Appendix 4. Examination of relationship between recognizer recall and mean relative sound level (RSL).

We examined the relationship between recall and sound energy, which is a proxy for detection distance (Yip et al. 2020, Hedley et al. 2020), to ensure low recall of our recognizer was simply due to a lower effective detection radius than a human observer. We measured the sound energy, or loudness of each Eastern Whip-poor-will detection in our test dataset with two methods. First, we assigned a loudness ranking (Table A3.1) to each calling bout during manual counting of Eastern Whip-poor-will calls in our test dataset. We defined a calling bout as a duration of consecutive "whip-poor-will" calls with less than one second in between successive calls. Second, we measured the relative sound level (RSL) in each detected call with Song Scope software while running the recognizer.

Table A3.1. Rankings used to describe the loudness of Eastern Whip-poor-will calling bouts during manual counting of calls in audio recordings.

Loudness ranking	Description	Total # of calling bouts	Total duration of calling bouts (minutes)	Total number of calls
1	Barely perceptible	16	4.75	240
2	Barely perceptible to quiet	13	8.10	426
3	Quiet	29	8.23	455
4	Quiet to medium	10	13.17	777
5	Medium	3	1.67	113
6	Medium to loud	2	4.80	270
7	Loud	23	16.58	191
8	Very loud	4	3.25	191
	Not assessed	1	0.63	38

We then visually evaluated the relationship between recall and each sound energy measurement across the range of possible score thresholds. One bout of 38 calls was missed during loudness ranking (Table A3.1) and was eliminated from analysis of loudness ranking. We also tested the relationship between score threshold and each of the sound energy measurements for true positive detections using a linear mixed effects model with file name as a random effect. We compared each model to a null model using Akaike's information criterion (AIC).

There was a clear relationship between recall and both sound energy metrics, suggesting that recall is related to detection distance (Figure A3.1). Furthermore, the model including sound energy had an AIC weight of 1.0 for both metrics (loudness ranking: ΔAIC = 283.1, RSL: ΔAIC = 57.8).

We conclude that the recall of our recognizer is simply due to a lower detection radius than a human observer, which is concordant with previous studies (Knight and Bayne 2018). We suggest the recording-level recall of the recognizer is much higher, however, due to the high number of calls that Eastern Whip-poor-wills produce (Mulaik et al. 1953).

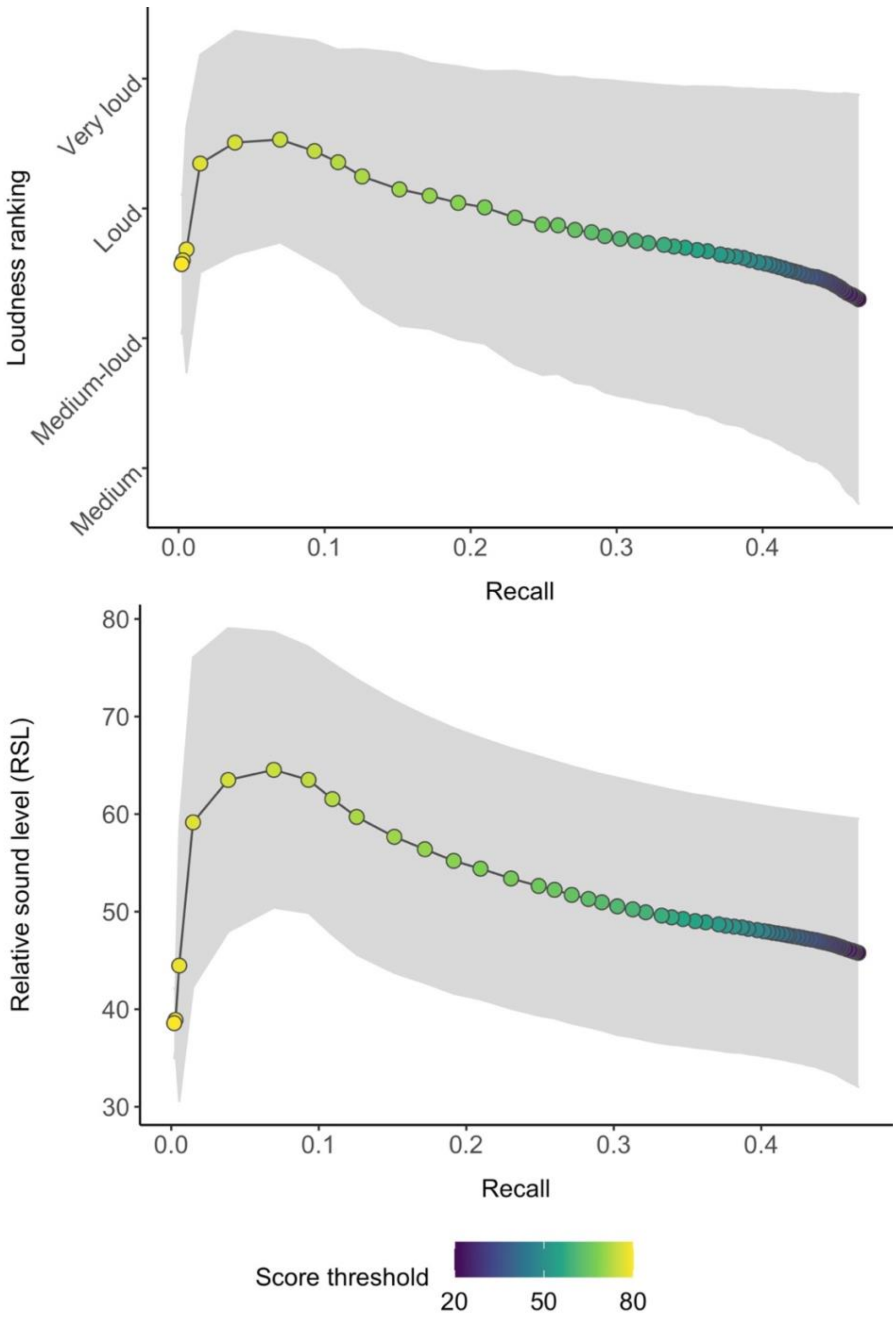

Figure A3.1. Relationship between recognizer recall and two metrics of sound energy (loudness ranking, relative sound level) across a range of score thresholds for an Eastern Whip-poor-will recognizer built in Song Scope software. Points are the mean value for each score threshold and grey shading is the standard deviation.

Literature Cited

Hedley, R. W., S. J. Wilson, D. A. Yip, K. Li, and E. M. Bayne. 2020. Distance truncation via sound level for bioacoustic surveys in patchy habitat. Bioacoustics 30:1–21. doi: 10.1080/09524622.2020.1730240.

Knight, E. C., and E. M. Bayne. 2018. Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. Bioacoustics 28:539-554. doi: 10.1080/09524622.2018.1503971.

Mulaik, S., W. Lee, and J. E. Stillwell. 1953. Analysis of the Call of the Whip-Poor-Will. Auk 70:367–368.

Yip, D. A., E. C. Knight, E. H. Audet, S. J. Wilson, C. Charchuk, C. D. Scott, P. Solymos, and E. M. Bayne. 2019. Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations. Remote Sensing in Ecology and Conservation 71:2759–15. doi: 10.1002/rse2.118.