The Relative Contributions of Bird Availability and Observer Detection Rates to Repeatability of BBS Results

D. Archibald McCallum, Ph.D.
Applied Bioacoustics
P. O. Box 51063
Eugene, OR 97405
(541) 221-2112
mccalluma@qwest.net

Final Report to
Patuxent Wildlife Research Center
Breeding Bird Survey
12100 Beech Forest Road
Laurel, Maryland 20708-4038

Under Purchase Order No. 02ERSA0327

Keywords: Monitoring, point count, BBS, detection probability, availability, detectability, correction factor, detection rate, observer effects, similarity index, temporal variation.

TABLE OF CONTENTS

ABSTRACT

A Breeding Bird Survey (BBS) route was simulated on the Patuxent National Wildlife Research Center, and four observers used standard BBS methods to survey 30 stops on this route eight times (twice per each observer) in a 11-day period in late summer of 2002. I transcribed their data and analyzed them for repeatability. Only the first 18 stops could be used for this analysis because of slight deviations in route after Stop 18. Qualitative and quantitative Sorensen similarity indices revealed substantial differences among days in the results of the surveys. Only a small amount of this variation was due to differences among observers. The two indices were highly correlated, indicating that differences in species detections, rather than differences in abundance estimates, were the major source of among-day differences. Species-specific detection probabilities (P) varied from .125 to .625, i.e., at best only about one-half of individuals (i.e., males) were detected on any given day. When these estimates were broken down into two components, availability and detection rate, it was found that the latter was approximately twice the former. These data suggest that it is most likely that natural variation in availability, the endogenous activity of the birds, that is the major source of negative bias in BBS data.

INTRODUCTION

Assessing the conservation status of wildlife populations requires monitoring. Most of the many monitoring techniques in use involve sampling rather than enumeration. Both the precision and bias of such samples must be estimated (Thompson 2002). Typically, a count, C, of individuals is made. The number of individuals in the survey area is N, which is estimated as 𝟙 = C / 𝟛, where 𝟛 is a correction factor for individuals that are present but not detected. Density, D, is estimated as 𝟙 * A, where A is the area surveyed.

Such estimates of N have three major sources of bias: accuracy, availability, and detectability. If an individual of one species is misidentified as another species, then the count C is erroneous. If not all individuals present are counted during the survey period, then 𝟛 < 1 and must be estimated. P, often referred to as detection probability, is conceptually and empirically more tractable if treated as the product of two independent probabilities, availability and detectability (Farnsworth et al, 2002, McCallum in press). Availability, a characteristic of the population being surveyed, is the probability that a bird that is present gives a cue during the survey period. Detectability is the probability that the observer detects at least one cue produced during the count period. These parameters are discussed more fully by McCallum (In Press). Detection probability is an area of active research, and many methods for detecting it have been published in recent years (). Estimating the precision and bias of C is not an area of active research. The major approach to improving accuracy is training.

Although it seems likely that estimates of N are biased in various ways, there is little empirical support for this conjecture. Moreover, it is not known which of the three major sources of error, accuracy, availability, and detection rate, contributes most to any bias of N. This study uses repeated realtime surveys of the same points (stops on a simulated BBS route), augmented with post-processing of concurrently recorded audio files, to estimate the relative importance of these factors. Specifically, I sought answers to the following three questions:

Question 1. What percentage of birds detected by realtime observers was misidentified, and how much did observers vary in accuracy?

Question 2. Are all birds resident at a stop available for detection during a three-minute BBS stop, and if not, what proportion of the population does emit a detectable aural cue during such a time period?

Question 3. What percentage of birds available for detection on the BBS is not detected (or ignored), and how much do detection rates vary?

METHODS

EXPERIMENTAL DESIGN

The same 30-stop simulated BBS route (see below) was run twice each by four observers. This design permits partitioning of variation into an inter-observer effect and a intra-observer, between-days effect. Data can be analyzed across stops, or pooled over all stops.

STUDY SITE

The study was conducted on the Patuxent National Wildlife Research Center, Laurel, Maryland, from 25 June 2002 through 7 July 2002.

THE ROUTE

Keith Pardieck of USGS laid out a pseudo-BBS route of 30 stops on the PWRC in June 2002. The route started at the station's main entrance, followed the left bank of the Patuxent River, crossed the river at Duvall Bridge, and followed Duvall Bridge Road to its junction with theWildlife Loop. Thence the route followed the Wildlife Loop eastward to Bailey Bridge, over the Little Patuxent River. The route contined northeastward from this point, paralleling the Amtrak Electric tracks to a point (Stop 18) 17.5 miles from the starting point. At this point the route was interrupted. It resumed 0.5 miles northwest of the junction of Duvall Bridge Road and theWildlife Loop on the Wildlife Loop. The route turned left on South Road, following it until it turned right on Combat Road. It then followed various roads along the northern boundary of the station, terminating near the Tipton Army Airfield (see map).

On 28 June 2002 one of the observers drove past stops 13-17 to the end of the first segment of the route at stop 18 by the Amtrak tracks. He then ran geographic stop 18 as his nominal stop 13, 17 as 14, etc. This departure would have had no effect on results except that his stop 19, noted on his data sheet as the last of the reversed stops, was clearly not at the same location as stop 19 on other days. First, running stop 18 as 13 would have made nominal 18 (at location 13), not 19, the last of the reversed stops. Second, spectrographic and aural comparison of the soundscapes at "stop 19" on June 28 and stop 19 on other dates indicated that the former was a woodland location, while the latter was in open habitat. This is indicated by the greater reverberation of sounds in the former location, as well as the identity of the species recorded at the locations.

Bioacoustic analysis of recordings made on each day confirmed that the following identities for June 28 stop numbers and those on other days: 13 = 18, 14 = 17, 15 = 16, 16 = 15, 17 = 14, 18 = 13. Stereotypy of song types of rare species, e.g., American Redstart, were used to confirm these identities. On the other hand, the observer on 28 June seems to have "frame-shifted" the final portion of the route by inserting a stop, as 19, that was not used on other dates. Moreover, this observer is known to have taken a different path toward the end of the route than other observers, and the survey terminated at stop 25 once (25 June) and stop 28 twice (27 June and 03 July). Considering that point counts at the same locations are needed for redundant samples to be used for estimating detectability, stops beyond 18 will not be useful for this analysis. Acoustic analysis of stops past 18 did not lead to high confidence that the nominal stops were at the same physical locations on all or even most days of the study.

