The Pros and Cons of Audio-taping Point Counts: Equipment Considerations

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

Keywords: Monitoring, point count, frequency response, sensitivity, recording equipment.

TABLE OF CONTENTS

ABSTRACT

Monitoring of protected terrestrial birds is intensive for species at risk, including demographic analysis, but only population trend is estimated for the hundreds of low-risk species. Point counts are the main source of data on these lower-risk species, and aural cues are the main source of detections on point counts. Recording point counts presents a number of potential advantages over simple one-time audition in the field, but also poses new challenges. This study documents differences in the sensitivity of very low-cost ($60) consumer electronic equipment and low-end ($1500) professional equipment. On average, the professional equipment produced recordings with more target information, as reflected by the amplitudes of calibrated test sounds on the tapes. But, results were so variable from trial to trial that definitive recommendations of equipment cannot be made. As expected, recordings made in a forested environment were lower in amplitude than comparable recordings from an open field environment, presumably because vegetation in the forest absorbed or reflected more acoustic energy that did vegetation in the field. Nonetheless, recordings made with a professional microphone during a field test of auditory acuity of seven human subjects were at least as sensitive as the humans' ears, i.e., a higher percentage of test sounds was locatable on the tapes than the human subjects detected. Because these results are somewhat contradictory, more testing is needed. But, the potential usefulness of recordings in bird monitoring remains high.

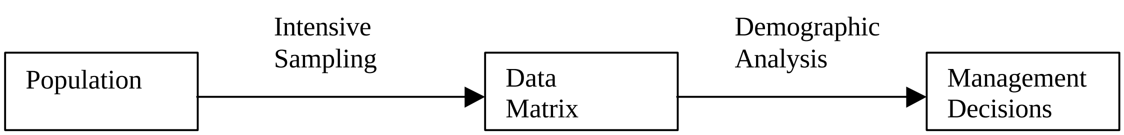
INTRODUCTION

THE NEED FOR SAMPLING

Decisions whether to and how to manage wildlife populations must be based on information. Data must be gathered on the population, or surrogate population, and then organized into a data matrix for analysis and subsequent decision-making. In the diagrams that follow, the boxes represent states of the system, the arrows and associated labels indicate transfer functions. Importantly, the data matrix, although a mere artifact of the population at a moment or brief interval in time, can be stored and retrieved for later re-analysis. Accuracy of this artifact can have a critical effect on the success of management decisions (see below).

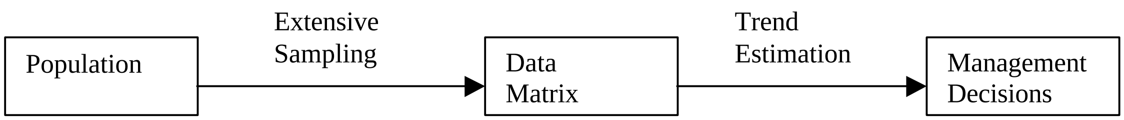
Intensive Sampling

Intensive management requires demographic data, obtaining which requires intensive sampling, which is time-consuming and economically expensive. A *de facto* two-tiered approach to management of nongame bird species in North America is a consequence of this fact. Intensive sampling, involving nest-finding, color-banding of individuals, and perhaps radio-tracking of dispersing inviduals, can produce a reasonably representative data matrix, although at a cost that is ordinarily only undertaken when trying to save a species or population from extinction. Although the presence of the investigators may indeed change the dynamics of the population itself, the data themselves are largely unbiased.



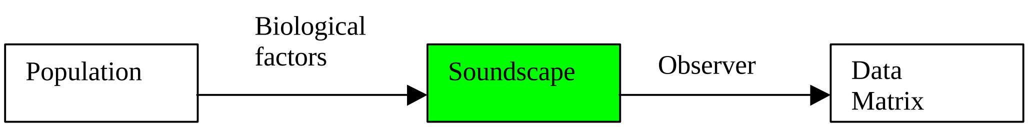
Extensive Sampling

Other species are monitored for decreasing population trends that may place them in the imperiled category, which would then lead to intensive study and management. Estimates of trend are based mostly on brief surveys conducted over wide geographic areas. Because of the briefness of these surveys, investigators are unlikely to bias population dynamics, but the potential for the resulting data matrix to be biased is high. The main source of trend data in the United States and southern Canada is the Breeding Bird Survey (BBS), which is based on 50 3-min point counts repeated once annually. The BBS, other point-count protocols, and indeed most survey methods applied to territorial land birds, rely heavily on songs and other vocalizations to detect the birds present in the population of interest. Approximately 90% of detections in such surveys are based on sound (Buckland et al. 2001).



THE SOUNDSCAPE

A variety of biological factors (e.g., time of day, pairing status, number of conspecifics) influences which individual birds sing and at what time. The resulting soundscape has objective reality, but because it is a transform of the population itself, any data matrix estimated from it becomes an artifact of the soundscape rather than the population. In the worst case, the soundscape may be a misrepresentation of the demographic vitality of the population. This unwelcome possibility could be reality if, for example, males with mates are the least likely males to sing. In polygynous species and those socially monogamous species in which mated males pursue extra-pair fertility, attractive males may continue singing throughout the season, and using the soundscape as a proxy for the population may be less problematic. Regardless, these biological factors are independent of any effort to sample the soundscape, i.e., the resulting concerns would pertain even if the soundscape were known without error. If there are lawful relationships between population characteristics and the soundscape, e.g., a seasonal change in the relationship between number of females present and number of males singing, then it may be possible to devise correction factors to improve the picture of the population that can be gained from the soundscape.

