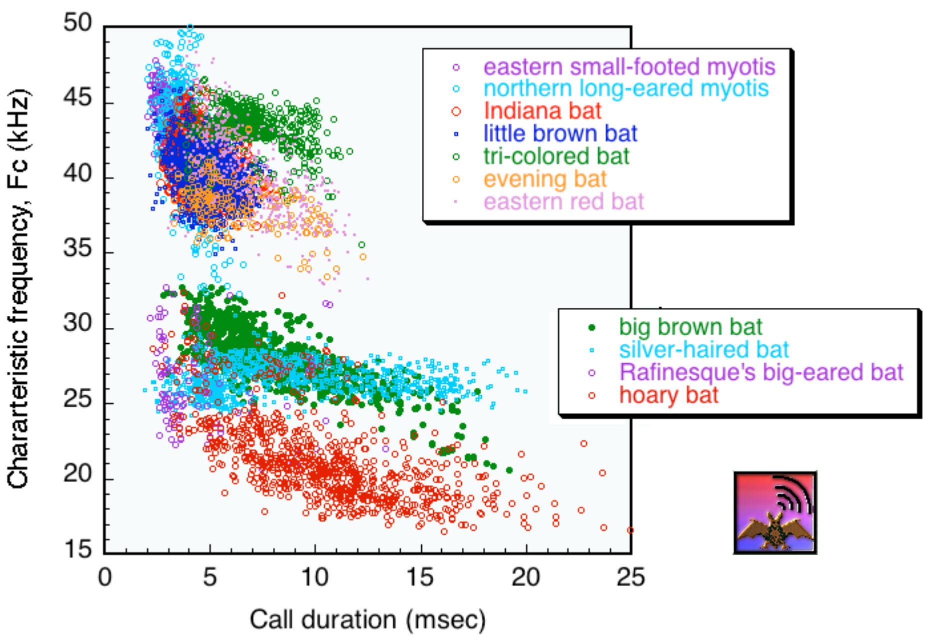
SonoBat 3.1 NE classification notes

Users should consider the species decisions generated by SonoBat as suggested classifications. Any final conclusions regarding species presence should require confirmation by a qualified biologist with knowledge of bat echolocation call characteristics and the limitations imposed by species having similar call characteristics. Although some species have distinctive call types that facilitate confident identification, other species exhibit many overlapping call characteristics that reduce the reliability of using bat echolocation calls as a sole indicator of presence. In some instances irrefutable species confirmation may require a "bat in hand."

Successful classification of the many bat species having overlapping acoustic characteristics depends upon discerning subtle nuances in their calls, and this depends exquisitely on clear, strong, and undistorted signals that rise above the background noise level. The reference data used to generate the SonoBat classifiers are based on recordings from electret condenser microphones and electrostatic condenser microphones (e.g., the types used in Pettersson, Binary Acoustic Technology, and Avisoft detectors) as these produce recordings having good contrast between the bat echolocation signals and background noise, both external and internal to the microphone (i.e., the microphone's noise floor). Generally, the more expensive the microphone or detector, the better it will retain and reveal lower amplitude call details from bats at greater distances from the microphones, and so provide a larger volume of airspace from which to acquire species-discriminating recordings that classify without error. SonoBat classification performance will decrease and the number of misclassifications will increase with degraded signals that cannot reveal clearly low amplitude components of call structure. The comments that follow provide guidance for interpreting call classification results, and recommendations for recording.



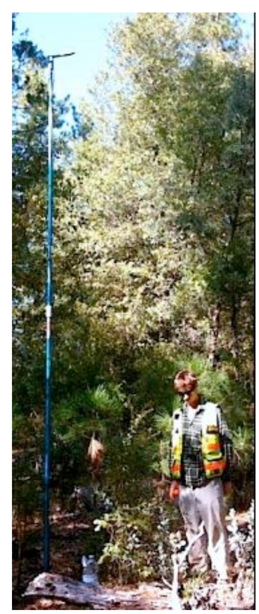
The SonoBat automated species classification algorithms are based upon several thousand species-known recordings (sample size varies from species to species) from specific sites within each geographic region covered. While derived from a robust data set acquired from a variety of environments and conditions, the data set nevertheless encompasses a finite set of vocalizations from each species covered and can not fully represent the repertoire of bat vocalizations that likely occur in nature. Bats exhibit considerable plasticity in their vocalizations, and considerable overlap in call parameters among species; this coupled with complications from noise and weak signals (as from bats at a greater distance from the detector) can potentially result in a recording from one species exhibiting parameters that fall into and match the expected parameter space of another species, resulting in a misclassification. The substantial overlap in the echolocation call characteristics of many species (see example plot above) often means that only a small portion of some species' repertoires will have a tendency toward discriminating characteristics, even with perfectly recorded call sequences.

