Using Observer-recorded Tapes to Replicate a BBS Survey: Agreement with Realtime Observer Results

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

Preliminary 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, observer effects, temporal variation, tape recording, post-processing.

TABLE OF CONTENTS

ABSTRACT

The North American Breeding Bird Survey (BBS) is an important tool for monitoring the health of territorial landbird populations. Although the sampling protocol is well-designed to produce consistency of results across years and observers, improvements that (1) reduce identification errors and/or (2) increase detection ratio of vocalizing birds would be welcome. Post-processing of observer-recorded audio-tapes is one technique with the potential to improve results in one or both of these ways. To evaluate the potential of this technique, an experienced point-counter was engaged to post-process tapes recorded by BBS observers on four replicates of a simulated BBS route at Patuxent National Wildlife Research Center in Maryland, USA. The interpreter used three protocols, in the following sequence: (1) Simply listening to the tape one time through without pausing, (2) Visually observing a realtime (moving) spectrogram while listening to the tape one time through without pausing, and (3) Unlimited audition and spectrographic visualization of the tape, with pausing and rewinding. The analysis for a stop-day was done at one sitting, and putative individual birds were synonymized across protocols. The count of putative individuals detected in this way was 920. Agreement between the first and final passes (Protocols 1 and 3) was 0.626, calculated as the proportion of individuals identified to species in the third pass also identified in the first pass. Similarly, agreement between passes 2 and 3 was 0.740. The results suggest that post-processing should not be limited to a single pass through the data, even with spectrographic visualization.

The results of post-processing were compared with those obtained by the realtime observers. Comparison of total individuals detected, by species, suggested that post-processing detected most of the birds within 50 m of the observer's location, plus some farther away, but failed to detect the more distant individuals. When a more sensitive analysis was conducted, it showed some to substantial disagreement between post-processing and the observers' data. Overall, the post-processing interpreter missed 54.6% of the individuals detected by observers, while the observers missed 25.9% of the individuals detected by the interpreter. Decomposition of these discrepancies into misidentifications and nondetections is possible with further, independent, analysis of audio tapes. Overall, the study so far suggests that low-technology post-processing of short samples of recorded sound is a poor substitute for a well-trained observer in the field. This does not mean that recordings are a totally inadequate substitute for a live observer. It does suggest that recordings are likely to be more useful for protocols other than rapid counts of individuals.

INTRODUCTION

Assessing the conservation status of wildlife populations requires monitoring. For territorial land birds in temperate North America the joint U.S.-Canadian Breeding Bird Survey (BBS) is the single most important source of information on population trends. Though an extremely valuable program, the BBS suffers from several problems (O'Conner et al. 2000), including the following ones intrinsic to point transects in general (Buckland et al. 2001), i.e., transects of "point counts" (Ralph et al. 1995, Huff et al. 2000) :

1. The BBS relies on short 3-min point counts for all data. The potential for non-detection of some of the birds during this 3-min period is high. Incorporating a correction factor for undetected birds would greatly increase the reliability of these data.
2. The BBS is a largely volunteer effort. Although volunteers are typically seasoned amateur birders, if not professionals, variability among observers in terms of experience on the route, auditory acuity, and knowledge of bird song is probably high. A method for adjusting for these differences would greatly increase the comparability of results.
3. The BBS, like most other counting techniques used with land birds, relies heavily on vocalizations for species identification. It is probably fair to say that almost all field ornithologists are more proficient at visual identification than auditory identification. Some very common birds that are very easy to distinguish visually can sound remarkably similar at times. Observers probably rely on other cues, such as habitat type, to identify many of the birds they hear on BBS routes, which may increase the probability of correctness, but can never be definitive. Indeed, such practices make the data useless for assessment of species/habitat relationships.
4. Each BBS route is run from slightly before dawn until the 50 3-min counts are completed. Over this span of time the soundscape has gone from the cacophony of the dawn chorus, during which it is difficult to hear all the singers, to the almost eerie quietness of mid-morning. As a result of its long duration, the BBS is subject also to a decrease in detection probability that is correlated with the linear arrangement of the stops.

All these problems exist for all rapid, voice-based counting schemes, so solutions developed for the BBS would be easily and profitably transferred to other protocols. This study investigates the use of concurrent sound recordings to improve the accuracy and precision of the BBS.