DATA COLLECTION AND RECORDING

Observers

Four observers participated in the study, each running the route on two separate days (Table 1).

Table 1. Assigment of observers to dates in this study.

Date	Observer	Round
6/25/2002	2	1
6/26/2002	4	1
6/27/2002	3	1
6/28/2002	4	2
6/29/2002	2	2
7/1/2002	1	1
7/3/2002	3	2
7/5/2002	1	1

Observers wrote their data on blank paper on the first iteration. On the second interation they mapped all birds onto circular paper diagrams.

Recording

I recorded the ambient soundscape during the 3-min BBS stops from a fixed antenna array mounted above the survey vehicle. Figure 1 shows an observer collecting data in the conventional BBS manner, while the sounds were simultaneously recorded on a computer (Fig. 2) in the vehicle. The recordist remained in the vehicle during the three minutes the observer was collecting data.

Up to six microphones (Table 2) were mounted in this array (Fig. 3).

Table 2. Assignment of recording channels to microphones. The first three channels are of interest here. They were saved on all dates. Channels 4-6 were recorded on some dates as backup data, but these were not needed for this study.

Channel	Microphone	Direction	Mic Power Supply	Recorder / Pre-amplifier	
1	Sennheiser ME88	180	5.7 v	TCD5-ProII	
2	AT815b	300	1.5 V AA	TCD5-ProII	

3	AT815b	60	1.5 V AA	TCD5-ProII	
4	CVX	Omni	9V	TCD5-ProII	
5	Sennheiser ME62/K6	Omni	1.5 V AA		
6	Optimus	Omni	1.5 V AAA		

The three directional microphones (two Audiotechnica and one Sennheiser) were directed away from the recording location, horizontally (parallel to the ground), with an angular separation of 120 degrees. Output from these microphones and the CVX omnidirectional microphone were routed through two Sony TCD5-ProII analog cassette tape recorders, serving as pre-amplifiers, and thence via a custom-made junction box and cable provided by Engineering Design (Berkeley, California) into a National Instruments 6062E DAQCard implemented in a Toshiba 2595CDT notebook computer. Input was digitized directly with NIDisk software (Engineering Design, Berkeley, California) at a sample rate of 25,000 points per second. Each channel at each stop was saved as a separate file, each of slightly more than 3-min duration. Channels 1-4 were also recorded on audio-tape in the Sony cassette recorders, as backup.

Delineation of Individual Birds

To evaluate the accuracy of the observer's identifications, a means of linking the assignments of the observer, i.e., each individual tallied on the standard BBS data sheet, to the sounds, preserved on the audio recordings, upon which the assignment was based. The following options were presented and discussed with the observers:

- "No-linkage" option. The person analyzing the tape makes an independent count, and attempts to infer which assignments are erroneous. For example, if the tape should reveal 3 Red-eyed Vireos and 1 Philadelphia Vireo, but the observer's record lists 4 Red-eyed Vireos and 0 Philadelphia Vireos, it would be inferred that one of 4 REVI was misidentified as a PHVI, but it would not be known which of the observer's 4 REVI was misidentified.
- "Map" option. Observer maps relative location of all detected birds on a circular map of the area inside a circle of radius 50 m, centered on the van. The person analyzing the tape attempts to equate sounds on the tape with singing locations on the maps.
- "Voice-over" option. Each time the observer makes an assignment, he/she calls it out loudly enough to be audible on the concurrent audio recording.
- "Dictation" option. Each time the observer makes an assignment, he/she will whispers it into a lip-microphone connected to a belt-mounted portable recording device (e.g., mini-disc recorder).

Although the "Voice-over" option is equivalent to the Double Observer method (Nichols et al. 2000) in the amount of speech that is produced, the observers felt that it disrupted normal procedures, and would result in survey results that are not applicable to the BBS. It and the "Dictation" option were therefore abandoned as unworkable. The remaining options provide much lower probability of accurately linking a sound on the recording with a bird on the observer's data sheet, but at least the data sheet itself would not be compromised. The "Map" option was implemented on each observer's second survey of the route, i.e., on 28 and 29 June, and 3 and 5 July (Table 2).

DATA ANALYSIS

I transcribed the observers's hand-written data to computer files, using annotations made at the time of collection to assign each detection to song, call, juvenile begging, nonvocal sound, visual subcategories. Data were collated in Excel. All quantitative analyses were performed in SAS. The resulting data matrix contained one row for each species at each stop on each day.

Spectrographic Post-processing

For selected stops and dates (Table 3), I reviewed channels 1-3 (Table 2) of the audio files recorded while the observers were collecting conventional BBS data. This analysis was not independent of the observers'. Rather, I attempted to find every species and individual detected aurally by the observer. The process is actually analogous to the "Double Observer" method (Nichols et al. 2000). The resulting data matrix contained all information in the observer-only data matrix, and was augmented in two ways. Additional variables (columns) were added to accommodate new fields (variables) describing birds detected via acoustic post-processing. Additional observations (rows) were added to accommodate species not detected by the realtime observers. This data set will be referred to as the "post-processed" data set.

Table 3. Stops subjected to spectrographic post-processing. In all cases all four observers' data for the round were examined.

Stop	Round 1	Round 2
1		x
3	x	x
6		x
7		x
8		x
10	x	x
11	x	x
14	x	x
18	x	x

Methodology for post-processing was as follows: Using custom macros written for SIGNAL^{tm} 4 software, I displayed time-aligned spectrograms for simultaneous segments of the three tracks, usually of 10-s duration (Fig. 4). The macro allowed me to listen to any track or part thereof repeatedly, and a subroutine allowed me to consult a library of identified sounds visually and aurally and then return to the segment being studied.

Because each of the three channels had been recorded with a directional microphone, only part of the soundscape was recorded on each channel. This reduced the number of birds on each channel, and thereby reduced "overprint" of spectrographic traces, where one bird sings at the same time and frequency as another (Fig. 4).

Because the microphones were always oriented in the same directions (Table 2), the channels also encoded directional information, relative to the direction the vehicle was headed. The observers' field data maps for round two were similarly calibrated. The idea of the "map option" for delineating individuals was that individuals on the recordings could be synonymized with individuals on the maps by their bearing from the recording array. This would make it possible to identify which individuals on the map had been misidentified. For example, if the observer mapped Pine Warblers in directions 1 and 3, but the only trill on channel three of the recordings was deemed a Chipping Sparrow, one of the observer's Pine Warblers would be deemed a misidentification. In the data matrix, observer attributions that I deemed erroneous were flagged as such.

