

Testing Automated Call-recognition Software for Winter Bird Vocalizations

Andrew Wolfgang^{1,2} and Aaron Haines^{1,*}

Abstract - Automated recording devices and call-recognition software are technologies available to survey vocal fauna. We evaluated the Song Scope[®] call-recognition software designed by Wildlife Acoustics to automatically identify vocalizations of 3 North American winter bird species. We used a Wildlife Acoustics SM-2 automated recorder to record winter avian vocalizations and then screened these field recordings using the Wildlife Acoustics Song Scope software programmed with recognition models, or recognizers, we created. Song Scope correctly identified an average 39% of target vocalizations to species using featured recognizers, with some recognizers performing better than others (accuracy range = 20–59%). Screening a 10-h field recording with Song Scope took an average of 7 minutes per recognizer. Call-recognition software can be used to survey vocal species; however, when biologists use this software to determine species presence or density, they need to be aware of potential bias in survey results because some species-recognizer models perform better than others.

Introduction

The ability to assess bird abundance and diversity by recording and identifying vocalizations is a valuable tool in avian ecology (Blumstein et al. 2011, Kroodsmma and Budney 2011). A number of studies have experimented with surveying populations using automated recording devices and recognition software to identify species by vocalizations (Brandes 2008, Buxton and Jones 2012, Holmes et al. 2014, Lopes et al. 2011, Venier et al. 2012, Waddle et al. 2009). Automated identification requires less time than manual scanning of recordings, but has been reported to lack the accuracy of trained surveyors (Swiston and Mennil 2009). However, automated detection of vocal species can be more efficient and improve detection probability because modern recording devices can be left in the field to survey either continuously or during programmed time-intervals (Acevedo and Villaneuva-Rivera 2006, Holmes et al. 2014, Venier et al. 2012). Seasonal changes, time restrictions, and weather events often limit efforts for surveyors, but the limitations can be reduced by using automated recorders (Bas et al. 2008, Bridges and Dorcas 2000, Diefenbach et al. 2007). Sound-recognition software programs such as Raven[®] (Duan 2013), XBAT (Brandes 2008, Swiston and Mennil 2009), and Song Scope[®] (Buxton and Jones 2012, Duan 2013, Holmes et al. 2014) can be programmed to attempt to distinguish wildlife species by their recorded vocalizations. These recordings can provide a permanent, biologically important library,

¹Department of Biology, Applied Conservation Lab, PO Box 1002, Millersville University, Millersville, PA 17551. ²Current address - 34001 Mill Dam Road, Wallops Island, VA 23337. *Corresponding author - aaron.haines@millersville.edu.

where digital sound files are available for review and re-analysis (Blumstein et al. 2011, Kroodsma and Budney 2011).

Audio signals from field recordings have many characteristics, frequencies, and syllables (Brandes 2008), and there are various methods for isolating and classifying them (Blumstein et al. 2011, Brandes 2008). Song Scope software analyzes an entire audio-signal structure using Hidden Markov Models (HMM) (Duan 2013), where coefficients record changes in patterns within the harmonic structure as the signal is processed (Brandes 2008). HMMs and other signal-processing techniques have been used to effectively identify various birds in the family *Thamnophilidae* (antbird species), as well as *Setophaga cerulea* (Wilson) (Cerulean Warbler), *Empidonax vireescens* (Vieillot) (Acadian Flycatcher), and *Passerina cyanea* L. (Indigo Bunting) songs (Holmes et al. 2014, Kirschel et al. 2009, Kogan and Margoliash 1998, Trifa et al. 2008).

However, accuracy of sound-recognition software to identify particular species calls has varied. Duan (2013) reported accuracies of only 37%, while Buxton and Jones (2012) reported accuracies of greater than 50%. Both of these studies used Song Scope software. In our study, we used Song Scope (v. 4.1.1) recognition software and Wildlife Acoustics hardware (automated recording device SM-2) (Wildlife Acoustics 2011a). These tools have been used to verify presence of avian woodland species including Cerulean Warbler and Acadian Flycatcher (Holmes et al. 2014). Our objectives for this study were to assess the accuracy of Song Scope software, document presence of 3 winter woodland species, and determine whether Song Scope is a tractable option for field biologists with no experience using signal-processing software. We believe that validating the accuracy of call-recognition software is important in order to avoid bias in species-detection rates.