Recording sounds encountered in censusing and monitoring schemes presents at least the following advantages:

1. Voucher specimens (acoustic) for hard-to-identify and/or rare species.
2. Correction of false positives and false negatives through re-listening and inspection of spectrograms.

3. Correction of miscounts through visual examination of spectrograms.
4. Potential for sampling over a larger time window, and therefore potential for sampling at more nearly optimal times of day.
5. Possible use of observers who are not expert in field identification of vocalizations to conduct counts (make recordings)(Rempel et al. 2005).
6. Possibility of remote monitoring schemes.

It is often complained that recordings cannot be used to estimate numbers of individuals present, only presence-absence. This is not strictly true. Individuality is encoded for the realtime auditioner by directionality, variation in amplitude, and variation in song-type. Variation in amplitude and song-type are both recorded faithfully on recording media, and the information on individuality they encode is available in multiple passes after the fact, which can offset to considerable degree the loss of information from directionality.

Recordings and post-processing are now used with increasing frequency (e.g., Haselmayer and Quinn 2000, Daw and Ambrose 2003, Scott et al. in press, and especially Rempel et al. 2005; see Gaunt and McCallum 2004 for review of pre-2003 literature), but direct comparisons of the results obtained from recordings and realtime counting of the same soundscape are few. Haselmayer and Quinn (2000) conducted standard point counts in Peruvian Amazonian rainforest, and simultaneously recorded the soundscape with a rotating directional microphone. In their study the interpreter of the recording listened noted presence-absence. They did not attempt to count individuals. Individual counts in tropical rain forest tend to be low anyway, so probably little information is lost. Daw and Ambrose (2003) compared the species count from longplay recording vs. standard point counts. Here the contrast was the much longer coverage available with recordings compared to point counts, not the efficiency and accuracy of post processing vs. realtime audition for brief surveys. Not surprisingly the longplay recordings produced many more species than the short-duration realtime surveys, even with visual detection. Rempel et al. (2005) evaluated consistency of results from post-processing by audition (no spectrographic visualization) on their data set of recordings. They have not conducted conventional point counts or intensive surveys (*sensu* Bart and Earnst 2002) at their recording sites that might be used to validate results obtainable through post-processing.

In the study of which this report is a part, several axes of variation in survey results were examined. Conducting simulated BBS surveys on eight (8) days permitted an assessment of the effect of random variation in availability on results, an analysis like that of Daw and Ambrose (2003). Single channel omnidirectional recordings permitted analyses like those of Haselmeyer and Quinn (2000) and Rempel et al. (2005). Multichannel directional recordings permitted usage of all four indicators of individuality, and also permitted checking of the accuracy of species identification and individual counts in the other data sets.

This report treats the usage of single-channel, omnidirectional recordings for estimation of presence-absence and abundance by post-processing. I address the accuracy and detection rate when data are taken from recordings, compared to data taken in the traditional way, in real time in the field. The design permitted assessment of the contribution of multiple revisitations of the data, as well as spectrographic visualization, to the species and individual count. In a forthcoming report, the accuracy of these counts will be addressed.

METHODS

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 the Wildlife Loop. Thence the route followed the Wildlife Loop eastward to Bailey Bridge, over the Little Patuxent River. The route continued 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 the Wildlife 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.

DATA COLLECTION AND SOUND RECORDING

Four observers participated in the study, each running the route on two separate days, as follows:

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

Observers wrote their data on blank paper on the first iteration. On the second interation they mapped all birds relative to the listening point. The standardized maps they used were inscribed with a circle, which represented the locus of a 50-min radius from the listening point. Estimated distance and direction to each detection were thereby recorded. The observer indicated for each individual on the map whether it was detected aurally or visually. In this study, I used only the aural detections, because that is the only sensory modality available on audio-recordings.

Each observer, in addition to collecting standard BBS data, made two audio recordings on cassette tape at each stop. Two recordings were made to compare the results from microphones of two quality levels. The recorders sat on the passenger seat of the observer's vehicle. Mic cables ran through the passenger window to the microphones, which were taped onto the radio antenna, just in front of the passenger window. At each stop, after preparing for conventional data collection, the observer started recording on both decks, spoke "start" as well as the time of day, date, and stop number, and then spent 3 minutes conducting the normal survey. At the end of the three minutes, the observer spoke "stop," then stopped the recorders.