Similarity Indices

To assess the effect of observer, date, and other possible causitive factors on agreement in census results, I calculated similarity indices (Rempel et al. 2005) for all possible pairwise comparisons at each of the 18 stops. With two different dates for each of four observers, the full matrix was 8 x 8. Qualitative and quantitative Sorensen indices were used.

The Qualitatitive Soronsen Index (Southwood and Henderson 2000) measures agreement in presence/absence of species on a scale of 0 to 1. The formula (Southwood and Henderson 2000) is

$$C = 2a / (2a + b + c)$$

Where a is the number of species held in common between two dates, and b and c are the number of species found on only one of these two dates.

The Quantitative Sorensen Index was adapted from the Qualitative Sorensen by Bray and Curtis to include abundance data (Southwood and Henderson 2000). As two samples at the same place at different times (only a few days apart) would ideally record not only the same species roster but also the same number of individuals, this index is an appropriate measure for the null hypothesis that the samples are identical. The formula (Southwood and Henderson 2000) is

$$C = 2jN / (aN + bN)$$

Where aN is the total individuals sampled at time a, bN is the total sampled at time b, and jN is the sum of the lesser count for each species common to both times.

Detection Probability and its Components

Detection probability has been discussed at length in recent years (e.g., Buckland et al. 2001, Thompson 1992, Farnsworth et al. 2002, in press, McCallum in press). It is the product of "availability," the probability that an average bird, when present, will produce a cue, and "detectability," the probability that a cue will be detected by an observer (Farnsworth et al. 2002, McCallum in press). It is perhaps better to call this measure a "correction factor," as it also may include the probability that a bird that is resident in the count area is present when the survey is conducted (Bart and Earnst 2002).

When detection probability or its components, availability and detectability, are used to estimate N, they must be estimated from survey data with estimation models derived *a priori* from statistical theory. In this study, I estimated these parameters directly from raw data, as simple ratios, without the benefit of such theoretically-based estimation methods. Doing so is permissable, for two reasons. First, these estimates will not be used to estimate N. Second, the estimates will be used heuristically to evaluate the relative contributions of availability and detectability to detection probability, and thereby to judge the relative merits of different estimation models, which make different assumptions about these two components. Hereafter, all mention of these parameters refers to such heuristic estimates and uses. I will refer to detectability as "detection rate," and composite detection probability, P, as the "correction factor."

I estimated the true number of resident birds, i.e., N, of each species in this study as follows: I excluded birds detected only once per stop, as likely transients. I excluded birds detected twice at a stop, but only by one observer, as likely misidentified. (This procedure was independent of post-processing, and in fact did not lead to any disqualifications of data.) Otherwise, the maximum number of birds vocalizing at a stop, in any of the eight visits to that stop, was taken as the true number of birds at that stop. The expected number of birds in the data set was thus the maximum at each stop * the number of dates upon which stop was surveyed, summed over all stops in the data set. For the complete observer data set this translates to the sum of all maxima * 8 dates * 18 stops. The post-processed data set has fewer stops and dates (Table 3), but is augmented with post-processing data. Total expected birds is calculated accordingly.

The correction factor, P, is the ratio of total individuals detected by the observer to the total expected individuals. It was calculated for the full observer data set and for the post-processed data set. Availability was calculated, for the post-processed data set, as the ratio of total individuals detected by the observers and through post-processing to the total expected number of individuals, i.e., it had the same denominator as P. The numerator was the maximum of the count made by the observer and the count made in post-processing. Because P = availability * detectability, observer detection rate (detectability) could therefore be calculated from these results as P / availability. A raw estimate of detection rate was also calculated as the number of individuals found by the observer divided by the maximum found by either the observer or post-processing. This measure was independent of the estimate of number of birds resident at a stop.

RESULTS

A total of 2447 individuals of 77 species was detected by the observers. At least one bird was detected at all 144 stops. The total number of species-stops was 1647, i.e., if only presence-absence were compiled, this would be the sample size.

I studied a total of 62 date-stops spectrographically and aurally (56 listed in Table 3 plus 6 others). Observers detected 998 individuals and 736 total species at these stops. I detected 162 additional individuals, which added 111 species to the aggregate count of species for all 62 date-stops. For the five stops at which I studied all eight samples (Table 3), I detected an 119 individuals not counted by the observers, comprising 76 species-stops. In other words, post-processing revealed nearly two species not detected by observers at each stop on each day, on average.

QUESTION 1. WHAT PERCENTAGE OF BIRDS DETECTED BY REALTIME OBSERVERS WAS MISIDENTIFIED, AND HOW MUCH DID OBSERVERS VARY IN ACCURACY?

I identified 11 likely misidentifications out of the 998 individuals identified by observers in the post-processed data set. In some cases I searched repeatedly for species, particularly warblers, without success. I could not interpret these cases as misidentifications, because I could not find the vocalization that may have been misidentified. On the other hand, if misidentification had been rampant, it would have been apparent. In point of fact, with sufficient effort, e.g., one hour devoted to a three-min time segment, I could find almost every bird on the observer list.

QUESTION 2. ARE ALL BIRDS RESIDENT AT A STOP AVAILABLE FOR DETECTION DURING A THREE-MINUTE BBS STOP, AND IF NOT, WHAT PROPORTION OF THE POPULATION DOES EMIT A DETECTABLE AURAL CUE DURING SUCH A TIME PERIOD?

QUESTION 3. WHAT PERCENTAGE OF BIRDS AVAILABLE FOR DETECTION ON THE BBS IS NOT DETECTED (OR IGNORED), AND HOW MUCH DO DETECTION RATES VARY?

These questions are addressed together, because availability and detection rate are estimated from the same data. First, the level of agreement among repeated surveys of the same BBS route is addressed with similarity indices. Then, heuristic estimates of availability and detection rate are reported.

Similarity Indices

A matrix of indices (both Sorensen and Bray-Curtis) was calculated for each stop, then the 18 matrices were averaged. The grand mean Sorensen index was .506, with a minimum of .389 and maximum of .582 for the mean matrix. The grand mean Bray-Curtis index was .425, with a minimum of .336 and a maximum of .504 for the mean matrix. In both cases, the cell means were normally distributed.

These very great differences in birds detected on different days at the same stops suggest that not all birds present at a stop are detected during a three-min BBS survey. This interpretation is supported by numerous cases in which spectrographically distinctive song-types, presumably specific to individuals, were found on some but not all days at a stop.

