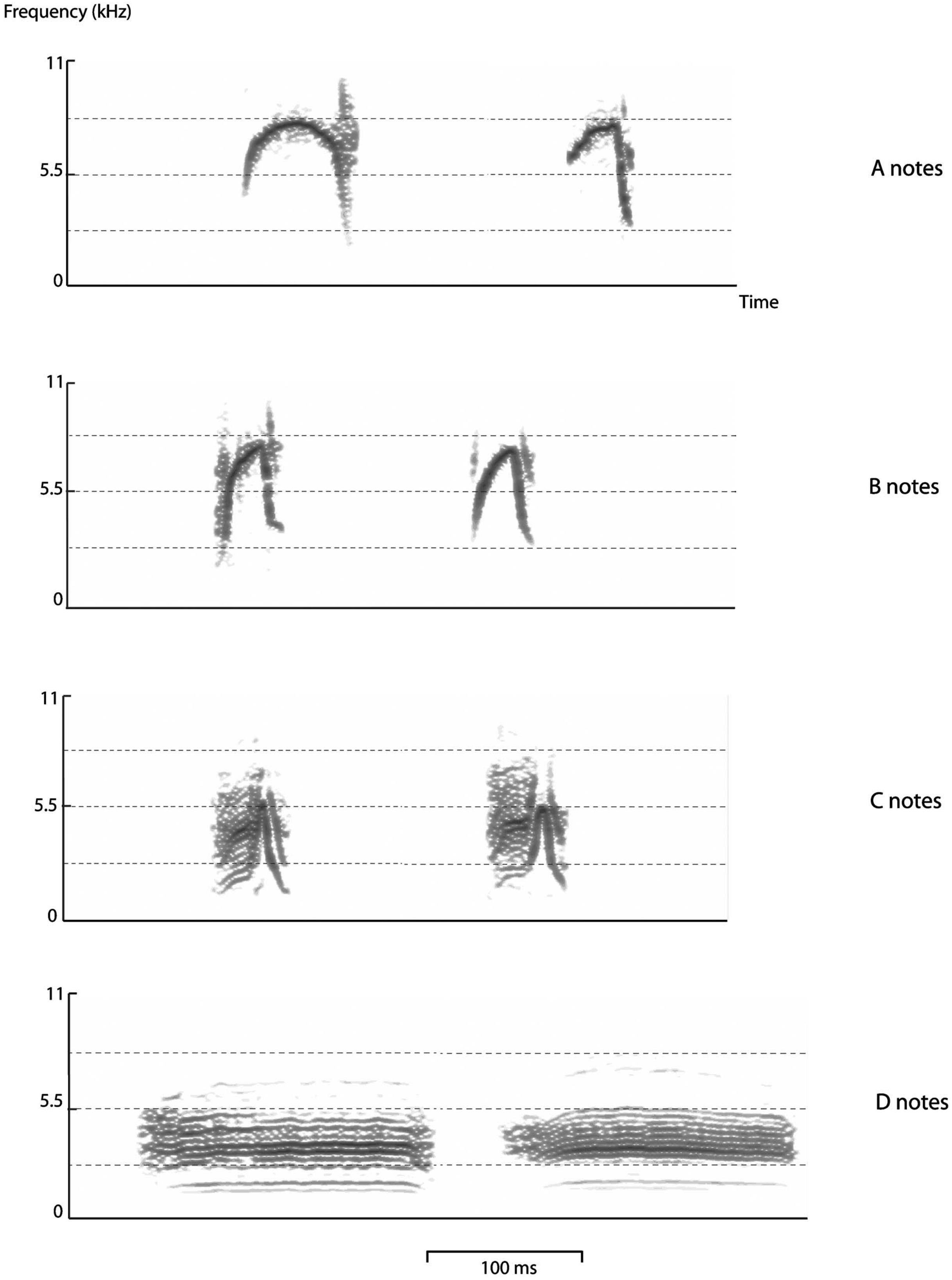

FIG. 2. Sound spectrograms of the four different note types found in black-capped chickadee chick-a-dee calls: A, B, C, and D notes. Exemplars for each note type from 2 different birds. (Hamming window size=1024 points, frequency precision=21 Hz).

accurate perception of individual note types in the chick-a-dee call is a crucial and fundamental building block of chickadee's acoustic communication system.

The fact that the constituent notes of the chick-a-dee call have important functional roles provides strong motivation for studying the classification of acoustic components of the call into one of the four different note types. Nowicki and Nelson (1990) describe such note classification as identifying natural categories in the acoustic signal. They point out that one critical research question is to determine how black-capped chickadees themselves identify such natural categories. A related and equally crucial question, studied in detail by Nowicki and Nelson, is how researchers themselves can identify these natural categories to aid in studying the properties of the call.

The traditional approach to this classification problem is exemplified by bioacousticians, who examine the visual representation of a vocalization (e.g., a chick-a-dee call) in the form of a sound spectrogram, and use visual features of the call to identify and classify its component note types. This technique has been used by several researchers to classify the notes of vocalizations from several songbird species [e.g., mountain chickadee, *Poecile gambeli*, chick-a-dee call notes, Bloomfield *et al.* (2004); Carolina chickadee, *P. carolinensis*, chick-a-dee call notes, Bloomfield *et al.* (2005); Black-capped chickadee chick-a-dee call notes, Charrier *et al.* (2005); Ficken *et al.* (1978); and Nowicki and Nelson (1990); zebra finch, *Taeniopygia guttata*, song notes, Sturdy *et al.* (1999b)]. Some of these visual classifications have been verified with the species that produces them via operant conditioning discrimination tasks (e.g., Sturdy *et al.*, 1999a, 2000), providing a necessary validation to the visual classification. That said, not all visually-guided classifications of animal vocalizations have been verified empirically, and depending on the species under study, many laboratory-based verifications may be impractical or impossible to conduct.

A less subjective approach to identifying natural categories of notes is to take spectrograms as raw data for traditional statistical analyses. Nowicki and Nelson (1990) provided one example of this approach. They took spectrograms representing 240 different notes (60 notes from each of 4 different birds), and cross-correlated all possible pairs of notes from this dataset. They then took the resulting correlation matrix and analyzed it with multidimensional scaling (MDS). MDS is a technique that provides a graphical representation of data, where data instances are represented as points in a multidimensional space (Kruskal and Wish, 1978). The more similar two instances of data are to one another, the nearer are their corresponding points in the space. Nowicki and Nelson found that a two-dimensional space provided an excellent account of their data. The MDS plot revealed a good segregation of each note type into different clusters with little overlap. This result was particularly encouraging because Nowicki and Nelson included D notes in addition to A, B, and C notes in their analysis. Because D notes are dramatically different than the other three in terms of acoustic properties, their inclusion possibly made differences between A, B, and C notes less apparent.