Field-site Description

We mounted the Song Meter[®] recording device (SM-2, Wildlife Acoustics, East Lansing, MI) facing northeast at a height of 2 m on a tree in a mixed deciduous forest on the Millersville University Biological Preserve west of the Conestoga River in Millersville, PA (39°99'5"N, 76°34'63"W). The habitat around the device was characterized by large *Liriodendron tulipefera* L. (Yellow Poplar) and *Platanus occidentalis* L. (American Sycamore). The mid-story was primarily *Acer rubrum* L. (Red Maple) with a thick shrub layer of *Rosa multiflora* Thunb. (Multiflora Rose). The proximity of the Conestoga River and a nearby road created an edge habitat utilized by birds in the winter months.

Methods

We tested Song Scope using 3 focal winter bird vocalizations: the “jay” calls of *Cyanocitta cristata* L. (Blue Jay), the basic “tea-kettle” songs of *Thryothorus ludovicianus* (Latham) (Carolina Wren), and the “chick-a-dee-dee-dee” calls of *Poecile carolinensis* (Audubon) (Carolina Chickadee) (Fig. 1). We chose these target vocalizations because they are consistently heard during the winter season in eastern temperate forests of the US.

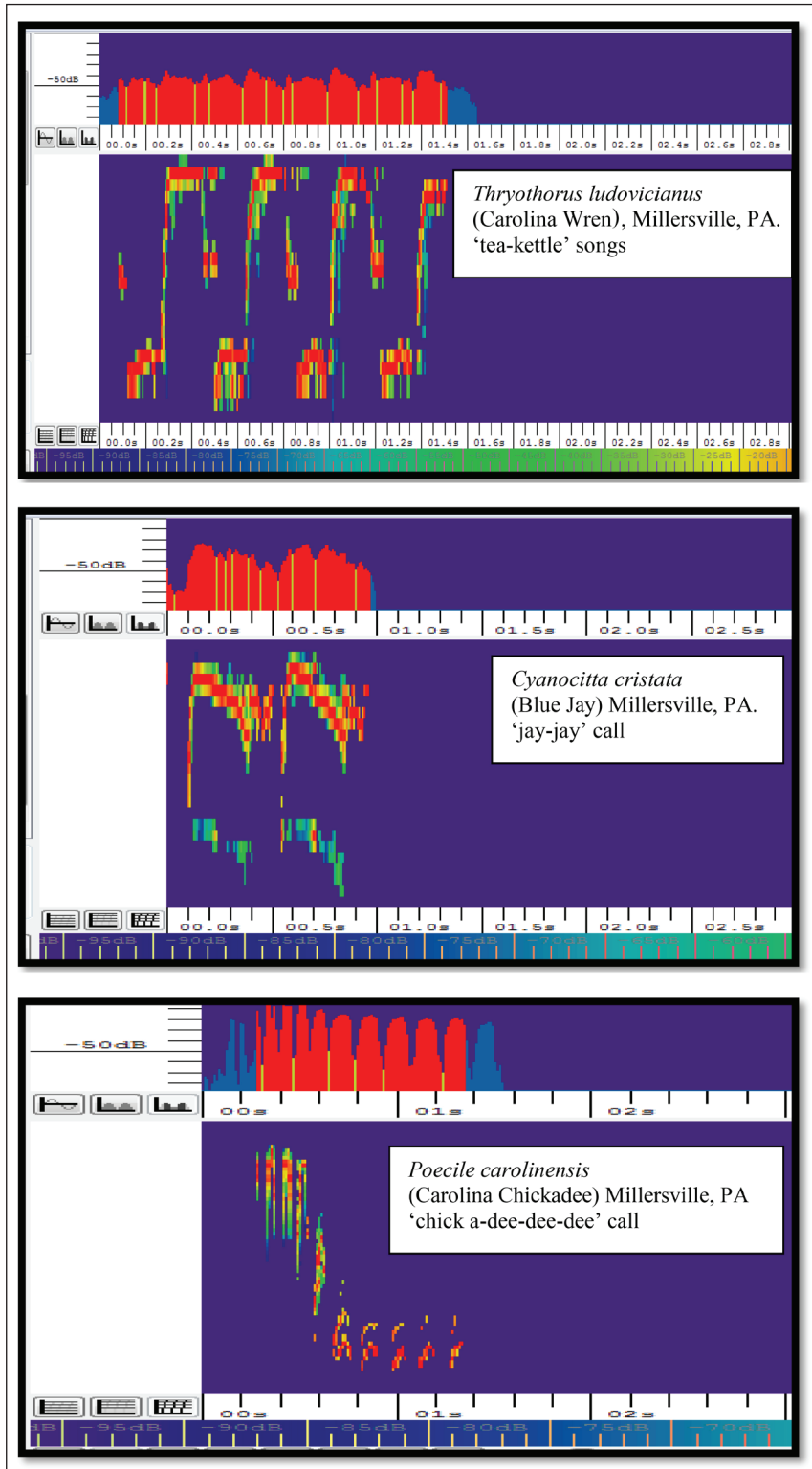


Figure 1. Spectrograms from Song Scope annotations showing the diversity of call structure between target vocalizations of 3 focal species.