For example the simple call shapes of shorter *Eptesicus fuscus* and *Lasionycteris noctivagans* present calls that overlap in data space with sufficient ambiguity to result in misclassification. Follow the rubric on the acoustic characteristic table for vetting shorter calls of these species. The longer curved variants shared by Epfu and Lano share even more characteristics that render them particularly difficult to discriminate with confidence. However, the longer flat calls of Lano provide a confident discriminating call variant for this species.

Because bats vary the amplitude through their calls, the farther a bat flies from the detector the more the call becomes truncated to just its strongest portions. In some cases these fragments of fully formed calls can mimic other species, e.g., the body fragment of a *Myotis lucifugus* may render as a simple curved call that mimics a *Lasiurus borealis*. SonoBat performs a number of signal quality checks to reject poorly formed calls, overloaded calls, or those with distorted signals or too much noise, but it is still limited in the information that it has, and because of the variability in bat calls (e.g., intraspecifically, social calls, feeding, inspection calls, etc.) classification remains a probabilistic process, so generally if a classification result seems unexpected, check it or reject it. The quality of call recordings strongly affects the performance of the SonoBat autoclassifier. Recording from the ground, near flat surfaces, or through tubes will render distorted signals. Signal distortion inhibits call trending and the recognition of call parameters essential to perform accurate classifications. In summary: *garbage in, garbage out*

Recommendations for quality recording: Avoid recording with a detector placed directly on the ground. Simply elevating a detector one or two meters above ground level can dramatically improve recording quality by reducing surface echoes, avoiding thermal layering, or near-ground air convection currents, all of which can distort ultrasound signals. In general, the longer duration calls that many species produce in open air flight, i.e., away from clutter, provide more information content and greater species-discrimination confidence. Bats flying in confined spaces or near roosts will generally provide shorter, less discriminating and perhaps ambiguous call variants. If you need to identify bats in such situations, try to record them on approach to such a space or follow them out and away from a roost to acquire longer and more representative search phase calls.

To record search phase call sequences of bats along a flyway, place detectors *out* of the flyway as bats may investigate the novel object resulting in many recorded sequences of short "inspection calls." Where possible, place detectors to blend in with vegetative clutter (but clear from it) and listen out into a flyway. Avoid placing detectors near large echo-producing surfaces: asphalt, building facades, bridge structural surfaces, flat water, etc. When you must record near such surfaces, attempt to position the detector to listen *away* from these surfaces rather than toward them. When possible, use a handheld detector to acoustically sample the potential detector placement site to reveal sources of ultrasonic noise before a recording session. Many things that seem quiet to our human ears can emit overwhelming ultrasonic noise, e.g., dried leaves or other vegetation rustling in a breeze, insects, loose cables and other windblown components, or metal structures cooling in the evening.



Detectors with microphones on cables separate from the detector electronics provide the best options for placement and best results. For ground-based monitoring, consider extending the microphone horizontally from a pole or other means at about 4-6 m to listen out into flight space and down toward the ground and up from there, rather than just listening up from the ground. This will increase your chances of acquiring distortion-free recordings. See the recording advice page on the SonoBat website for additional recommendations.

Even among the known species of the library reference data, the rate of correct classification varies by species, situation, recording quality, and settings. SonoBat allows the user to control call discrimination settings and in general, more discriminating settings increase the rate of correct species classification (up to a point) but decrease the percentage of files that SonoBat accepts and reports as confidently classified. ***The results reported below represent idealized classification performance based on good quality recordings. Users should expect their results to vary commensurate with recording quality as described above.*** Also, while derived from a robust data set acquired from a variety of environments and conditions, the data used to construct the classifier nevertheless encompasses a finite set of vocalizations from each species covered, and recording in nature will provide a virtually unlimited variety of vocal variants with some that will exceed that covered by the classifier. Each SonoBat regional classifier only “knows” the data and call types used to build it, and many spurious signals may generate a parameter set that will fall into one of the known data spaces that SonoBat may (incorrectly) recognize as a species.

