

Recommendations for using automatic bat identification software with full spectrum recordings

Robust information on bat species distribution and activity is lacking. With developments in passive full spectrum bat detectors and software packages for automating the analysis of sound files, there is the potential to analyse large volumes of acoustic data and thus inform a better understanding of bat ecology and distribution. However, for anyone making use of such tools, it is essential to understand the limitations and likely biases of the software, in order to make an informed interpretation.

Bat activity

Defining bat activity (as recorded by detectors)

Relative bat activity can be measured from the search-phase echolocation calls of bats or, more commonly, from 'bat passes/sequences' – where a pass/sequence is a series of calls belonging to an individual bat.

- Unless manually determined by a human observer, the way bat calls are divided into bat passes is defined by the automatic identification (Auto-ID) software selected.
- 'Bat pass' is therefore a parameter that needs to be defined for each study.
- The number of bat calls or bat passes does not directly relate to the number of bats in a location, although recent statistical approaches may now be able to model this directly¹. It is important to consider carefully (and possibly ground truth) how relative bat activity will be interpreted.

Defining bat passes

- Bat passes are often defined by a minimum number of calls in a series and/or the time between calls or a series of calls (the inter-pulse interval).
- For example, a bat pass could be *any call or series of calls separated by more than one second from another call or series of calls*, although other definitions state that a bat pass comprises at least two calls in a series.
- The definition of bat passes therefore needs to be determined before analysis, clearly stated, and kept consistent throughout a project.
- Some software (e.g. iBatsID using SonoBat-generated parameters) classifies each call rather than a series of calls, so further processing is required to determine the number of bat passes and to evaluate the most likely species based on how these are assigned over the entire series.

Understanding how the Auto-ID software works

What does Auto-ID software do?

In the simplest of terms, Auto-ID software compares an unknown object to a set of known objects, and returns the closest match. In this case, spectrograms of bat calls or passes that have been recorded are compared to spectrograms of calls/passes from known or 'expert-labelled' species, and the software identifies the species the calls are most similar to. This is also how human voice (or face) recognition software works.

GOOD PRACTICE SUMMARY

Describe your dataset:

- define your bat pass
- state how many passes were recorded
- state how many were identified as bats (without further identification)
- state how many were identified to a species or species group

Give adequate details for your recordings:

- your equipment
- the settings used

Specify the software:

- version
- settings chosen

Outline any manual verification undertaken; explain the rationale behind your species groupings

Process the data as it comes in; do not wait until the season is over

Understand the impacts of your choices when interpreting the data

Consider the distribution of your data when applying statistical analyses – it is very unlikely to be normally-distributed

Understand (and record) the limitations of recording bats in the field

Understand (and describe) the limitations of the software used

In order for matches to be found, the Auto-ID software must first determine ('learn') how to recognise the known objects from each other. There are many statistical methods that can be used to 'train' Auto-ID software, but all methods need to find ('detect') and then parameterise the object in some way (e.g. finding the call/pass in a long recording and then measuring some aspects of it - like the peak call frequency or call duration). The software then determines the best combination of those parameters to separate the objects. Part of the known set of objects is held back from this 'training' in order to 'test' the results; this gives an indication of how good the software is at distinguishing between objects (in this case, species).

Different Auto-ID software packages may be better than others at distinguishing between particular objects (i.e. between different bat species) because these packages (and even different versions of the same software package) not only differ in the object library used (in this case, species covered and number of calls/passes used), but also in the particular parameters selected for training and testing, and the statistical model applied.

Auto-ID software is ultimately dependent on: how feasible it is to distinguish objects from each other; whether the library covers all the objects that need to be identified; and whether it captures all the variation seen in those objects. If the object of interest is not in the library, then the software can only return the closest match to the objects on which has been trained. Thus, if you are recording in Europe and your Auto-ID software has been trained on UK bat species, it will not recognise species that occur only in Europe but will classify these to the most similar UK species.

Inherent problems with bat species acoustic classification

Bat echolocation calls are complex; they can vary within species, by sex and body size, and by activity and habitat. Some species have more distinctive calls than others and are therefore more readily recognised by either human or Auto-ID software, but there can be considerable overlap between the calls of different species. Similarly, call quality and 'typicality' can affect the reliability of identification.

