Rydell, J., Nyman, S., Eklöf, J., Jones, G., & Russo, D. (2017). Testing the performances of automated identification of bat echolocation calls: A request for prudence. *Ecological Indicators*, *78*, 416-420. https://doi.org/10.1016/j.ecolind.2017.03.023

# Testing the performances of automated identification of bat echolocation calls: a request for prudence

Jens Rydell[a], Stefan Nyman[b], Johan Eklöf[c], Gareth Jones[d], Danilo Russo[d,e]

Addresses:

a. Biology Department, Lund University, SE-223 62 Lund, Sweden; jens.rydell@telia.com

b. Skarpskyttevägen 30D, SE-226 42 Lund, Sweden

c. Krokdalsvägen 88, SE-51734 Bollebygd, Sweden

d. School of Biological Sciences, Life Sciences Building, University of Bristol, 24 Tyndall Avenue Bristol BS8 1TQ, U.K.

e. Wildlife Research Unit, Laboratorio di Ecologia Applicata, Sezione di Biologia e Protezione dei Sistemi Agrari e Forestali, Dipartimento di Agraria, Università degli Studi di Napoli Federico II, Via Università 100, Portici (Napoli), Italy

Communicating author

Jens Rydell, Biology Department, Lund University, SE-223 62 Lund, Sweden

jens.rydell@telia.com

**Abstract**

Echolocating bats are surveyed and studied acoustically with bat detectors routinely and worldwide, yet identification of species from calls often remains ambiguous or impossible due to intraspecific call variation and/or interspecific overlap in call design. To overcome such difficulties and to reduce workload, automated classifiers of echolocation calls have become popular, but their performance has not been tested sufficiently in the field. We examined the absolute performance of two commercially available programs (SonoChiro and Kaleidoscope) and one freeware package (BatClassify). We recorded noise from rain and calls of seven common bat species with Pettersson real-time full spectrum detectors in Sweden. The programs could always (100%) distinguish rain from bat calls, usually (68-100%) identify bats to group (*Nyctalus/Vespertilio/Eptesicus, Pipistrellus, Myotis, Plecotus, Barbastella*) and usually (83-99%) recognize typical calls of some species whose echolocation pulses are structurally distinct (*Pipistrellus pygmaeus*, *Barbastella barbastellus*). Species with less characteristic echolocation calls were not identified reliably, including *Vespertilio murinus* (16-26%), *Myotis* spp. (4-93%) and *Plecotus auritus* (0-89%). All programs showed major although different shortcomings and the often poor performance raises serious concerns about the use of automated classifiers for identification to species level in research and surveys. We highlight the importance of validating output from automated classifiers, and restricting their use to specific situations where identification can be made with high confidence.

**Key words**

Biosonar, Methodology, Software, Species identification, Ultrasound.

## *1. Introduction*

Acoustic methods of species identification represent a powerful approach to studying the distribution, ecology and behaviour of animals that broadcast sound for communication or echolocation (Towsey et al., 2014). In many cases, such as for bird- and cricket songs (e.g. Briggs et al., 2012; Lehmann et al., 2014), this approach involves reliable species identification. Bats are not birds or crickets, however, and, more importantly, echolocation calls, which generally are used to identify bats, are not songs (Barclay, 1999). While songs have the primary objective of announcing the identity of the singer, echolocation calls provide information for tasks such as orientation or prey detection, and therefore vary dramatically depending on task (Obrist, 1995). Moreover, different species often solve similar tasks using similar calls (Jones and Holderied, 2007), which means that considerable overlap in call structure is expected. Hence, species recognition based on echolocation calls is not nearly as straightforward as recognition based on e.g. bird or cricket songs and often leads to substantial challenges (Russo and Voigt, 2016).

