Effects of visual noise on the performance and robustness of multivariate classifiers in EEG

P. Cheng Lim, Karl M. Kuntzelman, Lauren N. Bandel, Mei Grace Behrendt & Matthew R. Johnson

BACKGROUND

- EEG data often has varied amounts of noise, even within a single dataset. How well do classifiers generalize across these different levels of noise?
- Used deep multivariate pattern analysis (dMVPA) classifiers, implemented via DeLINEATE, a deep learning toolbox[1]
- Same convolutional neural network (CNN) model used in all dMVPA analyses

DATASET

- Used an EEG dataset of a working memory task
- N = 16 young, healthy subjects
- Recorded with 32-channel low-impedance EEG cap, sampled at 1000 Hz
- Participants shown and remembered 4 images: 2 faces and 2 scenes
- On each trial, images presented 3 times in same order with increasing levels of visual noise, accompanied by short auditory tone
- 2 further cycles presented with only visual noise and auditory tones (*rehearsal phase*; used for only a subset of analyses). During this phase, participants visualized images in step with auditory tones
- 800ms presentation, 200ms ISI
- Filtered from 0.1 Hz – 50 Hz
- 17,660 epochs total; linearly detrended after binning
- Artifact rejection: blinks/saccades removed if peak-to-peak amplitude >20μV in any EOG channel; epochs rejected if peak-to-peak amplitude >100μV in any non-EOG channels