The advantage of Auto-ID software is that the set of rules used to distinguish between species can be applied consistently. Human observers vary in experience and capability; it is therefore difficult to ensure consistency across observers - and even within the same observer over time. Unlike Auto-ID software, human errors can be reinforced over time in a process called 'concept drift'.

It is important to appreciate that the bat species that are easiest for humans to distinguish acoustically are also often easiest for the Auto-ID software to distinguish; conversely, their limitations are similar.

Choosing the Auto-ID software you want to use

Choosing the right software

To understand the limitations of Auto-ID software, we recommend testing each of the Auto-ID software packages that are available or are being considered against a standard set of bat calls/passes. This could be a set of 'expert-labelled' data, or a library of known calls. Both have their limitations (the first relying on human interpretation; the second often comprising calls recorded in atypical situations to ensure the species is correctly identified), but either is a good starting-point at the current time.

Note that the 'expert labels' should only identify each call (or sequence of calls) to the level that is appropriate (i.e. not all calls can be identified to species).

The reference data should be run through the software option(s) and their results compared to the 'correct' results. This step would not be necessary if information on how each Auto-ID software package works or the libraries used were published by the authors, but this is currently not available for most software packages. Related to the libraries used, it is important to be aware that Auto-ID software can perform poorly, or differently, if the library underlying the Auto-ID software was not made on the same make of detector.

In the absence of clear guidance at the current time, we recommend that three steps are followed in determining the most appropriate software to choose for a particular study or scenario:

Step 1: assess the detectability of the software you are using – how many calls/passes is the software failing to recognise as 'bat' and is there a species (or frequency range) bias to detectability? Can you alter the settings to reduce the calls/passes 'lost'? Some Auto-ID software (iBatsID) currently rely on other software to 'detect' and parameterise calls in recordings, whilst other Auto-ID software integrate this altogether with identification.

Step 2: identify how close the software results are to your 'correct' classifications (i.e. to the expert-labelled or known data set). What percentage of calls/passes are correctly identified? Is there a species bias in classification errors? Will any species bias impair the interpretation of your results, or require high levels of manual verification? Are there anomalous results?

The software you choose would ideally give a level of 'identification certainty' between 0-1, which relates to the probability of a correct identification.

You might have to alter the trade-off between certainty and level of classification: i.e. you might opt for more bat calls/passes to be identified (but at a lower level of certainty), or opt for a higher level of certainty of ID (in which case, fewer bat calls/passes might be assigned an ID).

Step 3: Use the results of the tests iteratively to modify the settings to improve call/pass detection and classification until they are acceptable for at least one of the Auto-ID software package options being considered.

These steps should be followed even if you are considering one type of software only.

Other considerations

In addition to determining if the software is *detecting* enough calls (or passes), and *classifying* them adequately for the needs of your study, it is also important to consider the following.

- Will the output need post-processing?
- Can the Auto-ID software handle the volume of data at an acceptable speed?
- Can the Auto-ID software identify more than one bat of the same or different species in a sound file? Is this important?
- How well does the Auto-ID software handle bat social calls and feeding buzzes?
- Does the software give enough value for money?

Once the field work starts

Check your field set-up. Do not wait until the end of the season to process your data; analyse as you collect. This will enable you to:

- trial real data against the assumptions you made based on the test data, to ensure the software is functioning as you expected;
- check your field and analysis settings;
- ensure your hardware is working;
- modify/augment your survey methodologies (for example, if you pick up rarer species and need to increase survey frequency or trap).

Using the Auto-ID software

Checking the data

With a subset of your data, test again the implications of changing the level of certainty of classification. If you accept a lower certainty you will get more bat calls/passes classified but at the expense of accuracy. Explore what options the software gives you for lower certainty—for example, classifying to family, genus or other group level rather than species.

Once you have a setting you are happy with, retain those settings for the whole of the analysis. The settings used should be stated in any report: they are an important part of your methodology. Explain the rationale behind the species you have grouped.

Use the results of the trials to drive your verification. Firstly, manually check a proportion of the noise files to

check the proportion of calls/passes that are not being detected by your Auto-ID software.

If you know, for example, that the Auto-ID software is likely to fail to detect 20% of all calls/passes from your data, you will be analysing a smaller proportion of the bat activity you recorded (itself a sample), but this might be acceptable if it does not materially change your interpretation.

If, however, detectability of bat calls within recordings is so biased that 40% of the calls/passes of a particular species are missed or misclassified, you may need to: adjust your interpretation of the results to reflect that; undertake a higher level of manual analysis (to look for missing passes); or use different software.