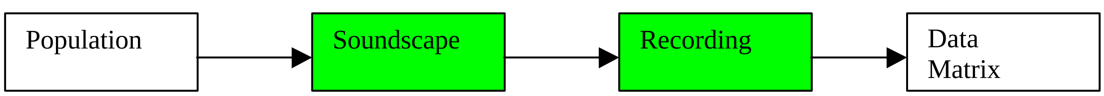


Even when the information available in the soundscape is positive, i.e., the amount of singing is correlated with the number of birds present, sampling the soundscape introduces additional, but tractable, problems, as follows: (1) Because many species of birds have repertoires of song-types and because some species sound similar to many observers who are competent to distinguish them visually, the potential for identification error is nontrivial. (2) When the sample is very brief, as in a 3-min stop on the BBS, additional statistical factors related to sampling error may lead to bias in the data matrix. (3) Point-count theory requires only one detection of an individual for it to be “counted” on the survey. Many birds sing at once, and the brief period available may not afford the observer sufficient time to hear all of them, lowering detectability. Moreover, when the sample is taken from a single point, as is usually the case, the fall-off in detectability owing to attenuation of sound with distance must also be accounted for. These three factors will be referred to as “accuracy,” “availability,” and “detectability,” respectively.

Factors 2 and 3, subsumed under the rubric of "detection probability," are currently an area of active research. The goal is to estimate correction factors that can be applied to the raw data matrix to make it a more accurate reflection of the population. The need for such correction factors is demonstrated by a U. S. National Park Service study in southeastern Utah. Long-play digital recordings detected 45 species at one location where point counts detected 11 (Daw and Ambrose 2003). Included in the needed correction factors is one for "availability," the probability that a bird that is present will sing during the count period (Farnsworth et al. 2002, in press; McCallum in press). In models introduced to date, corrections for unavailability treat all males (i.e., potential singers) in the population as equivalent, and hence do not get at the difference in detectability of mated and unmated males. A well-developed methodology for correcting for distance effects on detectability exists (Buckland et al. 2001), and is now receiving wide attention in the bird-monitoring community.

For brief, sound-based surveys, it is now technically and economically feasible to make electronic recordings of the soundscape for post-processing. The amount of post-processing can vary from one-time audition (in an attempt to delay the time and relocate the place at which the observer experiences the soundscape, while minimizing any change in the amount of information available to the observer / interpreter) to detailed bioacoustic analysis.

Although an audio-recording is a more objective record of the soundscape than a field notebook or a point-count data sheet, it is undeniably, like them, an artifact, interposed between the soundscape and data matrix, and care must be taken to ensure that it is an accurate representation of the soundscape at the time of recording.



Making a recording of the soundscape requires a (1) a microphone system, (2) a recording platform, and (3) storage media. All microphones are electronic devices that transduce variation in pressure waves to variation in voltage. Recording platforms amplify the time-varying voltage that is supplied by the microphone system and either encode it on magnetic tape (analog recorders), or convert it to binary-number format (digital recorders) and store the number stream on one of several possible types of media (magnetic tape, mini-discs, memory chip). A brief discussion of alternative microphones, recording platforms, and storage media follows.

TECHNICAL OPTIONS FOR FIELD RECORDING OF POINT COUNTS

Accurate rendering of frequencies on analog recordings requires constant and repeatable tape speed on the recording device and all playback devices. Because such precision is costly, the accuracy of frequency response is one of the features that is typically sacrificed in lower-priced analog recorders and playback machines. On digital audio tape

(DAT), as on other digital media, time is encoded in the bit rate, not the speed of playback. Digital recordings generally reproduce frequencies much more accurately than analog recordings.

In the field, individuals of the same species are distinguished by differences in amplitude and/or by differences in the direction from which the sounds appear to emanate. Variation in amplitude is encoded with any microphone system. The sensitivity of the microphone(s) is the only source of variation in the capture of amplitude variation. Microphone sensitivity is correlated with cost.

Capturing directional information is more challenging. Human hearing is binaural (2-channel), and stereophonic sound recordings encode directional information through the same differences in amplitude and time-delays that humans use for localization of a sound source. If the microphones used to make a 2-channel recording are positioned farther apart than the inter-aural distance of a human head, the soundscape will appear more polarized to the auditioner of the recording than it actually is. This may actually be an advantage for the auditioner, facilitating discrimination of two birds that are in different hemispsheres of the soundscape. Field auditioners, however, can turn their heads to change the orientation of the two poles of their binaural soundscape, which is not possible in post-processing. Slowly rotating a directional microphone system (Haselmayer and Quinn 2000) can provide more information on bearing to individual birds (presuming they are stationary), but keeping up with the orientation of the microphones could be daunting during post-processing. Three or more recording channels would largely compensate for the head-turning advantage of field auditioners, but multi-channel (> 2) recording platforms are expensive and not readily available in contrast to stereo recorders. Playback also is much more complicated to implement with > 2 channels.

Realistically, 2-channel (stereophonic) recordings are going to be the upper technical limit for most monitoring applications. The question then arises whether to use omni-directional or directional microphones for recording the two channels. Playback is a simple matter with a stereo recording. Equipment is widely available, reasonably priced, and easy to use. In most cases the recording device can be used as a playback device, either directly to headphones, or via a line-out/line-in connection to a stereo amplifier and speaker system. The one technical consideration is insuring that the playback device has a line-out or headphone jack.

Choice of a recording platform entails several decisions: (1) whether directional information is needed (1-channel vs. 2-channel recording), (2) the accuracy of frequency response needed, (3) the dynamic range required of the recording, and (4) ruggedness required of equipment that is deployed in the field day after day under varying environmental conditions. The number of channels is a correlate of cost. In general, analog recorders, which are typically purely mechanical in operation, have poorer frequency response and dynamic range, but are more rugged than digital platforms, which contain a computer chip and are often sensitive to humidity.

The error in frequency response of even low-end analog recorders is not likely to be problematic for use of recordings in monitoring programs, but quantitative data on frequency response of such machines are needed to back up this presumption. Dynamic range refers to the range of amplitudes that can be encoded with the recording process. This is wider intrinsically on digital recordings, and the absence of tape "hiss" also enhances the effective dynamic range of digital recordings. On the other hand, analog recording platforms are likely to be less expensive and more dependable than digital recorders. The recording media available may therefore be the deciding factor in choosing a recording platform.