Another approach, also explored by Nowicki and Nelson (1990), is to take raw spectrograms and summarize them into a more tractable set of summary features. These summary features can then be used as input to a variety of statistical methods as an aid to note classification.

For example, Nowicki and Nelson (1990) described each of their non-D notes, that is A, B, and C notes (a sample of 185 stimuli), as a small set of different acoustic features that were derived from the spectrogram of the note. These features included such measurements as the total note duration, the starting, peak, and ending frequencies of the note, and so on. They created a matrix of correlations between every pair of notes on the basis of these features, and then performed principal components analysis on this matrix. They were able to account for over 68% of the variance in the correlation matrix with three principal components. They then described each of the 185 notes in terms of their loadings on these three factors, and processed this stimulus representation with *k*-means cluster analysis. They found that three clusters provided the most parsimonious account of the data.

A classification matrix based on the *k*-means solution indicated that this statistical approach generated a 77.3% agreement with visual classification of notes. That is, 143 of the notes were given the same classification that was assigned after a visual inspection of the spectrogram. With respect to misclassifications, 8 of the 71 A notes were classified as B notes, and 6 were classified as C notes. Similarly, 1 of the 65 B notes was classified as an A note, and 22 were classified as C notes. Finally, 5 of the 49 C notes were classified as B notes.

On the one hand, Nowicki and Nelson (1990) demonstrated that it is possible to use a variety of statistical analyses to classify notes in such a way that there is a high amount of agreement between statistical classifications and visual classifications. On the other hand, their research raises two questions that deserve additional study. First, is it possible to improve the already good fit between statistical and visual classifications by exploring alternative statistical methods? Second, is it possible to explore some form of statistical classification that might provide additional insight into how the birds themselves might perform note classification?

The purpose of this paper is to explore both of these questions by examining the ability of an artificial neural network to classify notes when these notes are described as a small set of features that have been derived from a spectrogram. The artificial neural network that we investigated is called a multilayer perceptron. In comparison to traditional statistics, this kind of network can provide classification power because it determines an optimal nonlinear combination of features. This kind of technique is also of interest because many researchers have argued that it is more biologically plausible than other techniques (Clark, 1989; McClelland *et al.*, 1986; Schneider, 1987; Smolensky, 1988). As a result, if one can produce an artificial neural network that is capable of classifying note types, this network might suggest general ways in which note classification is performed by the birds themselves.

This paper proceeds as follows. First we describe the methods used to obtain and summarize a set of A, B, and C notes from the chick-a-dee call. Second, we provide a brief overview of the multilayer perceptron, and examine its ability to classify this new set of notes. Third, to compare this approach to more traditional statistical approaches, we explore the ability of discriminant analysis to classify this same set of stimuli. Finally, we use the results of these two sets of analyses to discuss some issues related to the statistical and to the biological classification of chick-a-dee call notes.

## II. METHODS

The animals, recording procedure, acoustic measurements, and analyses have been published previously (Charrier *et al.*, 2004) and will be outlined only briefly here.

### A. Animals

Six male and four female adult black-capped chickadees were captured during the winters of 2002 and 2003 from two locations in Alberta. Birds were housed in individual cages, had *ad libitum* access to food and were maintained on natural day–night cycle typical for the season in Edmonton and at approximately 20 °C.

### B. Recording procedure, signal acquisition, and analysis

Each bird was recorded until we had a sample of at least 20 calls for each bird. Recordings (effective frequency range: 90–12 000 Hz) were digitized at 44 100 Hz, 16 bit samples/s using a 16 bit DartDisk Direct-to-Disk recorder (Engineering Design, Belmont, Massachusetts, USA). Calls were analyzed using SIGNAL version 4.0 (Engineering Design 2003).

From a sample of 100 calls, 370 non-D (i.e., A, B, and C) notes were visually classified and measured (205 A notes, 134 B notes, and 31 C notes) as was described in a previous paper (Charrier *et al.*, 2004). Nine acoustic features, partially based on the methods described in Nowicki and Nelson (1990), were measured including: start frequency (SF in hertz), peak frequency (PF in hertz), end frequency (EF in hertz) all of which were measured on a digital spectrogram [window size=1024 points, frequency precision=43 Hz; Fig. 2(a)] using a cutoff amplitude of −35 dB relative to the peak amplitude in the note. We also measured the loudest frequency ($F_{\max}$ in hertz) using a power spectrum [average window size=4096 points, frequency precision=11 Hz; Fig. 2(c)]. The duration measurements included were ascending duration (AD in milliseconds), descending duration (DD in milliseconds), and total duration (TD in milliseconds), and were measured on a digital spectrogram [window size =256 points, temporal precision=5.8 ms; Fig. 2(b)].

Two other measurements of frequency modulation were also made: the slope of the ascending frequency modulation [$\mathrm{FM}_{\mathrm{asc}}$ in hertz per milliseconds following the formula (PF–SF)/AD] and the slope of the descending frequency modulation [$\mathrm{FM}_{\mathrm{desc}}$ in hertz per milliseconds following the formula (EF–PF)/ DD].

### C. Data preprocessing

There is a tremendous range in the values of the different input features that were described earlier. This raised a concern about how the artificial neural networks would perform on these raw data, because the processing units of the networks must "squash" input values into a narrow range of from zero to one. We decided to preprocess the data in such a way that the range of the input features would be reduced, but the essential characteristics of the raw data would be unchanged. This was accomplished by normalizing the value of each feature. That is, for each feature we took the 370 features values and converted them into z-scores. This preprocessed data served as the input to both of the different classification tools that are described in the following.