Similarly, it is also important to **check for anomalous results**, and manually examine enough data to understand what the Auto-ID software is doing. Anomalous results (for example, species outside their known range) may be real (i.e. the unexpected species is really there). However, these may also result from an inadequate call library within the software used to determine species, or result from mis-interpretation (a noise at a particular frequency is identified as a bat species that emits a similar constant-frequency call). It is clearly important to **use a suitably-experienced bat acoustics analyst to undertake the verification.**

Note: *In the UK, it is sometimes suggested that all non-common and soprano pipistrelle calls should be manually checked. Accepting automatic classification of common and soprano pipistrelles is normally done to save time, as these species often comprise over 90% of recordings, but this is based on an assumption that all common and soprano pipistrelle calls will be correctly identified, which is not necessarily the case. **Verification should be based on the results of your trials**, and depending on this, may include checking a random sample of recordings of these common species as well as an appropriate sample of the less common species.*

Analysing the bat activity data

This is a big subject beyond the scope of this guidance note. However, the following points are important to consider when analysing the outputs of the Auto-ID software:

- The results will usually comprise a series of bat calls/passes collected over a number of nights. It should not be assumed this data is normally-distributed; bat activity data is often highly skewed.
- The distribution of the data will dictate how it should be analysed and which metrics are reported. If the data are highly skewed, reporting mean activity alone may be misleading.
- It is important to also analyse your data by species / species groups rather than analysing overall bat activity (i.e. all species combined), as the patterns

of distribution of rarer and quieter-calling species will be masked.

Interpreting the results

Again, this is a wide subject, but the following may be helpful.

- Use the initial trial and the early tests on real data to assist your interpretation.
- Understand the proportions of calls/passes that may be lost (i.e. not detected by the Auto-ID software).
- Understand where misclassifications are likely to be occurring (for example, when a particular species is misidentified, is there a pattern to that misclassification?).

The factors influencing which bats are recorded in the first place must also be considered. Calls/passes captured are not independent of species. For example:

- Some bat species issue quieter calls and will therefore be picked up over shorter distances than those of species with louder calls.
- Calls of some species vary significantly in volume according to activity (e.g. some species reduce call loudness to catch prey).
- Bats also vary their calls in different habitats (e.g. according to the degree of clutter) and for different activities (e.g. fast flight, foraging, or social calls).
- Calls may also vary in directionality (which affects the ability of microphones to capture them).
- Sound attenuates differently with changes in temperature and humidity.

It is therefore important that your interpretation does not compare relative activity between species/species groups. Species which are easily detectable and common might dominate the activity recorded, but only because they are more detectable.

It is also important to consider the survey methods used when interpreting bat activity data. It is far easier to produce a standardised measure of bat activity that can be compared within and between sites if detectors are left out to record over a complete night. Activity in the first hours after sunset may not necessarily provide a good reflection of activity over a whole night, and activity at a particular point along a walked or driven transect will be dependent on the time relative to sunset. It is also important to consider night length when comparing activity data from different times of year.

Finally, results can only be directly compared across sites that have used the same equipment, settings, software and version of the software. It is possible to compare sites which have not sampled in the same way, but this requires calibration of the differences.

We would advise that all recordings are archived, to be able to reanalyse these in the future. This is particularly important if there is an interest in monitoring the same sites over time, but is also advisable given that the Auto-ID software is improving all the time.

The future

It is recommended that a 'gold standard' set of calls (ideally, a library of known calls) is collated and made freely-available for all software producers to test their products against. It is also recommended that the software producers publish their results so that the performance of different packages is clear.

Acknowledgements We would like to thank Kathryn Skinner (Arcadis), who assisted with the early trials and was meticulous in recording the steps taken; also Jon Davies (Arcadis) and Elizabeth Rowse (University of Bristol) who provided valuable comments on this note.

Authors' contact details

Paola Reason: Arcadis Consulting (UK) Limited.
paola.reason@arcadis.com

Stuart Newson: British Trust for Ornithology
stuart.newson@bto.org

Kate Jones: University College London
kate.e.jones@ucl.ac.uk

¹ Lucas T.C.D., E.A. Moorcroft, R. Freeman, J.M. Rowcliffe, and K.E. Jones. (2015). A generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution* **6** (5): 500-509.