Discrepancies could be attributable both to observer differences and to differences in the birds present and singing on different days. The similarity values for the same observer on two days should therefore be greater than those between observers. Table 4, displaying Bray-Curtis indices, shows that they were. (The quantitative and qualitative matrices were very highly correlated, see below)

Table 4. Matrix of mean Bray-Curtis (= "quantitative Sorensen") similarity indices for every possible comparison of eight different days. Each cell is the mean of separately calculated indices for each of the 18 stops. The Bray-Curtis index compares numbers as well as species count. The matrix of qualitative Sorensen indices is similar. Rows and columns are designated by a numeral for the observer, followed by a numeral indicating the first or second run of the route by that observer. The cells highlighted in yellow are comparisons of the same observer's results on different days.

	1-1	1-2	2-1	2-2	3-1	3-2	4-1
1-2	0.504						
2-1	0.374	0.369					
2-2	0.451	0.428	0.434				
3-1	0.438	0.395	0.336	0.416			
3-2	0.480	0.429	0.385	0.438	0.447		
4-1	0.473	0.467	0.408	0.453	0.431	0.415	
4-2	0.427	0.429	0.377	0.455	0.393	0.387	0.472

I regressed the 28 Bray-Curtis values on the matching Sorensen values. The regression was highly significant (F = 137.68, df = 1,26, P = .0001, and explained 84% of the variation. The very clean relationship between the quantitative and qualitative (presence-absence) indices is also indicated by the regression equation: Bray-Curtis = .755 * Sorensen + 0.043. The intercept, in fact, is not significantly different from 0. In other words, most of the discrepancy among the data sets is due to differences in the species detected, not the abundance estimates.

Average similarity in the various counts at a stop increased when observer data were augmented with acoustic detections from multi-channel recordings (Table 5). Because observer differences should have been reduced in the augmented data set, the intra-observer indices are not expected to exceed others. This is borne out by Table 5.

The difference in mean similarity index between the augmented data and aural data collected in the field for the same stops was highly signficant (paired comparison t-test) for both Sorensen (t = 6.88, df = 27, P = 0.0001) and Bray-Curtis indices (t = 7.51, df = 27, P = 0.0001).

Table 5. Matrix of mean Bray-Curtis similarity for augmented data. See Table 1 for organization of table. Values below the diagnonal are indices based on augmented data for all days at five stops. Notice that they are consistently higher than comparable cells in Table 4. This means that detecting additional birds improved agreement between days, which implied that the additional birds were residents that were detected on other days, not totally new finds for those stops, in which case the indices would have diminished. This contention is supported by the similarity of song types among days. Sorensen indices were similar but higher. Above the diagonal are the numeric differences between the indices from augmented data and strictly comparable indices from unaugmented data, i.e., based on aural detections at the same five stops. These differences are highly significant (see text). Differences > 0.1 are highlighted. Note that four of five highlighted cells pertain to the data for observer-day 2-1, i.e., the first round for observer 2, and that all differences for 2-1 > 0.07. This suggests a relatively lower detection rate on that day than on other days of the study.

	1-1	1-2	2-1	2-2	3-1	3-2	4-1	4-2
1-1		0.039	0.119	0.082	0.054	0.042	0.000	0.050
1-2	0.527		0.120	0.050	0.019	0.078	-0.007	0.058
2-1	0.449	0.404		0.090	0.118	0.087	0.073	0.114
2-2	0.535	0.479	0.463		0.129	0.030	0.092	0.153
3-1	0.540	0.404	0.399	0.533		0.040	0.006	0.042
3-2	0.548	0.541	0.423	0.476	0.450		0.004	0.030
4-1	0.458	0.452	0.370	0.458	0.440	0.391		0.022
4-2	0.464	0.436	0.406	0.491	0.503	0.462	0.453	

Although the per-stop assessment of similarity is appropriate for assessing the amount of daily variability in the results obtainable with the BBS protocol, BBS data are usually analyzed by pooling over all stops. If the data from this study are pooled over stops, the agreement among days is expected to improve. It does, as shown in the Table 6, the matrix of similarity values of pooled data for each day. This means of course that misses on different days are different but that they are somewhat offsetting when pooled over stops. It does not mean that any of these counts is complete.

Table 6. Matrix of Bray-Curtis (= quantitative Sorensen) similarity indices for every possible comparison of eight different days. Each cell is based on data pooled from all 18 stops. The Bray-Curtis index compares numbers as well as species count. The matrix of qualitative Sorensen indices is similar. Rows and columns are designated by a numeral for the observer, followed by a numeral indicating the first or second run of the route by that observer. The cells highlighted in yellow are comparisons of the same observer's results on different days.

	1-1	1-2	2-1	2-2	3-1	3-2	4-1
1-2	0.769						
2-1	0.675	0.631					
2-2	0.778	0.723	0.682				
3-1	0.766	0.687	0.670	0.721			
3-2	0.763	0.735	0.686	0.745	0.764		

4-1	0.767	0.740	0.735	0.763	0.741	0.749	
4-2	0.710	0.682	0.723	0.743	0.677	0.685	0.804

Estimates of Detection Probabilities

Species-specific correction factors, P, based on estimates of residency and the observers' counts, were similar for the raw data set of 18 stops and for the post-processed data set of nine stops. The latter data are presented in Table 7. Availability, estimated from counts augmented with post-processing data and the estimates of residency, was < 0.5 for most species. Detection rate exceeded availability in every species but one (Prairie Warbler, in which both probabilties are high). Indeed, detection rate doubled availability in the average species. Overall detection rates for the four observers were 0.760, 0.836, 0.900, and 0.905, all well above most availability rates.

Table 7. Estimates of detection parameters. Two estimates of each parameter were based upon (1) the five stops for which I assessed all eight visits ("8 only") amd (2) those five stops plus four more at which I assessed the second visit of each observer ("8 & 4"). The "Raw Data" estimate of detection rate is not limited to individuals deemed to be resident (see Methods).

