Supplementary Information for

Decoding naturalistic affective behavior from spectro-spatial features in multiday human iEEG

Maryam Bijanzadeh[1], Ankit N. Khambhati[1], Maansi Desai[2], Deanna L. Wallace[3], Alia Shafi, Heather E. Dawes[1], Virginia E. Sturm[4], and Edward F. Chang[1]*

[1] Department of Neurological Surgery, University of California, San Francisco, USA
[2] Department of Communication Sciences and Disorders, Moody College of Communication, University of Texas at Austin, Austin, TX, USA
[3] Departments of Mechanical Engineering, Psychology and Neurology, University of Texas at Austin, Austin, TX, USA
[4] Department of Neurology, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, USA

* Corresponding author name: Edward F. Chang
Email: Edward.Chang@ucsf.edu

Supplementary Information

Supplementary Results

Behavioral Annotations

Participants exhibited a wide range of positive (range: 42-164), negative (range: 34-133), and neutral (range: 277-499, Table S4, for example see Subject 3) behaviors that aligned with clean neural signals that were free from epileptic activity (Extended Data Figure 2). Overall, our dataset included more instances of positive (mean ± sem = 112 ± 17, n = 10 participants) than negative (mean ± sem = 61 ± 19, n = 5 subjects) affective behavior. While smiling and laughing occurred frequently, pain-discomfort and negative verbalizations were less common (Extended Data Figure 1).

Clustering Analyses

We conducted hierarchical clustering in each participant (See “Clustering” in Methods and Supplementary Figure 8), an objective way to map the features that characterized the positive and negative affective behaviors from each decoder. While we observed common changes in spectral power across mesolimbic regions during positive and negative affective behaviors, clustering the features allowed us to control for possible collinearity between features (e.g., increased high gamma activity in multiple brain structures during positive affective behaviors might have driven our previous results, Figure. 3-A). This clustering analysis identified two clusters—a “gamma” cluster and a “low-frequency” cluster (Extended Data Figure 5)—from the positive and negative decoders that separated affective from neutral behaviors based on spectral bands rather than regions (Supplementary Figure 8). These results suggested that, at an individual level, simultaneous increases in gamma activity and decreases in low-frequency activity across the mesolimbic network characterized both positive and negative affective behaviors when compared to neutral behaviors. In general, affective behaviors were separated from neutral behaviors along a spectral rather than spatial distribution (in which specific regions, not frequency bands, had predominant roles in certain behaviors). There were some exceptions to this pattern, however, in individual participants (Participant 5, Supplementary Figure 9).

We next investigated whether the spectral patterns that we observed for affective behaviors at the individual level (Supplementary Figures 9 & 10) were found across the sample, regardless of each participant’s spatial coverage. Proceeding to extraction of the difference scores using the feature medians in each affective class and populating these scores across participants (See Methods “Feature Normalization for group level analyses” & “Clustering” & Supplementary Figure 11), we found that consistent with the results from the clustering analyses conducted at the individual level, positive affective behaviors were characterized by higher median values in the gamma cluster (Figure 3-E) and lower median values in the low frequency cluster (median of gamma cluster = 1.12 vs. low frequency cluster = -1.43, ranksum test, $p < 0.0001$) than neutral behaviors. A similar pattern was found for negative affective behaviors when they were compared to neutral behaviors (Figure 3-F, median of gamma cluster = 0.68 vs. low frequency cluster = -0.92, ranksum test, $p < 0.0001$). In sum, simultaneous increases in high frequency activity and decreases in low frequency activity within the mesolimbic network may be a common network signature of both positive and negative affective behaviors.

Feature importance from binary decoders

To determine which of the selected features played a dominant role in the decoder models of each participant, we pooled the feature importance, which was generated by the RF models, for the gamma and low-frequency clusters from both the positive and negative decoders (Extended Data Figure 5-C & D). Although at the population level, the selected spectro-spatial features in all frequency bands (except alpha) were significantly different between positive affective behaviors and neutral behaviors (Figure 3-C), the gamma cluster (n=149, median = 0.36) was significantly more important than low-frequency cluster (n = 124, median = 0.29) for the positive decoder models (i.e., larger feature importance value, ranksum test, $p = 0.0017$, Extended Data Figure 5-C left). In distinguishing negative affective behaviors from the neutral behaviors, high gamma band, alpha and beta bands activity were significantly different between the two behaviors (Figure 3-D). In line with this observation, both the gamma (n = 62, median = 0.43) and low-frequency (n = 45, median = 0.38) clusters were equally important for the negative decoder’s successful decoding (ranksum test, $p = 0.16$, Extended Data Figure 5-D left). These findings suggest negative affective behaviors may be more heterogenous than positive affective behaviors and may rely on both types of spectral signatures to distinguish them from moments lacking affect.

Stability of features from binary decoders

To assess the robustness of the important features being selected with a likelihood better than chance, we counted the number of times each feature was selected across 100 bootstrapped runs of each RF model for each participant. We refer to the proportion of runs in which the features were selected in the positive and negative decoders as the "feature stability" (Extended Data Figure 5-C&D, right panels). We found that features within the gamma cluster of the positive decoder were more stable than features within the low-frequency cluster (87% of runs vs 80% of runs, p= 0.0015, ranksum test). We also observed greater stability of features within the gamma cluster compared to the low-frequency cluster for the negative decoder (median value of 79% of runs vs 68% of runs, p = 0.003, ranksum test). Also, as expected, stability and feature importance were significantly correlated across all features (r = 0.65, p <0.0001, n = 273; and r= 0.82, p <0.0001, n = 107, for positive and negative vs. affectless decoders, respectively, spearman correlation). This confirms that more important features were also more reliable features for decoding.

Differences between the positive and negative decoders

To assess whether the differences between the features that contributed to the positive and negative decoders were due to the feature selection method (i.e., the kneedle algorithm), we also compared the top 10 features from each decoder type regardless of the objective threshold (Supplementary Tables 5 and 6); the negative decoders were more likely to consist of features from the low-frequency cluster than the positive decoders (i.e., Subject1: 3/10 vs. 1/10, Subject 2: 6/10 vs. 4/10, and Subject 3: 7/10 vs. 0/10). Meanwhile, Subject 6 demonstrated a greater likelihood for important low-frequency features for the positive decoder (3/10) than the negative decoder (2/10). This finding was consistent with the objective feature comparison method used in our primary analyses (Extended Data Figure 5-C&D), which showed no significant difference in feature importance of the selected features between low-frequency and gamma clusters for the negative decoder but greater feature important for the gamma cluster than the low-frequency cluster for the positive decoder. For example, the selected and clustered features for Subject 1 (Supplementary Figure 6-A & B) show more low-frequency features selected for the negative decoders than the positive decoders. Thus, these findings are robust to the methods that were used.

