Supplementary Information

Supplementary Information 1: Playback preparation.

We built three different playbacks for each species to reduce pseudoreplication risk. The playbacks were built with either our personal files or files from public sound libraries such as Xeno-canto and Macaulay library. We selected the best quality recordings (high signal-to-noise ratio) for which the identity of the calling birds was clear. We acknowledge that recordings from libraries lack precision regarding the context in which the calls were recorded. We selected the recordings based on the personal knowledge we have of mobbing calls thanks to personal recordings (that were not clean enough to be added as playback, but clear enough to determine the characteristics of the sequence such as note composition and call rate). For all recordings, we used a low pass filter at 2000 Hz to reduce background noise.

We chose to combine several individuals in our final playbacks, for two reasons. First, our goal was to simulate a mobbing chorus that should be efficient in attracting conspecifics to mob. Secondly, these choruses contained a strong variability in their note composition and possible note combination. By combining three individuals that had different call rates and call organization, we created sequences of overlapping calls with variation in the note type and rates. We believe this method diminish the risk of recording responses stimulated because of one specific characteristic in the playback and allow the generalization of our results. In addition, several recent studies have shown that, when mobbing, several Parid species rely more on the general duty cycle of the call sequence than on the specific call organization (Landsborough et al. 2019, Salis et al. 2022). In the end, all playbacks had similar duty cycles (*B. bicolor*: 36.5 ± 6.82 sec, *C. caeruleus*: 37.57 ± 4.71 sec, *L. cristatus*: 37.53 ± 3.52 sec, *P. major*: 45.9 ± 1.77 sec, *P. ater*: 38.23 ± 2.80 sec, *P. carolinensis*: 39.9 ± 1.68 sec, *P. palustris*: 38.2 ± 10.90 sec).

For *P. major, P. ater, P. caeruleus* and *L. cristatus*, we used previous recordings from the authors in the same region. The birds were simulated with either predator calls (Eurasian Pygmy owl, *Glaucidium passerinum*) or with a chorus of conspecific calls. We detected no difference in call rate, call composition nor duty cycle between the two methods of recording. For all playbacks, the method consisted in walking into forest paths, then placing a loudspeaker at 1m height into a tree and retreating at a 15 m distance from the tree. We then launched either predator calls or a chorus of birds for 1 to 3 min until one bird approached with stereotypical behavior (wing flicks, calls, and repeated hops). We then stopped the playback and recorded the calling bird. We only kept recordings in which one bird was clearly distinguishable from other calling birds. We recorded until the bird stopped uttering mobbing calls (never more than 10 min).

For *B. bicolor*, *P. carolinensis* and *P. palustris*, we used a combination of Xeno canto, Macaulay library and personal recordings to create our playbacks, following the table below. We combine either two or three call sequences based on whether one or two individuals were calling in the recording and on the final duty cycle of the playback. The three personal recordings were obtained by *Author* in *Us State*, recorded by a stationary microphone on a tripod (~15 meters distance from the birds) while mobbing birds responded to owl playbacks.

Final ID	**Original ID**	**Classification from the recorder**	**Year**	**Country/ US State**
P. carolinensis 1	XC111140	call	2012	Tennessee
	XC391431	call	2017	Virginia
	XC391432	call	2017	Virginia
P. carolinensis 2	XC15190	call	2006	Tennessee
	XC33598	call	2009	Arkansas
	XC309487	call to human intruder	2016	Mississippi
P. carolinensis 3	Personal Recording	Mob to predator calls	2020	Florida
	Personal Recording	Mob to predator calls	2020	Florida
B. bicolor 1	ML84676	Mob	1996	Maryland
	ML303499	NA	2017	Viriginia
B. bicolor 2	ML110375	Scold	2001	West Virginia
	ML199010	NA	2010	New York
	Personal Recording	Mob to predator calls	2020	Florida
B. bicolor 3	ML23870	Call	1981	Louisiana
	XC333285	Call	2016	Alabama

	XC391445	Call	2017	Virginia
P. palustris 1	XC559277	Alarm call	2020	Sweden
	XC153824	Alarm call	2013	France
P. palustris 2	XC122187	Alarm call	2013	Poland
	XC213842	Alarm call	2015	Germany
P. palustris 3	XC114812	Alarm call	2011	Germany
	XC303443	Alarm call	2016	Poland
	XC210689	Alarm call	2015	Germany

Supplementary Information 2. Additional information about the recordings.

For the three last lines, we provide mean and standard deviation.