DATA PROCESSING

I transcribed their data to computer files, using annotations made at the time of collection to assign each detection to the following categories: song, call, juvenile begging, nonvocal sound, and visual. For the second iteration by each observer, all five categories were subdivided into within-circle (i.e., the estimated distance was less than or equal to 50 m) and outside circle data. Data were collated in Excel. All quantitative analyses were performed in SAS. The second iteration for each observer was chosen for post-processing because it allowed an assessment of the effect of distance to bird on detection via post-processing. For this study I excluded the visually-based data. If a bird was detected both visually and aurally, it was so indicated on the field data sheet by the observer. Such birds were included in the data set for this study on the basis of the aural detection.

POST-PROCESSING

An experienced point-counter,Glenn Johnson of Eugene, Oregon, performed the post-processing. He understands point-count methodology and is expert in identification of western North American species by sound, but previously had little experience with Eastern birds. Before interpreting the recordings, he trained with recordings of the species found on the route by listening to commercial recordings and simultaneously viewing spectrograms of these sounds.

The interpreter used the original tapes recorded by the observers at Patuxent. Quality of some tapes was poor, so he was given the option of using the better of the two tapes for each stop. This made comparison of performance of the two microphones impossible, but the microphone used for interpretation of each stop was noted in the data base, making it possible to compare overall performance, although not on the same samples.

Each 30-min observer tape contained nine stops. The entire set of tapes was randomly ordered, and an entire tape was interpreted before moving on to the next one. Interpretation was performed in late 2002 and early 2003. Spectrographic viewing was performed with Syrinx, a free program available at that time. Interpretation proceeded as follows:

1. Load observer tapes for GE1 and GE2 in L and R sides of Onkyo cassette deck. Review tapes for quality. Select better tape.
2. Load selected tape in Sony TCD5-ProII deck.
3. Play one stop (3min) audio only, straight through. Fill in column 1 during / after playback.
4. Rewind tape to beginning of current stop. Play straight through with Syrinx. Fill in column 2 during / after playback.
5. Save Syrinx buffer to disk. View/listen to any portion of tape as much as desired. Fill in column 3.
6. Go on to next stop.
7. Go on to next tape.

Columns 1, 2, and 3, mentioned above, will be referred to hereafter as Pass 1, Pass 2, and Pass 3, for the first, second, and third passes through the tapes. The data sheet for interpretation was similar to a BBS data sheet. Species were on the rows, and the number of individuals detected for each species was written in the column for Pass 1-3. When an individual was reinterpreted in a later pass as another species, an arrow was drawn connecting the cells on the data sheet. This kind of information cannot be recorded in a spreadsheet, so the following approach was taken to preserving this very important information. Each putative individual in the data set was placed on a different line in the spread sheet. The estimated specific identity of the individual was placed in the column appropriate for the Pass. The data were tallied in SAS.

DATA ANALYSIS

I considered using similarity indices from community ecology to compare the results of the five estimates that were made for each observer at each stop. But, although this approach is useful when comparing the results of different observers at the same stop, for example, it is not the best way to compare these data. The two samples, or communities, compared with a similarity index are assumed to be of equal rank, so a matrix of comparisons would be identical on the two sides of the diagonal. In the present study, the goal is to compare the performance of one protocol with that of another, which is considered a criterion. A matrix of such indices would not be symmetrical about the diagonal.

For each species at each stop on each date of the field work (i.e., the four separate runs by the observers), I tabulated whether a species detected in the criterion sample was also detected in every other sample. When a species was detected I calculated the proportion of individuals found in the criterion sample that were counted in every other sample. This procedure resulted in a 5 x 5 matrix of similarity values for each stop, species, and observer.

RESULTS

The interpreter detected a total of 920 birds by voice at the first 18 stops on the observer tapes. Total detections increased from 613 on the first pass to 638 on the second pass to 788 on the final pass. Fourteen percent of total detections were not confirmed in the final pass, and may have represented false positives on earlier passes, i.e. vocalizations that belonged to other detected individuals rather than different individuals. Of the 788 birds detected on the final pass, 476, 60.4% of the total, were detected on all three passes.

COMPARISON OF POST-PROCESSING RESULTS

Sixty-five (8.3%) of the 779 birds detected on the final pass could not be identified by the interpreter. If the identity ascribed to a vocalizing bird in the third pass is taken as correct, then the percent correctly identified on the first pass was 62.6% and the percent correctly identified on the second pass was 74.0 %. These results are subdivided by species in Table 1.