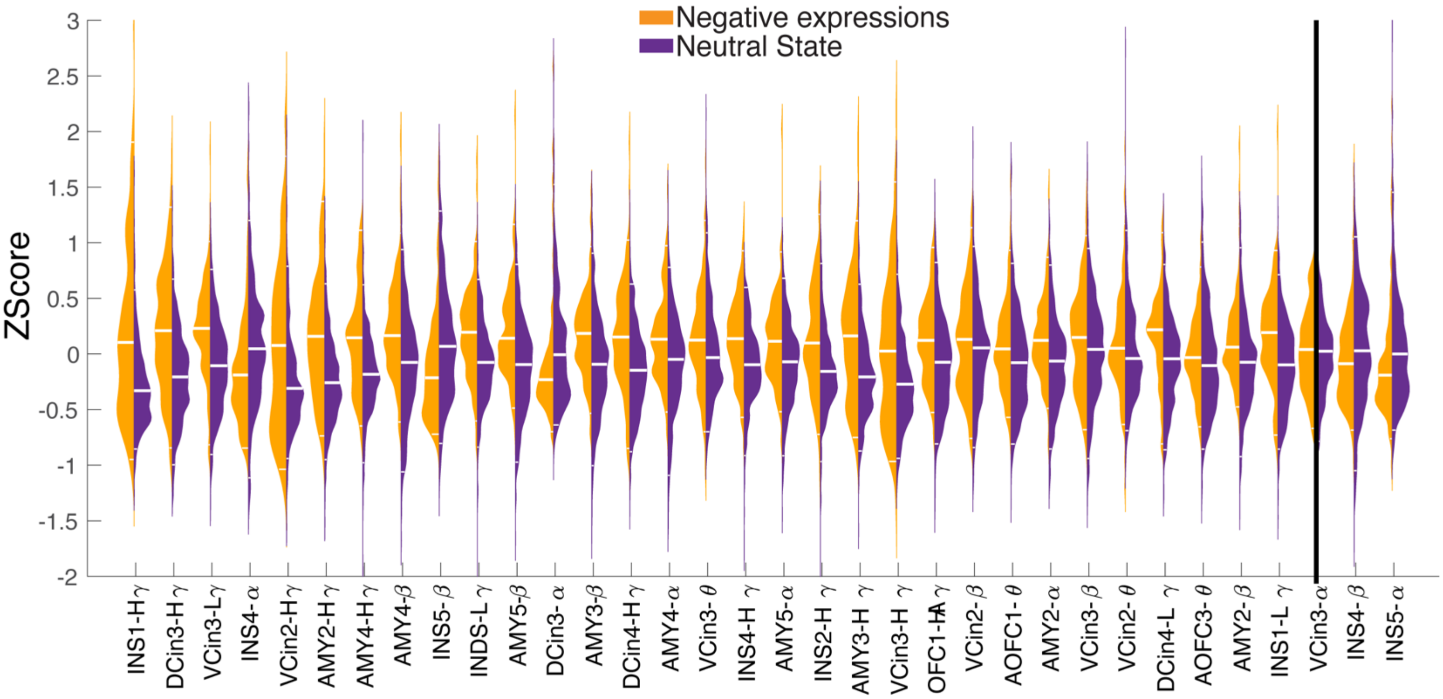
Additional tests for the feature selection

We applied other techniques to certify the robustness of the selected features. We extracted the t-statistics of the feature distributions between the behavioral classes for two participants and sorted the features (Supplementary Figure. 4). The results showed that the top 10 features with *t*-scores larger than the critical *t*-score were also selected features by the RF models. Moreover, we trained personalized linear support vector machine (SVM) models and sorted the features based on the absolute value of the feature weights (See Methods, "SVM Model Classification: Linear SVM", Supplementary Figures 5 & 6) for both the positive and negative decoders. The results uncovered similarities between the selected features of the linear SVM and RF models, but the RF models performed better in 7/10 and 4/5 participants. We also trained nonlinear SVM classifiers (with rbf kernel) using the selected feature sets that were derived from the RF models (See Methods, "SVM Model Classification: Non-linear SVM"). The resulting non-linear SVM models showed a similar performance as the RF models (Supplementary Figure 7), which further confirmed the robustness of the feature selection method.

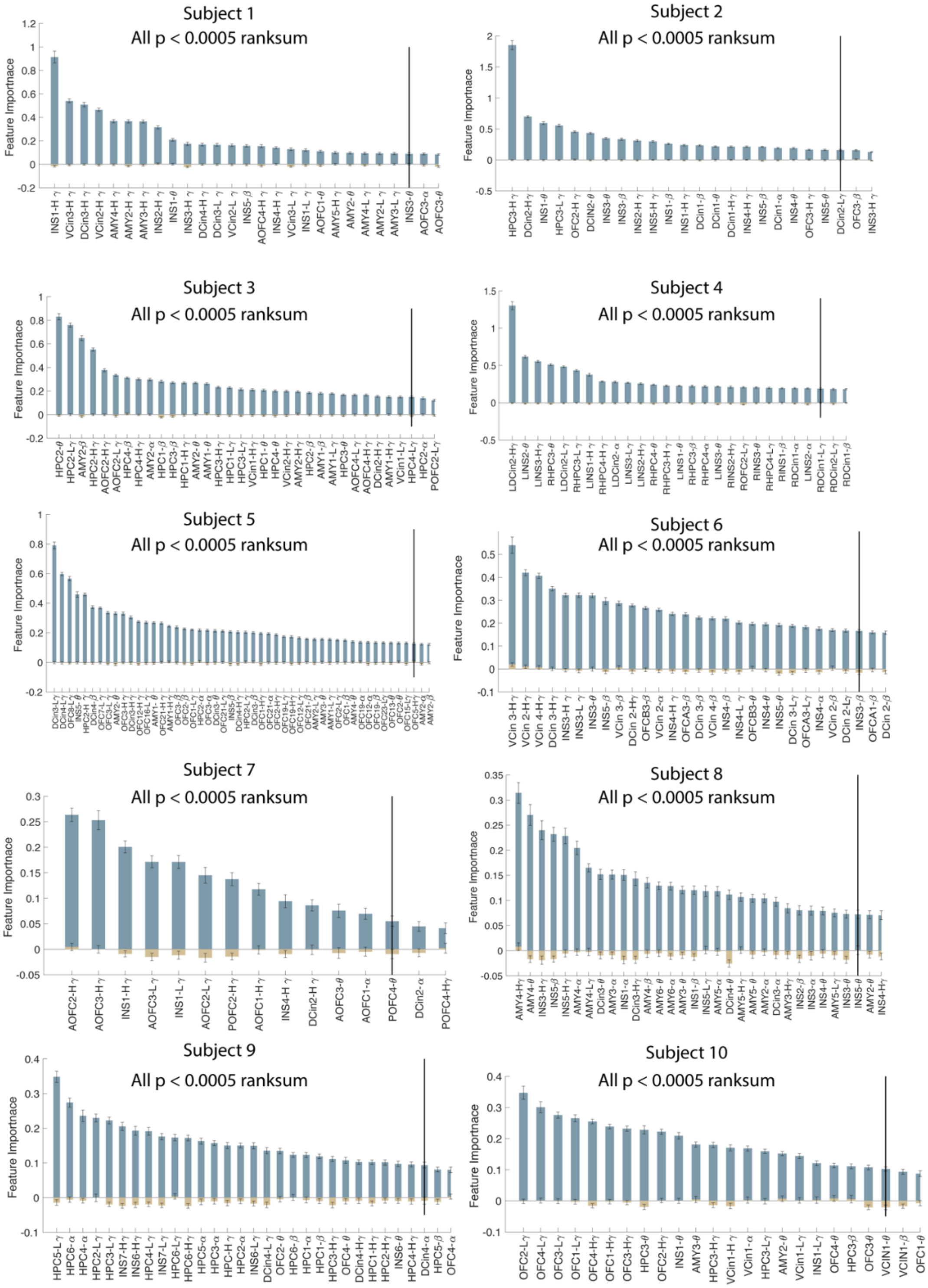
Regional variability and feature importance from multiclass decoders

