Automatic Detection of Cerulean Warblers

Using Autonomous Recording Units and Song Scope Bioacoustics Software

Ian Douglas Agranat, President, Wildlife Acoustics, Inc., http://www.wildlifeacoustics.com/, November 19, 2007

1. Abstract

Wildlife Acoustics'® Song Scope™ software (version 1.9) was used to automatically locate and detect Cerulean warbler vocalizations from over 250 hours of field recordings made in 9 different known and potential Cerulean warbler *(Dendrocia cerulea)* territories.

After developing a Cerulean warbler "recognizer", Song Scope found 4,328 potential Cerulean vocalizations of which 1,552 could be confirmed with only one hour of human labor. The automatic scan took 12 hours to complete on a 3.2GHz Pentium 4 under RedHat Linux.

The effort to develop the Cerulean warbler recognizer took less than 3 days, including a "first pass" using only high quality recordings of known Cerulean warblers for training data, and a "second pass" using only a few selected field recordings for training data. The recognizer can now be re-used to scan additional recordings without further development effort.

The human effort to review some 250 hours of audio recordings to count Cerulean warbler vocalizations by a person experienced in the use of Song Scope software is reduced by approximately 250 times using Song Scope software compared to physically listening to the recordings, and by approximately 90 times by an experienced Song Scope software user compared to visually inspecting spectrograms of the recordings.

2. Introduction

The US Forest Service, San Dimas Technology and Development Center (SDTDC), in cooperation with the Cornell Lab of Ornithology, Bioacoustics Research Program (CLO, BRP) and Wildlife Acoustics, Inc. (WLA) is investigating means to promote bioacoustic technology as a cost-effective tool for the inventory and monitoring of forest wildlife.

Researchers in the eastern US are actively engaging in extensive inventory and monitoring efforts of the Cerulean warbler, as past surveys indicate the population of this species has undergone a steady decline over the past 30 years. In fact, Cerulean populations have declined by over 70% since 1966 based on Breeding Bird Surveys.

In the spring/summer field season of 2007, the CLO, BRP and SDTDC cooperated in a "demonstration deployment" of 3 of CLO, BRP's Autonomous Recording Units (ARUs) by research biologists at the USFS' Northern Research Station, Forestry Sciences Laboratory (Irvine, PA) who have been actively engaged in the monitoring of Cerulean warblers on and near the Allegheny National Forest. The 3 ARUs were placed in different known and highly suitable habitats for Cerulean warblers to at least confirm/document the presence/absence of this species.[1]

The 3 ARUs each made autonomous, continuous recordings between 6:00-11:00AM for several days at a time at each location. They were then all redeployed two more times to cover a total of 9 known and potential Cerulean territories. About 251 hours of audio recordings were collected as detailed in Table 1. Wildlife Acoustics' Song Scope software was then used to automatically scan these recordings for Cerulean vocalizations.

In the "first pass", a Song Scope "recognizer" was developed using several high-quality recordings of known Cerulean warblers from the archives of the Macaulay Library at the Cornell Lab of Ornithology. Then in a "second pass", a recognizer was developed from selected Cerulean vocalizations in the ARU recordings detected in the first pass.

3. Creating a First Pass Recognizer

The Song Scope software automatically builds a classifier (known as a "recognizer") from training data consisting of known vocalizations and a number of input parameters. The resulting recognizer is a statistical model that considers both spectral and temporal characteristics of the vocalization as well as its variability across the training data. Specifically, the recognizer considers how syllables (short bursts of sound such as syllables in human speech) are formed by changing frequency components through time, and how groups of

---

1. Details on the monitoring effort by the Northern Research Station for Cerulean warblers can be obtained from Dr. Scott Stoleson at SStoleson *at* fs.fed.us.

syllables are combined into the vocalization (e.g. to form a song).

Given a set of recordings containing known vocalizations of Cerulean warblers, a human operator can annotate these recordings to indicate where the Cerulean warbler vocalizations are located in time and frequency.

Table 1: Deployments of Autonomous Recording Units

ARU-Deployment	Dates Deployed	Hours Recorded	Location	Cerulean Territory
ARU09-A	Jun 22-Jun 25	20.33	State Game Lands 86	Yes
ARU09-B	Jun 30-Jul 6	35.58	Forest Road 488	Unknown
ARU09-C	Jul 12-Jul 17	28.67	Forest Road 139	Unknown
ARU18-A	Jun 15-Jun 19	23.58	Bug Hollow Adjacent Site	Yes
ARU18-B	Jun 20-Jun 26	34.75	Forest Road 116	Yes
ARU18-C	Jul 11-Jul 16	30.42	Forest Road 160	Unknown
ARU20-A	Jun 19-Jun 26	36.58	Wolf Run	Yes
ARU20-B	Jul 6-Jul 8	13.25	Forest Road 256 & 256A	Unknown
ARU20-C	Jul 11-Jul 17	27.92	Forest Road 364	Unknown

Table 2: First Pass Training Recordings of Cerulean Warblers from the Macaulay Library at the Cornell Lab of Ornithology

MLNS #	Date	Location	Recordist	# Annotations
9182	May 12, 1964	New York	Stein, R.C.	9
26267	April 24, 1976	Ohio	Little, R. S.	8
77293	April 27, 1996	W. Virginia	Hershberger, W. L.	6
77296	April 28, 1996	W. Virginia	Hershberger, W.L.	19
85100	June 6, 1997	W. Virginia	Hershberger, W.L.	5
85116	June 7, 1997	W. Virginia	Hershberger, W. L.	3
85122	June 8, 1997	Maryland	Hershberger, W.L.	29
85129	June 9, 1997	Maryland	Hershberger, W.L.	10
105649	May 28, 1995	Indiana	Keller, G. A.	11
110206	May 27, 2001	W. Virginia	Hershberger, W.L.	12
113505	May 13, 1998	New York	Marantz, C. A.	25