We placed the Song Meter at the site to ensure that microphone settings and automated recording times were optimal for recording a large number of bird vocalizations. We then deployed the SM-2 with Wildlife Acoustic default audio settings and SMX-II microphone settings. All channels were stereo, left and right microphones were set at +40.0 dB, sample rate was 16,000 hertz and all files were stored in .wav format. We used these settings for the duration of the study. We first deployed the SM-2 device on 27 September 2013 and last retrieved it on 23 December 2013. During that interval, we sampled for reference calls and test vocalizations on a total of 33 days.

We collected reference vocalizations to use as training data to build song recognizers. Each species had a minimum of 15 high-quality reference vocalizations. Buxton and Jones (2012) built models using 2–5 high-quality reference vocalizations, while Holmes et al. (2014) used over 100 various vocalizations. Song Scope uses a “recognizer” created from reference vocalizations to isolate and classify signals within a field recording. Recognizers are recognition models that search field recordings for a match based on the features of the reference vocalizations from which they were created. Success of recognizers is dependent on the purity of pre-recorded reference vocalizations and correct model parameters (Wildlife Acoustics 2011b). Reference vocalizations must first be annotated in Song Scope, and then annotations can be grouped into a usable recognizer.

We recorded reference vocalizations with the SM-2 automated recording device from 25 October to 7 November 2013 and 26 November to 4 December 2013. The SM-2 was programmed to record from 0845 to 0900 when the target birds were most vocal during the winter. We obtained other reference vocalizations from Thayer Birding Software (2012) and hand-held recordings of target birds. For hand-held recordings, we used a Tascam DR-05 device (TEAC, Montebello, CA) from 0800 to 1000 on 22 December 2013 at Pinchot State Park (PSP), located ~40 mi west of Millersville in York, PA.

After annotating reference vocalizations, we adjusted Song Scope recognizer parameters to create the best recognizer using a “cross-training score” (see Wildlife Acoustics 2011b) for each recognizer based on its ability to identify vocalizations in trial recordings. We created recognizers with selected model parameters set to match the target structure of a reference vocalization for a specific species (Buxton and Jones 2012). Spectrograms show the diversity of our targets (Fig. 1); therefore, recognizer parameters had to be species-specific for our study. Recognizer-model parameters have values that can be manipulated based on the characteristics of a target vocalization (Table 1); a more in-depth explanation is outlined in the Song Scope manual (Wildlife Acoustics 2011b).

We tested recognizer parameters using 15-min field recordings of target vocalizations (trial recordings) from our study site. We changed parameters by trial and error until the recognizer identified all (100%) of the trial-recording target vocalizations. These targets were then assimilated into the recognizer as additional reference vocalizations to create an “improved recognizer.” We tested the improved recognizer for 100% accuracy on a second 15-min field recording of target vocalizations. We used 3 sets of reference vocalizations (i.e., SM-2 recordings, Thayer software, hand-held

recordings) to build 3 separate models for each target species. Different combinations of calls can be combined into a single recognizer model, allowing the model to capture and correctly identify a greater variation of species vocalizations (Wildlife Acoustics 2011b). Thus, we ran all 3 reference-vocalization models simultaneously for each target species. These combined models performed better than any solitary recognizer for each target species, and we used them for our analysis.

We tested improved recognizers on 30-min field recordings taken at 0800 and 1530 from 12 to 23 December 2013. We obtained a total of 10 hours of field recordings, which we later screened for target vocalizations with the Song Scope program. We considered complete vocalizations that stood out over filtered background noise (registering 20–40 dB) to be quality vocalizations and incomplete vocalizations or those that lacked the volume (registering 0–20 dB) to stand out over filtered noise at any given point in the recording as non-quality vocalizations. Song Scope has the ability to screen many hours of recordings at once by “batching” the recordings from separate recording times together; our 10 usable hours of field recordings were batched. We evaluated the Song Scope recognizer models by manually reviewing all 10 h of audio. If the software correctly matched a quality target vocalization, we defined this as a true positive, a non-target vocalization that was recognized as a target was a false-positive, and if the software missed a quality target vocalization which we identified, it was labeled a false-negative.

