2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Decoding Happiness from Neural and Video Recordings for Better Insight Into Emotional Processing in the Brain

Emil Azadian[1], Gautham Velchuru[2], Nancy Wang[3], Steven Peterson[4], Valentina Staneva[5], Bingni W. Brunton[4]

***Abstract*— Gaining a better understanding of which brain regions are responsible for emotional processing is crucial for the development of novel treatments for neuropsychiatric disorders. Current approaches rely on sparse assessments of subjects' emotional states, rarely reaching more than a hundred per patient. Additionally, data are usually obtained in a task solving scenario, possibly influencing their emotions by study design. Here, we utilize several days worth of near-continuous neural and video recordings of subjects in a naturalistic environment to predict the emotional state of happiness from neural data. We are able to obtain high-frequency and high-volume happiness labels for this task by first predicting happiness from video data in an intermediary step, achieving good results ($F_1 = .75$) and providing us with more than 6 million happiness assessments per patient, on average. We then utilize these labels for a classifier on neural data ($F_1 = .71$). Our findings provide a potential pathway for future work on emotional processing that circumvents the mentioned restrictions.**

I. Introduction

Identifying which brain regions are responsible for emotional processing is crucial for the development of novel brain-machine approaches to treating neuropsychiatric disorders [1], as for some disorders like major depression a significant portion of patients is not responding to current treatment options [2]. To date, most studies that aim to identify emotions from brain activity are conducted in a non-naturalistic environment where test subjects are asked to execute some form of task, possibly altering their emotional state. Such a setting also leads to a constraint on the amount of data that can be collected, as subjects can be engaged in a test scenario only for so long. This can in turn result in increased variability and uncertainty in the obtained findings. Examples are studies where subjects are asked to view photographs with differing emotional valences and intensities inside an fMRI scanner [3] or alongside EEG recordings [4]. In a review of 64 studies [5] attempting to decode emotion from human intracranial electrophysiology, thirty-nine studies examined neural activity during the experience of negatively or positively valenced emotional states, twenty-one while subjects perceived emotional states in others, and eleven during the elicitation of facial expressions of emotional states, with some studies appearing in more than one category. None of these 64 studies assessed emotion in a naturalistic environment. A more recent work [6] utilizes continuous ECoG recordings over several days in a naturalistic environment, very similar to the neural data used here, but relies on sparse mood data obtained by subjects filling out mood questionnaires twice per day.

In our study, we propose a novel approach to emotional assessment that avoids sparse emotional sampling and allows for the assessment of emotion in a naturalistic environment. We do so for the example emotional state of happiness. For five hospital patients, we use an average of five days of near-continuous neural and video recordings with a resolution of 1000Hz and 30Hz, respectively. Brain recordings were obtained through intracranial electrodes implanted via the invasive electrocorticography (ECoG) method. Our method consists of 1.) calculating facial features from video data on a frame-by-frame basis with the OpenFace 2.0 software [7], 2) using these features to predict the happiness of a subject for each frame, with manual happiness annotations as ground truth, and 3) using these high-volume, high-frequency happiness predictions to generate labels for a subsequent classifier that infers happiness from neural data. Happiness is therefore assessed on the basis of facial expressions [8]. This two-step approach allows us to obtain good results with standard machine learning techniques due to the amount of available data. Finally, we reconstruct which brain regions were most implicated in the classifier's decision making.

II. Data

We obtained video and ECoG recordings from five subjects at Harborview Medical Center (Seattle, WA) as part of standard patient care for intractable epilepsy. All subjects provided written, informed consent to use their data for research purposes. The experimental procedures involving human subjects described in this paper were approved by the UW Office of Sponsored Research. ECoG electrodes are primarily placed on the cortical surface. In two subjects, few electrodes were also implanted beneath the surface. The subjects then spend several days in the hospital under additional video monitoring, yielding the near-continuous video and ECoG recordings. For each subject, a multitude ($mean = 98.33$, $std = 10.23$) of intracranial electrodes provide measurements of neural activity sampled at 1000Hz. Coverage of the brain is limited (see figure 1), and electrode positions are inconsistent across subjects. This means that for some subjects, regions implicated in emotional processing

*This research was supported by funding from the Defense Advanced Research Projects Agency (FA8750-18-2-0259), the National Science Foundation (1630178 and EEC-1028725), the Sloan Foundation, and the Washington Research Foundation.