Figure 1: Spectrogram of Cerulean vocalization from New York (MLNS # 9182)
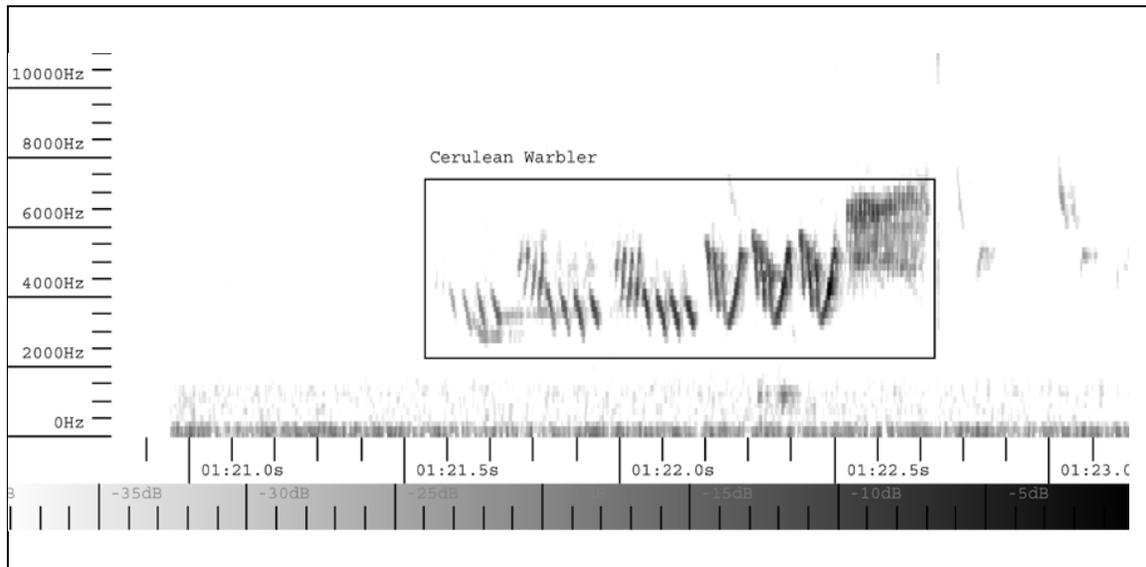
Figure 2: Spectrogram of Cerulean from West Virginia (MLNS #110206)
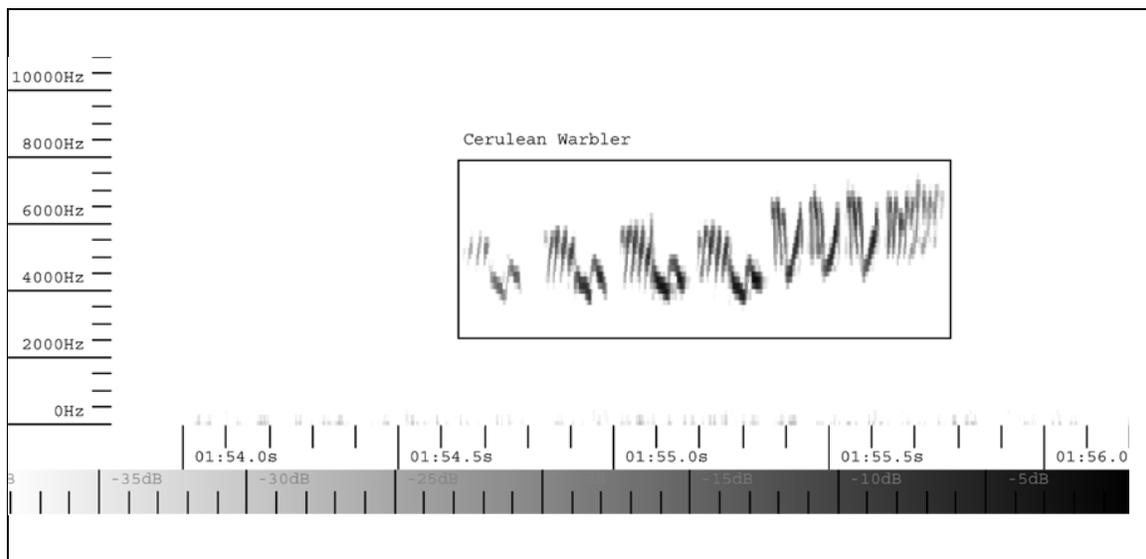
3.1. Training Data

To begin, eleven high-quality known Cerulean warbler recordings from the archives of the Macaulay Library at the Cornell Lab of Ornithology[1] were chosen as a source of training data representative of the range of song variation among individuals of the species. Most of

these recordings were made with parabolic reflectors resulting in a very high signal-to-noise ratio (SNR). Table 2 lists the recordings that were used.

Representative spectrograms from two of these recordings are shown in Figure 1 and Figure 2 for illustration purposes. Notice that these two songs are very similar in many ways, and yet are also different illustrating the fact that no two individuals will produce exactly the

1. See http://www.animalbehaviorarchive.org/

same song just as no two human beings sound exactly alike. It is essential to provide a wide range of training data to Song Scope so that the software can accurately model the expected range of variation across the species in order to detect the vocalizations of new individuals that were not present in the training data.

A total of 137 vocalizations were manually identified and annotated from these eleven recordings taking a human operator about 3 hours using Song Scope.