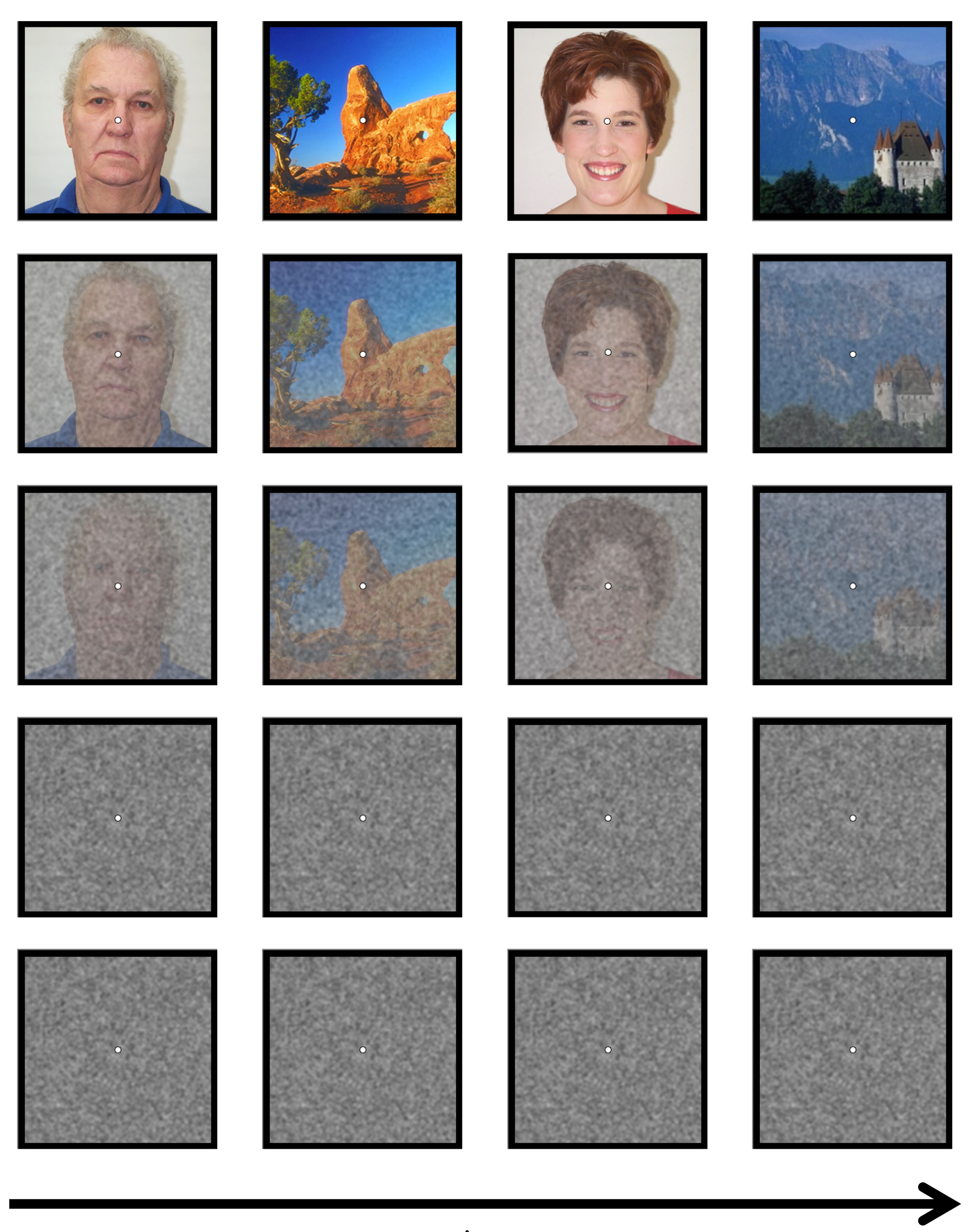
Cycle 1
Zero noise

Cycle 2
Medium noise

Cycle 3
High noise

Cycle 4
Noise only

Cycle 5
Noise only

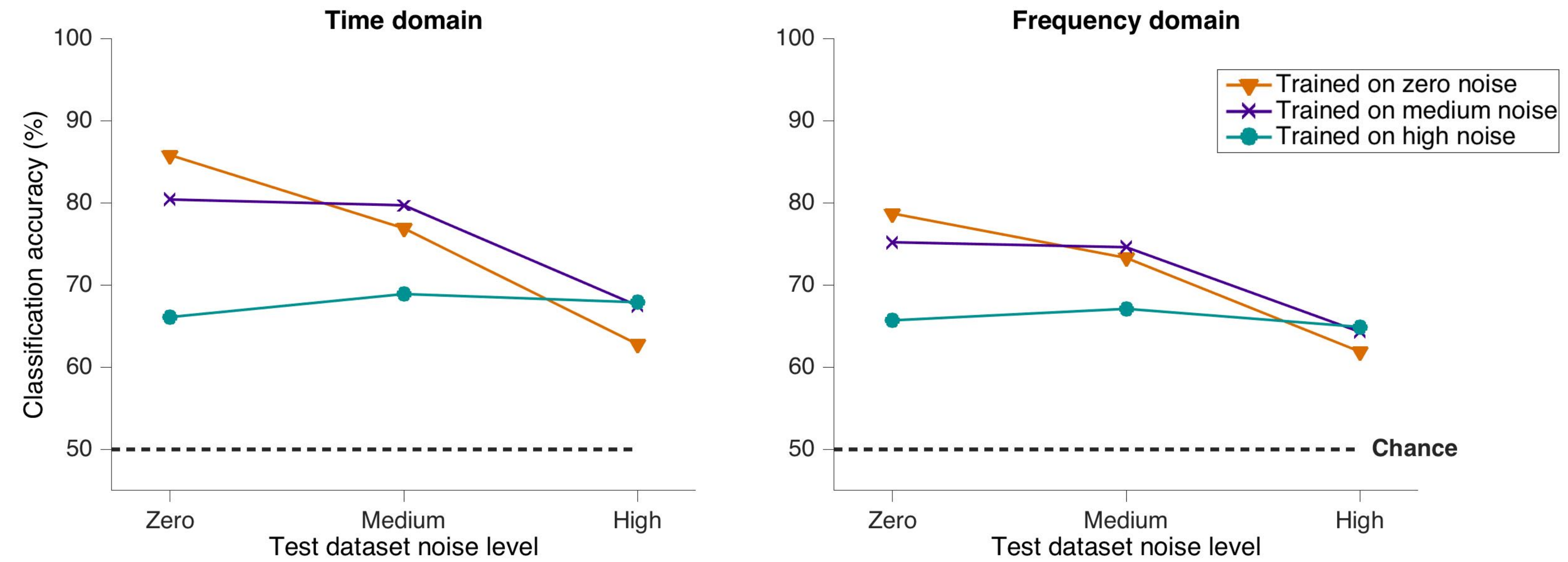


time

MULTIVARIATE PATTERN ANALYSES

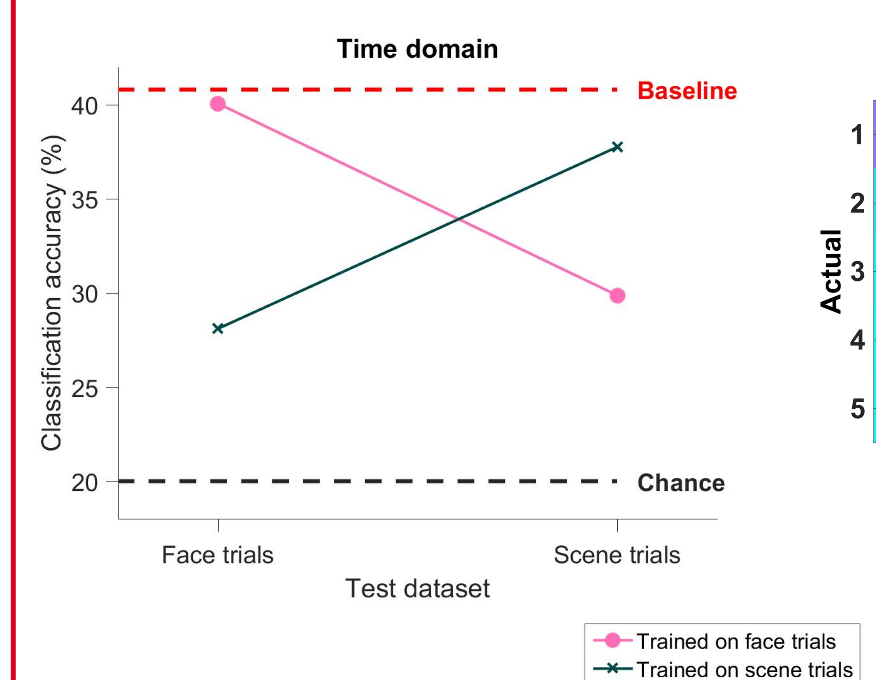
Analysis 1 (Stimulus Categories)

- Run on zero, medium, and high noise levels (Cycles 1-3)
- First, trained classifiers to decode stimulus category for each noise level (i.e. one classifier trained on zero-noise trials, one trained on medium-noise trials, one trained on high-noise trials)
- Then, applied pre-trained classifiers to trials for each noise level (i.e., zero-noise-trained classifier used to decode medium- and high-noise trials, medium-noise-trained classifier used to decode zero- and high-noise trials, etc)



Analysis 2 (Noise Levels)

- As Analysis 1, but decoded noise levels instead of stimulus categories, and run on all noise levels (Cycles 1-5)
