

# Applying Artificial Intelligence Methods to Detect and Classify Fish Calls from The Northern Gulf of Mexico

Emily E. Waddell <sup>1,\*</sup>, Jeppe H. Rasmussen <sup>1,2</sup> and Ana Širović <sup>1,3</sup>

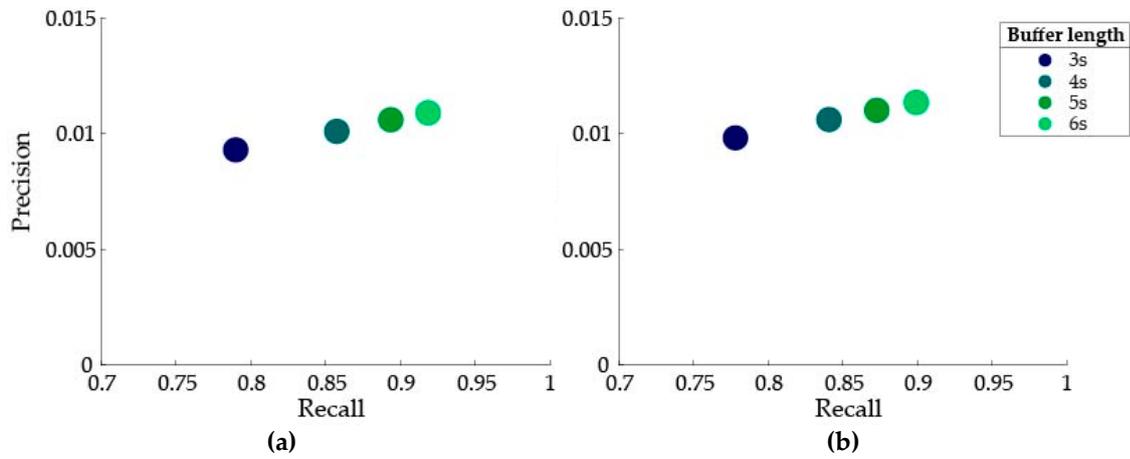
## Supplementary Materials

Table S1. Average confusion matrix of the three train/test dataset trials.

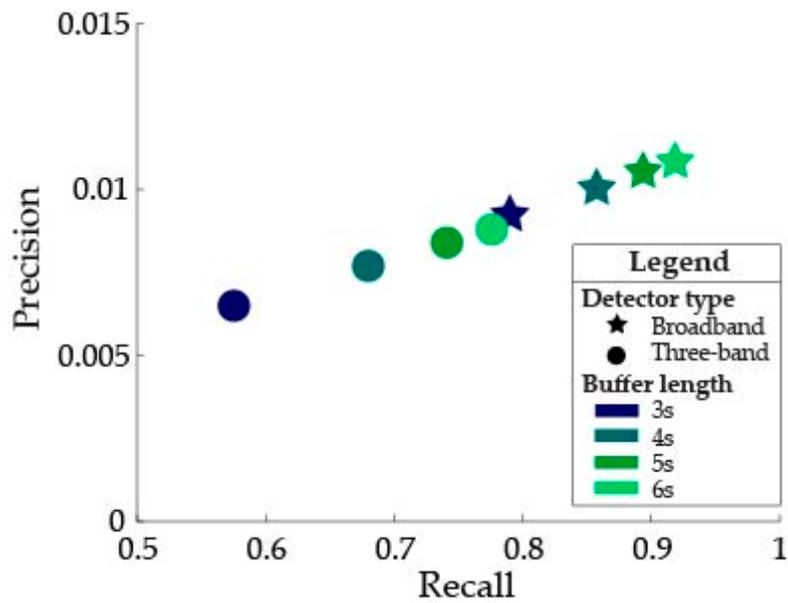
		Output												
Target		Beats	Buzz	Croak	Click Train	Disk Write	Down-sweep	Jet ski	Pulse Train	LF Noise	No Noise	Ra Noise	Sum	Recall
	Beats	529	0	0	0	6	0	0	1	9	68	12	626	85%
	Buzz	39	107	1	0	25	0	0	0	15	100	34	322	33%
	Croak	0	0	490	3	18	0	0	2	0	16	6	534	92%
	Click Train	5	0	12	919	57	0	0	2	3	498	17	1512	61%
	Disk Write	39	15	211	103	28,096	22	1	10	776	2,280	303	31,857	88%
	Down-sweep	0	0	0	0	14	138	0	8	1	19	0	180	76%
	Jetski	0	0	3	0	31	19	337	0	0	104	2	497	68%
	Pulse Train	0	0	4	0	44	18	0	221	1	40	3	332	66%
	Noise	288	23	182	387	4,855	131	5	98	1,994	41,633	5,930	55,527	89%
	Sum	902	146	903	1,412	33,146	328	343	341	2,801	44,759	6,308	91,387	
	Precision	63%	73%	56%	65%	85%	43%	98%	66%	76%	93%	94%		87.97%