3.2. Parameters

In addition to training data, several parameters are also used by Song Scope to generate a recognizer. These parameters must be chosen by a human operator. The following is a somewhat technical discussion of the parameters used to develop the recognizer. A general discussion on acoustics and more details relative to the parameters used in Song Scope are available in the documentation for the software (see *http://www.wildlifea-coustics.com/songscope_web_help/).*

Sample Rate

The sample rate parameter determines the sample rate at which the ARU field recordings will be analyzed by the recognizer when scanning for matching vocalizations. A sample rate of 20,000 samples per second was chosen because this is the sample rate used by the ARUs. The fundamental components of Cerulean vocalizations are generally under 7500Hz and require a sampling rate of at least two times, or 15,000 samples per second, to resolve.

Fast Fourier Transform (FFT) Parameters

By default, Song Scope uses a 256-point FFT with 50% overlapping windows. The spectrograms produced with these settings at 20,000 samples per second for Cerulean warbler vocalizations appears to provide a good compromise between temporal and spectral resolution, so the default values were chosen for the recognizer

Band Pass Filter

Song Scope recognizers are configured to consider only a specific band of frequency bins from the FFTs. The band should be as narrow as possible to screen out as much interference from noise and other sources, yet be wide enough to capture the identifiable characteristics of the desired vocalization across a range of natural variation among individuals of the species.

In the eleven MLNS recordings of Ceruleans, none of the vocalizations appeared to have significant fundamental frequency components below 2800Hz or above 7500Hz. The Frequency Minimum parameter was set to bin 36 (36 * 20000 / 256 = 2812.5Hz) and the Frequency Range parameter was set to 60 bins (up to bin 36 + 60 = 96; 96 * 20000 / 256 = 7500Hz).

Background Filter

Song Scope has a simple filter (based on a Wiener filter) to de-emphasize stationary noise (noise that is somewhat constant in frequency and time such as a distant highway). The filter can improve the signal-to-noise ratio by as much as 6dB against stationary background noise.

A noise estimate is made by tracking the rolling average power level in each frequency bin over a fixed window of time. For most applications, a one second window size is recommended and this value was chosen for the recognizer.

Dynamic Range

Song Scope normalizes signals to a fixed dynamic range (in decibels) relative to peak signal levels. This value is used in two different ways:

First, the dynamic range parameter is used in a signal detection algorithm to look for candidate vocalizations (i.e. vocalizations of interest) with peak signals exceeding the dynamic range above background noise. In other words, the Song Scope detector will look for candidate vocalizations with a peak-to-peak signal-to-noise ratio exceeding the dynamic range value. Smaller values will allow for weaker signals (e.g. quieter and/or more distant vocalizations) to be considered. Note that this value should not exceed the expected signal-to-noise ratio of candidate vocalizations or they will not be considered.

Second, the dynamic range parameter is also used to normalize the spectral features relative to the peak value within the frequency band for a given slice of time. Normalization is a term used in signal processing and is used to adjust values to a common scale suitable for comparison. In this case, the power level of the strongest frequency component is normalized to the dynamic range value, with other frequency bins adjusted accordingly. Any values that fall below 0dB are set to 0dB. This normalization allows the recognizer to compare two signals with different power levels (e.g. one louder than the other). Larger values of dynamic range result in the consideration of more spectral detail, but also leave the recognizer vulnerable to contamination from other (quieter) noise sources in the same frequency band, such as from other birds. Again, this value should not exceed the expected signal-to-noise ratio of candidate vocaliza-

tions or spectral detail will be considered that is below the noise floor and unattainable.

A value of 15dB was chosen for the dynamic range because it is at the lower end of the range of the expected signal-to-noise ratio for Ceruleans in open omnidirectional microphone recordings made by ARUs in the field (15-20dB), and values below 15dB would begin to lose important spectral features necessary for classification.

Max Song

The Max Song parameter indicates the maximum expected duration of a given vocalization. The MLNS training data has vocalizations generally in the 1.5 second range, so this setting was rounded up to 2.0 seconds for the recognizer.

Max Syllable

The Max Syllable parameter is an indication of the duration of a given syllable within a vocalization. For high-quality recordings such as the MLNS recordings, individual syllables can be easily segmented by Song Scope and a value of 0.40 seconds appears to cover the syllables observed in the training data. However, the target environment is a much lower signal-to-noise environment and individual syllables will get blurred together in the presence of noise and as the sound is reflected off trees in the dense forest (see Figure 5). Song Scope uses this value as a look-ahead when attempting to find the boundaries of a candidate vocalization, and there is generally little harm in using larger numbers up to the duration of the vocalization. A value of 1.50 seconds was chosen as this is the observed typical duration of vocalizations in the training set.

Max Syllable Gap

The Max Syllable Gap parameter is an indication of the longest gap between any two syllables of a vocalization. Song Scope uses this value to find the end of the vocalization. When a low-signal period exceeds this duration, Song Scope assumes that the vocalization has ended. If the value is too long, it is possible for Song Scope to incorrectly find that a subsequent rise in signal is part of the same vocalization, when this may in fact be a vocalization from a different individual or species. And if this value is too short, Song Scope will incorrectly terminate the vocalization before it is completed resulting in a partial vocalization against which to match.

A value of 0.20 seconds was chosen based on the observed maximum syllable gap in the training data.

Maximum Complexity

The maximum complexity parameter sets the maximum number of Hidden Markov Model (HMM) states (N) and the maximum number of syllable classes (N/4) to be considered by Song Scope when generating a recognizer.

The Song Scope recognizer is essentially an HMM designed to capture not only the spectral and temporal features of each syllable, but also the syntax of how syllables are connected to form a vocalization. To do this, Song Scope automatically segments syllables from the training data into classes and assigns initial HMM states to each class. A more detailed explanation of the Song Scope algorithms can be found in U.S. Patent application 20050049877.

When generating a recognizer, Song Scope iterates through different numbers of syllable classes from one through (Maximum Complexity / 4) and builds a different HMM for each iteration. The model that best generalizes across excluded training data is automatically chosen.

The Cerulean warbler vocalization is relatively complex as there are a number of different syllables. The maximum value 48 was chosen, though some trial and error may be helpful to find models that perform better.