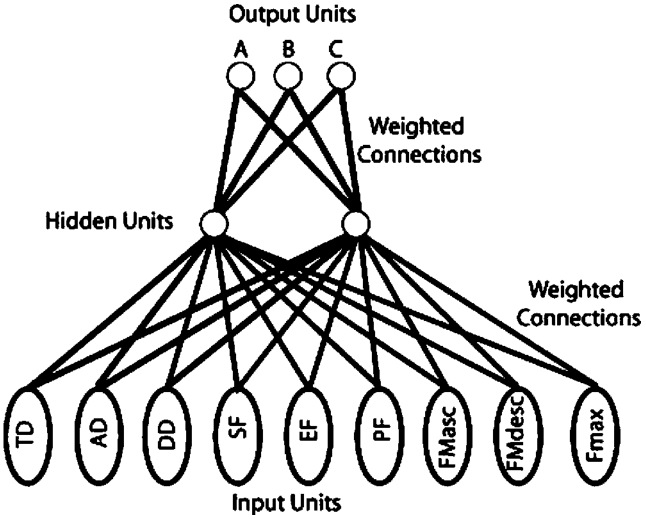

FIG. 3. An example multilayer perceptron. The network trained in the current manuscript is illustrated in which two hidden units processed the values of nine input features in order to assign a note to one of three different output categories.

### D. Note classification by artificial neural networks: A brief introduction to artificial neural networks

An artificial neural network is a computer simulation of a "brain-like" system of interconnected processing units (see Fig. 3). In general, such a network can be viewed as a multiple-layer system that generates a desired response to an input stimulus. The stimulus is provided by the environment, and is encoded as a pattern of activity in a set of input units. The response of the system is represented as a pattern of activity in the network's output units. Intervening layers of processors in the system, called hidden units, detect features in the input stimulus that allow the network to make a correct or appropriate response.

The behavior of a single processing unit in this system can be characterized as follows: First, the unit computes the total signal being sent to it by other processors in the network. Second, the unit adopts a particular level of internal activation on the basis of this computed signal. Third, the unit generates its own signal, which is based on its level of internal activity, and sends this signal on to other processors.

The signal sent by one processor to another is transmitted through a weighted connection, which amplifies or attenuates a numerical signal being sent through it. This is accomplished by multiplying the signal's value by the weight associated with the connection. The weight defines the nature and strength of the connection. For example, inhibitory connections are defined with negative weights, and excitatory connections are defined with positive weights. Strong connections have strong weights (i.e., the absolute value of the weight is large), while weak connections have near-zero weights.

The pattern of connectivity in an artificial neural network (i.e., the network's entire set of connection weights) defines how signals flow between the processors. As a result, a network's connection weights are analogous to a program in a conventional computer (Smolensky, 1988). However, in contrast to a conventional computer, an artificial neural network is not given an explicit program to perform some desired task. Instead, the network is taught to do the task.

Multiple layer networks (i.e., networks that contain one or more layers of hidden units), are frequently trained with a learning procedure called the generalized delta rule (Rumelhart *et al.*, 1986). One starts with a network that has small, randomly assigned connection weights. The network is then taught by presenting it a set of training patterns, each of which is associated with a known correct response. Each pattern is presented to the network's input units, and (on the basis of its existing connection weights) the network generates a response to it. An error term for each output unit is calculated by measuring the difference between the desired response of the unit and its actual response.

Error is used to modify connection weights by sending it backwards through the network. Once the error term for each output unit has been calculated, the weights of the connections directly attached to each output unit are modified. Then the output units send their error as a signal through the modified connections to the next layer of hidden units. Each hidden unit computes its overall error by treating the incoming error signals as net input (i.e., a hidden unit's total error is the sum of the weighted error signals that it is receiving from each output unit). Once a hidden unit has computed its overall error, then the weights of the connections that are directly attached to it can be modified. This process can be repeated, if necessary, to send error signals to the next layer of hidden units, and stops once all of the connections in the network have been modified. By repeating this procedure a large number of times for each pattern in the training set, the network's response errors for each pattern can be reduced to near zero. At the end of this training, the network will have a very specific pattern of connectivity (in comparison to its random start), and will have learned to perform a particular stimulus/response pairing.

### E. Network architecture

The network trained in this study was a multilayer perceptron with 9 input units, 2 hidden units, and 3 output units (see Fig. 3). Each input unit was used to represent one of the 9 spectrogram features; the input data for network training is identical to that used as predictors in the discriminant analysis reported later. Each input unit was connected to each hidden unit, and each hidden unit was connected to each output unit. There were no direct connections between input units and output units. The hidden units and the output units used the sigmoid-shaped logistic function to convert their incoming signals into internal activity that ranged between 0 and 1.

The network was trained to activate an appropriate output unit to classify a 9-feature input pattern. If the input pattern was an "A" note, then the network was trained to turn the first output unit on (and to turn the other two output units off). If the input pattern was a "B" note, then the network was trained to turn the second output unit on. If the input pattern was a "C" note, then the network was trained to turn the third output unit on.

TABLE I. Classification of three types of input notes (first column) into three note classes by an artificial neural network.

| | Note type according to artificial neural network | | | |
|---|---|---|---|---|
| Note type from visual examination | A | B | C | % correct |
| A | 203 | 3 | 0 | 98.5 |
| B | 1 | 132 | 1 | 98.5 |
| C | 0 | 0 | 30 | 100 |
| Total | 195 | 144 | 31 | 98.6 |

### F. Training

The ANN was trained using generalized delta rule. Initially, all connection weights were randomly assigned values between 0.1 and −0.1. The biases of the output unit and the two hidden units (which are analogous to neuronal thresholds) were initialized to a value of 0. The network was trained with a learning rate of 0.1 and zero momentum. Weights and biases were updated after the presentation of each training pattern. Each input pattern was presented once during a training epoch. Prior to each epoch, the order of presentation of the individual patterns was randomized. Pilot studies had indicated that the ANN never converged to a perfect solution to this problem—that is, it never learned to classify all of the notes with 100% accuracy. As a result, the achievement of perfect performance could not be used as a criterion to terminate training. Instead, the network was trained to the point that total network error (i.e., the sum of squared errors, with the sum taken over all of the output units and all of the training patterns) reached an acceptably low level. For the network reported in this paper, this was accomplished after training it for 2000 epochs. At the end of this amount of training, network SSE was 6.02. As is shown in the following, this level of SSE is associated with a very satisfactory degree of classification performance.