Results

After manual evaluation, we classified 20% of all recorded vocalizations (317/1280) as quality vocalizations (Carolina Wren [83 of 135], Carolina Chickadee [54 of 661], and Blue Jay [180 of 484]). We used these quality vocalizations to test the Song Scope software. The mean total accuracy for our featured recognizer models using Song Scope was 39%, based on true positives among quality calls. The highest percentages of true positive calls identified were 59% for the Carolina Wren, 39% for the Blue Jay, and 20% for the Carolina Chickadee (Table 2). In addition,

Table 1. Song Scope parameters needed to develop a recognition model (Wildlife Acoustics 2011b).

Song Scope parameter	Definition
Maximum song duration	The expected duration of a song or call
Maximum syllable duration	The expected duration of a syllable in a song or call
Maximum syllable interval	The expected duration of the time between a syllable in a song or call
Minimum frequency	The lowest frequency used in the song or call
Frequency range	The span of frequency for the song or call
Dynamic range	This sets a level of decibels to normalize song scope with peak song or call levels
Maximum complexity	The maximum number of Hidden Markov Models used by Song Scope in a recognizer
Maximum resolution	Maximum number of feature vectors used by a Song Scope recognizer
Background filter	Puts emphasis on song or call and cuts stationary noise
Sample rate	The rate at which the song or calls are analyzed by the recognizer
FFT	Fast Fourier Transform window size

Carolina Wren had the lowest number of false positives (13), while Carolina Chickadee had the highest (61).

Manually screening the field recordings for our focal species took 12 h, or an average of 4 h per species. Automated identification for these species using Song Scope took a total of 15 h and 21 min, or an average of 5 h and 7 min per species, including 4.5 h to create a featured recognizer for each species (Table 3). However, after recognizers were created, Song Scope, on average, processed the 10-h field recording in 7 minutes, and it took us 30 min to evaluate Song Scope's findings after the program displayed results. Not including the time needed to create a recognizer, identification of audio vocalizations using the Song Scope program took on average 37 mins per species for a 10-h recording (Table 3).

Discussion

We used percent of accurate identifications to evaluate Song Scope's ability to make correct identifications from quality vocalizations. Song Scope was not able

Table 2. Accuracy of Song Scope automated recognition software from Wildlife Acoustics identifying target vocalizations from 3 bird species in Millersville, PA. True positives = the number of quality calls of the species correctly identified. False positives = number of misidentifications of sounds that were not calls of the species as calls of the species. False negatives = number of undetected quality calls of the species. Accuracy = the percentage of quality calls of the species correctly identified.

Species	Quality calls (>20 dB)	True positives	False positives	False negatives	Accuracy
<i>Cyanocitta cristata</i> (Blue Jay)	180	71	29	109	39%
<i>Poecile carolinensis</i> (Carolina Chickadee)	54	11	61	43	20%
<i>Thryothorus ludovicianus</i> (Carolina Wren)	83	50	13	33	59%

Table 3. Time required to identify 3 winter bird species from 10-h field recordings in Millersville, PA, using manual identification and the Song Scope automated detection recognition software available from Wildlife Acoustics. The amount of time is variable dependent on skill level, number of calls used to make up each recognizer, and density of calls within field recordings.

Method Used	Total time required (h)	Average time required per species (h)
Manual identification	12.00	4.00
Automated identification, using Song Scope with newly created recognition models		
Gather reference calls	1.50	0.50
Prepare/annotate reference calls	3.00	1.00
Toggle recognizer settings	4.50	1.50
Verification trials for recognizer	4.50	1.50
Song Scope screening	0.36	0.12
Process Song Scope results	1.50	0.50
Total	15.36	5.12
Automated identification, using Song Scope with pre-made recognizer models		
Song Scope screening	0.36	0.12
Process Song Scope results	1.50	0.50
Total	1.86	0.62

to isolate faint or fragmented calls over background noise; therefore, it could not classify a large percentage of targets that a trained listener could both isolate and classify. However, our definition of accuracy was the software's ability to identify the vocalizations that it was trained to identify, not identify partial songs, interrupted songs, or faint calls just barely audible over background noise. Identifications of these cryptic vocalizations would require the development of an infinite number of models for partial songs and calls with various sources of background noise.