- First, trained classifiers to decode noise levels for each category (i.e. one classifier trained on all trials, one classifier trained on face trials, one trained on scene trials)
- Then, applied pre-trained classifiers to trials for the opposite category (i.e., face-trained classifier used to decode medium-scene trials, scene-trained classifier used to decode face trials)



Train all, test all (Time domain; rows: Actual, columns: Predicted)

Actual \ Predicted	1	2	3	4	5
1	.517	.253	.115	.065	.051
2	.148	.391	.294	.099	.068
3	.064	.239	.431	.158	.108
4	.046	.103	.198	.356	.297
5	.050	.085	.156	.363	.346

Train face, test face

	1	2	3	4	5
1	.457	.250	.111	.089	.093
2	.150	.416	.205	.121	.107
3	.049	.215	.393	.192	.151
4	.039	.096	.152	.351	.362
5	.041	.101	.125	.346	.386

Train face, test scene

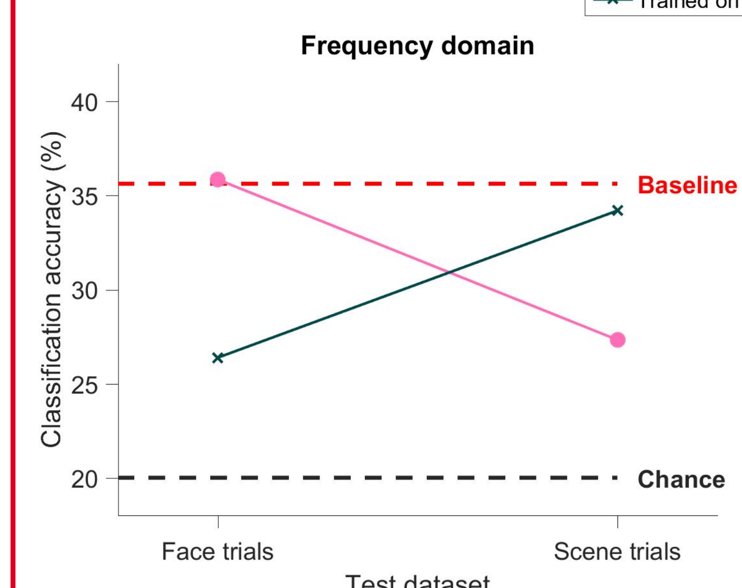
	1	2	3	4	5
1	.270	.149	.174	.198	.209
2	.098	.197	.264	.234	.207
3	.065	.158	.269	.282	.226
4	.040	.104	.148	.357	.352
5	.040	.087	.125	.344	.405