## III. RESULTS

### A. Network performance

The first question of interest at the end of network training concerns the amount of agreement between the visual classification of note types and the classifications performed by the network. In order to make this comparison, we first converted the activation patterns in the network's three output units into a specific note classification (A, B, or C). This was easily accomplished, because for each of the input notes, one of the output units generated activity of 0.9 or higher, while the other output units generated activity of 0.1 or lower. We classified the network's response by assigning it the note name associated with the output unit that was producing high activity. Table I provides the relationship between the visual classifications and the network classifica-

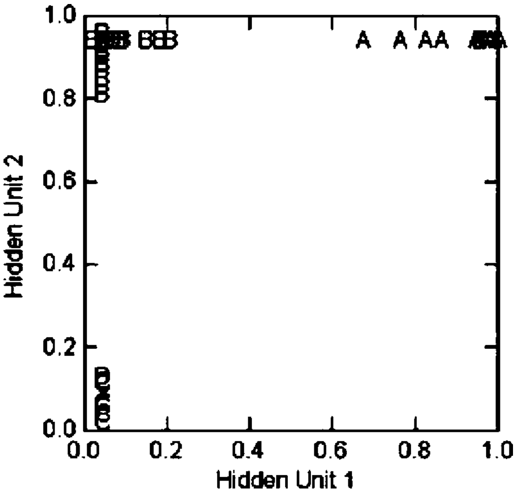

FIG. 4. The distribution of note types as a function of the activity produced in each hidden unit. See the text for details.

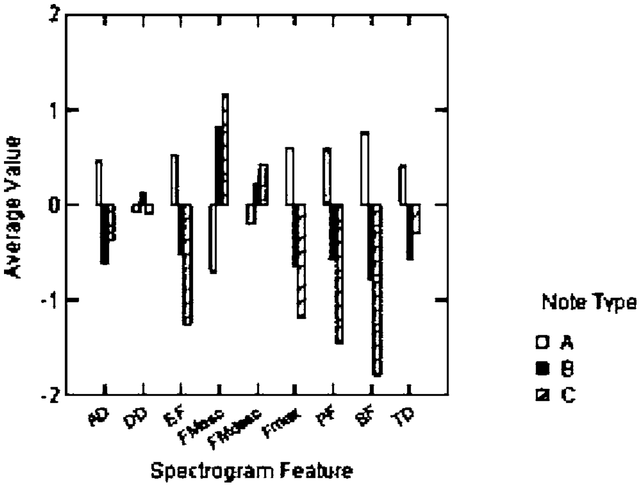

FIG. 5. Average values of the normalized features for each of the three note types. See the text for details.

tions. As can be seen from that table, the network was extremely accurate, misclassifying only 5 of the 370 stimuli for an accuracy level of 98.6%.

How is the network capable of performing at this high level? Some information relevant to answering this question can be obtained by examining the responses of the hidden units to all of the input patterns. In this network, the role of the hidden units is to map each pattern from a nine-dimensional space defined by the input features into a two-dimensional space defined by hidden unit activity. That is, each pattern can be described as a point in a two-dimensional space in which the coordinates of the point are provided by the activities produced by the pattern in the two hidden units.

A graph of this space is provided in Fig. 4. In this graph, each input pattern is represented as a letter that provides the note's type according to visual classification. By examining this graph, it appears that the two hidden units have very well defined functional roles. Hidden unit 1 is an "A-note detector," producing high activity to notes that belong to the A category, and producing little activity to other notes. In contrast, hidden unit 2 is a "not-C note detector," producing high activities to notes that do not belong to the C class, and producing near zero activities to notes that do belong to this class. These two hidden units classify notes as follow, then: if hidden units 1 and 2 are both on, the note is assigned to class A. If hidden units 1 and 2 are both off, then the note is assigned to class C. If hidden unit 1 is off, and hidden unit 2 is on, then the note is assigned to class B. This classification scheme works for 365 of the 370 notes in the training set.

How do these two hidden units respond in this fashion to discriminate between the different note types? One approach to answering this question requires an examination of the average feature profile of each note type in relation to the size of the connection weights that feed into each hidden unit.

Figure 5 provides the first type of information that is required for this interpretation. It is a graph of the average feature values for each of the three different note types in the training set. Two important observations follow from this graph. First, the feature profile for the average A note is markedly opposite to the feature profiles for the other two note types—the bars for each of its features point in an opposite direction than do the corresponding bars for the other two note types. Indeed, when the profile of the A note in Fig. 5 is correlated with the B note profile, the result is −0.99. When the A note profile is correlated with the C note profile, the result is −0.94.

The second important observation from this graph is the similarity of the profiles for B and C notes. For almost all of the features, the bars for these two notes presented in Fig. 5 point in the same direction. The correlation between these two profiles is 0.89. The difference between the two notes, as indicated in Fig. 5, would appear to be the fact that for many of the features the average C note has a more extreme value than does the average B note.

In order for a hidden unit to differentially respond to the various note types, the connection weights that feed into the hidden unit must transform the input features in such a way that the total signal is high for some notes (causing the hidden unit to activate) and is low for other notes (preventing the hidden unit from activating). Consider hidden unit 1, which generates high activity to A notes, and which generates little activity to B or C notes. The connection weights for this unit are presented in Table II. The bias for this unit, which is analogous to the unit's threshold, is also presented in Table II. The total signal coming into this unit from a pattern is equal to the sum of the input values of the features, after these input values have been multiplied by the weights. This signal is then combined with the bias, and passed as an input into the logistic equation $[f(\mathrm{net}_i)=1/(1+\exp(-\mathrm{net}_i+\theta_j))$, where $\mathrm{net}_i$ is the total signal and $\theta_j$ is the bias$]$ to convert net input into an activation value that falls in the range from 0 to 1.

The connection weights for hidden unit 1 are arranged in such a way that A note features tend to generate high positive net inputs, while features from the other two note types generate strong negative net inputs. For example, if the average feature values of an A note from Fig. 5 are converted into a net input using the weights in Table II, the result is 31.36. In contrast, the average feature values of a B note produce a net

TABLE II. Connection weights from the 9 input units (features) to the two hidden units of the network. The bias of each hidden unit—analogous to a threshold—is also provided.

| Input unit | Hidden unit | |
|---|---|---|
| | 1 | 2 |
| Unit bias | 6.55 | 28.33 |
| TD | −2.20 | −4.08 |
| AD | −0.05 | −0.57 |
| DD | −4.24 | 10.82 |
| SF | 33.88 | 29.80 |
| PF | 8.73 | −9.36 |
| EF | −16.20 | 1.94 |
| $\mathrm{FM}_{\mathrm{asc}}$ | −7.11 | 4.41 |
| $\mathrm{FM}_{\mathrm{desc}}$ | 1.96 | −6.06 |
| $F_{\max}$ | −2.80 | 0.04 |

input of −19.08, and the average feature values of a C note produce a net input of −49.18. Clearly, if one views the weights feeding into this hidden unit as a vector, then the orientation of this vector is far more similar to the orientation of the vector an average set of A note features than to the orientation of the vectors representing B and C notes.