Table 1. Comparison of the three passes through the observer recordings by the post-facto interpreter, by species. Individuals is the number counted on the third pass, when unlimited audition and visual inspection were allowed. Treating this estimate as correct, the percentage "correct" for Passes 1 and 2 were calculated at their respective counts divided by the number in the "Individuals" column of this table.

Species	Individuals	Pass 1 Correct	Pass 2 Correct
ACFL	43	0.605	0.698
AMCR	41	0.829	0.927
AMGO	29	0.655	0.690
AMRE	5	0.200	0.800
AMRO	7	0.429	0.714
BEKI	3	0.333	0.333
BGGN	37	0.757	0.811
BHCO	6	0.667	1.000
BLGR	7	0.571	0.857
BLJA	3	1.000	0.333
BRTH	1	0.000	0.000
CACH	24	0.667	0.750
CARW	51	0.627	0.804
CEDW	2	0.500	0.500
CHSP	2	1.000	1.000
COGR	3	0.333	0.667
COYE	16	0.688	0.750
DOWO	12	0.583	0.833
EABL	6	0.167	0.000
EAKI	11	0.091	0.636
EATO	10	0.300	0.400
EAWP	17	0.647	0.588
FISP	3	1.000	1.000
GCFL	1	0.000	0.000
GRCA	7	0.571	0.714
HAWO	1	0.000	0.000
HOFI	1	0.000	1.000
HOWA	6	0.500	0.500
INBU	21	0.429	0.476
KEWA	2	0.000	0.500
LOWA	1	0.000	0.000
MODO	8	0.875	0.875

NOCA	28	0.607	0.679
NOPA	7	0.286	0.714
OVEN	20	0.800	0.900
PIWA	3	0.667	0.667
PIWO	4	1.000	1.000
PRAW	6	0.500	0.500
PROW	1	1.000	1.000
PRTH	2	0.000	0.000
RBWO	4	0.750	0.750
REVI	60	0.750	0.783
RHWO	1	0.000	1.000
ROPI	1	1.000	1.000
RSHA	7	0.857	0.857
RWBL	6	0.667	1.000
SCTA	16	0.750	0.813
SUTA	2	0.000	0.000
TUTI	31	0.581	0.774
WBNU	3	1.000	1.000
WEVI	7	0.571	0.857
WEWA	1	1.000	1.000
WOTH	88	0.591	0.761
YBCH	11	0.545	0.636
YBCU	3	1.000	1.000
YTVI	15	0.600	0.733

Assuming that the identifications of the third pass are accurate (an assumption that will be tested in a later study), the amount of improvement resulting from a third pass is substantial and worthy of the time expended. For that reason, further analysis will be restricted to a comparison of the third pass of post-processing with the two estimates from the observers' field data (samples 4 and 5).

COMPARISON OF POST-PROCESSING WITH REALTIME RESULTS

To assess the performance of the interpreter, I compared his count of each species to the counts of the realtime observers. Results of such comparisons are in Table 2.

Table 2. Comparison of five estimates of total number of vocal individuals, by species, from stops 1-18, round 2 (i.e., second run by each observer). The first three columns are the three successive passes by the interpreter using only the observer tapes. Pass 1 (N1) was a single audition of the tape, Pass 2 was a second audition, with streaming spectrographic visualization, Pass 3 was unlimited audition and visualization. N4 and N5 are the grand totals from the observers' realtime counts, as transcribed from the maps on which they recorded their data. N4 is all vocal registrations with the 50-m radius of the observer. N5 is the total, inside and outside the 50-m circle. Notice that N3, the best estimate by the interpreter, typically falls short of N5, the total count by the realtime observer, while N3 exceeds N4 in numerous species. This seems to show clearly that many of the more distant birds were audible on the recording.