We examined whether regional variability existed across the channels in the participants in whom the multiclass decoder was applied. These analyses enabled us to investigate whether certain regions made more important contributions to positive or negative affective behaviors than others. In each participant, we visualized the median distribution of spectral power during each type of affective behavior (Supplementary Figure 14). These graphs suggested that high gamma power discriminated positive and negative affective behaviors from neutral behavior in 2/3 participants (Subjects 1 and 6) and generally showed a similar stratification as in Figure 6B (positive, then negative, then neutral). In 1/3 participants (Subject 2), however, there were clear divisions but in a different order (negative, then positive, then neutral), The graphs also indicated that, although a given subregion within the insula could exhibit stronger high-gamma activation during negative than positive affective behaviors (INS1 and INS3 in Subject 2 and INS5 in Subject 6), other subregions may be more tuned to positive affective behaviors (INS1 in Subject 1 and INS3 in Subject 6). Although high-gamma band activity had a significantly larger feature importance than the theta, alpha, and beta bands together (Supplementary Figure 15), we did not observe a significant difference between these spectral bands across different regions. Thus, different electrodes within same brain region may have made different contributions to the decoding performance and may have played distinct functional role in the neural representation of affective behaviors.

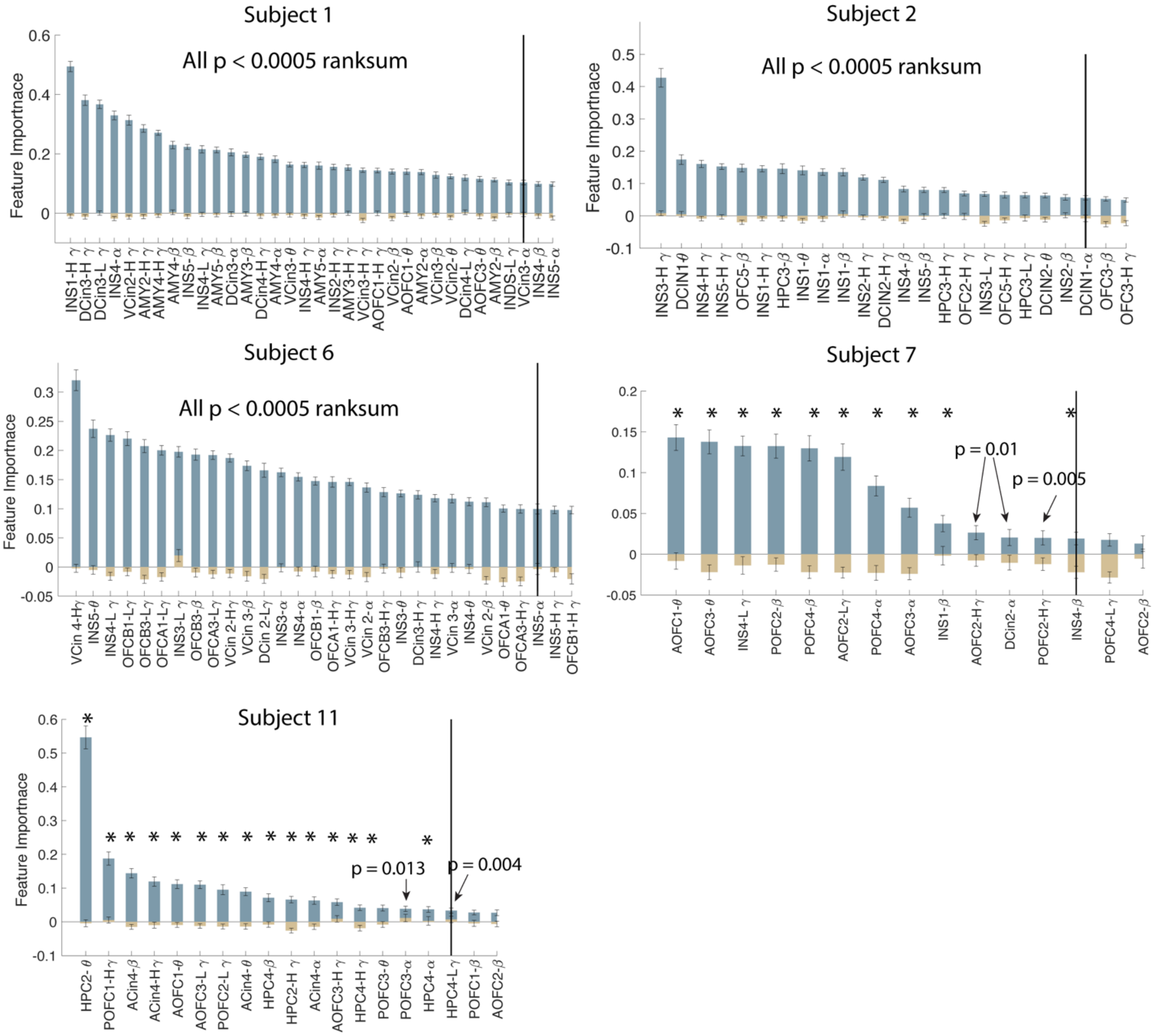
Supplementary Figure 1. *Sample distribution of selected features for example participant (Subject 1) that contributed to negative decoders.*