The software correctly identified our focal species an average of 39% of the time using quality vocalizations, and we found the identification accuracy of Song Scope was species dependent. We assume that this species-dependent accuracy is a function of the different categories of sound used by each species in their vocalizations (Brandes 2008). The Carolina Wren song has spectrogram components similar to frequency-modulated whistles with little harmonic structure (Brandes 2008), and of the 3 species tested, was most successfully identified by SongScope (Table 2). It has been reported that HMMs often fail to identify birds with strong harmonics such as the call of the Carolina Chickadee and especially the Blue Jay (Fig. 1; Brandes 2008). The variable accuracy of species-recognizer models creates the potential to underestimate or overestimate species presence when surveying for different vocal species because one species model may have lower accuracy in correctly identifying target calls than another (e.g., for Blue Jays or Carolina Chickadees vs. for Carolina Wrens in our study) or a higher rate of false positives (e.g., for Carolina Chickadees vs. for either Blue Jays or Carolina Wrens in our study). When comparing presence/absence and/or population density between species, we recommend that field biologists compare the accuracy of the different recognizer models they use to account for potential bias in detection rates.

Song Scope can detect diverse targets, and performs best with lower call volumes. True positives were instances when the software correctly identified targets, and false negatives were instances when the targets were missed or skipped. These false negatives occurred due to background noise and microphone limitations, a known problem with automated identification (Blumstein et al. 2011, Brandes 2008, Buxton and Jones 2012). Call identification is heavily influenced by background noise; thus, it is important to ensure that future studies incorporate multiple survey sites to account for various sources of background noise.

Buxton and Jones (2012) reported higher accuracy (56–69%) in audio-call identification using Song Scope than the 39% accuracy we report here with the same program. Buxton and Jones (2012) surveyed seabirds with simple calls on isolated islands and had a specialist from Wildlife Acoustics help develop their recognizer models. Duan (2013) recommended that an expert in audio-signal interpretation should manipulate model parameters to improve accuracy and reduce time in developing recognizers using Song Scope. Wildlife Acoustics acknowledges that patience and time are needed to create optimum parameters for recognizers (Wildlife Acoustics 2011b). Therefore, the use of acoustic-identification software is only a tractable option to the average field biologist if they are able to dedicate significant effort in developing a recognizer model. We recommend that field biologists take a

training workshop on the use of sound-recognition software if they wish to develop more reliable species-specific recognizer models.

The time needed to create recognizers is substantial (5.12 h per species; Table 3); however, when a good recognizer model is developed, it can be used repeatedly for multiple survey efforts (Holmes et al. 2014), making the method very time-efficient, even considering the initial time investment in model development. Waddle et al. (2009) suggested that biologists develop a database to archive accurate recognizers developed by experts utilizing programs such as Song Scope. Researchers could benefit from a robust online library full of recognizer-like files created by experts using standard software parameters. After creating species-specific recognizer models, we found that running these recognizers using Song Scope's batch feature can be done quickly. With the use of developed recognizers, Song Scope was able to screen and process results for each species in an average of 37 minutes (Table 3). Comparatively, manual screenings took 4 h per species. An online database of recognizers would make automated identification simpler and faster.

A shared database of recognizer models for sound-recognition software should be a multi-regional collaborative effort and include many hours of recordings. In addition, the database should cover the needs of users working with various sound-recognition software programs; thus, it should archive different types of "recognizer"-like files and be comprised of a diverse library of vocalization files. Once this database is created, researchers intending to survey for a diversity of species with automated software and equipment will save hours of time because they will not have to develop their own identifiers. As the use of remote recording devices and call-identification software increases, the community of researchers who employ the technology will grow. This trend will create a demand for the development of an online database to provide species-specific recognizer models for different sound-recognition software packages.

Song Scope has been successfully used for presence/absence surveys of vocal species (Buxton and Jones 2012, Holmes et al. 2014) and performs best when background noise is minimal. However, biologists must consider the accuracy of recognizer models. When using multiple species-recognizer models, field biologists should be aware that a model with low detection ability may bias survey results and under-represent the presence or density of species in the field in comparison to species for which available models are more accurate.

Acknowledgments

We acknowledge Millersville University and the Millersville University Student Grant for Independent Research for funding this project. We thank B. Horton for reviewing this manuscript.

Literature Cited

Acevedo, M.A., and L.J. Villanueva-Rivera. 2006. Using automated digital-recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34:211–214.