This is not to say that all of the input features have equal importance in determining the response of hidden unit 1. The following regression equation includes only those input features that were found to be significant predictors of hidden unit 1's activity, and accounted for 80.1% of the variance in this activity ($F_{5364}=293.83, p<0.0001$):

$$\mathrm{H1}_{\mathrm{activity}} = 0.375\ \mathrm{SF} - 0.114\ \mathrm{EF} - 0.114\ \mathrm{FM}_{\mathrm{asc}} + 0.059\ F_{\max} - 0.039\ \mathrm{TD} + 0.550.$$

From Fig. 5, it can be seen that all of these features correspond to average feature values that demonstrate the largest differences between A notes and the other two note types. The one exception is peak frequency (PF). However, this feature has a very large correlation with $F_{\max}$ ($r=0.964$), and because of this redundancy the inclusion of PF in the regression equation would not provide a significant increase in predictive power.

A similar analysis can be provided for hidden unit 2, whose connection weights and bias are also presented in Table II. This hidden unit has a different function than hidden unit 1: it generates low responses to C notes, and high responses to the other two note types. Again, this is because the orientation of its weight vector is more similar to the orientation of the average feature vectors for A and B notes, and much less similar to the orientation of the average feature vector for a C note. When average features for A and B notes are combined with the Table II weights to produce hidden unit 2 net input, the results are 42.19 and 15.84, respectively. In contrast, when the average features of a C note are used to determine net input, the result is −10.54.

Again, not all nine input features are equally important for determining the activity of this hidden unit. The following regression equation includes only those input features that were found to be significant predictors of hidden unit 1's activity, and accounted for 47.0% of the variance in this activity ($F_{6363}=53.60, p<0.0001$):

$$\mathrm{H2}_{\mathrm{activity}} = 0.332\ \mathrm{SF} - 0.214\ F_{\max} + 0.152\ \mathrm{PF} - 0.098\ \mathrm{TD} + 0.087\ \mathrm{FM}_{\mathrm{asc}} + 0.042\ \mathrm{DD} + 0.916.$$

From Fig. 5, it can be seen that many of these features correspond to average feature values that demonstrate the largest differences between A notes and the other two note types. Indeed, there are numerous features shared in both regression equations. Interestingly, this equation includes PF and not EF, these two features have a high correlation with each other ($r=0.672$), and EF also has a high correlation with $F_{\max}$ ($r=0.68$). This equation also includes an additional variable not found in the first regression equation, DD.

### B. Note classification using discriminant analysis

The previous section demonstrated that a neural network is able to classify chick-a-dee call notes, described in terms of a small set of spectrogram features, with a very high degree of accuracy. This was accomplished by having two different hidden units separate one note type from the other two (in one case, A from B and C; in the other, A and B from C), and by then combining the activities of these hidden units to segregate all three note types. The hidden units operated by weighting input features in such a way to maximize the distinction between note types, and appeared to favor features that seem to distinguish note types (Fig. 5). Therefore it may not be surprising, though it is interesting, to find that both hidden units are sensitive to an overlapping set of note features, though they weight these features differently, and that one hidden unit is also sensitive to a handful of features that the other unit is not sensitive to.

While these results are encouraging, it is important to compare them to those produced by alternative classification methods. For instance, can traditional statistical methods perform this classification task as well as the neural network did? If so, do these methods use the same sets of features, or do they use an alternative method to classify notes? In this section, we investigate these questions by using discriminant analysis to categorize notes using the identical data that were presented to the neural network that was described previously.

### C. Discriminant analysis

When researchers are faced with data cases that can be assigned to a set of predefined classes, they often classify the data using discriminant analysis (e.g., Klecka, 1980), although Nowicki and Nelson (1990) did not study this method. Discriminant analysis is similar in many ways to multiple regression. Discriminant analysis determines a set of discriminant functions that can be used to categorize each instance. Each discriminant function has the form of a regression equation, $L=b_1x_1+b_2x_2+\cdots+b_nx_n+c$, where $L$ is the value of the function, each $b_i$ is a discriminant coeffi-

TABLE III. Classification of three types of input notes (first column) into three note classes by discriminant analysis.

| | Note type according to discriminant analysis | | | |
|---|---|---|---|---|
| Note type from visual examination | A | B | C | % correct |
| A | 193 | 13 | 0 | 94 |
| B | 2 | 129 | 3 | 96 |
| C | 0 | 2 | 28 | 93 |
| Total | 195 | 144 | 31 | 95 |

cient, each $x_i$ is a predictor variable, and $c$ is a constant. If there are $k$ different classes that cases can be assigned to, then discriminant analysis will compute a set of $k$ different discriminant functions. The coefficients of these functions are selected in such a way that classification errors are minimized. A case is classified by providing its input features as the predictors for each of the discriminant functions. The case is assigned to the class whose discriminant function generates the maximum value.

We used discriminant analysis to classify the 370 notes into the three predefined classes of A, B, and C on the basis of the 9 input features that were described earlier. Table III provides the classification matrix that was produced from this analysis. In general, while discriminant analysis was slightly poorer at classifying the note types than was the artificial neural network, its performance was still excellent: it correctly classified 95% of the 370 notes.

As noted earlier, in order to make the classifications that are summarized in Table III, three different discriminant functions—one for each note type—had to be computed. Each of these functions is similar to the connection weights that were presented in Table II, insofar as each discriminant function is a set of coefficients that are used to weight an input pattern's features to make a classification decision. The constants and discriminant coefficients for the three discriminant functions are provided in Table IV.

As was the case for the two hidden units of the neural network, not all of the input features are equally important contributors to the discriminant functions. Table V provides the F-test of the contribution of each predictor variable to the set of discriminant functions. As can be seen from Table V, six of the nine note features were statistically significant predictors. Table V reveals a striking similarity between the discriminant analysis of the notes and the classifications made by the artificial neural network. All six of the statistically significant contributors in Table V also appear in at least one of the regression equations that were reported earlier; four of the six appear in both of the regression equations. This suggests that the classification power of both techniques may be rooted in processing similar sets of features. Only one feature (DD) was not a significant contributor in the table below, but was a significant contributor in the regression equation that predicted the activity of hidden unit 2.