Magnetic tape housed in cassettes is the only viable recording medium for analog recording platforms, as reel-to-reel field recorders, once the mainstay of bioacoustics, are prohibitively expensive for this application, and also are not widely available (www.birds.cornell.edu/MacaulayLibrary/equip). In addition to audio-cassettes, video cassettes are possible analog recording media. Video cassettes, which record in only one direction) can hold two or more hours of material in "long play" mode (lower tape speed, or sampling rate in the case of digital recording), whereas audio cassettes of more than 55-min duration (on one side) are not widely available. Video cameras are sensitive to humidity and are generally not as rugged as analog audio-cassette recorders, but when long-play recordings are needed, video cassettes provide much more capacity than audiocassettes.

Tape speed for analog audio-cassette recorders is universally a slow 1 7/8 ips (inches per second), yielding far-less accurate frequency response than the 7.5 ips reel-to-reel machines of the past. As mentioned above, tape speed poses no serious problem for using recordings to identify species and individuals present, for monitoring purposes. Cassettes are easy to obtain, inexpensive, and easy to play. They degrade with usage and time, but for simple audition in the same year surveys were conducted, they have few physical liabilities beyond the hiss that is inherent to analog recording.

Digital recording offers a variety of media. Indeed, the number of formats available for digital recording presents a problem for bioacoustics because the technical details of many formats are not widely appreciated and in some cases are proprietary. The major issue is compression. Professional DAT recorders do not compress data, and offer the most accurate recording of the soundscape available. Professional DAT recorders, however, cost several thousand US dollars, and are beyond the reach of the budgets of most monitoring programs. Consumer digital recording devices were designed for recording music and human speech, and for the most part use compression techniques to reduce data-storage requirements. Although compression may eliminate acoustic information that is important to birds, and therefore recordings made in these format should not be used for playback or analysis of the sounds of birds, it may be innocuous when recordings are used only for monitoring. This is a viable research question.

PROS AND CONS OF RECORDING POINT-COUNTS

Recording the soundscape has numerous advantages over simply listening to it and taking notes in real time, as is currently done with the BBS and other point-count protocols. The most important of these is that an objective representation of the soundscape can be stored and retrieved later for re-analysis. This is the standard practice in science. Data that cannot be independently verified violate a basic principle of science. As the consequences of management decisions based on survey data become more and more costly for powerful economic interests, the potential for survey results to be challenged in court increases. Recording the soundscape rather than simply taking notes on it produces a record that can be examined independently by contesting and neutral parties, thereby potentially averting harm to imperiled wildlife populations from legal challenges to the data matrices used to justify their protection.

Recording the soundscape also makes reanalysis possible, even decades later, for different research objectives from those intended at the time the recordings were stored. The usefulness of old bird-egg collections for demonstrating the negative effects of pesticide poisoning, and of old collections of study skins for molecular genetic analyses are but two examples of valuable but unintended uses of archived samples of biological material.

Recording also mitigates the logistic problem of an insufficient number of field-workers able to identify all bird species in the study area, by ear, in real time (R. Rempel: http://flash.lakeheadu.ca/~rrempel/CVX/). Even the simplest post-processing (listening to the tape) can reduce identification errors through use of an expert on the sounds of the birds in the study area in lieu of inexperienced field workers. Using more than one such expert permits estimation of the level of confidence in the data matrix. Alternatively, a less experienced observer can listen to a tape repeatedly until convinced of his/her identifications, or use visualization techniques (e.g., spectrograms) to increase the likelihood of correct indentifications.

When post-processing is not limited to a single audition of the tape, it increases detectability by allowing for independent examination of contemporaneous sounds from different individuals. Post-processing is especially advantageous when the sounds are visualized with spectrograms, which separate the sounds’ frequencies.

Because the amplitude of a sound is attenuated as the square of the distance from the source to the microphone, distance from the bird to the microphone is encoded in the recordings. Deriving the distance to singers from tapes would make it possible to use the distance method (Buckland et al. 2001) to estimate detectability, a major step toward decreasing bias in the population estimates derived from point counts (Rosestock et al. 2002). Decoding this information requires a distance-amplitude function, a curve that gives the true distance as a function of the amplitude on a calibrated tape. Such functions are likely to differ for species, call-type, direction the bird is facing (relative to the microphone), and vegetation type (because vegetation absorbs or reflects acoustic energy), but the potential of this method is great.

The potential liabilities of recording all fall under the headings of “availability” and “detectability.” Clearly, visual information is not available in a recording. Moreover, several potential auditory short-comings of recordings must be minized. In typical point-counts, the number of individuals is counted. When individuals are detected with aural cues, counters typically rely on differences in amplitude and bearing of the sounds to distinguish them. Amplitude differences are likely to be recorded faithfully, but it is an open question whether low amplitude sounds are as detectable on a tape as they are in the field. This depends primarily upon the sensitivity of the microphone system used (see below). Potentially a more serious problem for recordings is the loss of directional information in single-channel (“monaural”) recordings. This can be mitigated with multi-channel recordings (see above), but the ability of the field observer to rotate the head and localize the bearing of an aural cue is technically unfeasible in a recording. (But, attending to a single bird, the standard practice in point-counting, lowers the likelihood of detection of other birds, so replacing an on-site observer with a post-processor yields gains as well as losses in detectability.)

In summary, post-processing of recordings is likely to increase the accuracy of identifications made in the field, while reducing the count of species and individuals through the sacrifice of visual information. The numbers of individuals detected can be biased positively or negatively, depending primarily upon the sensitivity of the recording system. The number of individuals detected is likely to be biased negatively through the loss of directional information in single-channel recordings, but corrections for lowered detectability owing to distance can be obtained from tapes, albeit with much expenditure of time.

Sensitivity of the recording system is the easiest of these issues to investigate, and to rectify if necessary. The question is whether all the birds audible to an average observer in the field can be detected from a recording made with inexpensive recording equipment. If not, does more expensive equipment ameliorate the problem? This is the question investigated in the present study. I used the simplest possible set-up: a single-channel omni-directional microphone, analog recorder, and Type I cassette tape. The low-end microphone and recorder are widely available for a total of approximately $60. The high-end equipment used costs a total of $1,300. I also investigate frequency response, to test the assumptions that (1) frequency response is poorer in low-cost equipment, and (2) that the inaccuracies resulting from poor frequency response are innocuous when identification of species is the only use of the recording.