Species	N1	N2	N3	N4	N5	N3 / N4	N3 / N5	N4 / N5
ACFL	33	35	43	58	71	0.741	0.606	0.817
AMCR	39	41	41	15	63	2.733	0.651	0.238
AMGO	25	24	29	28	29	1.036	1.000	0.966
AMRE	3	5	5	9	10	0.556	0.500	0.900

AMRO	4	5	7	7	13	1.000	0.538	0.538
BAOR	0	0	0	0	1	.	0.000	0.000
BARS	0	0	0	1	1	0.000	0.000	1.000
BAWW	1	0	0	0	0	.	.	.
BEKI	2	2	3	2	3	1.500	1.000	0.667
BGGN	32	35	37	41	41	0.902	0.902	1.000
BHCO	8	8	6	15	18	0.400	0.333	0.833
BLGR	7	7	7	9	14	0.778	0.500	0.643
BLJA	7	2	3	0	1	.	3.000	0.000
BRCR	0	0	0	1	1	0.000	0.000	1.000
BRTH	0	0	1	2	2	0.500	0.500	1.000
CACH	21	19	24	37	41	0.649	0.585	0.902
CAGO	0	0	0	1	1	0.000	0.000	1.000
CARW	39	44	51	51	59	1.000	0.864	0.864
CEDW	2	1	2	15	16	0.133	0.125	0.937
CHSP	2	2	2	1	1	2.000	2.000	1.000
CHSW	2	1	0	1	1	0.000	0.000	1.000
COGR	2	2	3	7	12	0.429	0.250	0.583
COYE	16	16	16	21	26	0.762	0.615	0.808
DOWO	8	11	12	14	23	0.857	0.522	0.609
EABL	1	0	6	3	6	2.000	1.000	0.500
EAKI	4	9	11	10	13	1.100	0.846	0.769
EATO	11	5	10	6	7	1.667	1.429	0.857
EAWP	15	16	17	5	23	3.400	0.739	0.217
EUST	0	0	0	3	3	0.000	0.000	1.000
FICR	0	1	0	0	0	.	.	.
FISP	3	3	3	5	6	0.600	0.500	0.833
GBHE	0	0	0	3	5	0.000	0.000	0.600
GCFL	0	0	1	1	1	1.000	1.000	1.000
GRCA	5	6	7	12	12	0.583	0.583	1.000
GRHE	0	0	0	1	1	0.000	0.000	1.000
HAWO	1	0	1	2	2	0.500	0.500	1.000
HOFI	0	1	1	3	3	0.333	0.333	1.000
HOME	0	0	0	3	3	0.000	0.000	1.000
HOSP	0	1	0	0	0	.	.	.
HOWA	4	4	6	4	5	1.500	1.200	0.800
HOWR	0	0	0	0	1	.	0.000	0.000
INBU	12	14	21	24	32	0.875	0.656	0.750
KEWA	2	4	2	2	3	1.000	0.667	0.667
KILL	0	0	0	0	1	.	0.000	0.000
LOWA	2	1	1	1	2	1.000	0.500	0.500
MODO	7	8	8	13	23	0.615	0.348	0.565
NOCA	24	24	28	24	31	1.167	0.903	0.774
NOMO	1	0	0	0	1	.	0.000	0.000
NOPA	4	6	7	12	15	0.583	0.467	0.800
OROR	0	0	0	1	2	0.000	0.000	0.500
OVEN	24	24	20	25	30	0.800	0.667	0.833
PIWA	2	2	3	8	10	0.375	0.300	0.800
PIWO	4	4	4	3	11	1.333	0.364	0.273
PRAW	3	3	6	4	4	1.500	1.500	1.000
PROW	1	1	1	2	2	0.500	0.500	1.000

PRTH	0	0	2	0	0 .	.	.	
PUMA	1	0	0	2	3	0.000	0.000	0.667
RBWO	3	3	4	11	18	0.364	0.222	0.611
REVI	53	48	60	71	107	0.845	0.561	0.664
RHWO	1	2	1	0	0 .	.	.	
ROPI	1	1	1	0	0 .	.	.	
RSHA	6	7	7	4	8	1.750	0.875	0.500
RTHA	0	0	0	1	1	0.000	0.000	1.000
RTHU	0	0	0	3	3	0.000	0.000	1.000
RWBL	5	8	6	9	24	0.667	0.250	0.375
SCTA	14	15	16	16	28	1.000	0.571	0.571
SUTA	3	1	2	1	2	2.000	1.000	0.500
SWSP	0	0	0	0	1 .		0.000	0.000
TRES	0	0	0	10	10	0.000	0.000	1.000
TUTI	20	25	31	40	51	0.775	0.608	0.784
WBNU	5	3	3	12	14	0.250	0.214	0.857
WEVI	5	7	7	9	14	0.778	0.500	0.643
WEWA	1	1	1	0	1 .		1.000	0.000
WITU	0	0	0	1	2	0.000	0.000	0.500
WOTH	56	67	88	51	87	1.725	1.011	0.586
YBCH	8	10	11	18	20	0.611	0.550	0.900
YBCU	3	3	3	3	7	1.000	0.429	0.429
YSFL	0	0	0	0	2 .		0.000	0.000
YTVI	9	12	15	11	17	1.364	0.882	0.647
YTWA	0	0	0	1	1	0.000	0.000	1.000
totals	577	610	714	790	1127	0.894	0.630	0.701