[1]Technical University of Berlin, Dept. of Computer Science, Germany emil.azadian@gmail.com
[2]Microsoft, Seattle, WA
[3]Amazon, Berlin, Germany
[4]University of Washington, Dept. of Biology, Seattle, WA
[5]University of Washington, Paul G. Allen School of Computer Science and Engineering

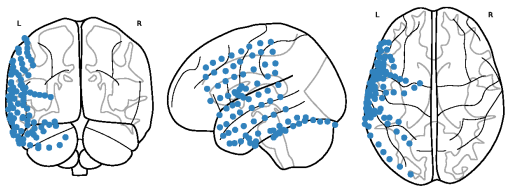

Fig. 1. Neural recording electrode coverage for subject S1. Coverage of the brain is limited, with the prefrontal cortex and most of the deeper layers of the brain not covered.

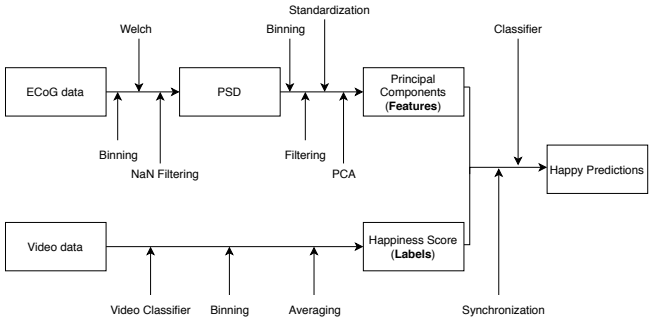

Fig. 2. ECoG classifier pipeline. Happiness labels are generated with the video pipeline, which are classified using features from processed ECoG recordings.

are not directly recorded. The video data was sampled at 30FPS throughout several hundred videos per subject at a length of about 120 seconds each. Videos were censored for privacy-sensitive content.

III. Method

We first trained a *video classifier* to predict happiness from the videos on a frame-by-frame basis using manual annotations that were available on a subset of frames across all subjects. Then, we used these predictions to generate labels for a subsequent *ECoG classifier* inferring happiness from neural data. The complete workflow is depicted in figure 2.

Video Classifier To extract the happiness state from video, we made use of the OpenFace 2.0 open-source toolkit to calculate 709 facial features on a frame-by-frame basis for all videos of all subjects. These features included eye gaze [9], facial landmarks [10], head pose, and facial action units [11]. Each feature is calculated using deep neural networks that have achieved state-of-the-art performance [7]. These calculated features served as the features for a single Random Forest classifier that we trained across the data of all subjects. As labels, we used manual binary happiness annotations on a subset of 299 videos pooled across all subjects. Within a given video, frames and thus facial features were fairly similar to each other. To avoid these similar data points ending up in both train and test set, possibly leading to an overestimation of the classifier performance, we applied grouping during the train-test split, ensuring that all data points belonging to a single video ended up in either the train or the test split. We also tried a linear SVC in combination with a Nyström kernel approximation, but achieved inferior results. We applied the trained classifier to all unlabeled video frames and thus obtained on average 6 million happiness expression predictions to be used in the follow-up stages.

ECoG Classifier We preprocessed the ECoG data by bandpass filtering between 1–200 Hz, removing high amplitude artifacts, downsampling to 500 Hz, and discarding individual electrodes whose standard deviation or kurtosis laid several magnitudes (5 and 10, respectively) outside the median range of all electrodes' respective values. We synchronized in time the neural recordings and the happiness predictions coming from the video classifier, taking into account filtered or missing data points on either side. To account for the fact that emotions can last up to several minutes [12], we applied a window of 100 seconds to the data. Within each window, we calculated the fraction of frames predicted as happy by the video classifier and converted it to a binary happy label via a cutoff at a ratio of .3. This subsuming approach helps alleviate some of the inaccuracy in the video classifier predictions.

For the neural data, we calculated the power spectral density using the Welch method [13] for each electrode in each window. Since there were on average 91 retained electrodes per subject ($std = 13$) and the Welch method returned power spectrum densities up to 250Hz with a resolution of 2Hz, this resulted in $91 \times 125 \approx 1.1 \cdot 10^4$ feature points. In order to reduce that number, several measures were taken. First, power densities for frequencies above 150Hz were cut off, as based on current research we assume the frequencies relevant to be below that threshold [14]. Second, the remaining densities were binned according to their frequencies on a logarithmic scale.