Table S2. Average confusion matrix of the three evaluation dataset trials.

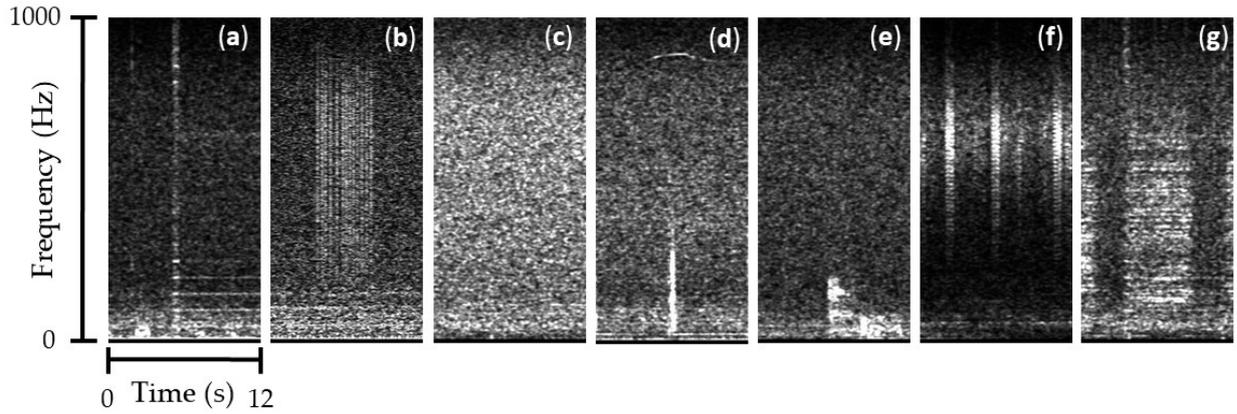
		Output												
Target		Beats	Buzz	Croak	Click Train	Disk Write	Down-sweep	Jet ski	Pulse Train	LF Noise	No Noise	Ra Noise	Sum	Recall
	Beats	422	1	0	0	6	0	0	2	23	31	3	488	86%
	Buzz	16	95	0	0	0	0	0	0	20	24	58	213	44%
	Croak	1	0	261	0	4	0	0	0	8	10	6	290	90%
	Click Train	4	0	3	39	4	0	0	7	2	15	3	77	51%
	Disk Write	3	8	84	0	4,521	43	4	3	421	1,255	178	6,521	69%
	Down-sweep	0	0	0	0	2	87	0	2	2	3	1	96	91%
	Jetski	0	0	5	0	4	60	525	2	14	215	22	847	62%
	Pulse Train	2	0	12	0	5	62	0	230	24	53	13	400	58%
	Noise	426	398	1,410	49	2,898	817	59	196	19,660	72,206	21,888	120,006	95%
	Sum	873	502	1,775	88	7,444	1,069	589	441	20,173	73,812	22,173	128,938	
	Precision	50%	27%	18%	44%	62%	10%	90%	53%	97%	98%	99%		93.02%