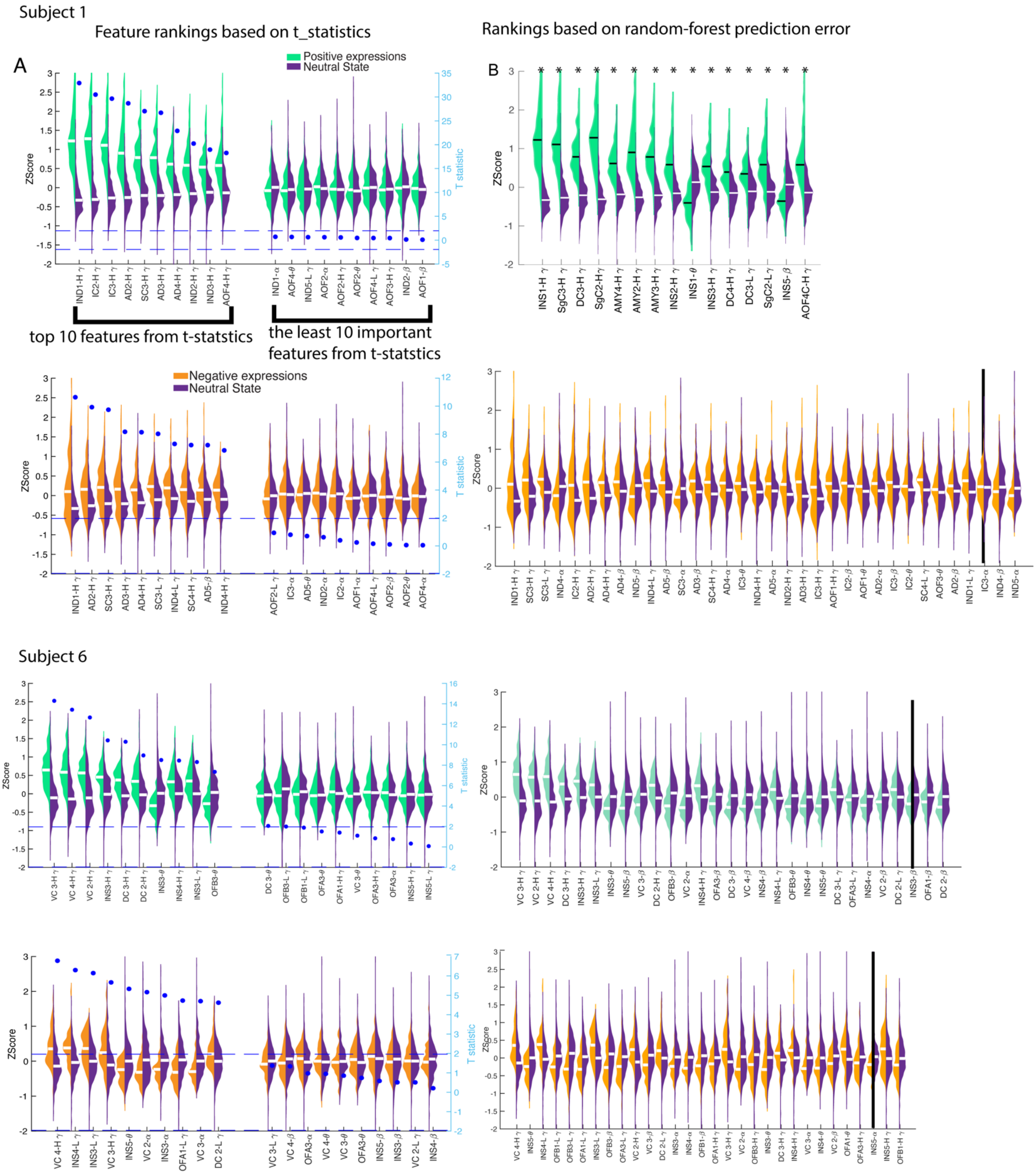
Supplementary Figure 2. *Selected features from the RF models that were trained on positive affective behaviors and neutral behaviors (gray shading).* The feature importance of the selected features from the shuffled models are shown in yellow. Bar charts represent mean values +/- SEM across 100 datasets of the selected features. INS: insula, VCin = Ventral cingulate, DCin = dorsal cingulate, AMY: amygdala, OFC = Orbitofrontal cortex. All statistics are reported by two-sided pairwise ranksum test between 100 runs of RF models and the shuffled models for each feature. All statistics are reported by two-sided pairwise ranksum test between n= 100 runs of RF models and the shuffled models for each feature. All p values are less than 0.0005 except the two those that are noted in the figure. the comparison is between feature importance of main models and permuted models in which the labels are shuffled, thus the significance level = 0.0005 (refer to the Methods section “Statistical Analyses”).