For each frequency bin, we summed up the power spectral densities, leaving us with eight features per electrode instead of 75 (a reduction of $\sim 90\%$). To remove segments with noisy data caused by an electric hum usually noticeable at around 120Hz, we took the high-frequency bin containing the 120 Hz frequency and calculated the standard deviation and median across time for each electrode. We then discarded all sample points where for at least one electrode the high-frequency bin value deviated further than four standard deviations from the median. Additionally, we standardized each feature across time. As the last step towards lowering the number of features, we applied principal component analysis (PCA) to the data, retaining the components that cumulatively explained $> 90\%$ of the total variance.

Electrodes are prone to temporary failures and artifact occurrences. Thus, different days might yield a different set of good channels. To ensure congruence over multiple days, we took the intersection of channels that were non-faulty throughout the entire time span. We split the data into train and test sets, paying special attention to not let any information leak from the test set into the training set and vice versa. For example, we ensured that the standardization

of the data, the artifact detection, and the principal directions are not in part determined by the test data. These steps are therefore taken after the shuffling of the data. Similar to the first pipeline, we made use of a Random Forest classifier to infer happiness from the data. Here, we trained a separate classifier for each subject, as inter-comparability is limited due to variability in electrode coverage. We also tested support vector classifiers with both polynomial and RBF kernels but obtained significantly worse results.

For both the video and the ECoG classifiers, we applied a random search over the number of estimators, the maximum depth per tree, and the maximum number of features considered at each split using 10-fold cross-validation. Performance for a given hyperparameter selection was measured by calculating the average precision score for each validation fold and then taking the median across the ten average precision scores. We retrained the classifier on the whole train set for the best combination of hyperparameters found during this procedure. For the video classifier, grouping was again applied when splitting the train data into folds.

IV. Results

The main goal of the video classifier was to provide us with high-volume happiness predictions to be used for the labels for the ECoG classifier. On average, we achieved an increase of 1676% from manual annotations to annotations derived through the video classifier. As for the reliability of our approach, we were able to obtain overall good results with the video classifier, achieving an F1 score of .89 (*precision* $= .87$, *recall* $= .91$, *accuracy* $= .92$) on the train set and an F1 score of .75 (*precision* $= .90$, *recall* $= .65$, *accuracy* $= .73$) on the test set, as depicted in table 3. While results are consistently good for almost all subjects, the classifier performs poorly on subject S1. Possible reasons are given in section V. Based on these results, we can assume the predictions of the classifier to be reliable enough for continued usage, with the exception of subject S1.

For the ECoG classifier, we achieved an average F1 score of .77 (*std* $= .11$) with an average accuracy of .78 (*std* $= .05$) on the train sets of the subjects. On the test sets, we obtained an average F1 score of *F1* $= .71$ (*std* $= .17$) with an accuracy of .73 (*std* $= .09$). The classifier trained on S1 performs the worst, which based on the poor video classifier performance for that subject was expected.

To obtain insight into the brain regions most involved in ECoG classification, we multiply the feature importances of the Random Forest classifier with the weightings used durixng principal component analysis, giving us an importance value for each recording site. We determine the nearest Brodmann areas (BA) for each subject's five most important electrodes (Table 5). This is appropriate since all important electrodes are located on the cortical surface. Since the metrics for subject S1 render their predictions unreliable, we have excluded them from the table to ensure better insight into commonalities between all other subjects. The Brodmann Areas for subject S1 were 1, 21 (2x), 22, 45.