TABLE IV. Discriminant coefficients for the three discriminant functions used to classify the 370 notes into classes A, B, and C on the basis of the 9 predictors listed in the leftmost column.

| Predictor | Discriminant Function | | |
|---|---|---|---|
| | A | B | C |
| Constant | −3.027 | −3.033 | −14.696 |
| TD | −0.846 | 0.547 | 3.350 |
| AD | −0.205 | 0.144 | 0.760 |
| DD | −0.092 | 0.241 | −0.462 |
| SF | 5.284 | −4.468 | −16.291 |
| PF | 1.737 | −0.870 | −8.007 |
| EF | −1.206 | 1.332 | 2.303 |
| $\mathrm{FM}_{\mathrm{asc}}$ | −0.965 | 1.376 | 0.463 |
| $\mathrm{FM}_{\mathrm{desc}}$ | 0.370 | −0.344 | −0.984 |
| $F_{\max}$ | −1.246 | 0.109 | 8.049 |

TABLE V. $F$ statistics for evaluating the contributions of each note feature to the discriminant functions. Each $F$ value was evaluated with degrees of freedom of 2367.

| Feature | $F$ To remove | $p$ value |
|---|---|---|
| SF | 89.59 | 0.0001 |
| $F_{\max}$ | 26.99 | 0.0001 |
| TD | 25.39 | 0.0001 |
| PF | 14.83 | 0.0001 |
| EF | 14.26 | 0.0001 |
| $\mathrm{FM}_{\mathrm{asc}}$ | 8.82 | 0.0002 |
| DD | 1.66 | 0.193 |
| $\mathrm{FM}_{\mathrm{desc}}$ | 1.17 | 0.312 |
| AD | 1.04 | 0.354 |

The similarities between the discriminant analysis and the artificial neural network, in terms of their sensitivity to input features, suggest that a more detailed comparison of the two is in order. Table VI provides such a comparison. It presents the correlations between the two sets of connection weights in Table II (excluding the biases) and the three sets of coefficients in Table IV (excluding the constants).

These correlations are very informative. First, consider the correlations among the three sets of coefficients. The discriminant function for A notes has a strong negative correlation with the other two discriminant functions; however, the discriminant functions for B and C notes have a strong

TABLE VI. Pearson product moment correlations between the two vectors of connection weights that feed into the neural network's hidden units and the three vectors of coefficients for the discriminant functions.

| | H1 weights | H2 weights | A function | B function | C function |
|---|---|---|---|---|---|
| Weights for hidden unit 1 | 1.000 | | | | |
| Weights for hidden unit 2 | 0.600 | 1.000 | | | |
| Discriminant function for A | 0.958 | 0.649 | 1.000 | | |
| Discriminant function for B notes | −0.979 | −0.663 | −0.966 | 1.000 | |
| Discriminant function for C notes | −0.860 | −0.583 | −0.958 | 0.851 | 1.000 |

positive correlation. This suggests that the discriminant functions are taking advantage of the same regularities in note properties that the artificial neural network also exploited. Recall from Fig. 5 that, on average, A notes have an opposite feature profile to both B and C notes, while B and C notes have similar profiles, but that C notes tend to have more extreme feature values. The relationships between notes in Fig. 5 are consistent with the correlations between discriminant function coefficients in Table VI.

Now consider the correlation between the connection weights and the discriminant function coefficients. Recall that in order to be classified as an A note, a note's features must produce higher activity in the A function than in the other two discriminant equations. In this sense, the discriminant function for A notes has exactly the same function as hidden unit 1 in the network, which only generates high activity to A notes. One can see in Table VI that this similarity in function is reflected in the strikingly high correlation between the weights for hidden unit 1 and the discriminant function for A notes. Convergent evidence for this is reflected in the extreme negative correlations between these weights and the other two sets of coefficients; this mirrors the relationship between the A function coefficients and the other two sets of coefficients. In short, the mechanisms feeding into hidden unit 1 are functionally equivalent to the discriminant function for classifying A notes.

In contrast, the relationship between the connection weights for hidden unit 2 and the discriminant functions is less clear. On the one hand, the correlations are all quite strong, and show the same general pattern as the weights for hidden unit 1: a positive relationship with the A function's coefficients, and a negative relationship with the other two sets of coefficients. This is to be expected, perhaps, because of the relationships between average note profiles in Fig. 5, which showed an opposite trend of features between A notes and the other two note types. This is also to be expected because there is a strong positive correlation between the two sets of connection weights, indicating that they should have similar relationships to the three discriminant functions.

On the other hand, because none of the correlations involving the weights of hidden unit 2 are as high as those for hidden unit 1, we are not in a position to declare any functional equivalence. However, this is completely consistent with our previous analysis of that unit (e.g., Fig. 4). Hidden unit 2 generates high activity to A notes or B notes, and does not respond to C notes. Therefore its behavior, unlike that of hidden unit 1, is much less likely to be directly related to any one of the discriminant functions.

## IV. DISCUSSION

It was noted earlier that there are two important and related issues concerning note identification in the chick-a-dee call. The first is the development of an objective research tool for identifying note types in the lab. The second is the development of a theory about how the birds themselves process the acoustic signal to identify the different note types.

With respect to the first issue, the above-reported results are significant in several respects. They show that two very different techniques—artificial neural networks and linear discriminant analysis—are capable of classifying notes with a very high degree of accuracy. This is the case even when the large amount of information contained in a spectrogram is reduced to a small, manageable set of descriptive features. For our stimuli, the artificial neural network achieved an accuracy level of over 98%, and discriminant analysis achieved an accuracy level of 95%, for a relatively large set of stimuli (370 notes). The fact that both techniques performed so well would indicate that there is little reason to prefer one technique over another. Either method would perform extremely well as a note classification tool in a research lab.

With respect to the second issue, our analysis of the methods by which both techniques classified notes provides some important information to be considered in the development of a theory of how the birds themselves process acoustic signals to classify notes.

First, both the neural network and the discriminant analysis were able to classify notes by focusing on a particular subset of acoustic features. From a statistical perspective, it would appear that these features are particularly useful for note classification. The empirical issue that this observation raises is whether black-capped chickadees also focus on these features when they process this call.