METHODS

STUDY SITE

The test recordings were made on the Patuxent National Wildlife Research Center, Laurel, Maryland. Tests were conducted in two “habitats,” forest and field. The field habitat test was conducted alonged the paved service road east of Hance Creek. The forest habitat test was conducted in the beech forest along unpaved River Road, paralleling the Patuxent River. The recording devices were set up at a stationary location. In the field habitat this location was at the side of the paved road, where it makes a right-angle turn. In the forest habitat this location was in the forest, so the sounds travelled through vegetation en route from the speaker which was on a road. The playback speaker was mounted on the roof of a motor vehicle, which was positioned at distances of 50, 100, 150, and 200 m from the recorders. These distances were measured with a tape measure.

The main sources of noise on the days the tests were performed, July 4-5, 2002, were jetliner overflights and wind. Traffic was minimal on the station on July 4, when the “field” test was performed. This made it possible to conduct the test on a paved station road without interruption or disruption by passing cars. One car did pass during a test, and the test was run again. The “forest” test site was isolated by distance and vegetation from automobile noise.

EXPERIMENTAL DESIGN

The goal of the study was to compare the sensitivity of the cheapest available equipment for recording ambient bird sounds with equipment that would be considered acceptable by a professional bioacoustician, yet still as cheaply priced as possible. The low-end equipment tested was the Optimus 33-1060 electret condenser micophone, which retails at approximately $29.00, and the GE 3-5025A portable cassette recorder, available at approximately $30 from consumer electronics retailers. The professional equipment tested was the Sennheiser ME-62 omni-directional microphone with K6 power supply, and the Sony TCD5-ProII stereo cassette recorder. These devices are available from professional supply houses, beginning at approximately $300 and $1000, respectively.

Each brand of microphone was paired with each brand of recorder, resulting in four combinations. For replication, two examples of each type of microphone and two examples of each type of recorder were tested. Each microphone was tested with each recorder, resulting in 16 tests. These 16 tests were performed in two different habitats, field and forest. All four recorders could be tested simultaneously, so four tests could be run simultaneously in a trial. The sequence of trials was identical in the two habitats. The four equipment combinations in each trial recorded an identical soundscape, so within-trial variation was limited to main effects and interactions. Comparisons across trials incorporated considerable environmental variation because of variation in ambient noise and other stochastic factors, as well as the main effect of habitat. Two other treatments were varied at the time of data collection, distance and speaker orientation. Every equipment combination was tested at distances of 50, 100, 150, and 200 m, and at speaker orientations of 0, 90, and 180 degrees. As these treatments were performed for another purpose, calculation of distance-amplitude functions, only the results from 50 m and 0 degrees orientation are reported here.

The test sounds were played with Signal 3.0 software (Engineering Design), implemented in a Toshiba Satellite 2595CDT laptop computer. Computer output was broadcast over a single Advent AV 570/570G speaker system, rated at 70 watts, mounted on the top of a mini-van. Speaker height was approximately 2 m above ground level. The volume control dial of the speaker was taped in an immobile position, and sound pressure at 1 m was confirmed to be within 2 dB of 56 dB SPL at the outset of each trial with a Radio Shack 83-2055 Sound Level Meter.

The replication of each microphone and recorder type allows complete exploration of microphone*recorder interaction. In the full model, environmental variation such as ambient noise is controlled statistically. Because each trial included four tests conducted under identical environmental conditions, each of these experiments can be analyzed with a reduced model containing no environmental term, albeit without the ability to assess interaction.

DIGITIZATION OF RAW DATA

All tapes were played for digitization on the same Onkyo TA-RW400 stereo deck. The Onkyo had two tape drives. Tests revealed that the right drive on the Onkyo produced tapes with very little harmonic distortion, so only the right drive was used for playback of all test tapes. Field tapes were digitized in their entirety with Engineering Design software, implemented on a Compaq 5080US desktop computer. The machine had a 800 mHz Intel Celeron processor with a 128 kB integrated L2 Pipeline Burst Cache. At the time of this project, Signal 4.02.04 (dated December 17, 2003) was the available version of Signal (Engineering Design). For input/output (e.g., digitization) version 4.02.04 was supported by WaveDisk, version 1.01 (Engineering Design), dated 01/09/2002. Wavedisk uses the computer’s native sound card for analog-to-digital (A/D) conversion. The sound card installed on the Compaq 5080 at the time of digitization was an Allegro PCI Audio (WDM) version 018, from ESS Technology, Inc.

Sound files were captured at a sample-rate of 22050 Hz, which means that 22050 measurements of the input voltage were taken every second. This is a standard low-fidelity sample for the Windows .wav digital sound format. Although a higher sample rate (44100 Hz) is available, the cost in doubled storage demands at that rate is not offset by any benefit. The maximum frequency of a digitized sound is one-half the sampling/sample rate, and the sounds of interest in this analysis are > 3000 Hz lower than 11025 Hz. The 22.05 kHz sample rate is therefore deemed optimal.

MEASUREMENT OF SOUND PARAMETERS

Using Sound Analysis Macros (SAM) (copyright D. Archibald McCallum), operating in Signal 4.02.04, a technician located the tests of interest on each digitized tapefile, and visually selected approximately 10 samples of the alerting sound that were not seriously degraded by ambient noise. The alerting sound was broadcast at a sound pressure level (SPL) of approximately 89 dB SPL.The intended test sounds were broadcast at approximately 56 dB SPL. Only a few of these test sounds were audible on the tapes.

This SPL was calculated to be natural for a variety of singing birds, but apparently it was considerably lower than normal, as a Tufted Titmouse and Indigo Bunting singing 150 – 200 m from the recorders were readily audible on several tapes. It is also possible that the SPL was appropriate, but that the rather low position of the speaker (2 m) relative to the ground, led to significant degradation of the test sounds owing to ground scatter. (In the field environment, test sounds were broadcast over a paved roadway, with no intervening obstacles between speaker and microphones. The forest environment was a beech forest, with tree boles and a few shrubs obstructing passage of sound waves. The groundcover in this environment was mostly leaf litter.) Regardless, the main purpose of the tests was to assess relative performance of the recording devices, and for this purpose the alerting sound was suitable. For the purposes of this comparison, those test sounds least attenuated by distance were likely to provide the most dependable results. Sounds broadcast at a distance of 50 m with a bearing of 0 degrees (speaker oriented toward microphones) were chosen *a priori* for this purpose.