Developments in ultrasonic technology have revolutionized the study of bats over the last few decades, and ultrasound detectors or "bat detectors" are now used routinely to study and survey bats in the field all over the world (Parsons and Szewczak, 2009). Over the years, researchers have moved from manual species identification, listening to heterodyned and/or time-expanded sound sequences (Ahlén, 1981), to analyses of displayed call sequences using various software (e.g. Russo and Jones, 2002). Recently, different automated approaches, typically employing multivariate sets of spectral and temporal variables of bat calls, have been attempted with variable results (e.g. Parsons and Jones, 2000; Walters et al., 2012; Zamora-Gutierrez et al., 2016). Freeware and commercial software used to speed up the screening of long recordings, select echolocation calls and identify species have recently appeared and are used extensively. The frequent use of automatic recorders triggered by bat calls and generating large audio data-sets when left unattended in the field for long periods have made such software welcome, because it saves time and facilitates analysis of large data-sets.

Software producers certainly make warning notes of the risk of misclassification of some species or under certain recording conditions, but the temptation of using automatic tools non-critically remains strong. This may be especially true for ecological consultants with little or no experience with bats. Another reason for concern is that the performance of automated classification software has not been sufficiently validated before their release into the market (Russo and Voigt, 2016). Although the limitations of automated classification have been highlighted by showing that different software packages identify calls of unknown bat species in different ways (Lemen et al., 2015), the reliability of identifying species of known identity remains little known.  A first step is to test the performances of some popular software in the field under normal working conditions. This would give users a better grasp of the possibilities and limitations of these tools. To help fill this gap, we tested the absolute identification performances of three popular packages by recording echolocation calls from free-flying bats of known identity (i.e. the recorded bats were identified beforehand based on several complementary criteria - their real identity was therefore accurate and not based solely on our own sound identification ability).

Hence, the aim of this work was to test the program performance in an "absolute" sense, i.e. against known bats, or at least as absolute as we could. We did not test it against subjective identifications of recordings made by ourselves, or by invited experts or volunteers (e.g. Jennings et al., 2008; Lewandowski and Specht, 2015). Indeed, we do not claim that we would be able to manually identify all species included here with sufficient accuracy based on the recordings alone. We also tested whether the software can distinguish environmental noise such as rain from bat echolocation calls. Finally, we provide some preliminary guidelines on how automated classifiers for the identification of bat species may and may not be used.

## ***2. Materials and methods***

### *2.1 Field recordings*

We made recordings of free-flying bats in Sweden in 2013-2016, using Pettersson D1000X and D500X bat detectors (Pettersson Elektronik AB, Uppsala, Sweden; www.batsound.com). The recordings were 3-4 s long full spectrum real time sequences with good signal to noise ratio sampled at 384 or 500 kHz and 16 bits. For each species we used 1-4 sets of recordings made in various parts of the country and under different conditions (specified in supplementary material 1). However, the identification performances of the programs turned out to be very similar across all the sets within each species, and we therefore pooled the sets before presentation. We used only recordings for which there was no doubt of the real identity of the bat being recorded.

The identities of the recorded bats were established as follows:

a) Individuals of *P. pygmaeus*, *Myotis brandtii* and *Plecotus auritus* were recorded as they were seen to emerge from or return to roosts where the bats had been identified beforehand, usually morphologically (captured individuals). We carefully avoided making recordings of bats that did not use typical search-phase echolocation pulses, i.e. those being < 20 m from the roost exit. The exception is *P. auritus*, which was also recorded inside roosts (churches), where colonies had been identified visually beforehand. *P. auritus* emerging from roosts could always be recognized on its large and diagnostic ears.

b) Individual *Eptesicus nilssonii* and *Vespertilio murinus* were recorded at specific feeding territories where they have been observed regularly over long periods during previous studies. In these cases the recordings were made at close range (< 10 m) and under good light conditions prevailing during the light nights of summer in Scandinavia. The bats´ identities were thus confirmed based on a combination of size, wing shape and colour and, in addition, echolocation calls, by use of the bat detector. The light bellies and smaller sizes distinguish the two species from *Nyctalus noctula*, although all three sometimes co-occur in the area where the recordings were made, and may emit similar echolocation calls (authors´ unpublished observations).