Note that there is a risk of "over training" if this parameter is set too high. When an HMM is over trained, the model may fit the training data too tightly and not generalize well to similar but different patterns and thereby missing some target vocalizations.

Maximum Resolution

The maximum resolution parameter sets the maximum number of feature vector dimensions used by the Song Scope recognizer. This is essentially the number of Discrete Cosine Transform (DCT) coefficients plus other features used to represent each time slice of a vocalization. More dimensions imply a greater level of spectral detail to be considered when comparing vocalizations.

Because all of the training data has significant spectral detail with strong frequency modulated narrowband signals, a large number of feature vectors can be used to successfully match against similar quality vocalizations such as those found throughout the training set. However, the low signal-to-noise ARU recordings will be relatively "fuzzy" and "blurred" because these are made from unattended omnidirectional microphones in the open field at some distance to the target through thick

Figure 3: Normalized View of New York Cerulean vocalization (MLNS #9182)
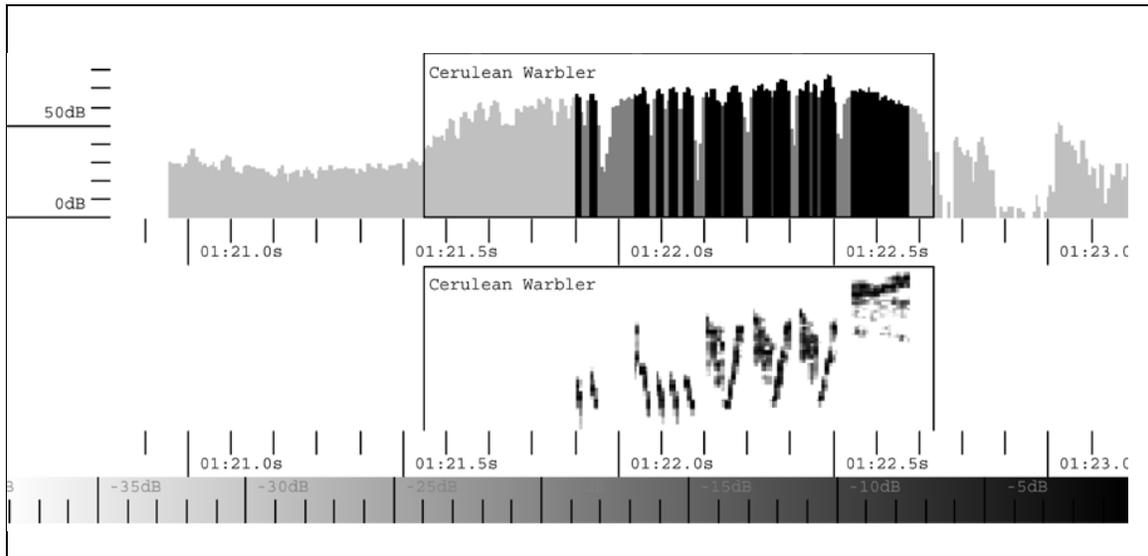
Figure 4: Normalized View of West Virginia Cerulean vocalization (MLNS #110206)
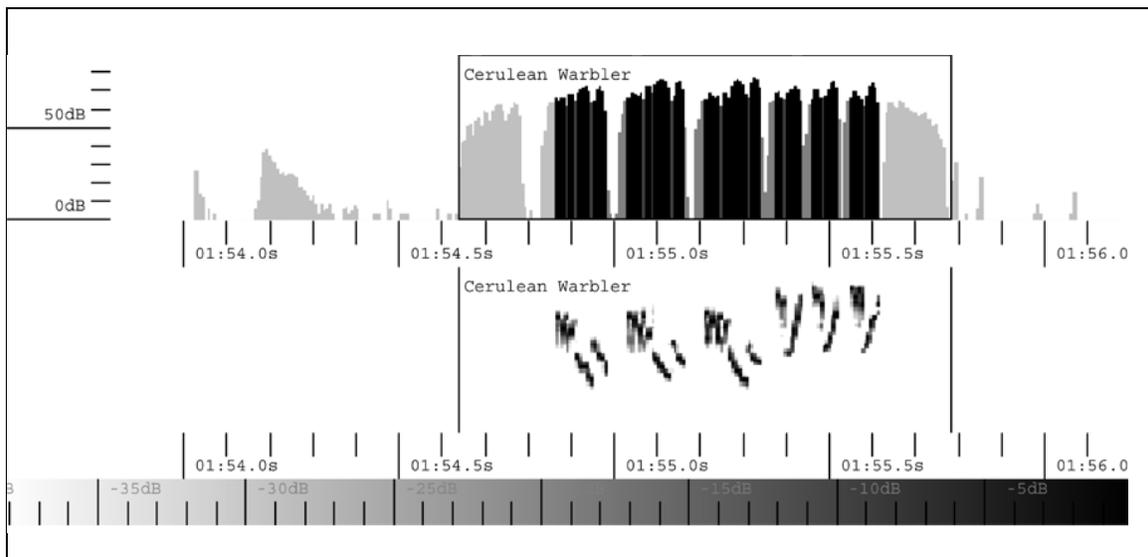
forest. Therefore, a small feature vector should be used to compensate for the difference between high quality training data and the low signal-to-noise ratio expected in the ARU recordings. Through some trial and error, we found that best results came from the minimum setting of 4 features.