For this reason, you should use the classifications that SonoBat provides as a guide, and vet any species result that may seem unexpected or unusual to confirm such classifications. Even with a 98% correct rate (and that's ideal- for good data), the similar calls from many species will bring up some misclassifications for any substantial data set. The overlapping acoustic characteristics of some species challenge the automated classifier just as they challenge us to manually discriminate them, although the classifier generally knows when to reject or not make a decision. For example Tabr/Laci can present a particular challenge to discriminate as both bats emit loud calls that travel far and can generate recordings of just the core part of the call when partially out of range, and that can lead to ambiguous or spurious results. It seems that Laci calls more often get misclassified as Tabr, rather than the other way around. So on batch runs where you might not expect many Tabr you should vet everything that comes out as Tabr and only accept ones with strong discriminating characteristics, e.g., a pronounced downward roll at the beginning or end of the call.

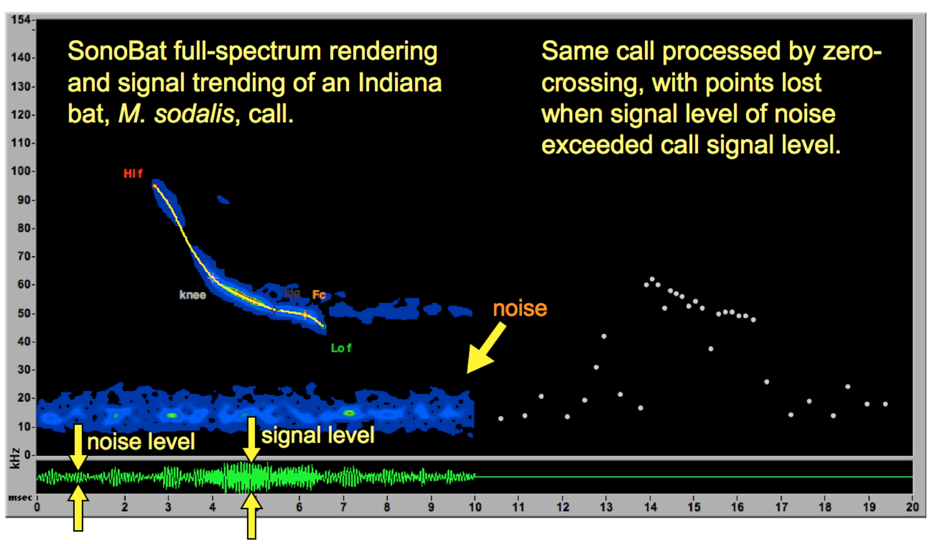
The Indiana bat, *Myotis sodalis* (Myso), and the little brown bat, *M. lucifugus* (Mylu), present substantial overlap in their echolocation call characteristics that render only a small portion of their repertoires with a tendency toward discriminating characteristics. To prevent outputting null species identification results, the SonoBat classifier uses this rubric: when a species decision for either of theses species does not exceed the threshold discriminant probability setting (DP, SonoBat uses 0.90 as the default setting), and if the second potential species comes out as the opposite of this pair, and their combined discriminant probability score meets or exceeds the threshold setting, then SonoBat will output this result using the ambiguous designation “MysoMylu” or “LuSo” on SonoBatch output. This will indicate the call or sequence probably came from one of these two species, but presented call characteristics within overlapping data space that prevented disambiguation. Refer to “SonoBat Discrimination of Myso *vs.* Mylu” for more information:
http://www.sonobat.com/download/MysoMylu_Classification_Note_NE_v306.pdf

The longer curved variants shared by Epfu and Lano are particularly difficult to disambiguate with confidence. Follow the rubric on the acoustic characteristic table for vetting shorter calls of these species. The longer flat calls of Lano provide a confident discriminating call variant for this species.