Species	Detection Rate			Availability		Correction Factor, P	
	Raw Data	8 & 4	8 only	8 & 4	8 only	8 & 4	8 only
ACFL	0.927	0.950	0.920	0.455	0.391	0.432	0.359
AMCR	0.934	0.934	0.976	0.412	0.375	0.385	0.366
AMGO	0.941	0.937	1.000	0.381	0.347	0.357	0.347
AMRE	0.900	0.900	0.833	0.313	0.250	0.281	0.208
AMRO	0.900	0.750	0.500	0.333	0.250	0.250	0.125
BEKI	0.500	0.500	0.500	0.250	0.250	0.125	0.125
BGGN	0.778	0.771	0.720	0.417	0.391	0.321	0.281
BHCO	0.783	0.839	0.882	0.383	0.354	0.321	0.313
BLGR	0.714	0.684	0.625	0.594	0.500	0.406	0.313
CACH	0.917	0.906	0.889	0.320	0.307	0.290	0.273
CARW	0.839	0.836	0.833	0.444	0.375	0.371	0.313
CEDW	0.786	0.769	0.769	0.325	0.325	0.250	0.250
CHSP	1.000	1.000	1.000	0.313	0.313	0.313	0.313
COGR	0.889	1.000	1.000	0.250	0.250	0.250	0.250
COYE	0.917	0.943	0.909	0.486	0.458	0.458	0.417
DOWO	0.762	0.750	0.786	0.417	0.350	0.313	0.275
EABL	0.667	0.750	0.500	0.333	0.250	0.250	0.125
EAKI	0.895	0.941	0.929	0.425	0.438	0.400	0.406
EATO	0.842	0.833	0.833	0.375	0.375	0.313	0.313
EAWP	0.611	0.583	0.583	0.500	0.500	0.292	0.292
FISP	0.882	0.882	0.882	0.531	0.531	0.469	0.469
GRCA	0.700	0.857	0.857	0.438	0.438	0.375	0.375
HOWA	0.800	0.800	0.800	0.313	0.313	0.250	0.250
INBU	0.885	0.882	0.889	0.637	0.625	0.562	0.556

MODO	1.000	1.000	1.000	0.361	0.313	0.361	0.313
NOCA	0.800	0.792	0.758	0.480	0.458	0.380	0.347
NOPA	0.500	0.462	0.462	0.406	0.406	0.188	0.188
OVEN	0.857	0.833	0.833	0.500	0.500	0.417	0.417
PIWA	0.882	0.882	0.882	0.531	0.531	0.469	0.469
PIWO	0.875	1.000	1.000	0.188	0.188	0.188	0.188
PRAW	0.714	0.667	0.667	0.750	0.750	0.500	0.500
RBWO	1.000	1.000	1.000	0.400	0.375	0.400	0.375
REVI	0.921	0.921	0.880	0.543	0.481	0.500	0.423
RTHU	0.750	0.500	0.500	0.250	0.250	0.125	0.125
RWBL	1.000	1.000	1.000	0.472	0.375	0.472	0.375
SCTA	0.895	0.882	0.882	0.354	0.354	0.313	0.313
SUTA	0.500	0.667	0.667	0.375	0.375	0.250	0.250
TUTI	0.768	0.820	0.775	0.474	0.455	0.389	0.352
WBNU	1.000	1.000	1.000	0.250	0.188	0.250	0.188
WEVI	0.636	0.600	0.611	0.556	0.562	0.333	0.344
WOTH	0.927	0.921	0.909	0.475	0.458	0.438	0.417
YBCH	0.920	0.909	0.909	0.687	0.687	0.625	0.625
YBCU	0.857	1.000	1.000	0.250	0.250	0.250	0.250
YTVI	0.773	0.879	0.729	0.350	0.300	0.308	0.219

DISCUSSION

POST-PROCESSING METHODS

Three-channel visual searching for species has both strengths and liabilities. Sound is “spread out” on both the temporal and frequency axes. Although the human auditory system is manifestly an extraordinarily fine frequency analysis device, spectrograms perform this function at least as well. Moreover, because time is frozen in a spectrogram, simultaneous frequencies can be examined in turn, a definite advantage over audition. This not only permits discrimination of different individuals that are vocalizing simultaneously, it reveals the spectral structure of individual sounds, which may be diagnostic of a species. My analytical setup allowed me to see or hear any sound. Typically I scanned a 10-s segment of sound visually, and listened to either the entire 10-s segment of one channel, or a shorter segment of sound, if it was visually interesting. After some training, it was not necessary to listen to most sounds to identify them to species and often to individual.

Distinguishing individuals is one of the important and heretofore unheralded strengths of spectrographic analysis. In many species individuals have their own repertoires of one or a few song types. Song-types are much easier to distinguish visually than aurally (pers. obs.). Thus, even when singing in the same channel, individuals can often be distinguished, as long as their signal are clear, by differences in amplitude and song-type.

Some species are especially easy to detect visually, even when embedded in other sounds. The Eastern Wood-Pewee’s song, with its long continuous “whistle,” is unlike any other sound one is likely to hear. I found a greater percentage of undetected individuals of this species than any other. As many of these birds seemed vocally conspicuous on the recordings, it may be that it is easy to overlook for auditioners as well as being conspicuous spectrographically. Another example is the Blue Grosbeak, which is easy to detect spectrographically because its song is confined to a narrow frequency band and the notes are packed together temporally.

Call notes have great potential because their fine structure, which is quite variable among species, is much more apparent on a spectrogram than aurally in real time. As it is unclear how many determinations were made by observers on the basis of call notes, it is impossible to evaluate the utility of call notes with this data set, but the potential for increasing detection of species through examination of the fine structure of call notes is significant.

The liabilities of post-processing are, like the strengths, related to physical factors. The main factor is the loss of directional information in even a multi-channel recording. As a rule of thumb, one probably needs the same number of independent channels as the number of individuals present to detect all the singing birds on a recording. With Blue Grosbeak and Indigo Bunting I often found more individuals than the realtime observer, but these species have spectrographically distinctive songs and seldom exceeded three in number. Red-eyed Vireo and Wood Thrush often exceeded three in number, and I was unable to distinguish that many individuals on the recordings. It may be that multi-channel playback and audition would reveal all the individuals. Increasing the number of channels examined visually would not solve the problem unless directional microphones with narrower “field of view” were also used.

SIMILARITY INDICES

Similarity indices are a good method for quantifying the degree of difference between samples when the expectation is that there is no difference, i.e., the null hypothesis is a matix of ones. Their utility is greater for diagnosis than analysis. It is interesting that indices increased when additional individuals and species were detected. This is not necessary. If the additional detections were transients, randomly distributed among dates, the indices would have decreased. The fact that they increased suggests that the additional birds were filling holes in the data matrix, i.e., they were individuals detected by observers on other days. The similarity of song-types, particularly of vireos, Indigo Buntings, and Eastern Towhees, among birds detected by observers and those missed by observers at the same stop, supports this interpretation.