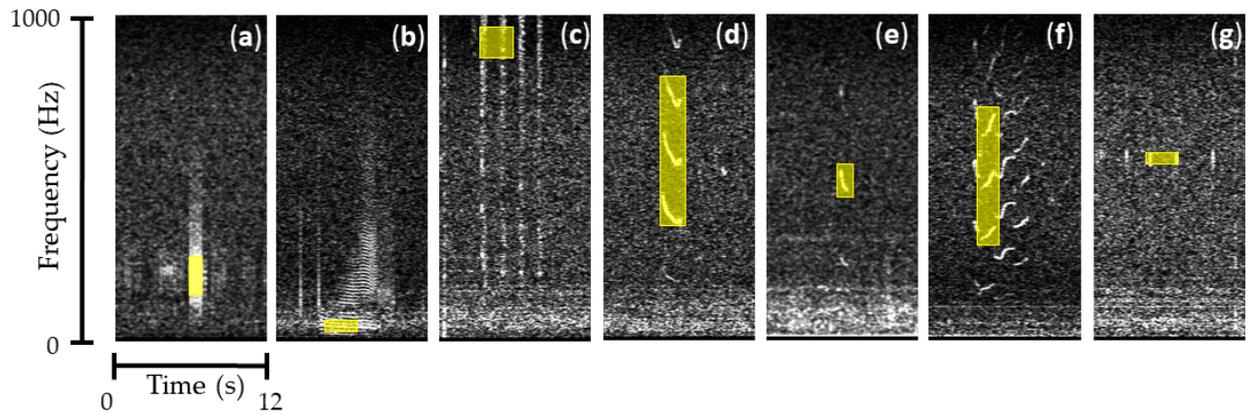
**Figure S1.** Energy detector precision and recall dependent on various buffer lengths (3 s, 4 s, 5 s, and 6 s) for: (a) the train/test dataset and (b) the evaluation dataset.



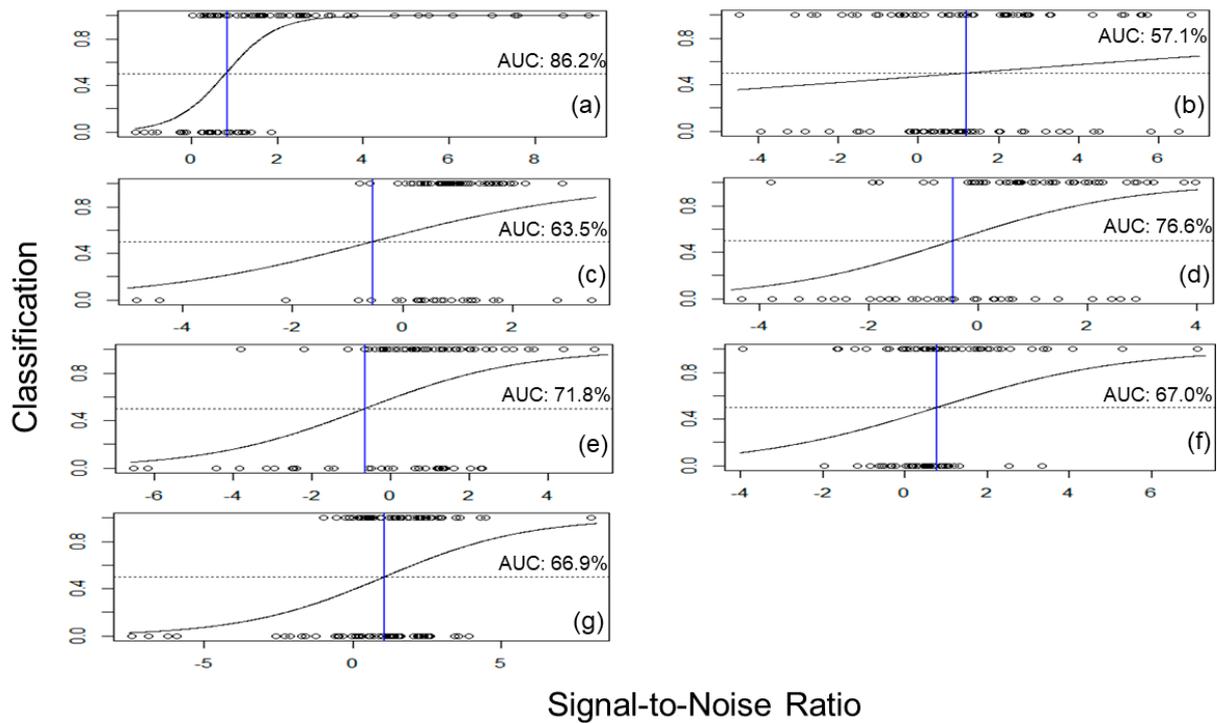
**Figure S2.** Comparison of precision and recall between the three-band detector (circles) and single, broadband (stars) energy detector based on various buffer lengths of 3 s, 4 s, 5 s, and 6 s.