Visualizing the Effect of Parameters

Song Scope has display modes that enable the user to immediately visualize how parameter settings affect the signal detection algorithm and normalization of the data. Figure 3 and Figure 4 show the signal detection and normalized spectrogram for the two MLNS recordings previously illustrated in Figure 1 and Figure 2 respectively using the parameter settings described above. The top part of each plot illustrates the signal detection algo-

rithm and shows overall power levels within the selected frequency band on a dB scale. The black color indicates syllables, dark grey indicates inter-syllable gaps, and light grey indicates signal not considered part of the vocalization. Notice that the first syllable of MLNS #9182 and the first and last syllables of MLNS #110206 are not detected (but are visible in the original spectrograms shown in Figure 1 and Figure 2). This is because the dynamic range is set low at 15dB and the missing syllables fall below this threshold relative to the strongest parts of the vocalization. Larger values of the dynamic range parameter would allow these syllables to be detected in the high quality training recordings. However, the recordings made by the ARU's omnidirectional microphones will have a substantially lower overall signal-to-noise ratio. The weakest syllables of a detected vocalization may in fact fall below the noise floor and therefore be undetectable (the noise floor is the power level of random background noise, and signals quieter than the background noise can not be detected). Thus, the dynamic range parameter of Song Scope helps to compensate between the high quality training data and the low signal-to-noise ratio typical of ARU field recordings.

3.3. Cerulean Classifier Statistics

After Song Scope generates the recognizer from the training data and parameters described above, the following statistics are reported:

Table 3: MLNS Recognizer Statistics

Parameter	Value	Comments
Model states	30	Size of HMM model in states
Syllable classes	10	Number of classes into which syllables were clustered to generate HMM.
Cross training	62.05 ± 2.96%	Average and standard deviation of model fit when testing against excluded recordings in the training set
Total training	62.37 ± 3.23%	Average and standard deviation of model fit when testing against entire training set.

Table 3: MLNS Recognizer Statistics

Parameter	Value	Comments
State usage	12 ± 5	Average and standard deviation of different HMM states traversed by training data
Symbols	66 ± 26	Average and standard deviation of spectral symbols found in training set.
Duration	0.95 ± 0.32 sec.	Average and standard deviation of the vocalization durations found in the training set.

4. Running the First Pass Recognizer

Song Scope was able to automatically (unattended) scan the 251 hours of field recordings using this recognizer in only 12 hours on a 3.2GHz Pentium 4 machine running Red Hat Linux ES-3 with an SMP kernel. (Three processes were used to scan recordings, one for recordings from each of the three ARUs in parallel to take advantage of the dual core processor).

4.1. Result Filter

For each detected vocalization candidate, Song Scope produces two values - Quality and Score - that are used to indicate the likelihood of a positive match.

Quality

The quality value is in the range of 0.00 to 9.99 and indicates where on a Gaussian bell curve the candidate falls with respect to duration, symbols, and state usage compared to the training data. A value of 5.00 would indicate that these three components in aggregate are average. A value of 2.00 indicates that the candidate is in the bottom 80%, and a value of 8.00 indicates that the candidate is in the top 20%.

A low quality score indicates that there is not likely enough raw information (e.g. in duration, symbols, or state transitions) to produce a meaningful HMM probability output.

Song Scope provides a control to set the minimum acceptable quality level for candidate vocalizations. If a candidate falls below this level, it will not be considered.

For the first pass, it was desirable to err on the side of letting through more false positives to be sure of finding Cerulean vocalizations in the training data. A minimum quality value of 2.0 was used.

Score

The score value is a percentage in the range of 0.00% to 100.0% and indicates the average probability (across all the state transitions) that the HMM recognizer model could have produced the candidate vocalization. In other words, the score is an indication of how well the candidate vocalization fits the model.

As mentioned above, this value will be meaningless if the candidate has a low quality value.

Song Scope provides a control to set the minimum acceptable score for candidate vocalizations. If a candidate falls below this level, it will not be counted.

Because we know the low-signal ARU field recordings will not closely resemble the high quality training data, scores well below the 62% mean training data score are expected. So a minimum score value of 50% was used.

4.2. First Pass Results

Song Scope detected a total of 9,145 candidate vocalizations with a quality of at least 2.0 and a score of at least 50.00%.

The next step is a manual review of these vocalizations to confirm which are actual Cerulean vocalizations. A human operator was able to use Song Scope to scan through and visually inspect these 9,145 candidate spectrographs of the vocalizations in about 2 hours. Once visually inspected, these vocalizations were considered "confirmed" Cerulean vocalizations.

Of the 9,145 candidates, only 522 were Cerulean. Table 4 shows the distribution of these results among the nine monitoring locations.

Table 4: First Pass Results

ARU Deployment	Hours Recorded	Known Cerulean Territory	Confirmed Cerulean Vocalizations
ARU09-A	20.33	Yes	29
ARU09-B	35.58	Unknown	0
ARU09-C	28.67	Unknown	0

Table 4: First Pass Results

ARU Deployment	Hours Recorded	Known Cerulean Territory	Confirmed Cerulean Vocalizations
ARU18-A	23.58	Yes	78
ARU18-B	34.75	Yes	58
ARU18-C	30.42	Unknown	0
ARU20-A	36.58	Yes	357
ARU20-B	13.25	Unknown	0
ARU20-C	27.92	Unknown	0
Total			522

Note that all of the previously known Cerulean locations had confirmed detections of Cerulean vocalizations. This suggests that the recognizer was able to find vocalizations where they were expected. Additionally, note that no Cerulean vocalizations were detected in any of the unknown territories. While this suggests that there may not be Ceruleans in any of these territories, lack of evidence is not proof of absence.

4.3. False Positives

While the first pass recognizer was able to detect over 500 confirmed Cerulean warbler vocalizations, the number of "false positive" detections was high with over 9,000 detections.

Figure 5 shows a typical example from one of the confirmed Cerulean warbler detections, this from deployment ARU18-A on June 11 at 06:11:37AM. Figure 6 shows the normalized view of the same vocalization.