Proper interpretation of SonoBat classification results requires an appreciation that species discrimination by echolocation calls uses a probabilistic process. Although called a “discriminant probability,” a DP = 1.00 does not indicate 100% confidence of the species classification result. Rather, it indicates that the quantitative parameters measured from the unknown call or sequence under consideration fall completely at the centroid of the multi-dimensional data space of all the data known for that species. A species with similar call characteristics can occasionally (or often depending on the overlap) produce calls with data on the fringes of its parameter space that intrudes into the parameter space of another species, or even falls at the centroid of the other species’ parameter space. But, a DP = 1.00 *probably* indicates the classified species, and that confidence increases for species having more unique parameter space. Although SonoBat may report a result indicating a greater likelihood of one similar species over the other, e.g., 0.85 Myso versus 0.15 Mylu, such a result only indicates the relative distances from the centroid of the known multivariate data space for each species. Because these species have their centroids buried in the multivariate data clouds of the other species, they never clearly separate, and either species could just have well vocalized a call producing those results, despite lying closer to the mean values of one over the other.

SonoBat classifies calls and sequences using an expert system incorporating an ensemble consensus of redundant hierarchical decision algorithms and reports a single species decision when that result exceeds the discriminant probability (DP) threshold at each decision, *and* passes post-decision checks of known call characteristics. This expert decision path optimization approach outperformed tests using other standard machine intelligence systems (Artificial Neural Networks, Bayesian, etc.). SonoBat reports the DP of the final species decision if all hierarchical decision steps met or exceed the threshold. If any decision step does not meet or exceed the threshold, then SonoBat displays the species or hierarchical groups that sum to the threshold at that step, e.g., 0.775 MyleMyseMysoMylu, 0.225 PesuNyhuLabo.

Classifying an entire sequence (i.e., bat pass) typically provides more confident results than individual call classification as this method benefits from the combined information within the sequence. SonoBat reports two results for sequences, a consensus vote and a mean sequence decision. The consensus vote requires a minimum of two calls per majority species (except for *Lasionycteris noctivagans* and *Lasiurus cinereus*) and requires the majority species to have equal to or better than twice the number of calls as the sum of the second and third most prevalent species (if classified). The mean sequence decision calculates mean parameter values of the most prevalent hierarchical classification group (e.g., MyleMyseMysoMylu or PesuNyhuLabo) of accepted calls with a minimum of two calls (except for *Lasionycteris noctivagans* or *Lasiurus cinereus*) and sends those mean values through the decision engine.



To minimize misclassifications, SonoBat performs quality control by assessing a number of signal quality and reliability indicators. If calls fails accepted thresholds for any of these indicators, SonoBat rejects the result from automated classification as it can indicate a poor quality signal that can lead to misclassification. (During manual inspection, SonoBat will report the classification result of calls that fail any of the reliability tests but will gray out the display to indicate an unreliable result.)

The amplitude and multiple frequency content of full-spectrum data enables assessment of signal quality and evaluation of the acoustic environment of the recording site. For example, one such measure, the signal to noise ratio (SNR), measures the relative strength of a signal of interest (the call) to the strength of the background signal level (a measure unavailable in zero-crossing data that can only access the dominant frequency at any time). Calls with low SNR more often render poor data that can lead to misclassification. Multiple frequency content also enables more reliable tracking of call trends, and better call data extraction, through echo clutter and ambient noise (see figure). For more information, see
http://www.sonobat.com/download/FullSpect_and_Zero-Crossing.ppt

SonoBat based the 11 species US Northeast classifier on an exemplar reference library set of 1,444 recordings[1] that yielded 8,370 parameterized calls using a maximum of 8 calls considered per sequence, a quality acceptance threshold of 0.80, and discriminant probability threshold settings for acceptance of **0.90**. The classification algorithm based on this data yielded the following results for **individual call** classification:[2,3,4]

	%correct	*%accp*
All Species	**97.2**	*55.4*
Myle	**98.7**	*33.8*
Myse	**99.0**	*50.3*
Myso[3]	**91.0**	*12.6*
Mylu[3]	**90.6**	*14.3*
Pesu	**98.9**	*89.3*
Nyhu	**95.1**	*71.1*
Labo	**96.5**	*64.1*
Epfu	**98.6**	*84.5*
Lano	**98.1**	*83.2*
Laci	**96.1**	*85.7*
Cora	**99.1**	*70.0*
Myso/Mylu[3,4]	**100.0**	*51.8*