Table 4 shows the striking degree to which single point counts under-estimate presence and abundance. Table 6 demonstrated clearly that even with the statistical value of larger numbers, routes run on different days, even within the same week, differ greatly in results. The improvement obtained by pooling data suggests that increasing the number of stops analyzed up to 25, the maximum possible for this data set, might further increase agreement. Because the BBS has twice that many stops, it remains possible that the effect of different days demonstrated here is reduced to negligible levels with full-length routes. The decrease in detectability of most species in the second half of a standard BBS route may, however, largely cancel out the effect of larger sample size.

Agreement of pooled data, however, is only a statistical effect. If the indices approach 1 as number of pooled stops increases, it is because biases in the stop data, which must be stochastic, cancel out. It means all samples are equally biased. Precision increases as the number of pooled stops increases, not accuracy. It is detection probability not similarity among samples that matters.

QUESTION 1. WHAT PERCENTAGE OF BIRDS DETECTED BY REALTIME OBSERVERS WAS MISIDENTIFIED, AND HOW MUCH DID OBSERVERS VARY IN ACCURACY?

Post-processing of concurrent recording has high potential for estimation of rates of misidentification, if the basis upon which the observer made the identification is identified. The map method was not sufficient to this purpose. In practice, locations on the field maps often did not correspond with directions indicated by the three-channel spectrographic display. Directionality is indicated aurally and spectrographically by amplitude and clarity. Amplitude (roughly, loudness) is indicated on a spectrogram by the darkness of the trace. It is possible, for technical reasons, that apparent amplitude is misleading, but clarity, the sharpness of the spectrographic trace, is not likely to be biased.

Correspondingly, there are several sources of potential error in an observer’s estimation of bearing, including inequality of auditory acuity of the two ears and the well-known difficulty of determining distance and direction to a sound overhead. Regardless of cause, the map option did not make it possible to synonymize many individual birds, and so identification of misidentifications was possible in only a few special cases.

Estimating misidentification rates from tapes is unlikely to be successful unless the observer synonymizes specific sounds with the identification, i.e., a method such as “voice-over” or “dictation” (see Methods). This need not be done for every sound or even every bird. It can be done with a sample, as long as the sample is not biased toward the more easily identified nearby birds. With distant birds, detectability influences accuracy, because the signal is attenuated. This is an area that needs more theoretical study.

QUESTION 2. ARE ALL BIRDS RESIDENT AT A STOP AVAILABLE FOR DETECTION DURING A THREE-MINUTE BBS STOP, AND IF NOT, WHAT PROPORTION OF THE POPULATION DOES EMIT A DETECTABLE AURAL CUE DURING SUCH A TIME PERIOD?

The validity of the estimates of availability rests on the validity of the method used to estimate N from survey data. The maximum number of birds counted for a species at a stop on all eight visits was taken as N for that stop. In many cases the maximum number was achieved on more than one occasion, so is likely to be accurate. The most common maximum was 1, so statistically, the estimates of N are not likely to be biased positively. If these estimates are biased negatively, i.e., N was underestimated, it only decreases the estimates of availability and hence strenghthens the arguments that follow.

As this analysis is based entirely on sounds, the number of detected birds is much more likely to be the number of males present than the number of birds present. This is a problem with all point count data, so no new issues arise here. A few counts were perhaps increased by calling females, but this is not likely to have been a major source of bias. Counts of American Crows, Carolina Chickadees, and Tufted Titmice were swollen by vocal juveniles, so the analysis is not likely to be dependable for these species.

The values for availability are lower than those for detection rate, suggesting that the major concern for developers of estimation methods is a parameter that was not recognized in print until 2001 (Buckland et al. 2001) and not defined formally until 2002 (Farnsworth et al. 2002). All of the estimation methods currently available, except Farnsworth et al. (2002), focus on the component referred to here as detection rate.

Independent confirmation that availability is often low comes from several sources. Daw and Ambrose (2003) found a number of species on recordings that were not detected with conventional point counts, showing that not all birds present are active during a point count. Scott et al. (in press) showed that a Least Bell’s Vireo was silent during 33% of 1691 minutes, despite singing up to 22 songs per minute at times. The well-known fall-off in bird activity in late morning and especially in mid-afternoon is ample evidence that birds are not always active, and hence detectable.

QUESTION 3. WHAT PERCENTAGE OF BIRDS AVAILABLE FOR DETECTION ON THE BBS IS NOT DETECTED (OR IGNORED), AND HOW MUCH DO DETECTION RATES VARY?

Several assumptions were necessary to estimate detection rate, and they will be discussed here. I assumed that all aural detections recorded by the observers were correct, except for the 11 I excluded as likely errors. Some of the remainder undoubtedly were erroneous, but it is impossible to be sure given the methods observers used to record their data in the field. The most likely candidates for errors are those species I was unable to find on the recordings. Because it is possible that such detections were based on calls I didn’t recognize (quite possible) or sounds not present on the recordings (unlikely), I took the conservative approach of letting all of them stand. The number of identification errors is quite likely to be negligibly low, though, because I was able to find almost all observer species and individuals (up to two or three) with sufficient effort.

I may have misidentified a few birds that augmented the count for a stop, but after tabulating low-confidence identifications along with confident ones, I excluded the former from all analyses. The additions I made are therefore conservative, and the bias in my data is much more likely to be negative than positive. It follows that my estimates of the observers’ detection rates are conservative, i.e., the observers likely did no better and probably worse than these estimates suggest. The birds I missed, though, were mostly near the limit of detection, which is after all observer-specific. I omitted a number of songs that were visible on the spectrogram but too fuzzy to identify visually and inaudible upon playback. I omitted some calls that I could not identify because of lack of diagnostic features, despite studying fine structure of many calls and comparing them to voucher specimens from published sources. Many BBS observers may simply omit sounds they cannot identify, while others may misidentify some of the call notes they hear.

With these assumptions clear, I can say confidently that detection rate of the observers in this study was very high. The surprising and unexplained exceptions to this generality are Eastern Wood-Pewee, Blue Grosbeak, Northern Parula, and White-eyed Vireo. Other low values in the table may be due to small sample size. The wood-pewee had the lowest detection rate, despite many of the missed sounds being clearly audible on the tape. It may be that my detection rate was especially high for these species, because they all have spectrographically distinctive songs that are very easy to pick out, even when faint. This seems unlikely, because other species with spectrographically-conspicuous sounds, e.g., Acadian Flycatcher and American Crow, had detection rates approximating 0.9.