The simple comparisons in Table 2 can be improved on by treating each sample as the criterion and assessing the agreement of each other sample with it (Tables 3 and 4). Although the total number of individuals detected on pass 3 was 63% of the total at all distances by the observers, the interpreter actually detected only 43% of the individuals detected by the observers (Table 4). The reason agreement between observer totals and interpreter totals was better than this is that the realtime observers missed 26% of the individuals putatively identified by the interpreter (Table 4)

Table 3. Matrix of all "missed" species for samples 3-5 (the final pass through the recordings and the two observer estimates). Each cell is the proportion of species detected in the column sample that was missed in the row sample, accumulated over stops. Of greatest interest is the top right cell which shows that a species detected aurally by the realtime observer was missed 42.6% of the time by the post-processing interpreter. Row 4, column 5 shows that nearly 75% of species detected were in the 50-m circle. Column 3 shows that many species detected by the interpreter were "missed" by the realtime observers. Misses in either direction could of course be misidentifications. The extent to which misidentification is responsible for the discrepancies tabulated here is testable through independent examination of the recordings.

	3	4	5
3	0.000	0.369	0.426
4	0.383	0.000	0.258
5	0.243	0.000	0.000

Table 4. Matrix of all “missed” individuals for samples 3-5 (the final pass through the recordings and the two observer estimates). This matrix is analogous to the one above (Table 3). The higher values show that a few individuals were missed even when the species was detected.

	3	4	5
3	0.000	0.493	0.546
4	0.408	0.000	0.285
5	0.259	0.000	0.000

A similar analysis was conducted for each species. Four are presented here, Red-eyed Vireo (Table 5), Wood Thrush (Table 6), Carolina Wren (Table 7), and Acadian Flycatcher (Table 8). Data for all species are tabulated in Appendix 1. These tables are interpreted specifically in the table legends. General interpretations appear in the Discussion.

Table 5. Performance of samples 3 (post-processing), 4 (observer 50-m radius), and 5 (observer unlimited distance) with respect to the Red-eyed Vireo. The data are presented as a table rather than a matrix to allow inspection of raw data as well as percentages. The matrix index indicates the sample that is being compared with a criterion sample, e.g., “34” indicates that sample 4 is used as the criterion, and agreement of sample 3 with it is tabulated. “Column” refers to this criterion sample, “Row” to the sample being compared to it. “Row Species Misses” therefore means the number of stops on which the species was detected in the column sample but was missed in the row sample. “Row Individual Misses” is the number of individuals detected in the column sample that were missed in the row sample. The rows of greatest interest in these and subsequent tables are those indicated by matrix indices 34 and 35 (post-processing compared to realtime estimates) and 45 (50-m sample compared to entire sample).

These data show that post-processing (indices 34 and 35) was rather good (nearly 80% detection) at detecting the presence of the species, but much poorer at counting the individuals. This is probably because of the singing style of the Red-eyed Vireo. Relatively short gaps between rather short songs makes it difficult to hear distant birds during the gaps. One-third of individuals were outside the 50-m circle, which probably exacerbated the aforementioned source of difficulty.

Matrix Indices	Row Species Misses	Column Stops with Species	Percent Stops Species Missed	Row Individual Misses	Column Individuals	Percent Individuals Missed
33	0	48	0.000	0	60	0.000
34	9	42	0.214	33	72	0.458
35	13	56	0.232	56	108	0.519
43	15	48	0.313	21	60	0.350
44	0	42	0.000	0	72	0.000
45	14	56	0.250	36	108	0.333
53	5	48	0.104	8	60	0.133
54	0	42	0.000	0	72	0.000
55	0	56	0.000	0	108	0.000

Table 6. Performance of samples 3 (post-processing), 4 (observer 50-m radius), and 5 (observer unlimited distance) with respect to the Wood Thrush. See Table 5 for explanation of table format.