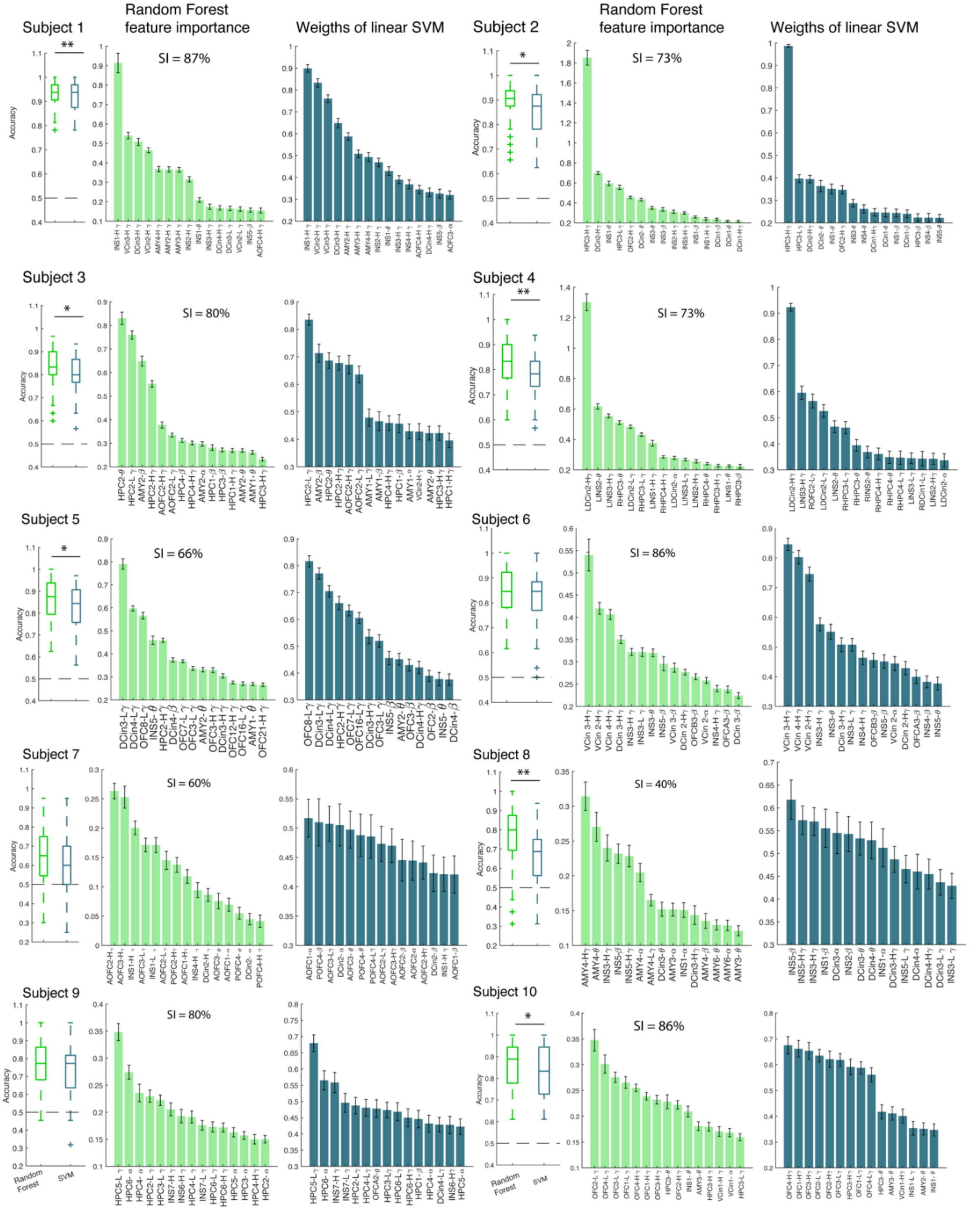
Supplementary Figure 3. *Selected features from the RF models that were trained on negative affective behaviors and neutral behaviors (gray shading).* The feature importance of the selected features from the shuffled models are shown in yellow. Bar charts represent mean values +/- SEM across 100 datasets of the selected features. INS: insula, VCin = Ventral cingulate, DCin = dorsal cingulate, AMY: amygdala, OFC = Orbitofrontal cortex. All statistics are reported by two-sided pairwise ranksum test between n= 100 runs of RF models and the shuffled models for each feature. All p values are less than 0.0005 except the two those that are noted in the figure. the comparison is between feature importance of main models and permuted models in which the labels are shuffled, thus the significance level = 0.0005. (refer to the Methods section "Statistical Analyses").



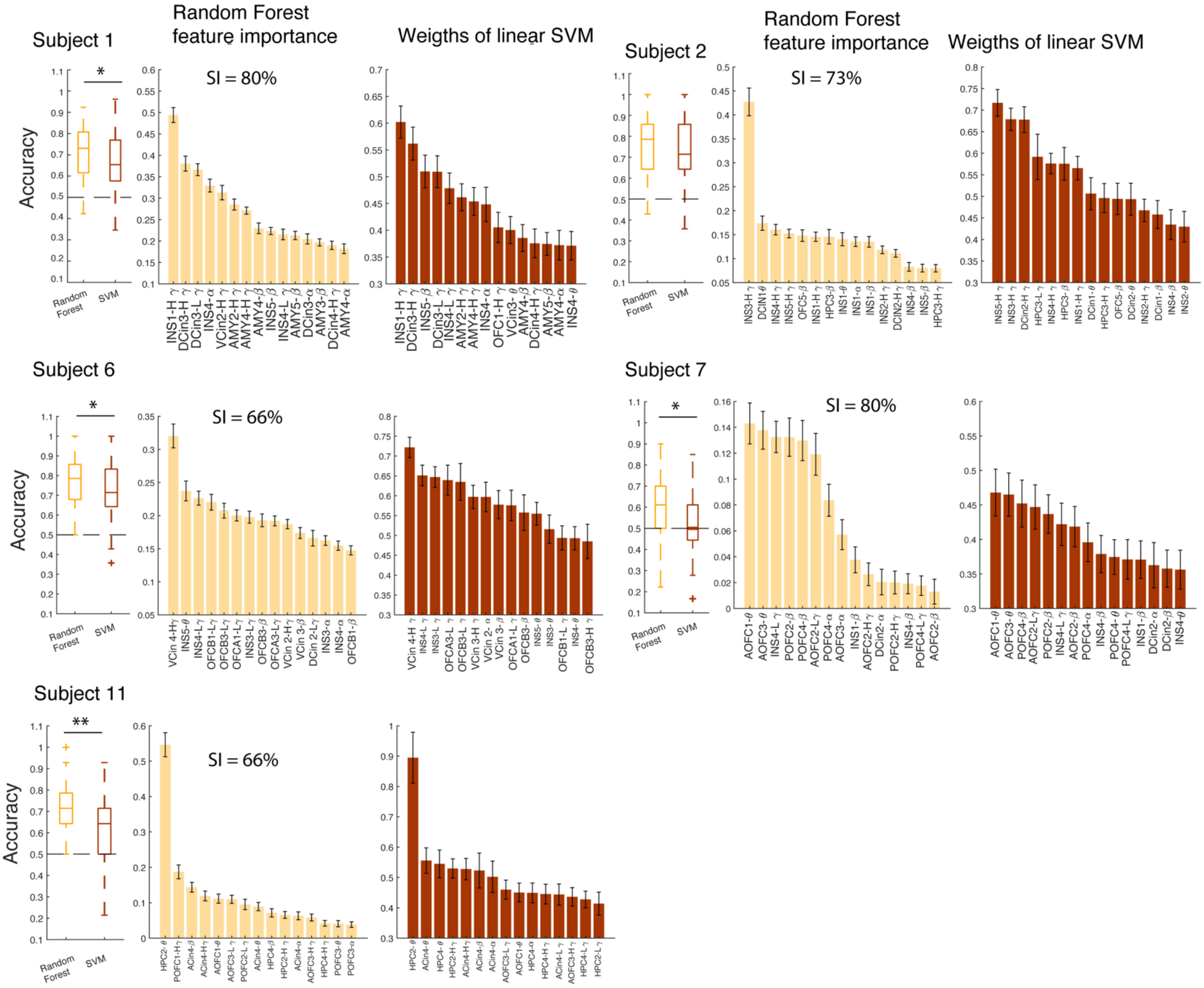
Supplementary Figure 4. *Comparison of selected features by sorted t-statistics (left) and random forest prediction error (right) for two participants.*