Myotis leibii (Myle)
Myotis septentrionalis (Myse)
Myotis sodalis (Myso)
Myotis lucifugus (Mylu)
Perimyotis subflavus (Pesu)
Nycticeius humeralis (Nyhu)
Lasiurus borealis (Labo)
Eptesicus fuscus (Epfu)
Lasionycteris noctivagans (Lano)
Lasiurus cinereus (Laci)
Corynorhinus rafinesquii (Cora)

[1] ***The results reported here represent idealized classification performance based on high quality recordings*** made with Pettersson D240X and D500X detectors, and with Binary Acoustic Technology AR125 detectors. ***Actual performance will decline along with recording quality*** (see "*Recommendations for quality recording"*).

[2] Values listed as %correct considered just those results that emerged from the classifier at or above a discriminant probability threshold of 0.90. The %accepted reports the proportion of the sample that met or exceeded the discriminant probability threshold, whether correct or incorrect.

[3] Refer to "SonoBat Discrimination of Myso *vs.* Mylu" for more information.

[4] Myso/Mylu indicates a result of MysoMylu, Myso, or Mylu, whether correct or incorrect for Myso (if Mylu) or Mylu (if Myso), i.e., the overall rate for correctly discriminating this species pair from other species.

As species adjust their call characteristics across their repertoires from short to long calls, some similar species will discriminate better or worse for different duration calls. Generally, Myotis species discriminate better at the longer end of their repertoires in which they present more robust features. In contrast, Pesu, Nyhu, and Labo that all have simple feature-thin calls discriminate better at the shorter end of their repertoires in which they present more differences in shape and amplitude distribution. At the longer end of their repertoires Pesu, Nyhu, and Labo all present more feature-thin flatter calls that do not discriminate as well. Refer to the special characteristics listed in the table of echolocation call characteristics for specific guidance, and use the results that follow for general guidance for classification performance for different duration calls to assess confidence in classification results.

Using the same 11 species US Northeast classifier on an exemplar reference library set of 1,444 recordings[1] using a maximum of 8 calls to consider per file, a quality acceptance threshold of 0.80, and a discriminant probability setting thresholds for acceptance of **0.90**, the classification algorithm based on this data yielded the following results for **sequence classification**:

		%correct	*%accp*
All Species	by vote:	**97.3**	*68.8*
	mean sqnc:	**97.1**	*85.9*
	agreement:	**98.4**	*62.7*
Myle	by vote:	**92.3**	*55.4*
	mean sqnc:	**94.4**	*52.3*
	agreement:	**100.0**	*33.8*
Myse	by vote:	**97.6**	*65.3*
	mean sqnc:	**95.8**	*73.4*
	agreement:	**100.0**	*55.6*
Myso[3]	by vote:	**93.9**	*21.7*
	mean sqnc:	**91.7**	*15.4*
	agreement:	**95.2**	*14.0*
Mylu[3]	by vote:	**100.0**	*27.4*
	mean sqnc:	**91.8**	*19.6*
	agreement:	**97.2**	*15.2*
Pesu	by vote:	**97.0**	*82.2*
	mean sqnc:	**100.0**	*93.1*
	agreement:	**100.0**	*81.2*
Nyhu	by vote:	**97.0**	*68.1*
	mean sqnc:	**90.7**	*83.0*
	agreement:	**100.0**	*61.7*
Labo	by vote:	**98.5**	*72.0*
	mean sqnc:	**98.6**	*78.5*
	agreement:	**98.3**	*62.4*
Epfu	by vote:	**97.5**	*93.7*
	mean sqnc:	**94.9**	*87.4*
	agreement:	**98.2**	*85.0*
Lano	by vote:	**92.2**	*94.4*
	mean sqnc:	**94.7**	*88.9*
	agreement:	**95.3**	*87.7*

Cora	by vote:	**100.0**	*83.7*
	mean sqnc:	**100.0**	*91.8*
	agreement:	**100.0**	*79.6*
Laci	by vote:	**100.0**	*91.1*
	mean sqnc:	**100.0**	*89.1*
	agreement:	**100.0**	*88.3*
MyluMyso	by vote:	**97.4**	*68.7*
	mean sqnc:	**96.7**	*68.4*
	agreement:	**98.3**	*60.2*