Notice how the ARU spectrogram appears to be a "smeared" or "blurred" image compared to the sharp and detailed training data recordings. This is typical for ARU recordings in a low signal-to-noise environment. As the Cerulean vocalization propagates through the forest, the sound waves are reflected and scattered by the trees before arriving at the omnidirectional microphone at slightly different times and directions. The resulting interference causes the "smearing" of the spectrogram. In addition, with a low signal-to-noise ratio, some of the spectral detail is lost below the noise floor. As a result, the training data does not closely match.

If the recognizer was built with more spectral resolution (i.e. a larger value for Maximum Resolution resulting in

Figure 5: Spectrogram of Cerulean Vocalization from ARU
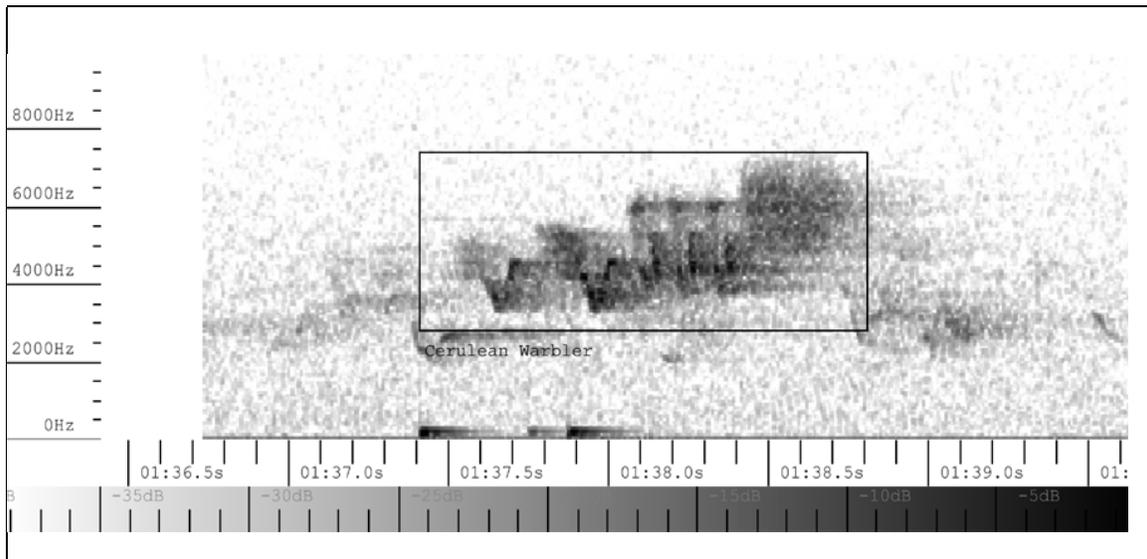
Figure 6: Normalized View of Cerulean Vocalization from ARU
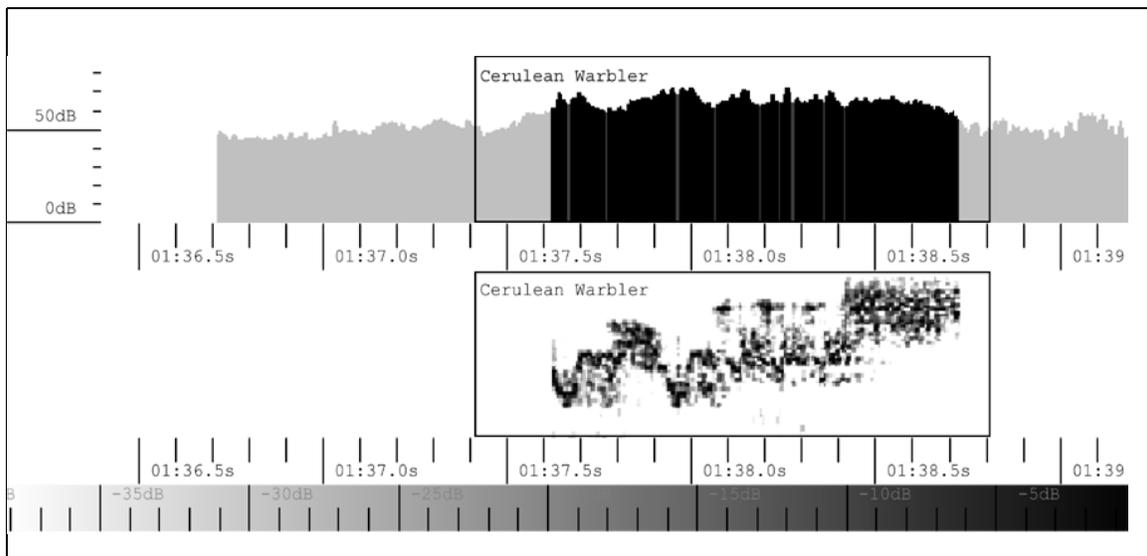
a larger feature vector), the model made from high quality training data would fail to match Ceruleans found in the ARU recordings. We deliberately set the Maximum Resolution to a small value to compensate for the difference and increase the chance for detection. However, by making the recognizer more flexible to detect Cerulean vocalizations in ARU recordings, it is also more forgiving of other non-Cerulean vocalizations appearing in the same frequency band resulting in a high number of false positives. Figure 7 illustrates one particular example of

a non-Cerulean vocalization that was mis-classified as a match. Notice how some of the syllables in this vocalization have similar characteristics in shape to those in known Cerulean vocalizations.