For measurement of each sound sample, a copy of the sample was normalized to an amplitude of 1 V RMS (root mean square). The onset of the sound was defined as the point at which the amplitude envelope exceeded 1.2 v. The sample to be measured was extracted from the un-normalized copy beginning 300 points after the onset as defined above. Although this process was automated in Signal, the time window selected in this manner was checked visually to ensure that a clean sample was obtained.

This sample was exactly 16,834 (2^{14}) points, approximately ¾ sec, in duration. Fourier Analysis in Signal is performed only on sound segments of length 2^N, where N is an integer < 16. If the sound segment does not have an exact power-of-2 number of points, it is zero-padded to the next highest power of 2. Although zero-padding is said to be harmless, I chose to make this question moot by selecting a sample length that would not require zero-padding.

The power spectrum was calculated for each sample segment. The spectrum was expected to have local peaks at 1, 2, 3, ... 8 kHz. The actual peak frequency was identified, and its amplitude was measured in 8 1000-Hz bands centered on the expected peaks. The deviation of the observed peak frequency from the expected peak frequency was calculated. In the following section, this parameter, in units of Hz, is referred to as "deviation," and the amplitude, in units of volts (V), is referred to as such.

RESULTS and DISCUSSION

FREQUENCY RESPONSE

The deviation of the observed frequency peak from the expected of course increased with frequency, so deviation was expressed as the ratio of the absolute value of the deviation to the expected value. The data were not normally distributed, so nonparametric tests of significance were performed. The analysis of variance (ANOVA) table below is for the ranks of the deviation ratios. ANOVA over ranks is equivalent to conventional nonparametric tests such as Willcoxon and Kruskal-Wallis. This gambit permits a full exploration of interactions. The ANOVA was highly significant for both main effects (microphone brand and recorder brand) and the interaction term (see Effect Tests below).

Summary of Fit

Rsquare	0.47244
RSquare Adj	0.471199
Root Mean Square Error	268.5861
Mean of Response	640
Observations (or Sum Wgts)	1279

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	82366772	27455591	380.5957
Error	1275	91976546	72138.467	Prob > F
C. Total	1278	174343318		<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
MicBrand	1	1	623512	8.6433	0.0033
DeckBran	1	1	72593715	1006.311	<.0001
MicBrand*DeckBran	1	1	7686441	106.5512	<.0001

These results confirm that the professional recorder outperformed the consumer deck, as expected. The significant effect of microphone brand on deviation was, on the other hand, not expected. Deviations from expected were almost all negative (deviation = observed – expected). Negative deviations indicate that the recorders were all recording at greater than rated speed. (The Onkyo deck on which these tapes were digitized was confirmed to play at the rated tape speed, so measured deviations were due to the decks on which the original recordings were made.) The summary statistics tabulated below are for percentage deviations, expressed as decimal fractions.

Microphone Brand

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Opt	0.02876807	0.00056962	0.028414
Sen	0.02587847	0.00056516	0.024155

Deck Brand

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
ge	0.03954989	0.00058831	0.039503
son	0.01509665	0.00054568	0.014877

MicBrand*DeckBran Interaction

Least Squares Means Table

Level	Least Sq Mean	Std Error
Opt,ge	0.03770731	0.00082137

Level	Least Sq Mean	Std Error
Opt,son	0.01982882	0.00078945
Sen,ge	0.04139247	0.00084249
Sen,son	0.01036448	0.00075354

Despite the signficant differences among equipment combinations, all combinations performed more than adequately for the purposes of identifying bird sounds to species. A 4% difference in frequency (the worst deviation from expected, Sennheiser microphone + GE recorder) is equivalent to less than one-half step on a musical scale, for example, the difference between C and C#. An error of this magnitude in the absolute pitch of the sound would likely be noticed only by an auditioner with "perfect pitch." Recognition of a musical melody or the pattern of frequency modulation of a bird song is based on the relative frequencies of the series of sounds, not their absolute frequencies. These ratios are preserved even when the absolute frequencies are shifted down or up, even extremely (e.g., by the 1970s recording group "The Chipmunks"). The poor frequency response of inexpensive analog recorders is therefore unlikely to present an operational problem for auditioners attempting to identify the species recorded with such machines. At least one species, the Black-capped Chickadee (*Parus atricapillus*), provides a natural test of this assertion. The widespread “Hey-sweetie” dialect of this species varies in frequency, but the two frequencies present in the song always stand in the same ratio, 1.134 (Weisman and Ratcliffe 2004). This song is one of the most familiar to birders precisely because the ratio approximates the musical interval known as a major second. The absolute frequency is largely irrelevant. At any rate, playing field tapes back on the recording machine should minimize this problem, assuming that recording and playback speeds are similar and precise, which they should be if fully-charged batteries are used in both instances. Moreover, affordable digital alternatives to analog recorders have recently entered the consumer market (see recommendations below).

SENSITIVITY

The higher the frequency of a sound, the greater the attenuation of it with distance. Therefore, although the eight frequency components (harmonics) of the test tone were all broadcast with equal energy, the energy received at the microphones was expected and confirmed to vary inversely with frequency. I present here results of analyses of the 8 kHz tone. Even though this tone was more attenuated by distance than lower frequencies, there is much less environmental noise at this frequency, and these results provide a clearer test of the equipment. Wind, aircraft noise, and insect noise were the major sources of ambient noise. Both wind and aircraft noise decrease with frequency and are minimal at the highest frequency studied, 8 kHz. Insect sounds were especially loud in the forest environment, although present also in the field environment. The frequency of their peak amplitude was between 5 kHz and 6 kHz. The 8 kHz test was therefore the “cleanest,” and most suitable for the purposes of this test. Results werer similar for other frequencies, and for all frequencies combined.