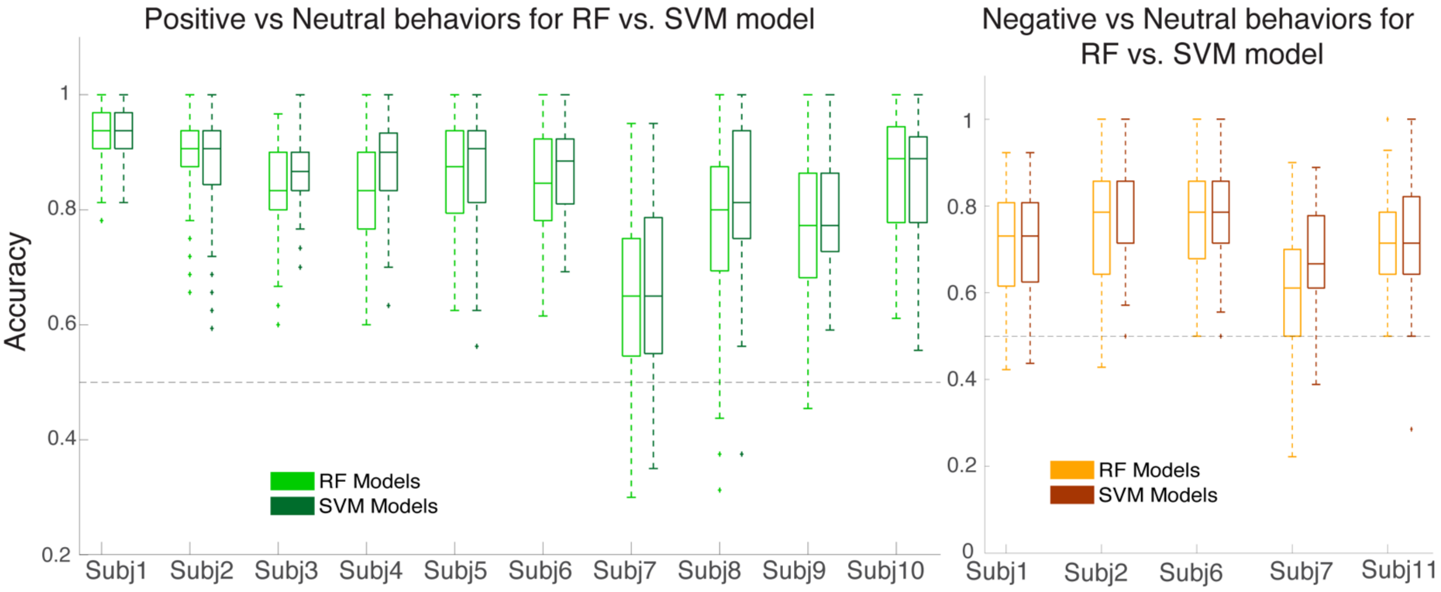
Supplementary Figure 5. *Linear SVM classifiers were trained on the positive affective behaviors and neutral behaviors.* The decoder performance, as well as the top 15 features, are contrasted with the RF models. Box plots represent distribution of accuracy for both models across n=100 datasets. Central lines represent the median and the two edges represent 25 and 75 percentiles, whiskers show the most extreme datapoints and outliers are shown individually (see MATLAB boxplot function). Bar charts represent mean values +/- SEM across 100 datasets of the 15 features for both RF (middle panels) and SVM(Right) models. The similarity index (SI) of the selected features from the two models is also stated in each panel. INS: insula, VCin = Ventral cingulate, DCin = dorsal cingulate, AMY: amygdala, OFC = Orbitofrontal cortex. *** signifies p < 0.0001, ** signifies p < 0.01 and * signifies p <0.05.



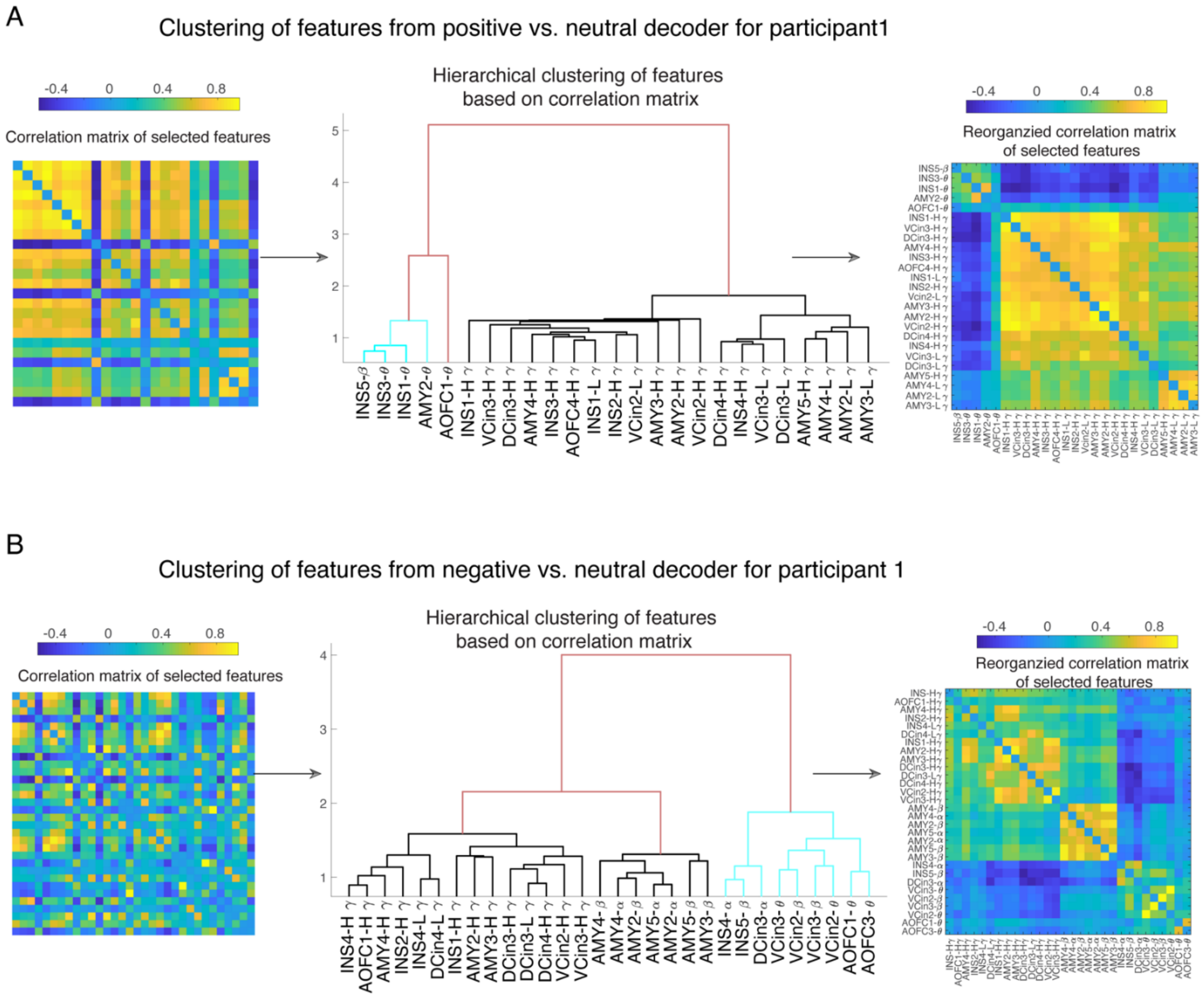
Supplementary Figure 6. *Linear SVM classifiers were trained on the negative affective behaviors and neutral behaviors.* The decoder performance, as well as the top 15 features, are contrasted with the RF models. Box plots represent distribution of accuracy for both models across n=100 datasets(i.e. runs). Central lines represent the median and the two edges represent 25 and 75 percentiles, whiskers show the most extreme datapoints and outliers are shown individually (see MATLAB boxplot function). Bar charts represent mean values +/- SEM across 100 datasets of the 15 features for both RF(middle panels) and SVM(Right) models. The similarity index (SI) of the selected features from the two models is also stated in each panel. INS: insula, VCin = Ventral cingulate, DCin = dorsal cingulate, AMY: amygdala, OFC = Orbitofrontal cortex. ** signifies p < 0.01 and * signifies p <0.05