Train scene, test face

	1	2	3	4	5
1	.269	.124	.145	.213	.249
2	.107	.160	.204	.259	.270
3	.073	.146	.232	.282	.268
4	.056	.077	.144	.366	.357
5	.065	.067	.120	.357	.391

Train scene, test scene

	1	2	3	4	5
1	.544	.190	.110	.063	.092
2	.210	.287	.244	.135	.123
3	.124	.215	.293	.193	.176
4	.053	.082	.137	.365	.363
5	.055	.062	.135	.348	.400



Train all, test all (Frequency domain; rows: Actual, columns: Predicted)

Actual \ Predicted	1	2	3	4	5
1	.446	.229	.148	.086	.092
2	.177	.328	.261	.124	.110
3	.113	.234	.334	.175	.144
4	.075	.118	.183	.328	.297
5	.067	.105	.153	.330	.346

Train face, test face

	1	2	3	4	5
1	.456	.203	.140	.111	.090
2	.200	.335	.247	.122	.096
3	.077	.200	.402	.183	.139
4	.065	.107	.214	.316	.297
5	.075	.111	.188	.342	.284

Train face, test scene

	1	2	3	4	5
1	.265	.129	.205	.245	.157
2	.124	.162	.293	.258	.163
3	.097	.152	.284	.280	.187
4	.073	.113	.203	.347	.264
5	.068	.106	.188	.328	.310

Train scene, test face

	1	2	3	4	5
1	.176	.161	.222	.221	.220
2	.061	.181	.271	.253	.234
3	.051	.170	.288	.260	.232
4	.046	.104	.214	.337	.299
5	.044	.087	.178	.346	.345

Train scene, test scene

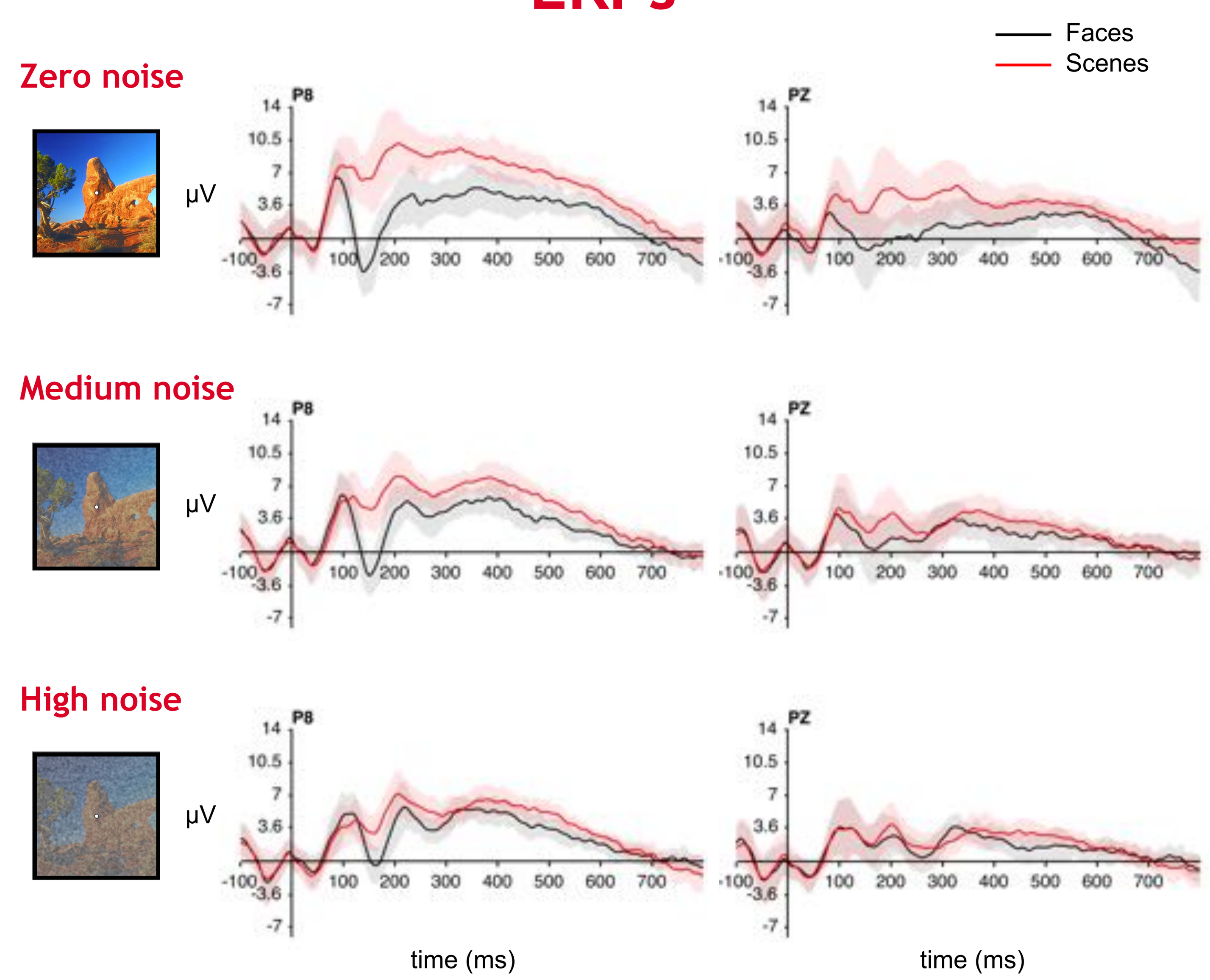
	1	2	3	4	5
1	.406	.258	.194	.081	.061
2	.176	.282	.322	.143	.076
3	.104	.260	.344	.162	.129
4	.047	.110	.202	.348	.293
5	.039	.102	.176	.353	.330