Post-processing was relatively effective for this species, despite its abundance, probably because the songs of an individual are separated in time by several song-lengths and are individually distinctive. The relatively high miss-rate for sample 4 (50-m) maximum shows that many detectable individuals beyond this radius (comparison 45), and that many of these distant birds were audible on the tape (comparison 43).

Matrix Indices	Row Species Misses	Column Stops with Species	Percent Stops Species Missed	Row Individual Misses	Column Individuals	Percent Individuals Missed
33	0	47	0.000	0	88	0.000
34	2	31	0.065	5	52	0.096
35	6	48	0.125	17	88	0.193
43	18	47	0.383	41	88	0.466
44	0	31	0.000	0	52	0.000
45	17	48	0.354	36	88	0.409
53	5	47	0.106	17	88	0.193
54	0	31	0.000	0	52	0.000
55	0	48	0.000	0	88	0.000

Table 7. Performance of samples 3 (post-processing), 4 (observer 50-m radius), and 5 (observer unlimited distance) with respect to the Carolina Wren. See Table 5 for explanation of table format.

The Carolina Wren shows relatively low miss-rates in all cells, revealing a species that is well-suited to monitoring with recordings. It has loud, distinctive songs that are widely-spaced in time. Repetition of a single phrase several times in a song is distinctive for identification and it enhances detectability.

Matrix Indices	Row Species Misses	Column Stops with Species	Percent Stops Species Missed	Row Individual Misses	Column Individuals	Percent Individuals Missed
33	0	42	0.000	0	51	0.000
34	3	36	0.083	13	52	0.250
35	5	41	0.122	18	60	0.300
43	9	42	0.214	12	51	0.235
44	0	36	0.000	0	52	0.000
45	5	41	0.122	8	60	0.133
53	6	42	0.143	9	51	0.176
54	0	36	0.000	0	52	0.000
55	0	41	0.000	0	60	0.000

Table 8. Performance of samples 3 (post-processing), 4 (observer 50-m radius), and 5 (observer unlimited distance) with respect to the Acadian Flycatcher. See Table 5 for explanation of table format.

The Acadian Flycatcher should be one of the easiest species in the area to detect and identify. Its song is loud and not easily confused with that of any other species present. Its song is brief, so temporal overlap should not lead to missing distant individuals. The high miss-rate for individuals, even within the 50-m circle, therefore is somewhat surprising, and calls for further analysis. Perhaps the missed individuals were detected or identified by calls rather than songs.

Matrix Indices	Row Species Misses	Column Stops with Species	Percent Stops Species Missed	Row Individual Misses	Column Individuals	Percent Individuals Missed
33	0	37	0.000	0	43	0.000
34	9	41	0.220	24	60	0.400
35	13	45	0.289	37	73	0.507
43	5	37	0.135	7	43	0.163

44	0	41	0.000	0	60	0.000
45	4	45	0.089	13	73	0.178
53	5	37	0.135	7	43	0.163
54	0	41	0.000	0	60	0.000
55	0	45	0.000	0	73	0.000

DISCUSSION

ADEQUACY OF POST-PROCESSING

Considering the interpreters results alone, the second 3-minute pass through the recordings agreed better with the final pass than did the first. Whether this is attributable merely to a second hearing or at least partially to realtime spectrographic visualization cannot be determined from these data. It does suggest that there is a functional relationship between the "accuracy" of the count and the amount of time the interpreter spends with the tape. If such a relationship does exist, it means that an agency employing post-processing of tapes could make an informed decision about how much "accuracy" to seek at a cost of how much interpreter time. This interpretation of the results rests on the assumption that the final pass by the interpreter is most accurate, an assumption that is independently testable in this study, as it would be in any protocol that used tapes.

I compared results from post-processing of audio-recordings with only the aural results of the realtime observers. Levels of agreement between post-processing and conventional BBS data (i.e., visual + aural detections) would therefore be lower for those species detected visually.

The data on the rarer species are included in Table 2 for interest. The data for the more abundant species show a serious shortfall between the final pass by the interpreter and the full acoustic data set of the observer. To assess the effect of distance to birds on detectability on tape, the column headed Pct34 shows the ratio of birds detected on Pass 3 to the birds detected within an estimated 50 m by the realtime observers. This number exceeds unity in many cases, suggesting that some species are detectable on tape at greater distances.