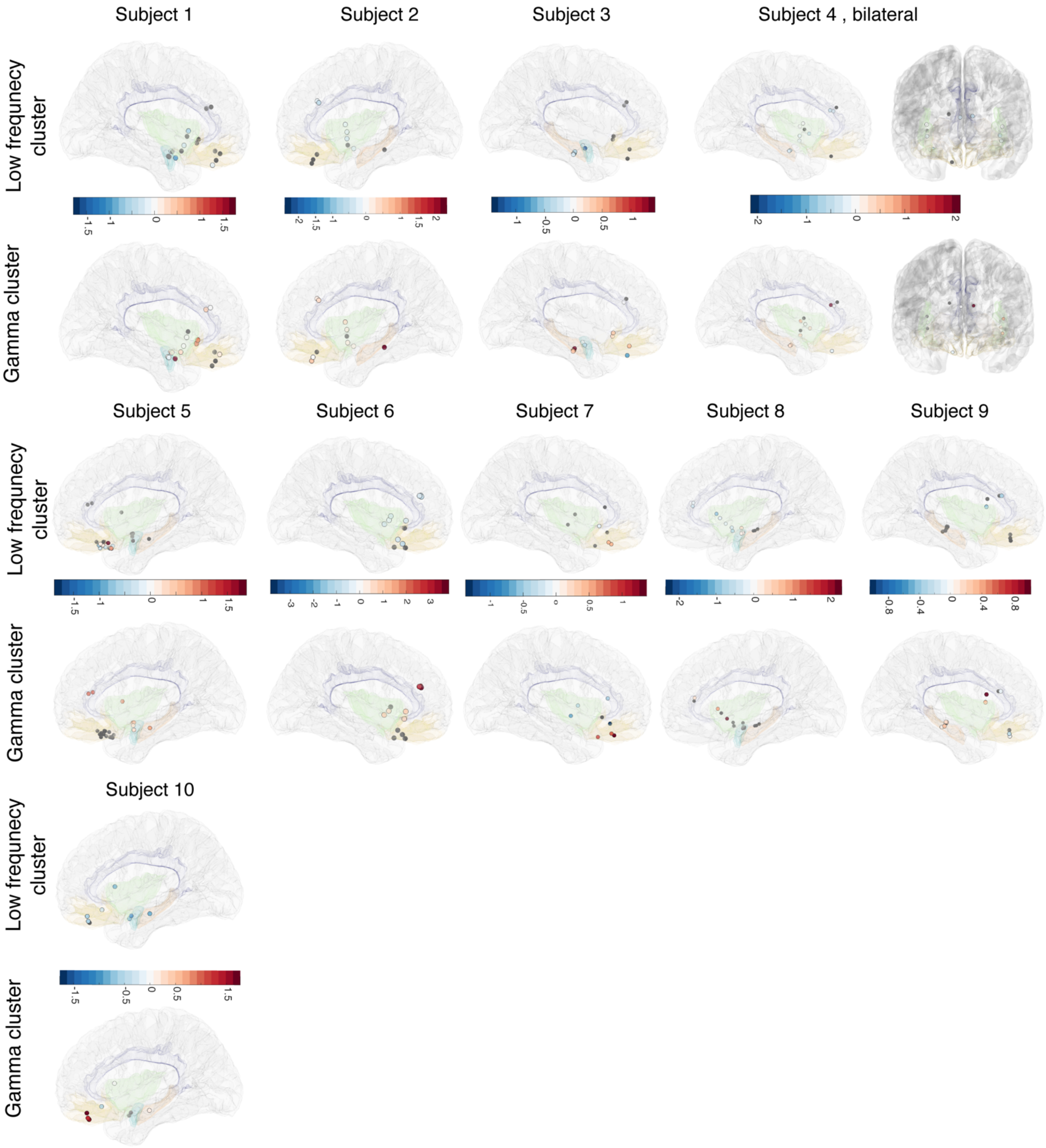
Supplementary Figure 7. *Comparison of the RF and nonlinear SVM models across n=100 datasets.* The SVM models were trained using the selected features from the RF models. The RF and SVM models had a similar accuracy, which indicates that the selected features were robustly identified. In the box plots central lines represent the median and the two edges represent 25 and 75 percentiles, whiskers show the most extreme datapoints and outliers are shown individually (see MATLAB boxplot function).