c) We used recording sequences containing intermittent and diagnostic social calls of presumed *V. murinus* to identify the bats unambiguously (Zagmajster, 2003). Search phase echolocation call sequences from these recordings were used in the test, but the social calls were excluded.

d) Recordings of *E. nilssonii* and *M. brandtii* were made in subarctic Lappland, where no other bats occur (Ahlén, 2011, author´s unpublished observations). In this case, the identifications were facilitated by very good visual views, sometimes in sunlight.

e) We recorded *M. daubentonii* at a locality in southernmost Sweden, where it is the only bat species foraging low over water (trawling). We only recorded these bats as they flew low over water and hence immediately and unambiguously were recognized to species.

f) *B. barbastellus* was recorded at a known feeding territory and near a hibernaculum, in both cases in places regularly used by several individuals over long periods. We made sure that all recordings used in the analysis included the unique alternating pulses diagnostic for this species (Görlitz et al., 2010).

g) To minimize the risk that the recorded rain (noise) files actually were from bats, they were recorded in Tärendö in northernmost Sweden, an area where no bats are known to occur.

To simplify the classification tasks as far as possible and make our analysis conservative, we excluded all files containing calls from more than one individual. We also excluded social calls and sequences emitted in close proximity (<20 m) to roosts or clutter. However, for *P. auritus* we included the short broadband sweeps typically used in cluttered situations, which is the normal foraging habitats of this species, and also sequences with its characteristic low frequency sweeps (Furmankiewicz et al., 2013), some of which were recorded inside the roost (a church loft). To simplify the task even further we made sure that all sequences used in the analysis were recorded under typical flight conditions for each species, e.g. *M. daubentonii* low over water, *E. nilssonii*, V. *murinus* and *P. pygmaeus* in more or less open space, *M. brandtii* and *B. barbastellus* in semi-open situations or ecotones, and *P. auritus* in clutter. Some typical sequences used in the test are provided as supplementary material 3.

*2.2 Software tested*

Using SonoChiro v. 3.3.3 (Biotope, France; www.biotope.fr ), Kaleidoscope Pro 3.14B (Wildlife Acoustics, U.S.A; www.wildlifeacoustics.com) and BatClassify version 2014-07-14 (Chris Scott and John Altringham, U.K.; https://bitbucket.org/chrisscott), we tested whether the software could correctly attribute recorded call sequences to species group, an output provided by two (SonoChiro and BatClassify) packages, and species, provided by all three packages. Species groups were *Nyctalus*/*Vespertilio*/*Eptesicus* ("NVE"), *Pipistrellus*, *Myotis*, *Plecotus* and *Barbastella*. All three programs were provided with the same sets of recordings, but some files were discarded by the programs or were attempted but without any species being suggested ("no id"). The remaining sequences were either identified correctly or erroneously.

We used the default settings; for SonoChiro - type of recorder, region (North Boreal), time expansion (x1), maximum call duration (0.5), and sensitivity (7), for Kaleidoscope - filter (filter noise files, keep noise files), signal of interest (8-120 kHz, 2-500 ms, minimum 2 calls), classifiers = bats of Europe 3.1.3 (-1 more sensitive). No setting choices were available for BatClassify. Neither *V. murinus* nor *E. nilssonii* files was used to test BatClassify at the species level as its reference libraries only cover species occurring regularly in the U.K. However, the recordings of these species were tested to group level.

All software classified the sequences (files) according to the echolocation calls they contain, so results were expressed as percent of files correctly or erroneously classified to species groups or species. The programs provided "probabilities" of correct classification and one of them (Kaleidoscope) also suggested alternative species. However, as we found no way to interpret and standardise this information we did not use it.

Following Jennings et al. (2008), we also calculated, for each species and program, two indices, namely Sensitivity and Positive Predictive Power (PPP). Sensitivity is the percentage of recordings that belong to a given species that were correctly classified, while PPP is the percentage of recordings classified as a species that were actually of it. This was done to facilitate comparison of performances across different identification methods, such as e.g. by professional field workers or volunteers (Lewandowski and Specht, 2015). The indices are presented as supplementary material 2.