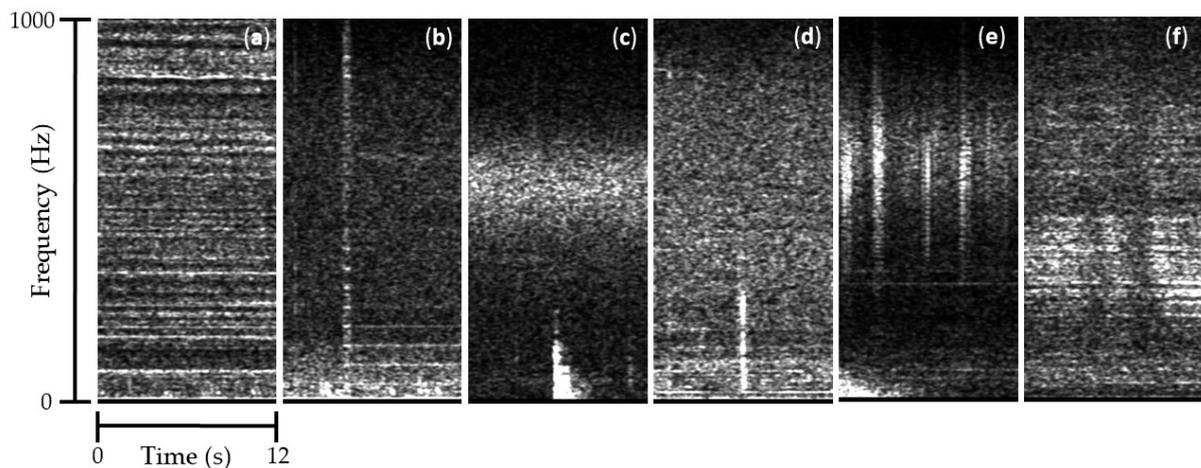
**Figure S3.** Images (created from spectrograms) of the five noise types that the classifier was trained with: (a) Disk write, (b) Click train, (c) Blank/ noise, (d) Low frequency noise, (e) Low frequency noise, (f) Random noise, and (g) Random noise.