Some missed birds were so conspicuous that they must have been heard. Either they were not noticed because of attention to other birds, or they were noticed but forgotten before they were written down. An unfortunate oversight in the methodology of observer data collection was that data were recorded on blank sheets rather than BBS field data sheets on seven of the eight days of this study. Official BBS sheets have species lists that may serve to jog the memory of the observer recording observations. The contribution, if any, of the use of blank sheets to missed birds cannot be assessed.

The observers differed somewhat in detection rate. These differences probably explain why the improvement in the similarity indices after post-processing (Table 5) was uneven.

RECOMMENDATIONS

The post-processing performed for this study was laborious and time-consuming. It would not be feasible to post-process BBS data in this way. Fortunately, the results of this effort suggest that observer error, both misidentification and nondetection, is not the major source of bias in BBS data. Rather, it is bird behavior itself that is the major reason birds that are present are not detected. Availability, heuristically estimated, was consistently lower than detection rate, and averaged half the value of detection rate. Resources should therefore be directed toward estimation of availability.

The best method currently available for this is the removal method of Farnsworth et al. (2002). Availability is estimated by dividing the sample period into two or more segments, and noting which birds are detected in either or both. This can be done in the field, but probably at the cost of a lowered detection rate for other individuals. It also can be done with recordings, preferably with the enhanced recognition of individuals afforded by spectrographic visualization, provided a model can be devised that estimates availability from a subset of individuals.

Detection rate varies among individuals partly because of differences in auditory acuity (Emlen and DeJong 1981). This is a function of distance to vocalizing birds. The distance method (Buckland et al. 2001) can be used to estimate a correction factor that should compensate for observer differences in hearing while also correcting for the fall-off in detectability with distance to bird. But, I found numerous birds that were singing conspicuously that were not noted by the observers. Such misses are the purview of the Double Observer method (Nichols et al 2000). Post-processing of a tape is functionally equivalent to having a second observer record misses, without several of the logistical shortcomings of that method. Observer-made tapes may be a viable way to estimate a correction factor for conspicuous singers that were missed. (See forthcoming report on observer tapes.)

Acknowledgments

I would like to thank Keith Pardieck and Bruce Peterjohn for their able administration of this contract, and Mary Gustafson, Sam Droege, Keith Pardieck and Bruce Peterjohn for participating as observers in this study, and Glenn Johnson for meticulously transcribing the somewhat inscrutable field data sheets.

LITERATURE CITED

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: Estimating abundance of biological populations. Oxford Univ. Press, Oxford, UK.

Daw, S., and S. Ambrose. 2003. Using digital recordings to supplement avian surveys, Canyonlands National Park, 2002. USDI, National Park Service, Southeast Utah Group, unpublished final report.

Emlen, J. T., and M. J. DeJong. 1981. The application of song detection threshold distance to census operations. Studies in Avian Biology No. 6:346-352

Farnsworth, G.L., K.H. Pollock, J.D. Nichols, T.R. Simons, J.E. Hines, and J.R. Sauer. 2002. A removal model for estimating detection probabilities from point-count surveys. Auk 119:414-425.

Farnsworth, G.L., J.D. Nichols, J.R. Sauer, S. G. Fancy, K.H. Pollock, S.A.Shriner, and T.R. Simons. In Press. Statistical approaches to the analysis of point count data: A little extra information can go a long way. *In* Ralph, C.J., and T.D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PNW-GTR-191.

McCallum, D. A. In Press. A conceptual guide to detection probability for point counts and other count-based survey methods. *In* Ralph, C.J., and T.D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PNW-GTR-191.

Rempel, R. S., K. A. Hobson, G. Holborn, S. L. Van Wilgenburg, and J.Elliott. 2005. Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory. Journal of Field Ornithology 76:1–11.

Scott, T. A., P.-Y. Li, G. C. Greene, and D. A. McCallum. In Press. Singing rate and detection probability: An example from the Least Bell's Vireo (*Vireo belli pusillus*). *In* Ralph, C.J., and T.D. Rich (Eds.). Bird Conservation Implementation and Integration in the Americas. USDA Forest Service Gen. Tech. Rep. PNW-GTR-191.

Southwood, T. R. E., and P. R. Henderson. 2000. Ecological methods. Blackwell Science, Oxford.

Thompson, W. T. 2002. Toward reliable bird surveys: Accounting for individuals present but not detected. Auk 119:18-25.

FIGURES

Figure 1. Field data collection for this study. The scene here is typical of the BBS, except for the antenna array (Fig. 3) attached to the rear left side of the vehicle.