Recent research from our laboratory suggests that in fact, black-capped chickadees do in fact attend to at least a subset of the acoustic features described here. Charrier *et al.* (2005) trained black-capped chickadees on four different note type discrimination tasks, A+B−, A−B+, B+C− and B−C+, and then conducted transfer tests with two sets of probe stimuli to determine to which features birds were attending in order to solve the task. Test notes were either pitch shifted, with the overall pitch of the note shifted up or down in frequency, or truncated notes, with either the ascending portion or descending portion presented. The results of these tests were quite clear. When notes were shifted in pitch toward the pitch space of another note type (e.g., when A notes had their overall note pitch lowered closer to that typical for B notes) the perception of that note also shifted from one type to the next. Further, birds could discriminate among note types using only the ascending portion of the note, and could generally do so better than using the descending portion of the note. This suggests that pitch cues contained in the notes such as SF, PF, EF and $F_{\max}$, as well as particular "dynamic" temporal/spectral cues such as $\mathrm{FM}_{\mathrm{asc}}$, are salient to the birds and control the perception of note types in the chick-a-dee call. Interestingly, these same features also appear to contribute to the functioning of the ANN reported here, based on the interpretation of the activity of hidden units 1 and 2, and, suggesting that perhaps the ANN is tapping into similar features when it sorts notes as when the birds themselves do an analogous task.

A second important piece of information, related to the first, is the fact that both statistical techniques took advantage of the fact that when notes are described as normalized sets of summary features, B and C notes are more related to one another (and therefore harder to discriminate from one another), while A notes seem quite different (Fig. 5). This was reflected in the patterns of correlations among the three sets of discriminant function coefficients, and in the fact that one of the functional components of the neural network was an A note detector. Again, this raises empirical questions about how birds process this call. Which note types do they appear to have more trouble discriminating? If they behave as if A notes are easier to differentiate than the other two note types, then this would suggest that the birds are processing a similar set of features as those represented in Fig. 5. However, if a different pattern of difficulties is in evidence, then this would suggest that the birds are processing different feature profiles.

Initial research on note type perception (Sturdy *et al.*, 2000) did in fact suggest that birds confused A, B, and C note types. When the pattern of errors was more closely examined, some interesting patterns emerged. Rewarded A note were confused with unrewarded B notes, rewarded B notes were confused with unrewarded A *and* unrewarded C notes, and rewarded C notes were confused with unrewarded B notes. These results, of an apparent perceptual continuum from A to B to C notes, mirrored the observation made earlier by Hailman *et al.* (1985) of a production continuum in these same notes types when the call is produced. Said another way, A notes, although sometimes confused, are only confused with one note type, B notes, whereas B notes are confused with both A and C note types. C notes are confused with B notes more than A notes, but Sturdy *et al.* also found that there was a trend for C notes to be confused with the next note type in the call sequence, D notes, not analyzed in the current paper. These findings provides some support, albeit indirectly, for the notion that both the birds and the ANN have difficulty in similar note type assignments, with A notes being misclassified less often than either B or C notes. This notion would have to be tested directly, perhaps by determining the number of notes of each type that a bird could discriminate, and to what level of accuracy, possibly in a choice task that required the discrimination and classification of all three note-types simultaneously, in a fashion similar to that of the ANN.

A third important point of information is that while the two techniques that we studied demonstrated similarities in features being processed, and also demonstrated some functional similarities, there were important differences between them too. The key similarity between the two techniques was that both included a distinct detector for A notes. By definition, the discriminant analysis produced distinct detectors for B and for C notes as well. Importantly, the artificial neural network did not develop these latter two types of detectors. Instead, its second hidden unit was an A or B note detector, which only turned off to C notes.

This difference between the two techniques demonstrates two different representational formats. In a local representation, each functional component of a system has a particular localized role, and can perform a meaningful task independently of the other functional components. The three discriminant functions created by the discriminant analysis are example of such local functional components. In contrast, artificial neural networks often develop distributed codes (Dawson, 2004; Dawson *et al.*, 2000a; Dawson *et al.*, 2000b; Dawson and Piercey, 2001; Hinton *et al.*, 1986; Leighton and Dawson, 2001). In a distributed code, different functional components of a system may not individually function in a fashion that can be directly related to the system's output. Instead, the meaningfulness of the functional components requires that their separate functions be considered simultaneously. The artificial neural network demonstrated this kind of representation. While the one hidden unit can be considered as being a local component (an A note detector), the behavior of the second hidden unit only makes sense when considered in the context of the behavior of hidden unit 1. That is, the representation is distributed because whether a note is an A, B, or C requires considering the activities of both hidden units at the same time (see the previous discussion of Fig. 4).

This is an important point for developing a theory of how birds process acoustic signals because it demonstrates that even when one focuses on the same set of features, there are different ways in which these features can be combined or processed to mediate note classification. The two techniques that we have described in this paper demonstrate two different approaches to using similar sets of features to make the same kind of classification. As theories of how chickadees process acoustic signals become related to the brain, one issue to address is whether acoustic features are represented in a local or distributed fashion. One advantage of artificial neural networks is that they provide examples of possible neural representations that can be used to guide this aspect of theory development (Dawson, 2004).

## ACKNOWLEDGMENTS

This research was supported by an NSERC discovery grant, start-up funding from the University of Alberta and an Alberta Ingenuity New Faculty Grant to C.B.S. and NSERC and SSHRC research grants to M.R.W.D. I.C. was supported by an Isaac Walton Killam Memorial Trust postdoctoral fellowship and an Alberta Ingenuity Fund postdoctoral fellowship. I.C. is now a member of the Bioacoustics Team, NAMC-CNRS UMR 8620, Université Paris Sud, Bat.446 F-91405, Orsay, France. This research was approved by the University of Alberta Biological Sciences Animal Care Committee (protocol No. 351201, originally approved on 14 November 2001) and the University of Calgary Life and Environmental Sciences Animal Care Committee (protocol No. BI2001-028 originally approved on 29 November 2001). Birds were captured under an Environment Canada, Canadian Wildlife Service Scientific permit (permit No. WSA-1-02, originally approved on 25 January 2002) and Alberta Fish and Wildlife Research permits (Nos. 4619 GP, 4621 GP, and 8734 GP) and collection licenses (Nos. 088 CN, 089 CN, and 147 CN) that were originally approved on 21 December 2001. Correspondence can be addressed to either C.B.S. or M.R.W.D.

Bloomfield, L. L., Phillmore, L. S., Weisman, R. G., and Sturdy, C. B. (**2005**). "Note types and coding in Parid vocalizations. III. The chick-a-dee call of the Carolina chickadee (*Poecile carolinensis*)," Can. J. Zool. **83**, 820–833.

Bloomfield, L. L., Charrier, I., and Sturdy, C. B. (**2004**). “Note types and coding in Parid vocalizations. II. The chick-a-dee call of the Mountain chickadee (*Poecile gambeli*),” Can. J. Zool. **82**, 780–793.

Charrier, I., Bloomfield, L. L., and Sturdy, C. B. (**2004**). “Note types and coding in Parid vocalizations. I. The chick-a-dee call of the black-capped chickadee (*Poecile atricapilla*),” Can. J. Zool. **82**, 769–779.