ERPs

Zero noise

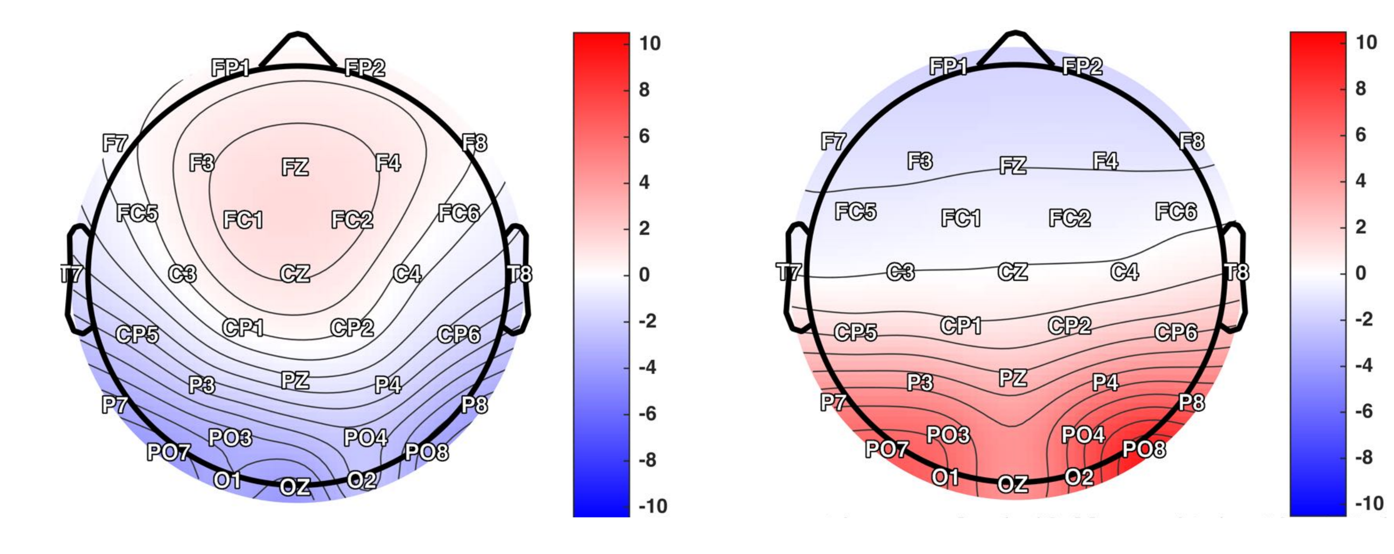
Medium noise

High noise



Faces, Cycle 1, @140ms

Scenes, Cycle 1, @140ms



CONCLUSIONS

- Classifiers trained and tested on similar data performed best (e.g., the same or similar levels of noise)
- Classifiers trained on higher noise levels were more robust when tested on data dissimilar to their original training set (e.g., zero-noise-trained classifier drops off in accuracy quickly, while medium- and high-noise trained classifiers retain similar levels of accuracy across all test datasets)
- Future work can explore this effect in different sources of noise (e.g., biological rather than visual)

REFERENCES & ACKNOWLEDGEMENTS

[1]DeLINEATE: A deep learning toolbox for neuroscientists. Kuntzelman KM, Williams JM, Samal A, Rao PK, Johnson MR. 2019. *Neuroscience 2019*, Poster 432.07. http://delineate.it

Supported by NSF/EPSCoR grant #1632849 and NIH P20 GM130461, awarded to MRJ and colleagues