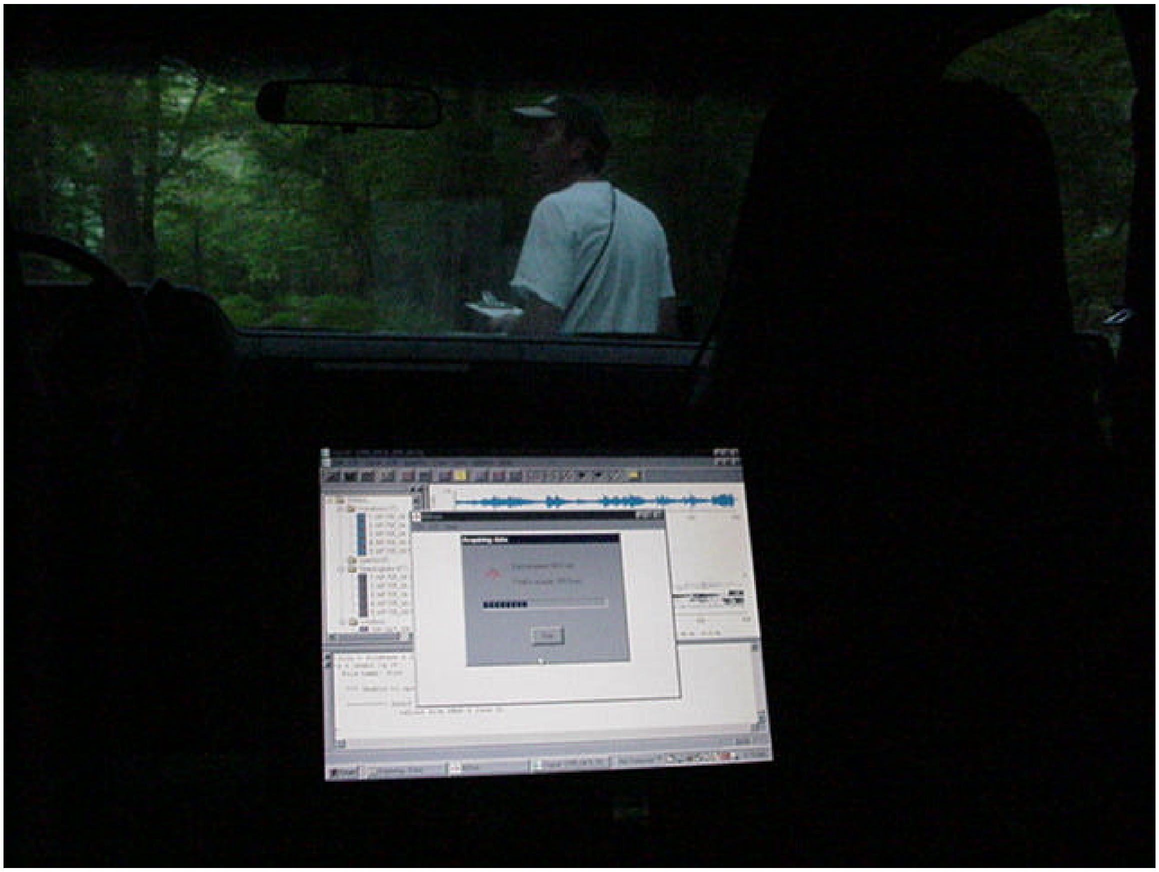

Figure 2. View of laptop computer in operation inside the vehicle while observer conducts three-minute survey. Sounds from the antenna array (Fig. 3) were routed by cables to a junction box and thence to the computer, which was capable of digitally recording up to eight channels simultaneously.

Figure 3. The antenna array used to capture ambient sounds during BBS surveys. See Table 2 for description of microphones. Most important are the three long, thin directional microphones pointing away from the center with an angular separation of 120 degrees. Because such microphones drop off in sensitivity away from their axis of orientation, the three channels recorded from these microphones emphasized different sections of the soundscape, facilitating localization of the sound sources during post-processing of recordings.

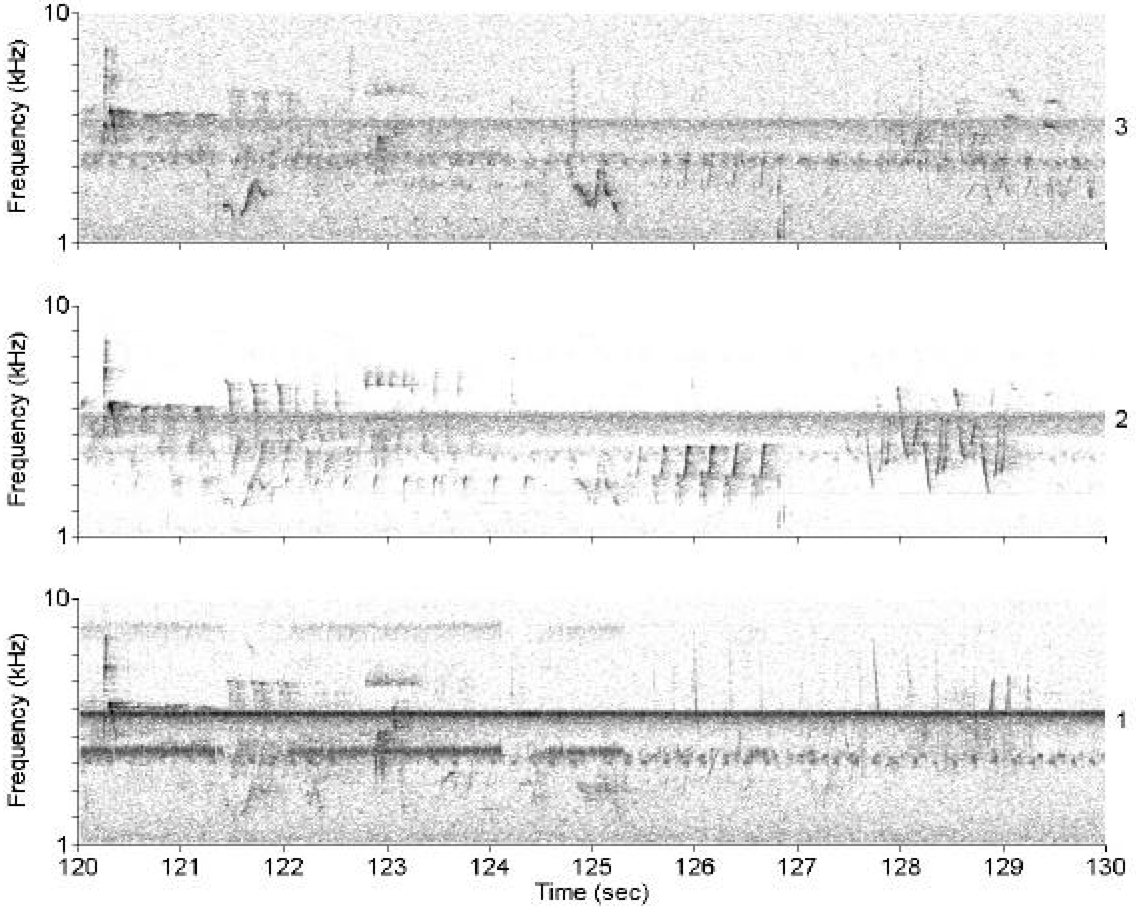

Figure 4. Example of a SIGNAL^{tm} screen display used to evaluate three-channel recordings of the BBS soundscape. The spectograms are numbered (right side) according to the directional microphone used to record the channel. Sounds emanating from overhead or close to the microphones appear on all channels (e.g., the Eastern Kingbird sound at the beginning of the segment). Others are localized. For example, the White-eyed Vireo at seconds 122 and 125 of channel 3 is faint on channel 2 and fuzzy on channel 1. This indicates that it is between microphones 2 and 3, nearly in line with 3. The Common Yellowthroat, clear near the end of channel 2, is barely visible on the other channels. Northern Cardinal is clearly visible on channel 2 and Acadian Flycatcher is at 123 on channels 1 and 3. An Indigo Bunting is faint at 121ff. on channel 2. Other sounds are visible but faint. Some can be identified by shape or by listening repeatedly to the sound track. Background insect noise includes a constant, loud drone at 6 kHz (most evident on channel 1) and most of the sounds at 4.5 kHz on channel 1.