Figure 7: Spectrogram of Mis-classified Vocalization
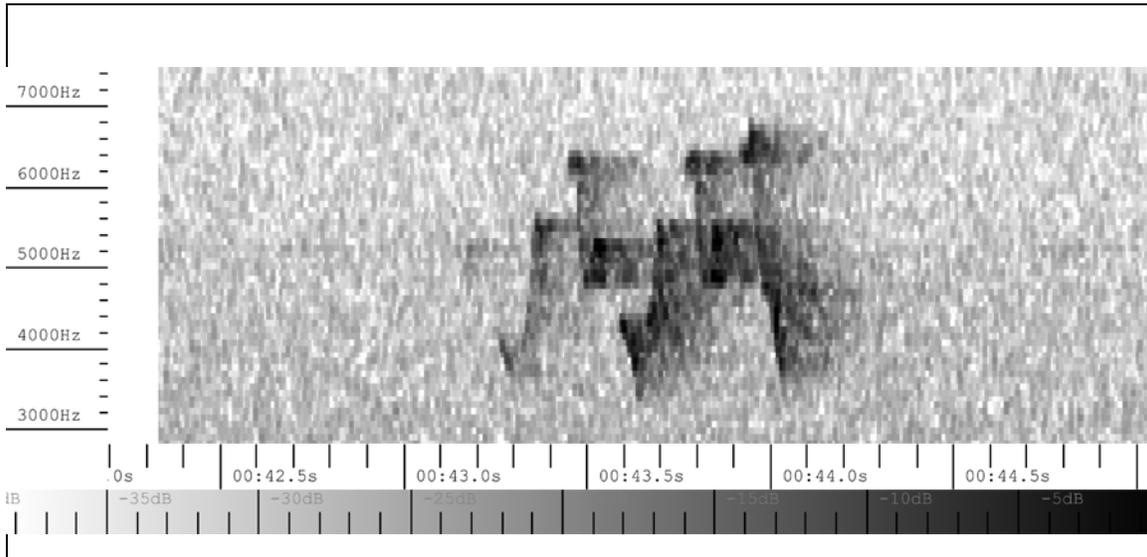
5. Creating a "Second Pass" Recognizer

With the help of our initial recognizer, we have located Cerulean vocalizations in the ARU recordings. Some of these vocalizations can now be used to build a new recognizer that more accurately reflects Cerulean vocalizations as they are likely to be observed in ARU recordings.

5.1. Second Pass Training Data

Four 5-minute long ARU recordings were chosen for this new training data, one from each of the four locations in which Ceruleans were detected (see Table 5). A total of 61 vocalizations were annotated with about one hour of human labor. The reason for choosing recordings from different locations was to make it more likely that the vocalizations would come from different individuals.

The Song Scope algorithms are designed to generalize on unknown training data by iteratively excluding one recording while building a model from the remaining training data, then observing how the model performs against the excluded recording. Our objective is to build a recognizer that is more likely to detect Ceruleans that have not been observed before.

5.2. Second Pass Parameters

The same recognizer parameters used in the first pass (See "Parameters" on page 4) were also used in this sec-

ond pass recognizer with the exception of the Maximum Resolution. Now that the training data is expected to more accurately compare against the ARU recordings (since it comes from ARU recordings), a Maximum Resolution value of 16 was chosen (the default) instead of 4.

Table 5: Second Pass Training Recordings from ARUs

ARU Deployment	Date	Time	# Annotations
ARU09-A	Jun 24	07:15-07:20	3
ARU18-A	Jun 16	06:10-06:15	24
ARU18-B	Jun 20	09:00-09:05	15
ARU20-A	Jun 19	10:50-10:55	19

5.3. Second Pass Classifier Statistics

Table 6 shows the generated recognizer statistics

6. Running the 2nd Pass Recognizer

Song Scope was able to scan the 251 hours of ARU recordings with the second pass recognizer in about the same time (twelve hours) as it did with the first pass recognizer.

6.1. Second Pass Result Filter

The generated recognizer cross training and total training statistics of around 55% are somewhat lower than the first pass recognizer statistics of around 62%. In order to allow most matching Cerulean vocalizations to be counted, the Minimum Score was set to 47%. This is approximately 3 standard deviations below the mean. As before, a Minimum Quality value of 2.0 was used.

6.2. Second Pass Results

In the second pass, Song Scope detected 4,328 candidate vocalizations. About one hour of human labor was required to manually review these candidates to count 1,552 confirmed Cerulean vocalizations as summarized in Table 7.

The second pass recognizer detected about twice as many Cerulean vocalizations as the first pass, and did so with about half as many total detections dramatically reducing the false positive rate.

Adjusting the filtering parameters (i.e. Minimum Quality and Minimum Score) will further reduce the number of false positives, but will also miss more confirmed vocalizations. For example, increasing the minimum score from 47% to 50% results in the detection of 758 confirmed Cerulean vocalizations out of 1,612 total candidates.

Just as in the first pass, the second pass recognizer only found confirmed Cerulean vocalizations in the previously known Cerulean territories and did not detect any in the unknown territories.

.

Table 6: Second Pass Recognizer Statistics

Parameter	Value	Comments
Model states	41	Size of HMM model in states
Syllable classes	12	Number of classes into which syllables were clustered to generate HMM.
Cross training	55.39 ± 2.51%	Average and standard deviation of model fit when testing against excluded recordings in the training set

Table 6: Second Pass Recognizer Statistics

Parameter	Value	Comments
Total training	56.23 ± 2.62%	Average and standard deviation of model fit when testing against entire training set.
State usage	24 ± 3	Average and standard deviation of different HMM states traversed by training data
Symbols	145 ± 36	Average and standard deviation of spectral symbols found in training set.
Duration	1.14 ± 0.20 sec.	Average and standard deviation of the vocalization durations found in the training set

.