Amplitudes of the harmonics of the recorded test tones were not normally distributed, so nonparametric tests of significance were performed. The analysis of variance (ANOVA) table below is for the ranks of the measured voltages. ANOVA over ranks is equivalent to conventional nonparametric tests such as Willcoxon and Kruskal-Wallis. This gambit permits a full exploration of interactions. The ANOVA was highly significant for all three main effects (microphone brand, recorder brand, and habitat) and for the three-way interaction term (see Effect Tests below).

Response: Rank of Amplitude(volts)

Response VoltRank
Summary of Fit

RSquare	0.392313
RSquare Adj	0.364328
Root Mean Square Error	220.7479
Mean of Response	392.25
Observations (or Sum Wgts)	160

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	4781785	683112	14.0184
Error	152	7406906	48730	Prob > F
C. Total	159	12188691		<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
MicBrand	1	1	1399146.5	28.7124	<.0001
DeckBran	1	1	1277246.5	26.2109	<.0001
MicBrand*Habitat	1	1	41861.5	0.8591	0.3555
DeckBran*Habitat	1	1	118398.5	2.4297	0.1211
MicBrand*DeckBran	1	1	158161.0	3.2457	0.0736
MicBrand*DeckBran*Habitat	1	1	1199303.4	24.6114	<.0001
Habitat	1	1	709222.2	14.5542	0.0002

For inspection of summary statistics, measurements of amplitude in dB-volts are more useful. The tables below are for the 8-kHz harmonic. The dB (deciBel) unit permits comparison of the levels of sounds. It is the difference in these values, not their absolute magnitudes, that matters. Decibels are measured on a logarithmic scale. The “Just Noticeable Difference” (JDL) for sound level has been measured at aproximately 0.6 dB. Ten times this amount, a difference of 6 dB, is equivalent to a two-fold difference in amplitude. The best way to appreciate these differences is to hear them. A useful demonstration is provided at http://www.phys.unsw.edu.au/~jw/dB.html. The dB values reported below are relative to a standard with an amplitude of 1 Volt-RMS (root mean square). Sounds of this intensity can be auditioned at http://www.phys.unsw.edu.au/~jw/hearing.html.

Response: Amplitude in dB-volts

Effect Details

MicBrand
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Opt	-54.337136	1.3757392	-54.6945
Sen	-42.504554	1.3662680	-42.7608

DeckBran
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
ge	-53.672675	1.4210484	-54.3359
sony	-43.169016	1.3190785	-43.7633

Habitat
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
field	-45.314019	1.4424098	-45.3395
forest	-51.527671	1.2956854	-51.3643

MicBrand*DeckBran
Least Squares Means Table

Level	Least Sq Mean	Std Error
Opt,ge	-62.305822	1.9693586
Opt,sony	-46.368451	1.9215253
Sen,ge	-45.039528	2.0491806
Sen,sony	-39.969581	1.8076538

MicBrand*Habitat
Least Squares Means Table

Level	Least Sq Mean	Std Error
Opt,field	-52.425635	2.0500175
Opt,forest	-56.248638	1.8352278
Sen,field	-38.202403	2.0296827
Sen,forest	-46.806705	1.8295195

DeckBran*Habitat
Least Squares Means Table

Level	Least Sq Mean	Std Error
ge,field	-52.713694	2.1191385
ge,forest	-54.631656	1.8938760
sony,field	-37.914345	1.9574053
sony,forest	-48.423686	1.7687386

MicBrand*DeckBran*Habitat
Least Squares Means Table

Level	Least Sq Mean	Std Error
Opt,ge,field	-67.381894	2.8574458
Opt,ge,forest	-57.229750	2.7108111
Opt,sony,field	-37.469376	2.9402875
Opt,sony,forest	-55.267525	2.4746207
Sen,ge,field	-38.045493	3.1301751
Sen,ge,forest	-52.033562	2.6454808
Sen,sony,field	-38.359314	2.5846570
Sen,sony,forest	-41.579848	2.5278444

Differences between Habitats

Vegetation severely attenuates sounds. As expected, sounds reaching the recording devices in the forest habitat were significantly lower in amplitude than those in the field habitat ($F = 14.55$, df = 1, 152, $p = 0.0002$) . The 6 dB difference corresponds to a factor of 2 in amplitude. While this problem is faced by human auditioners as well as recording devices, it does underscore the lower detectability of similar sounds in the two habitats. Estimating detectabilty with the distance method, via amplitude measurements from recordings, would require a separte set of distance-amplitude curves for each species for the two habitats. Because many BBS stops are mosaics of open and forested acoustic environments, the distance-amplitude functions are likely to be stop-specific, depending upon the amount and configuration of the two habitats in the area surrounding the stop. The 6-dB difference in level of the sounds in the two habitats suggests that these differences are not negligible.

This is at least as true, of course, for subjective estimation of distance by the observer as it is for estimation by means of amplitudes measured from recordings. The simplest application of the distance method to point counts (Farnsworth et al. 2002) continues the tradition of relying on training to solve all the problems of monitoring; i.e., there is no independent means of checking the validity of the distance estimates of the observers. Given the varying attenuation of sounds with vegetation, one wonders how accurate such estimates can be. One advantage of estimating distance from recordings is that distance-amplitude functions from open and very-densely vegetated habitats could be used to bracket the true stop-specific distance-amplitude function, yielding a range of detectability factors that likely encloses the parametric value. This approach may produce more accurate counts than observer estimates of distance.

Differences among Microphones

Overall, the Sennheiser ME-62 with K-6 power supply produced recordings of higher amplitude than the Optimus 33-1060 ($F = 28.71$; df = 1, 152; $P < 0.0001$). The overall difference of 12 dB-v in a mixture of conditions represents a four-fold difference in intensity. This is not a surprising result.

Differences among Recorders

Overall, the Sony TCD-5 Pro II deck produced recordings of higher amplitude than the General Electric 3-5025A ($F = 26.21$, df = 1,142 $P < 0.0001$). The GE, like most low-end recorders, has an automatic level control feature that cannot be disabled. High ambient noise, such as wind or a passing airplane, reduces the recording level temporarily. The Sony suffers from the opposite problem, when deployed unattended, as in this application. Sounds that “redline” the recorder overload the system, producing recordings with no information other than the presence of an extreme source of noise. Such sounds were manually excluded from the data set used for all analyses. The mean differences between the two models (10 dB-V) is nearly as great at that between microphone brands.