Overall Results

Set	Support	Pr	Re	F1	Acc
Train	19870	0.87	0.91	0.89	0.92
Test	10313	0.90	0.65	0.75	0.73

Train Results

Subject	Happy Ratio	Pr	Re	F1	Acc
S1	448 / 4690	0.75	0.91	0.82	0.97
S2	0 / 486				0.97
S3	5424 / 2454	0.92	0.94	0.93	0.90
S4	1170 / 4270	0.71	0.91	0.79	0.90
S5	320 / 608	0.94	0.49	0.65	0.81

Test Results

Subject	Happy Ratio	Pr	Re	F1	Acc
S1	15 / 909	0.03	0.13	0.04	0.90
S2	1292 / 262	0.92	0.59	0.72	0.62
S3	3716 / 1286	0.92	0.74	0.82	0.76
S4	1039 / 746	0.91	0.45	0.60	0.66
S5	381 / 667	0.78	0.51	0.62	0.77

Fig. 3. **Top**: Overall results of the video classifier. Support indicates the total number of labels in each set. **Middle**: Train Results broken down by subject. Happy Ratio signifies the count of happy/not happy labels. We only trained a single classifier across all subjects, explaining the imbalances in the happy ratio on a subject level. Remarkably, the videos that S2 contributed to the train set did not include any happy labels, but the classifier was still able to perform well on that subject's test happy labels, indicating that the classifier generalizes well across patients. **Bottom**: Classifier performance on the test set. Apart from S1, results are consistently good.

V. Conclusion and Discussion

Apart from S1, the video classifier generalizes well across all subjects, even with group splitting that rendered train and test sets to be more distinct. Remarkably, the classifier did not see a single happy instance for subject S2 during training, but still performed well on their portion of labels in the test set. Using qualitative insight into the video data, we could identify three possible reasons why performance on subject S1 is notably worse than on the rest. The subject was the only one to wear glasses, to have a beard, and to have a dark skin tone, which led to many incorrect facial feature detections. Thus, we may be seeing a result of disproportionate representation in the datasets OpenFace was trained on. With the constant expansion of existing training data for face detection and recognition tasks, we expect those problems overcome in future versions of the software.

The ECoG classifier performs well overall. The dropoff in performance for subject S1 was expected, as the labels are based on the output of the video classifier and therefore likely inaccurate for this subject. Some of the inter-subject variance in involved brain regions may be attributed to the deviations of recording sites across subjects, making it more difficult to pinpoint involvement of the same region for all subjects. We observe regions known to be explicitly involved with emotions, such as the prefrontal cortex (BA10) [15][16] and the temporal pole (BA38) [17]. Others are highly connected to areas involved in emotional processing (BA20) [18] or are likely involved due to correlative effects, such as the motor

Train Results

Subject	Happy Ratio	Pr	Re	F1	Acc
S1	479 / 1351	0.55	0.65	0.60	0.77
S2	844 / 582	0.81	0.89	0.85	0.81
S3	1355 / 406	0.85	0.97	0.91	0.85
S4	643 / 1016	0.57	0.85	0.68	0.69
S5	968 / 745	0.74	0.88	0.80	0.76
Average		0.70	0.85	0.77	0.78

Test Results

Subject	Happy Ratio	Pr	Re	F1	Acc
S1	110 / 348	0.43	0.53	0.48	0.72
S2	232 / 128	0.81	0.86	0.83	0.78
S3	337 / 103	0.86	0.98	0.92	0.86
S4	145 / 264	0.44	0.68	0.53	0.58
S5	235 / 197	0.71	0.85	0.77	0.73
Average		0.65	0.78	0.71	0.73

Fig. 4. Results of the ECoG classifier. Happy Ratio stands for the amount of happy/not happy labels present. As expected, the classifier trained on subject S1 performs the worst.

Subject	Brodmann Area									
	3	6	7	10	19	20	21	36	38	40
S2	2	2							1	
S3			1		2					2
S4				1		1			3	
S5						2	1	1	1	

Fig. 5. Five electrodes per subject most important for the happiness prediction across Brodmann areas, excluding subject S1.

areas (BA7, BA6) or visual and language processing areas (BA19, BA40).

The general trend of our results provides evidence that our approach to assessing emotions from video data in a naturalistic environment to then work with high-frequency, high-volume information about the subject's emotional state can be a viable alternative method for further research in the field, avoiding sparse emotion sampling or trials in non-naturalistic environments. This is further supported by the good performance metrics of the video classifier itself, making its predictions reliable enough for further usage. Notably, we were able to achieve our results using only standard machine learning tools, which facilitates interpretability and quick adoption by others but also leaves room for improvement in future works. Further, increasing the number of subjects will increase the coverage of the electrodes and will lead to more consistency in the observed brain regions to allow for quantitative analysis of emotional processing across patients. Investigating the kind of activities that trigger certain emotions and their corresponding brain responses would also be a possibility. Obtaining emotional data from video recordings could likewise be used to augment other methods of acquiring information on emotional states like surveys in order to achieve a comprehensive understanding of the relationship between subjects' emotions and their brain activity.