The comparisons in Table 2, however, are based on totals, and are not an assessment of how well one sample detected birds detected in another sample. For such an assessment I counted the number of individuals of each species present at a stop (on one day) as estimated in one sample (the criterion) that were estimated in another sample. Totaling these data led to similarity matrices, which could be subdivided in a number of interesting ways, e.g., by observer, stop, species. The grand matrices, over all data, are presented in Tables 3 and 4. Table 3 is for presence-absence, Table 4 is for quantitative counts. The similarity of Tables 3 and 4 results from the vast majority of counts being one individual. Missing the single individual tallied in the criterion sample would therefore result in both a missed species (Table 3) and a missed individual (Table 4).

It is clear from Table 4 that a portion of the rather good agreement displayed in Table 2 is spurious. Not only did the interpreter miss a substantial number of birds counted in realtime (top two rows in the matrix), he found a substantial number not counted in realtime (left two columns in the matrix). In other words, the high percentage of birds counted in sample 4 (50-m realtime) also found by the interpreter is due in part two birds from different stops and days. The accuracy of both estimates is therefore deserving of further scrutiny.

High abundance of individuals can cause confusion to a tape interpreter because of the overlap of songs. This is particularly the case with Red-eyed Vireo, because of its pattern of singing: short phrases repeated incessantly at approximately 1-sec intervals. This phrasing does not allow discrimination of individuals to the degree that the phrasing of emberizids, cardinalids, and parulids, for example, do. As a case in point, compare the low detection rates of Red-eyed Vireo (Table 5) with the higher rates for another abundant species, Wood Thrush (Table 5). Not surprisingly, a given method works better for some species than others. This merely underlines the importance of estimating species-specific detection probabilities in any kind of monitoring program.

FUTURE ANALYSIS

With a preliminary analysis now completed for temporal variability in observer results, coupled with this preliminary analysis of post-processing effectiveness, several questions arise that may be addressed through careful inspection of the multi-channel recordings made during data collection.

1. The a priori question of observer identification accuracy with respect to similar-sounding species. Three such groups merit examination:
   a. Pine Warbler – Chipping Sparrow – Worm-eating Warbler
   b. Common Yellowthroat – Carolina Wren
   c. Indigo Bunting – Blue Grosbeak
2. Disagreement between Table 2 and Tables 3-4. Were those individuals detected by the interpreter but not in the field spurious, or are the columns of the matrix tables unbiased estimates of the miss-rates of the field observers? If they are, post-processing of BBS tapes by persons other than the field observers could provide vital correction factors, even though post-processing does not detect enough individuals to substitute for field estimates. This interesting possibility will be evaluated in the analysis of multi-channel recordings.

Overall, the study so far suggests that low-tech post-processing of short samples of recorded sound is a poor substitute for a well-trained observer in the field. This does not mean that recordings are a totally inadequate substitute for a live observer. It does suggest that using recordings alone is likely to be more useful for protocols other than rapid counts of individuals, e.g., presence-absence assessment. Discriminating individuals is more difficult on a recording than in the field. Using recordings for presence-absence assessment is less tedious, and more time efficient, because a longer recording can be scanned for presence absence in the same time required to scan a shorter recording for abundance. Given the very poor repeatability between days documented in another phase of this study, extending the temporal coverage by conducting presence assessment over a longer period of time, with recordings, may be needed. Or it may be that the additional detections via presence absence on long recordings will no more than offset the lowered detectability of some species on recordings.

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. I would like especially to thank Glenn Johnson for the diligence, professionalism, and good nature with which he performed the post-processing.

LITERATURE CITED

Bart, J., and S. Earnst. 2002. Double sampling to estimate density and population trends in birds. Auk 119:36-45.

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.

Haselmayer, J., and J. S.Quinn. 2000. A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru. Condor 102: 887-893.

O'Connor, R. J., E. H. Dunn, D. H. Johnson, S. L. Jones, D. Petit, K. Pollock, C. R. Smith, J. L. Trapp, and W. Welling. 2000. A programmatic review of the North American Breeding Bird Survey. [Online] Available at www.mp2-wrc.usgs.gov/bbs/index.htm.

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.