### *3. Results*

Each program was given 2275 files containing bat calls and 190 containing only noise from rain. All rain files were distinguished from bat sound by all packages (table 1). The frequency of rejected or not identified files varied between the programs and even more so between species within each program. Files with *E. nilssonii* were rejected particularly often by SonoChiro (41%) and Kaleidoscope (44%), and the latter also rejected many (55%) *P. auritus* files. The two programs that classify to group (SonoChiro and BatClassify) did so correctly in most cases (87-100%), although *E. nilssonii* and *V. murinus* were only correctly classified to their group ("NVE") about half the time (55% by Sonochiro, for the two species combined), as many files (40%) were rejected (not attempted).

Identification at the species level was highly variable both among programs and bat species. This is also clearly shown by the heterogeneity in the values of sensitivity and PPP (supplementary material 2). Two species that employ either a unique frequency band (*Pipistrellus pygmaeus*) or unique alternating call sequences (*B. barbastellus*) were identified with good or at least reasonable accuracy by all software (97-99% and 62-95% correct, respectively). In contrast, *E. nilssonii* and *V. murinus*, belonging to the NVE group with several species using similar calls, were classified correctly only about half the time or less (49-54% and 16-20%, respectively). However, for *E. nilssonii* the low score was not primarily a result of errors, but of many rejected files.

Classification of *M. brandtii* and *M. daubentonii* was extremely variable and inconsistent (4-93% and 0-98% correct, respectively) with error rates as high as 96-100% in some cases (*M. brandtii* by SonoChiro and *M. daubentonii* by BatClassify, respectively; table 1). *P. auritus* was usually identified correctly by two programs (80% and 89% for SonoChiro and BatClassify, respectively), but not at all by the third (0% for Kaleidoscope). In the latter case the files (95%) were usually rejected, only three identification attempts were made, all resulting in errors (table 1). We double-checked Kaleidoscope's performances on *P. auritus* by trying various settings but always obtained the same result.

Misclassifications (false positives) occurred within genera (e.g. *M. brandtii* and *M. daubentonii* misidentified as other *Myotis* spp.) but also across genera and groups of genera (table 2). For example, *E. nilssonii* was identified as belonging to five different genera, including *Myotis* and *Barbastella* and was particularly often identified as *M. dasycneme* (82% of the misidentifications by SonoChiro) and *Nyctalus leisleri* (64% by Kaleidoscope). Likewise *V. murinus*, which is notoriously difficult to identify manually from sonograms because of its broad frequency overlap with other species (Ahlén, 1981), was misidentified as belonging to four genera, including *Nyctalus* spp. (88% of the misclassifications by Kaleidoscope) and *Eptesicus serotinus* (40% by SonoChiro), but also quite frequently as *P. auritus* (32% by SonoChiro).

### *4. Discussion*

Although the software packages that we tested showed inconsistent performances, some generalization can be made. For example, rain noise was distinguished from bat calls successfully by all programs, suggesting that they can be used to sort files containing bat calls from those containing only rain noise. However, environmental noises other than rain, such as sounds from rustling leaves, strong winds or running water, were not tested, so we cannot generalize across all sorts of environmental noise. We also caution that the high rejection rate of some bat calls, such as the short sweeps of *P. auritus* flying in clutter, suggests that there may be a risk that true bat calls were rejected as noise.