REFERENCES

[1] Maryam M. Shanechi, "Brain-machine interfaces from motor to mood," *Nature neuroscience*, vol. 22, no. 10, pp. 1554–1564, 2019.

[2] "Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a star*d report," *The American journal of psychiatry*, vol. 163, no. 11, pp. 1905–1917, 2006.

[3] Simone Grimm, Johannes Beck, Daniel Schuepbach, Daniel Hell, Peter Boesiger, Felix Bermpohl, Ludwig Niehaus, Heinz Boeker, and Georg Northoff, "Imbalance between left and right dorsolateral prefrontal cortex in major depression is linked to negative emotional judgment: an fmri study in severe major depressive disorder," *Biological psychiatry*, vol. 63, no. 4, pp. 369–376, 2008.

[4] Daren C. Jackson, Corrina J. Mueller, Isa Dolski, Kim M. Dalton, Jack B. Nitschke, Heather L. Urry, Melissa A. Rosenkranz, Carol D. Ryff, Burton H. Singer, and Richard J. Davidson, "Now you feel it, now you don't: frontal brain electrical asymmetry and individual differences in emotion regulation," *Psychological science*, vol. 14, no. 6, pp. 612–617, 2003.

[5] Sean A. Guillory and Krzysztof A. Bujarski, "Exploring emotions using invasive methods: review of 60 years of human intracranial electrophysiology," *Social cognitive and affective neuroscience*, vol. 9, no. 12, pp. 1880–1889, 2014.

[6] Omid G. Sani, Yuxiao Yang, Morgan B. Lee, Heather E. Dawes, Edward F. Chang, and Maryam M. Shanechi, "Mood variations decoded from multi-site intracranial human brain activity," *Nature biotechnology*, vol. 36, no. 10, pp. 954–961, 2018.

[7] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 15/05/2018 - 19/05/2018, pp. 59–66, IEEE.

[8] Paul Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[9] E. Wood, T. Baltruaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling, "Rendering of eyes for eye-shape registration and gaze estimation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3756–3764.

[10] A. Zadeh, T. Baltrusaitis, and Louis-Philippe Morency, "Convolutional experts constrained local model for facial landmark detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2051–2059, 2017.

[11] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2015, vol. 06, pp. 1–6.

[12] Philippe Verduyn, Pauline Delaveau, Jean-Yves Rotgé, Philippe Fossati, and Iven van Mechelen, "Determinants of emotion duration and underlying psychological and neural mechanisms," *Emotion Review*, vol. 7, no. 4, pp. 330–335, 2015.

[13] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[14] Kai Yang, Li Tong, Jun Shu, Ning Zhuang, Bin Yan, and Ying Zeng, "High gamma band eeg closely related to emotion: Evidence from functional network," *Frontiers in human neuroscience*, vol. 14, pp. 89, 2020.

[15] Turhan Canli, Heidi Sivers, Moriah E. Thomason, Susan Whitfield-Gabrieli, John D. E. Gabrieli, and Ian H. Gotlib, "Brain activation to emotional words in depressed vs healthy subjects," *Neuroreport*, vol. 15, no. 17, pp. 2585–2588, 2004.

[16] Jeffrey B. Henriques and Richard J. Davidson, "Decreased responsiveness to reward in depression," *Cognition & Emotion*, vol. 14, no. 5, pp. 711–724, 2000.

[17] Ingrid R. Olson, Alan Plotzker, and Youssef Ezzyat, "The enigmatic temporal pole: a review of findings on social and emotional processing," *Brain : a journal of neurology*, vol. 130, no. Pt 7, pp. 1718–1731, 2007.

[18] Khader M. Hasan, Amal Iftikhar, Arash Kamali, Larry A. Kramer, Manzar Ashtari, Paul T. Cirino, Andrew C. Papanicolaou, Jack M. Fletcher, and Linda Ewing-Cobbs, "Development and aging of the healthy human brain uncinate fasciculus across the lifespan using diffusion tensor tractography," *Brain research*, vol. 1276, pp. 67–76, 2009.