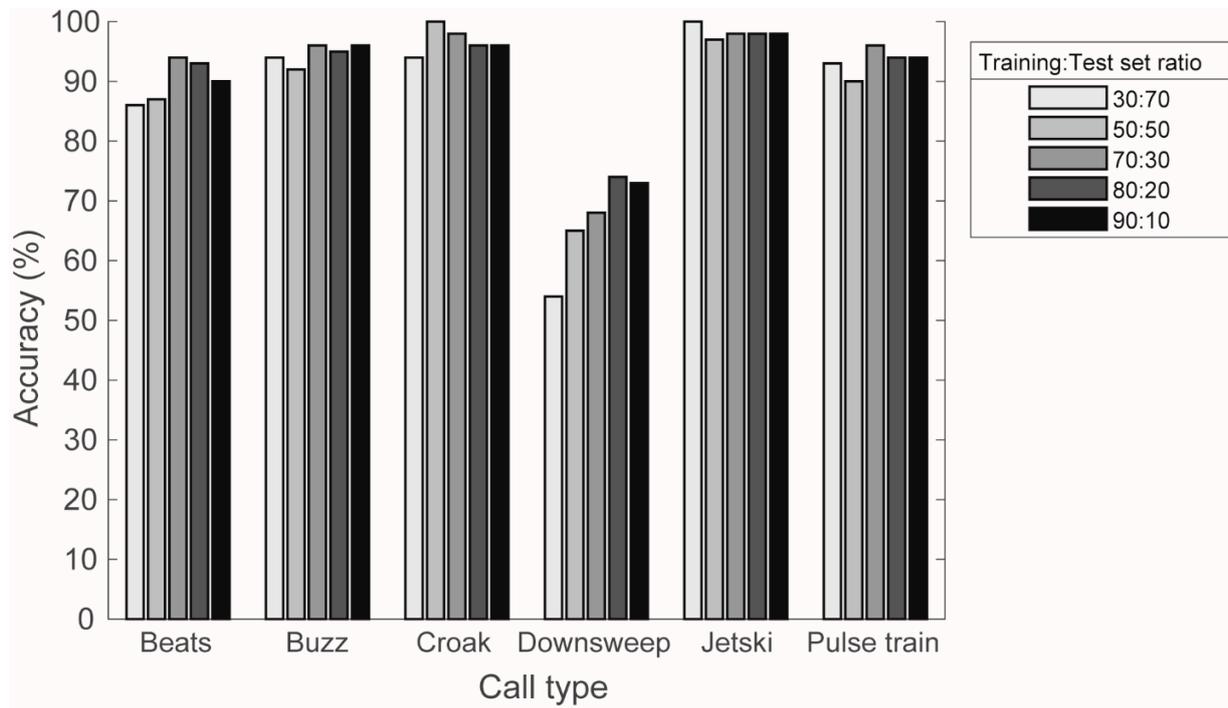
**Figure S4.** Images (created from spectrograms) of the six fish call types: (a) Beats, (b) Buzz, (c) Croak, (d) Downsweep full sweep, (e) Downsweep strongest sweep, (f) Jetski, and (g) Pulse train, with shaded bands (yellow) representing the frequency band and duration over which the signal sound pressure level was calculated for each call type. Two SNR calculations were made for each Downsweep call (indicated by d and e) because a high intensity Downsweep call had multiple downsweeps, but a less intense Downsweep call, which was more commonly observed in the data, had only one or two downsweeps.



**Figure S5.** Binomial logistic regression plots with a 0.5 threshold (black, dotted horizontal line; 0 = incorrectly classified, 1 = correctly classified) to determine the signal-to-noise ratio (SNR) threshold value (vertical blue line in each subplot), the SNR value above which a call should be correctly classified, for each fish call type: (a) Beats, (b) Buzz, (c) Croak, (d) Downsweep full sweep, (e) Downsweep strongest sweep, (f) Jetski, and (g) Pulse train. Area under the receiver operating characteristic curve (AUC) calculation included in each subplot for each fish call type.



**Figure S6.** Images (created from spectrograms) of commonly detected noises in the three datasets: (a) ship noise, (b) disk write, (c) airguns, (d) airguns, (e) unidentified noise, and (f) unidentified noise.



**Figure S7.** ResNet-50 classification accuracy for each of the six fish call types depending on the training:test set ratio (30:70, 50:50, 70:30, 80:20, 90:10).