Generally, the programs successfully classified bat calls into broad groups (genera or in one case a group of genera) or identified the species with the most characteristic echolocation calls such as *P. pygmaeus* and *B. barbastellus*. This suggests that the programs may be used to survey these particular genera or species. However, it must be stressed that our study took place in a country with relatively low diversity of bats (19 species, Ahlén, 2011), so that the task was much simpler than in more species-rich sites. It was also simpler than it would have been if we had included calls from atypical habitats, social calls or calls from more than one individual or even several species at the same time. Indeed, serious shortcomings were evident for most species, including *V. murinus* and *E. nilssonii*, as already discussed, and also *M. brandtii* and *M. daubentonii*. Generally, many *Myotis* species, including those that we included in this test, use similar echolocation calls which may be difficult to classify (Parsons and Jones, 2000). Therefore, by recording the two species only in their most typical habitats (*M. brandtii* in forest and *M. daubentonii* low over water, respectively), we gave the programs a chance to base the classification not only on the species but also on the variation that relates to habitat (Obrist, 1995). Since none of the programs could distinguish the two species nevertheless, it seems unlikely that any Scandinavian *Myotis* can be recognized reliably. It was also unexpected that *B. barbastellus* was so frequently misclassified. This species uses two unique call types that alternate at different frequencies, unlike any other bat in Scandinavia. Barbastelles were identified as *Myotis* spp., *Pipistrellus* spp., *E. nilssonii* and even *P. auritus* (table 2).

It is striking that basic discriminant analysis or neural network approaches attempted many years ago (Parsons and Jones, 2000; Russo and Jones, 2002) did better than the suite of algorithms currently used in the modern software that we tested. Our results raise serious concerns about the risk of making considerable identification errors by using automated identification of bat calls, and this may bring about potentially detrimental consequences for conservation and species management. Needless to say, identification performances ranking well below 100% of correct classification should not be used for mapping species distributions, but obviously other surveys would also be compromised by incorrect identifications. Overall, our work confirms the concerns expressed by Russo and Voigt (2016) on the reliability of automated identification software and calls for prudence in the adoption of such tools for acoustic surveys and research.

We recognise that automated identification of bat echolocation calls can be valuable for specific purposes and provided that certain caveats are met. For example, it may be effective for recognizing some particular easy-to-recognize species in particular areas such as *Pipistrellus pipistrellus* and *P. pygmaeus* in the U.K. (Rowse et al., 2016). Also our results suggest that it is sometimes preferable to classify bats to species groups rather than to species, as error rates were relatively low for the former, although pooling different species may sometimes be insufficient to provide the information needed.

## *5. Acknowledgments*

Funding was provided by the Swedish Energy Agency through the Vindval program (2016-000101) to JR.

## *6. References*

Ahlén, I., 1981. Identification of Scandinavian bats by their sounds. Swedish Univ. Agric. Sci. Dep. Wildlife Ecology, Uppsala 6, 1-56.

Ahlén, I., 2011. Fladdermusfaunan i Sverige. Arternas utbredning och status. Fauna flora Stockholm 106, 2-19.

Anderson, E., Racey, P. A., 1991. Feeding behaviour of captive brown long-eared bats *Plecotus auritus*. Anim. Behav. 42, 489-493. http://dx.doi.org/10.1016/S0003-3472(05)80048-X

Barclay, R.M.R., 1999. Bats are not birds - a cautionary note on using echolocation calls to identify bats. J. Mammal. 80, 290-296. https://doi.org/10.2307/1383229

Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J.K., Hadley, A.S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach. J. Acoust. Soc. Am. 131, 4640-4650. doi: 10.1121/1.4707424

Furmankiewicz, J., Duma, K., Manias, K., Borowiec, M., 2013. Reproductive status and vocalisation in swarming bats indicate a mating function of swarming and an extended mating period in *Plecotus auritus*. Acta Chiropterol. 15, 371-385. doi: http://dx.doi.org/10.3161/150811013X678991

Görlitz, H., Hofstede, H.M. ter, Zeale, M.R.K., Jones, G., Holdereid, M., 2010. An aerial-hawking bat uses stealth echolocation to counter moth hearing. Curr. Biol. doi:10.1016/j.cub.2010.07.046

Jennings, N., Parsons, S., Pocock, M.J.O., 2008. Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. Can. J. Zool. 86, 371-377. doi: 10.1139/Z08-009

Jones, G., Holderied, M.W., 2007. Bat echolocation calls: adaptation and convergent evolution. Proc. R. Soc. Lond. B 274, 905–912. doi: 10.1098/rspb.2006.0200