Interactions

Inspection of the summary statistics for the Microphone * Deck interaction reveals that the Sony-Optimus and GE-Sennheiser combinations produced recordings of equivalent amplitude. In practical terms, this means that one could gain equivalent improvement over the poor performance of the $60 GE-Optimus combinations with upgrades to either the Sennheiser micrphone ($350) or the Sony deck ($1200).

Despite the clear differences in performance of the “cheap” consumer products and “low-end” professional equipment, much of the variation in the results was not explained by these factors. Including in the analysis of variance a term for the specific recorder increased the coefficient of determination (r^2). This indicates that two examples of the same model did not perform equally well. In particular, one of the two Sony recording decks performed rather poorly. Despite these complications, the differences revealed by the main effects of the ANOVA are sufficiently great (in dB-v) to indicate an assessment of their practical consequences.

AUDITION TESTS

Perhaps the significant differences in results summarized above are not operationally significant, i.e., perhaps all recordings are more than loud enough to allow censuses to be conducted from them. The results reported above are based on the amplitudes of spectral peaks. One of the advantages of recordings is that the gain can be “turned up,” compensating in some cases for lower native amplitude on the recording. I performed a spot check of this possibility by means of spectrographic scanning of the raw tapes. In general, test sounds were spectrographically visible on the lower-amplitude tapes when the spectrographic “gain” was elevated (lower “gftlo” setting in Signal). This means that the information was indeed available on the tape, even though it was recorded there at lower amplitude. Two caveats pertain. First, this does not necessarily mean these low-amplitude sounds would be audible to an auditioner of the tapes. Tape hiss and/or sounds at other frequencies might mask the target sounds. Second, these test sounds are among the highest amplitude sounds on the tapes. Their spectrographic visibility does not necessarily mean that fainter sounds visible with more sensitive recording equipment would be visible on the tapes made with less sensitive equipment. Nonetheless, this exercise does suggest that spectrographic visualization can mitigate the effects of less sensitive equipment.

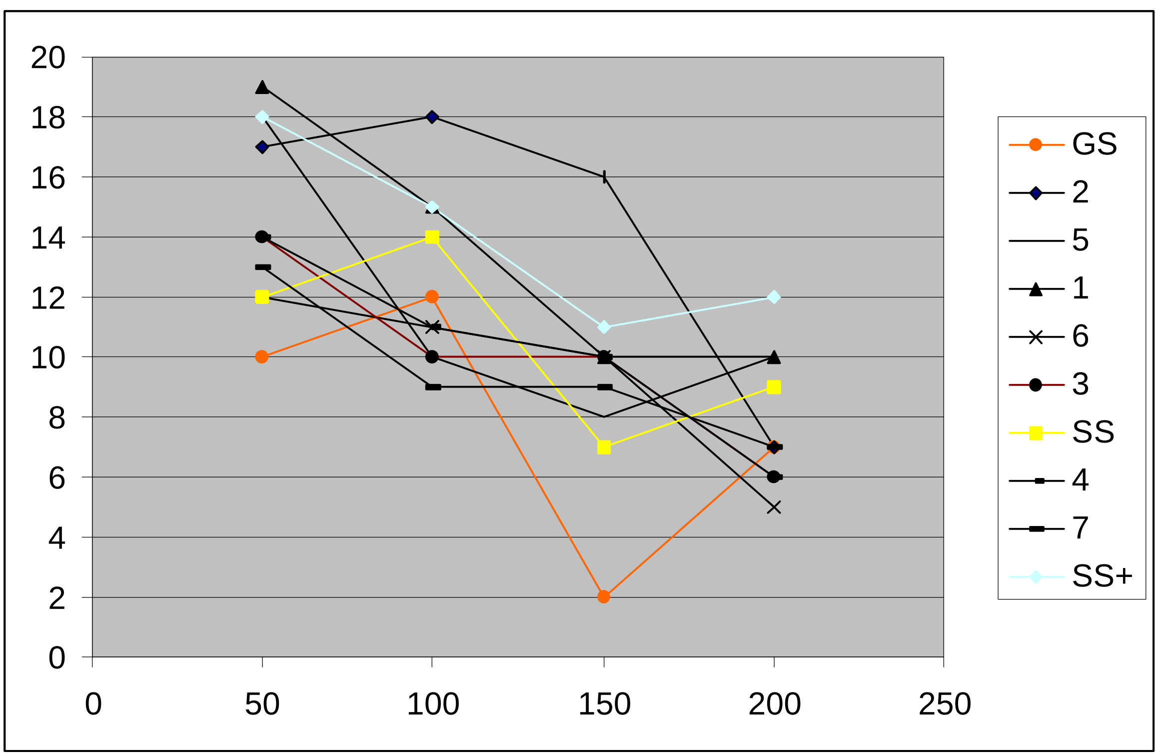
The best test of the adequacy of recordings is to compare detection rate from tapes with detection rate of human observers. An experiment performed on July 2, 2002 at Patuxent provides the data needed for such an assessment. Seven volunteers, all experienced at identifying birds by ear, were tested in the field for their auditory acuity. A battery of test sounds was played, in random order, at distances of 200, 150, 100, and 50 m, in that order. As in the equipment test described above, the speaker was mounted on top of a van, and the van and playback apparatus were positioned along a dirt road at the test distances, measured with a tape measure. The speaker was directed toward the subjects (bearing = 0) and perpendicular to them (bearing = 90) at each distance. The 0-degree placement was performed first at each distance. The 20 test sounds were randomly presented at each of the 8 combinations of distance and bearing.

Each test sound was preceded by a loud alerting tone. The subjects indicated on an answer sheet whether they detected a sound, and, if possible, what kind of sound it was. Nine sounds were tones at a single frequency (1-8 kHz, 10 kHz), one was a 1 kHz tone with 9 harmonics, and ten were recordings of birds from nature. The bird species were

Ruffed Grouse, Bicknell's Thrush, Varied Thrush, American Crow, Cedar Waxwing, Brown Creeper, Black and White Warbler, Brown-headed Cowbird, Dark-eyed Junco, and Northern Cardinal. These were chosen to cover a range of frequencies and note-durations.