- Bas, Y., V. Devictor, J.P. Moussus, and F. Jiguet. 2008. Accounting for weather and time-of-day parameters when analyzing count data from monitoring programs. *Biodiversity and Conservation* 17:3403–3416.
- Blumstein, D.T., D.J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J.L. Deppe, A.H. Krakauer, C. Clark, K.A. Cortopassi, S.F. Hanser, B. McCowan, A.M. Ali, and A. Kirschel. 2011. Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations, and prospectus. *Journal of Applied Ecology* 48:758–767.
- Brandes, T.S. 2008. Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International* 18:163–173.
- Bridges, A., and M. Dorcas. 2000. Temporal variation in anuran calling behavior: Implications for surveys and monitoring programs. *Copeia* 2000:587–592.
- Buxton, R.T., and I.L. Jones. 2012. Measuring nocturnal seabird activity and status using acoustic recording devices: Applications for island restoration. *Journal of Field Ornithology* 83:47–60.
- Diefenbach, D., M. Marshall, and J. Mattice. 2007. Incorporating availability for detection in estimates of bird abundance. *Auk* 124:96–106.
- Duan, S., J. Zhang, P. Roe, J. Wimmer, X. Dong, A. Trusking, and M. Towsey. 2013. Timed probabilistic automaton: A bridge between Raven and Song Scope for automatic species recognition. Pp. 1519–1524, *In* H. Muñoz-Avila and D.J. Stracuzzi (Eds.) *Proceedings of the 25th Innovative Applications of Artificial Intelligence Conference*, Bellvue, WA.
- Holmes, S.B., K.A. McIlwrick, and L.A. Venier. 2014. Using automated sound-recording and analysis to detect bird species at risk in southwestern Ontario woodlands. *Wildlife Society Bulletin* 38:591–598.
- Kirschel, A.N., D.A. Earl, Y. Yao, I.A. Escobar, E. Vilches, E.E. Vallejo, and C.E. Taylor. 2009. Using songs to identify individual Mexican Antthrush, *Formicarius moniliger*: Comparison of four classification methods. *Bioacoustics* 19:1–20.
- Kogan, J., and D. Margoliash. 1998. Automated recognition of bird-song elements from continuous recordings using dynamic time warping and Hidden Markov models: A comparative study. *Journal of the Acoustical Society of America* 103:2185–2196.
- Kroodsma, D., and G.F. Budney. 2011. Sound Recordings: An essential tool for conservation. *Conservation Biology* 25:851–852.
- Lopes, M.T., L.L. Gioppo, T.T. Higushi, C.A. Kaestner, C.N. Silla, and A.L. Koerich. 2011. Automatic bird-species identification for a large number of species. Pp. 117–122, *In* *Proceedings of the 2010 IEEE International Symposium on Multimedia*, Taichung, Taiwan.
- Swiston, K.A., and D.J. Mennill. 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed Woodpeckers. *Journal of Field Ornithology* 80:42–50.
- Thayer Birding Software. 2012. Thayer's Birds of North America, Version 4.5 (DVD). Thayer Birding Software, Naples, FL. Available online at <http://www.thayerbirding.com/>. Accessed 15 September 2013.
- Trifa, V.M., A.N. Kirschel, C.E. Taylor, and E.E. Vallejo. 2008. Automated species-recognition of antbirds in a Mexican rainforest using Hidden Markov models. *The Journal of the Acoustical Society of America* 123:2424–2431.
- Venier, L.A., S.B. Holmes, G.W. Holborn, K.A. McIlwrick, and G. Brown. 2012. Evaluation of an automated recording-device for monitoring forest birds. *Wildlife Society Bulletin* 36:30–39.

- Waddle, J.H., T.F. Thigpen, and B.M. Glorioso. 2009. Efficacy of automatic vocalization-recognition software for anuran monitoring. *Herpetological Conservation and Biology* 4:384–388.
- Wildlife Acoustics. 2011a. Song Scope: Bioacoustics Software, Version 4.1.1. Wildlife Acoustics, Inc. Maynard, MA. Available online at <http://www.wildlifeacoustics.com/product/analysis-software>. Accessed 15 September 2013.
- Wildlife Acoustics. 2011b. Song Scope User Manual: Bioacoustics Software (Version 4.0) Documentation. USA: Wildlife Acoustics, Inc. Maynard, MA. Available online at <http://www.wildlifeacoustics.com/images/documentation/Song-Scope-Users-Manual.pdf>. Accessed 15 September 2013.