	B. bicolor	*C. caeruleus*	*L. cristatus*	*P. major*	*P. ater*	*P. carolinensis*	*P. palustris*
Sample size	19	20	20	20	20	20	20
Total time analyzed (sec)	915	1130	921	974	1010	984	964
Total number of notes	1599	4317	4252	2347	2167	2462	1792
Total Number of calls	359	330	658	334	617	263	398
D notes analyzed for acoustic analyses	180	200	200	200	200	200	185
F notes analyzed for acoustic analyses	94	123	164	121	97	200	198
Call size	4.45 ± 2.68	13.08 ± 5.43	6.46 ± 3.26	7.02 ± 2.79	3.51 ± 3.82	9.26 ± 4.52	4.50 ± 3.37
Call per minute	24.10 ± 6.70	17.72 ± 4.93	43.40 ± 15.16	21.34 ± 7.19	36.68 ± 7.19	16.84 ± 6.06	25.14 ± 6.95
Duty cycle (ratio time with calls / total time of the recording)	0.34 ± 0.15	0.19 ± 0.08	0.28 ± 0.07	0.22 ± 0.7	0.34 ± 0.15	0.24 ± 0.10	0.18 ± 0.10

Supplementary Information 3. Details on the Machine Learning program.

Creation of the spectrogram images

First, each recording was opened in R studio (R version 4.1.1) and with the function *segment* (package *soundgen*, version 2.5.2., Anikin 2019), we detected each segment of sound (i.e. each note). The parameters of the segment were the following: samplingRate = 44100/10, propNoise = 0.01, shortestPause = 1, shortestSyl = 1. The Signal-to-noise ratio was adapted to each file (median = 12; Q1 = 12; Q3 = 20) to avoid the over detection (i.e. one note is detected as several notes) or the under-detection (i.e. no note detected although there is one in the recording). We visually checked the correct detection of notes with the segmentation plot produced by the function. We then extracted a .jpeg picture of each note detected with a unique ID (spectrogram image of width = 240 pixels and height = 480 pixels).

Building the machine learning program

We relied on convolutional neural networks (CNN) models, known to achieve the best performance in image classification since the 2010's (Krizhevsky et al. 2017) and now used in different fields including ecology (Christin et al., 2019.; Tuia et al., 2022). We relied on an EfficientNetB3 model (Tan and Le, 2019) pre-trained on ImageNet (*https://www.image-net.org*), with resolution 300x300 and developed our ML program using TensorFlow-Keras in Python. We first split the training data set of FME and D notes into an 80/20 partitioning, to obtain a validation data set on which we evaluated the performance of our CNN model. We then estimated the model parameters using an iterative algorithm, stochastic gradient descent, and stopped the estimation when the validation accuracy (i.e. how much the notes in the validation sets are well classified) did not increase and reached a maximum of 98%. This procedure was run on a Titan X GPU. Finally, our ML program was able to return predicted notes (FME or D) with a certainty score (between 50% and 100%).

Extraction of the combination results

For each recording, we extracted a dataset for which each line was either a note or the silence between two notes. For each note and each silence, we obtained its duration and the percentage of certainty from the machine learning program. Notes assigned with certainty score of < 90% (9.7% of the notes) were independently checked according to their spectrogram by three experimenters (AS and two trained assistants). If at least two out of the three experimenters disagreed with the machine learning program, the note was changed accordingly (1.67% of the notes). In 75% of the cases in which a note was changed, the three experimenters agreed.

From these files, we extracted calls and their composition. To determine the difference between inter-call and intra-call duration, we plotted a histogram of all durations between notes present in the recording. The intra call duration is a small and stable duration with high occurrence in the call sequence. We therefore were able to visually detect the duration threshold between intra and inter call distance by looking at a rapid and clear discrepancy in the distribution. Once we determined our threshold (one per species, ranging from 0.1 to 0.3 sec) we suppressed all silence lines with a duration inferior to that threshold. We then extracted the calls (e.g. FME-FME-D-D-D-D-D) and created variables to describe these calls qualitatively and quantitatively: number of notes FME, number of notes D, total number of notes, first letter of the call, and number of transitions FME→D or D→FME (e.g. FME-D-D-D-D-D has only one transition). If the first letter of the call was a FME and if only one transition occurred inside the call, the call was therefore labelled as an FME-D call (e.g. FME-D-D, FME-FME-FME-D-D-D-D, and FME-FME-FME-FME-FME-FME-FME-D were all considered as FME-D calls). For each recording, we were therefore able to calculate the percentage of calls that were FME-D calls compared to any other call organization (D notes only, FME only, D-FME calls, and any other combination). From this final dataset, we calculated the average number of notes in calls, call number per minute, and duty cycle (Appendix 2).

References

Anikin, A (2019) Soundgen: an open-source tool for synthesizing nonverbal vocalizations. Behav research methods 51:778-792.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017) glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal 9:378–400.

Christin, S., Hervet, É., & Lecomte, N (2019) Applications for deep learning in ecology. Meth Ecol Evol 10:1632-1644.

Krizhevsky, A., Sutskever, I., & Hinton, G. E (2017) Imagenet classification with deep convolutional neural networks. Communications of the ACM 60:84-90.

Tan, M., & Le, Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., ... & Berger-Wolf, T (2022) Perspectives in machine learning for wildlife conservation. Nature communications, 13:792.