In October, 2004, I took the same test, using tapes made on July 2, 2002 when the human subjects were tested. The recording apparatus was set up on tripods near the human subjects. All four combinations of microphone brand and recorder brand used in the equipment test were used in the acuity test. This design replicated brands but not combinations, unlike the main equipment test. The two recordings made with the Optimus microphones were too faint to use. I first tested myself with the tape from a Sennheiser microphone and a GE recorder. After completing that exercise, I took the test again using the tape from a Sennheiser microphone and Sony recorder, which according to the statistical analysis reported above should have preserved more of the soundscape. After scoring and graphing the results, I went back to the Sennheiser/Sony tape again with the answers at hand, and searched for the test sounds on the recordings.

Results are graphed below. Only results for the "head-on" speaker alignment (bearing = 0) are presented.The seven real-time subjects (indicated by numbers) had generally decreasing scores with distance, as one would expect from acoustical theory. Using the Sennheiser/GE tape (curve GS) I performed at a median level at 100 and 200 m, but well below real-time subjects at 50 and 150 m. Using the Sennheiser/Sony tape (curve SS) I was able to improve detections at all distances, but this was not an independent test, and the pattern of the first test re-occurred. With a directed search for the test sounds (curve SS+), I was able to outperform all but one of the real-time subjects.



A significant flaw in the design of this audition test was the presence of singing birds in the soundscape, despite the test's being conducted in mid-afternoon. Most noticeable among the species singing were Blue-gray Gnatcatcher, Carolina Chickadee, Eastern Wood-Pewee, American Goldfinch, and Northern Cardinal. This led to many false positives. For data analysis, I changed "detection" to "nondetection" when the type or species of sound indicated on the answer sheet was clearly indicative of a false positive (e.g, subject wrote "Northern Cardinal" when the test sound was a 7-kHz tone). This reduced but probably did not eliminate the false positives. In addition, persisent insect sound at 4 kHz rendered the 4 kHZ test tone inaudible under almost all test conditions.

Despite these flaws, the results of the audition test do support the conjecture that a recording can capture as many of the sounds as most humans can hear, even at 200 m, which is well beyond the distance to most birds detected on the BBS. My two attempts at detecting test sounds without knowing what they were was stymied by the absence of directional information. Real-time subjects were able to focus their attention in the direction of speakers. When an entire census is taken from a tape, this difference between real-time and post-processing will disappear.

CONCLUSIONS and RECOMMENDATIONS

Although main effects (quality of microphone, quality of recorder) in the equipment test were signficant, they explained a minority of the variation in amplitude of recorded test sounds. Moreover, the Sennheiser/Sony combination did not always outperform other combinations, as the main effects would lead one to expect. Under these sub-optimal circumstances, it seems wisest to trust the main effects. The tests were performed in the field, rather than in the laboratory, to assess the impact of field conditions on variation, and to produce data for distance-amplitude functions of selected species. The speakers was directed at 90 and 180-degree angles from the microphones to emulate the effect of variation in orientation of a singing bird.

Clearly, field conditions introduce considerable variation. Wind, insect sounds, and anthropogenic noise were all obvious on these tapes. A House Wren sang for hours in close proximity to the microphones for the field habitat, masking many of the intended test sounds (but not the alerting sounds that were used for analyses reported here). Despite the usage of identical set-ups in the forest and field tests, including cables as well as main equipment components, the patterns of variation in the two sites were not concordant, implicating stochastic variation in ambient conditions.

Using recordings in lieu of live observers can potentially reduce identification error, if the persons post-processing the recordings are more expert than the field workers they replace and/or spectrographic visualization is used. But, these results suggest, inter-observer variability in detection rate will simply be replaced by inter-machine variability. The Large Scale Ecology Program of the Ontario Ministry of Natural Resources, (http://flash.lakeheadu.ca/~rrempel/ecology/) has enthusiastically adopted recordings as a solution to the identification problem (http://flash.lakeheadu.ca/~rrempel/CVX/). Whether they have engaged the machine-variability problem is not known.

Recent improvements in consumer-grade equipment may make recording point counts more feasible economically. Digital recording has several technical advantages over analog recording, but until recently digital recording platforms were either very expensive, or used compression techniques that raised issues about the fidelity of the recording to the soundscape. The Macaulay Library (formerly Library of Natural Sounds) of the Cornell Lab of Ornithology has monitored technical developments in this arena for many years, and regularly publishes up-to-date assessments of field recording equipment on its website (www.birds.cornell.edu/MacaulayLibrary/equip). According to this site, two welcome improvements in mini-disc recording technology have occurred. First, improvements in the ATRAC compression technique make it less problematic for recording bird sounds. Second, and even more encouraging, a new mini-disc format, HiMD, can record 1.5 hr of material, in stereo (allowing some localization of sounds), on a single mini-disc, without compression. Other technologies (digital hard drive, digital solid state), although currently not cost effective for monitoring operations, may offer additional options in the near future.

As mentioned above in the “Soundscape” section of the Introduction, monitoring birds with point counts and other brief survey protocols faces the twin challenges of accuracy and detection probability. Improvements in accuracy from post-processing of recordings are assessed in the companion report. A pre-condition of using recordings to improve accuracy is a detection rate of sounds on recordings that is at least equal to that obtainable in real time in the field. The audition tests reported here suggest that recordings can “detect” sounds as acutely as human observers. This is the most encouraging part of the present report, and suggests that further exploration of equipment options is warranted.

Acknowledgments

I would like to thank Keith Pardieck and Bruce Peterjohn for their able administration of this contract, and Mary Gustafson, Mark Rimer, Sam Droege, Marshall Howe, Keith Pardieck and Bruce Peterjohn for participating in the auditory acuity tests.

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.

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.

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.

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.

Weisman, R., and L. Ratcliffe. 2004. Relative pitch and the song of Black-capped Chickadees. American Scientist 92:532-539.