Table 7: Second Pass Results

ARU Deployment	Hours Recorded	Known Cerulean Territory	Second Pass Confirmed Ceruleans
ARU09-A	20.33	Yes	11
ARU09-B	35.58	Unknown	0
ARU09-C	28.67	Unknown	0
ARU18-A	23.58	Yes	655
ARU18-B	34.75	Yes	122
ARU18-C	30.42	Unknown	0
ARU20-A	36.58	Yes	764
ARU20-B	13.25	Unknown	0
ARU20-C	27.92	Unknown	0
Total			1,552

7. Human v.s. Machine

There are a number of interesting observations we can make in the results such as the relatively small number of vocalizations detected at site ARU09-A compared to

Table 8: Song Scope performance compared to Manual Inspection of Baseline Subset

ARU Deployment	Manual Detection				Automatic Detection (by Song Scope "recognizer")	
	Total Confirmed	Weak Signal (<15dB SNR)	Corrupted	Good Signal	Pass 1 (Confirmed)	Pass 2 (Confirmed)
ARU09-A	12 (10.2%)	9	2	1 (2.9%)	0 (0.0%)	0 (0.0%)
ARU18-A	57 (48.3%)	5	33	19 (54.2%)	5 (35.7%)	15 (42.8%)
ARU18-B	10 (8.5%)	2	7	1 (2.9%)	0 (0.0%)	3 (8.6%)
ARU20-A	39 (33.0%)	3	22	14 (40.0%)	9 (64.3%)	17 (48.6%)
TOTAL	118 (100%)	19	64	35 (100%)	14 (100%)	35 (100%)

site ARU20-A, or the proportionally larger number of detections made in the second pass recognizer for site ARU18-A as compared to the first pass.

A baseline must be established in order to interpret these results in a meaningful way. For example, the results themselves do not tell us how many vocalizations were missed by Song Scope or why.

It is impractical to manually review over 250 hours of ARU recordings, so a representative subset of recordings were chosen to establish a baseline. A subset of recordings consisting of the 5-minute recordings made between 8:00-8:05AM each day from all three ARUs in all nine locations was chosen for a baseline. This includes 51 5-minute recordings, 4.25 hours in total.

Spectrograms of these recordings were manually inspected looking for Cerulean vocalizations. This process took about 90 minutes (note that by extrapolation, it would have taken over 88 hours to manually inspect all 250 hours of recordings using this method) and resulted in 118 confirmed Cerulean vocalizations. However, 70% of these vocalizations were either extremely weak (i.e. less than 15dB signal-to-noise ratio) or corrupted by overlapping vocalizations from non-Cerulean species or other noise sources in frequency and time. Additionally, while many of the weaker vocalizations were barely detectable in a spectrogram (with a Wiener filter applied to filter out stationary background noise), they could not be heard by a human listening to the recording.

Cerulean vocalizations were once again only found in the four known Cerulean territories and not in any of the unknown territories.

The pass one and pass two recognizers were then used to scan the same baseline subset of recordings and the results were manually verified to count confirmed Cerulean vocalizations detected.

The Song Scope recognizer performance is approximately proportional to the good quality (i.e. not too weak and not corrupt) vocalizations present, with the second pass recognizer detecting about the same number of vocalizations in each territory.

The second pass recognizer confirmed Cerulean detections across the four territories in the baseline subset is also approximately proportional to the detections across the entire set of recordings.

This suggests that Song Scope is able to automatically detect a majority of vocalizations that have sufficient signal strength and are not corrupted by collisions with other sound sources in frequency and time.

In the case of the ARU09-A territory, it appears that while there is detectable Cerulean activity, the vocalizations are much fainter, most likely because the ARU was placed at a greater distance from singing perches compared to the ARU18-A and ARU20-A territories. This data may give some indication of the effective range each ARU has in monitoring a given potential Cerulean habitat. This ARU was placed immediately below an active nest. There has been some anecdotal information suggesting that males tend to sing most on the margins

of their territories, well away from nests, and this result seems to confirm that idea. Additionally, this territory was unlike the ARU-18A and ARU-20A ones in being located high up on a very steep slope with a greater amount of foliage between the ARU and other singing males.

In the case of the ARU18-B territory, manual inspection suggests that the vocalizations are stronger than in the ARU09-A territory, but other closer/louder birds (in particular a persistent Black-throated Blue Warbler) are singing at the same time overpowering the Cerulean vocalizations such that they become corrupted and undetectable by the Song Scope recognizers.

8. Conclusions

The use of bioacoustics technology can dramatically reduce the time and labor required for the inventory and monitoring of vocal species such as songbirds.

The combination of Autonomous Recording Devices and signal processing software is capable of the collection, analysis and documentation of a substantial amount of presence/absence evidence with a fraction of the time and effort that it would take to manually listen to hours of recordings.

Without ARUs or the software to analyze their recordings, over 250 hours of human labor would be required to listen to the sounds from these 9 locations from 6-11AM daily for several days of monitoring.

With ARUs and simple signal processing software to view spectrograms of recordings, the amount of human labor can be reduced to approximately 1/3 by visually inspecting spectrograms instead of directly listening to the recordings. But this still represents over 80 hours of human labor and effort.

The use of advanced signal processing classifiers such as those available in Song Scope can reduce the human labor effort from 250 or 80 hours down to only 1 hour. This is possible because the software is capable of automatically scanning the recordings to find only the likely vocalizations of interest, and a human review of these likely vocalizations to confirm which ones represent the target species can be completed very efficiently, even with a high false-positive rate.

While automatic detection and classification can be adversely affected by weak (distant) signals and/or corruption by louder/closer noise (e.g. from other species), the same is true with a human listener, and the Song Scope algorithm was able to detect the majority of vocalizations that were not too faint or corrupted.

Manual review of spectrograms, at least for Cerulean warblers, appears to be the most accurate way to detect the largest number of vocalizations, even those too weak or too corrupted to be clearly heard by human ears or analyzed by classification software. However, this effort can be tedious and labor intensive.

ARUs and advanced signal processing classifiers can be used to efficiently survey potential territories with minimum effort for collecting and documenting presence/absence data, and the results can then be used to direct additional deployments (e.g. moving ARUs closer to singing perches) or more human labor if further study is desired or warranted.