Lehmann, G.U.C., Frommolt, K.-H., Lehmann, A.W., Riede, K. 2014. Baseline data for automated acoustic monitoring of Orthoptera in a Mediterranean landscape, the Hymettos, Greece. J. Insect Conserv. 18, 905-925. doi: 10.1007/s10841-014-9700-2

Lemen, C., Freeman, P., White, J. A. Andersen, B.R., 2015. The problem of low agreement among automated identification programs for acoustical surveys of bats. Western North American Naturalist 75: 218-225. doi: http://dx.doi.org/10.3398/064.075.0210

Lewandowski, E., Specht, H., 2015. Influence of volunteer and project characteristics on data quality of biological surveys. Conserv. Biol. doi:10.1111/cobi.12481.

Obrist, M., 1995. Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. Behav. Ecol. Sociobiol. 36, 207-219. doi: 10.1007/BF00177798

Parsons, S., Jones, G., 2000. Acoustic identification of 12 species of echolocating bats by discriminant function analysis and artificial neural networks. J. Exp. Biol. 203, 2641-2656.

Parsons S. Szewczak J.M., 2009. Detecting, recording, and analyzing the vocalizations of bats, in: Kunz, T.H, Parsons, S. (Eds.). Ecological and behavioral methods for the study of bats, 2nd ed., pp. 91-111. The Johns Hopkins Univ. Press, Baltimore, pp. 91-111.

Rowse, E.G., Harris, S., Jones, G., 2016. The switch from low-pressure sodium to light emitting diodes does not affect bat activity at street lights. PLoS ONE, 11, e0150884. http://dx.doi.org/10.1371/journal.pone.0150884

Russo, D., Jones, G., 2002. Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. J. Zool. Lond. 258, 91-103. doi: 10.1017/S0952836902001231

Russo, D., Voigt, C.C., 2016. The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound analysis. Ecol. Indicators 66, 598-602. http://dx.doi.org/10.1016/j.ecolind.2016.02.036

Towsey, M., Parsons, S., Sueur, J., 2014. Ecology and acoustics at a large scale. Ecol. Informatics. 21, 1-3. http://dx.doi.org/10.1016/j.ecoinf.2014.02.002

Walters, C.L., Freeman, R., Collen, A., Dietz, C., Fenton, M.B., Jones, G., Obrist, M.K., Puechmaille, S.J., Sattler, T., Siemers, B.M., Parsons, S., Jones, K.E., 2012. A continental-scale tool for acoustic identification of European bats. J. Appl. Ecol. 49, 1064-1074. doi: 10.1111/j.1365-2664.2012.02182.x

Zagmajster, M., 2003. Display song of parti-coloured bat *Vespertilio murinus* Linnaeus, 1758 (Chiroptera, Mammalia) in southern Slovenia and preliminary study of its variability. Natura Sloveniae 5, 27–41.

Zamora-Gutierrez, V., Lopez-Gonzalez, C., MacSwiney Gonzalez, M.C., Fenton, M.B., Jones, G., Kalko, E.K.V., Puechmaille, S.J., Stathopoulos, V. Jones, K.E., 2016. Acoustic identification of Mexican bats based on taxonomic and ecological constraints on call design. Methods Ecol. Evol. 7, 1082-1091. doi: 10.1111/2041-210X.12556

**Legends to tables**

Table 1. The performance of three automatic bat identification programs given as % of files submitted to the programs (N). SonoChiro and BatClassify identified to species group and species, Kaleidoscope only to species. “No id” files were not attempted by the programs or attempted but not resulting in any identification. Dashes mean that the species were not included in the package, because they are not recognized members of the U.K. fauna. Asterisk means that the output actually was *Myotis brandtii*/*Myotis mystacinus*.

Table 2. Misclassifications at the species level, where n is the number of misclassifications of the species by the program in question. In addition to those shown in the table, misclassifications also occurred frequently at the group level. Dashes mean that the species were not included in the package, because they are not recognized members of the U.K. fauna.