Charrier, I., Lee, T. T-Y., Bloomfield, L. L., and Sturdy, C. B. (**2005**). “Acoustic mechanisms of note-type perception in black-capped chickadee calls,” J. Comp. Psychol.. **199(4)**, 371–380.

Charrier, I., and Sturdy, C. B. (**2005**). “Call-based species recognition in black-capped chickadees,” Behav. Processes. **70**, 271–281.

Clark, A. (**1989**). *Microcognition* (MIT, Cambridge, MA).

Clucas, B. A., Freeberg, T. M., and Lucas, J. R. (**2004**). “Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*),” Behav. Ecol. Sociobiol. **57**, 187–196.

Dawson, M. R. W. (**2004**). *Minds And Machines: Connectionism And Psychological Modeling* (Blackwell, Malden, MA).

Dawson, M. R. W., Boechler, P. M., and Valsangkar-Smyth, M. (**2000a**). “*Representing space in a PDP network: Coarse allocentric coding can mediate metric and nonmetric spatial judgements*,” Spat. Cog. Comp.. **2**, 181–218.

Dawson, M. R. W., Medler, D. A., McCaughan, D. B., Willson, L., and Carbonaro, M. (**2000b**). “Using extra output learning to insert a symbolic theory into a connectionist network,” Mind. Mach. **10**, 171–201.

Dawson, M. R. W., and Piercey, C. D. (**2001**). “On the subsymbolic nature of a PDP architecture that uses a nonmonotonic activation function,” Mind. Mach. **11**, 197–218.

Doupe, A. J., and Kuhl, P. K. (**1999**). “Birdsong and human speech: Common themes and mechanisms,” Annu. Rev. Neurosci. **22**, 567–631.

Ficken, M. S., Ficken, R. W., and Witkin, S. R. (**1978**). “Vocal repertoire of the black-capped chickadee,” Auk **95**, 34–48.

Ficken, M. S., Hailman, E. D., and Hailman, J. P. (**1994**). “The chick-a-dee call system of the Mexican chickadee,” Condor **96**, 70–82.

Freeberg, T. M., and Lucas, J. R. (**2002**). “Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*,” Anim. Behav. **63**, 837–845.

Gaddis, P. K. (**1985**). “Structure and variability in the vocal repertoire of the mountain chickadee,” Wilson Bull. **97**, 30–46.

Goldstein, M. H., King, A. P., and West, M. J. (**2003**). “Social interaction shapes babbling: Testing parallels between birdsong and speech,” Proc. Natl. Acad. Sci. U.S.A. **100**, 8030–8035.

Hailman, J. P. (**1985**). “Ethology, zoosemiotic and sociobiology,” Am. Zool. **25**, 695–705.

Hailman, J. P., and Ficken, M. S. (**1996**). “Comparative analysis of vocal repertoires, with reference to chickadees,” in *Ecology and Evolution of Avian Vocal Communication*, edited by D. E. Kroodsma and H. H. Miller (Cornell University Press, Ithaca, NY), pp. 136–159.

Hailman, J. P., Ficken, M. S., and Ficken, R. W. (**1985**). “The ‘chick-a-dee’ calls of *Parus atricapillus*: A recombinant system of animal communication compared with written English,” Semiotica **56**, 191–224.

Hinton, G. E., McClelland, J., and Rumelhart, D. (**1986**). “Distributed representations,” in *Parallel Distributed Processing*, edited by D. Rumelhart and J. McClelland (MIT, Cambridge, MA), Vol. **1**, pp. 77–109.

Hughes, M., Nowicki, S., and Lohr, B. (**1998**). “Call learning in black-capped chickadees (Parus atricapillus): The role of experience in the development of ‘chick-a-dee’ calls,” Ethology **104**, 232–249.

Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage, Beverly Hills, CA).

Kruskal, J. B., and Wish, M. (**1978**). *Multidimensional Scaling* (Sage, Beverly Hills, CA).

Leighton, J. P., and Dawson, M. R. W. (**2001**). “A parallel distributed processing model of Wason’s selection task,” Cog. Sys. Res. **2**, 207–231.

Mammen, D. L., and Nowicki, S. (**1981**). “Individual-differences and within-flock convergence in chickadee calls,” Behav. Ecol. Sociobiol. **9**, 179–186.

McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (**1986**). “The appeal of parallel distributed processing,” in *Parallel Distributed Processing*, edited by D. Rumelhart and J. McClelland (MIT, Cambridge, MA), Vol. **1**, pp. 3–44.

Nowicki, S. (**1983**). “Flock-specific recognition of chickadee calls,” Behav. Ecol. Sociobiol. **12**, 317–320.

Nowicki, S., and Nelson, D. A. (**1990**). “Defining natural categories in acoustic-signals—Comparison of 3 methods applied to chick-a-dee call notes,” Ethology **86**, 89–101.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (**1986**). “Learning representations by back-propagating errors,” Nature (London) **323**, 533–536.

Schneider, W. (**1987**). “Connectionism: Is it a paradigm shift for psychology?,” Behav. Res. Methods Instrum. Comput. **19**, 73–83.

Slater, P. J. B. (**2003**). “Fifty years of bird song research: A case study in animal behaviour,” Anim. Behav. **65**, 633–639.

Smith, S. T. (**1972**). *Communication and Other Social Behavior in Parus Carolinensis*, Publi. No. 11 (Nuttall Ornithological Club, Cambridge, MA).

Smolensky, P. (**1988**). “On the proper treatment of connectionism,” Behav. Brain Sci. **11**, 1–74.

Sturdy, C. B., Phillmore, L. S., Price, J. L., and Weisman, R. G. (**1999a**). “Song-note discriminations in zebra finches (*Taeniopygia guttata*): Categories and pseudocategories,” J. Comp. Psychol. **113**, 204–212.

Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**1999b**). “Note types, harmonic structure, and note order in the songs of zebra finches (*Taeniopygia guttata*),” J. Comp. Psychol. **113**, 194–203.

Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**2000**). “Call-note discriminations in black-capped chickadees (*Poecile atricapillus*),” J. Comp. Psychol. **114**, 357–364.

Vicario, D. S., Raksin, J. N., Naqvi, N. H., Thande, N., and Simpson, H. B. (**2002**). “The relationship between perception and production in songbird vocal imitation: What learned calls can teach us,” J. Comp. Physiol., A **188**, 897–908.