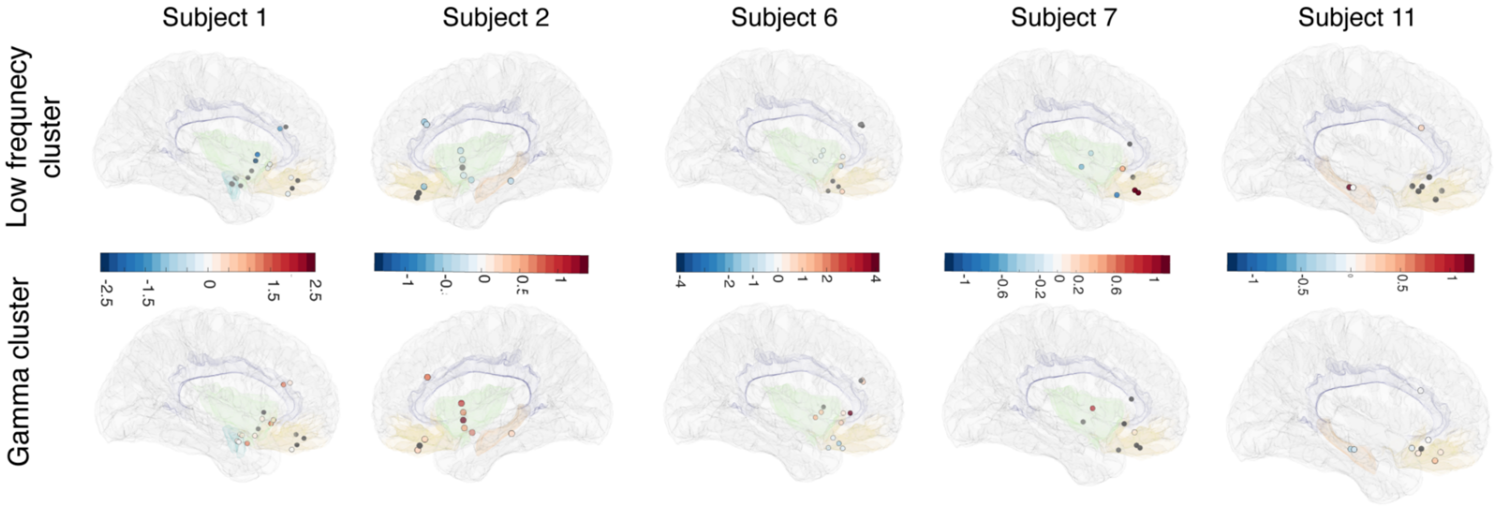
Supplementary Figure 8. *Clustering results from an example participant.* left: Correlation matrix across samples for the selected features, middle: dendrogram results from hierarchical clustering, right: similar correlation matrix as in the left but reordered based on the dendrogram. Example results for Subject 1 from the A) positive decoder and B) negative decoder.



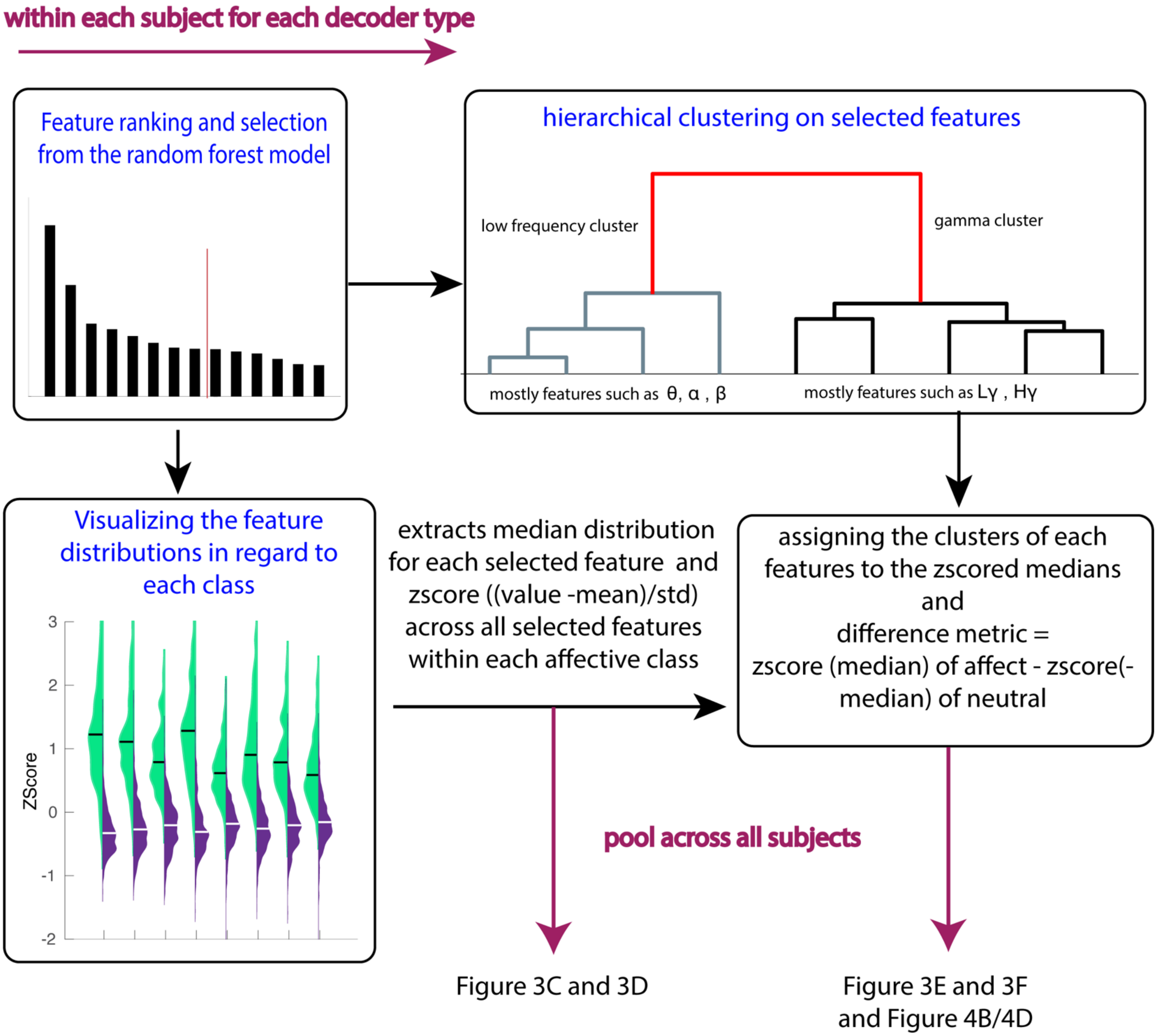
Supplementary Figure 9. *Personalized neural features from the positive decoders in 10 participants.* The values, that are illustrated on MNI brain template (Methods, section “Electrode localization”), are the median difference (positive affective behavior distribution minus neutral behavior distribution) scaled by the feature importance (a positive value) of the selected features that comprised the low-frequency (top row) and gamma (bottom row) clusters. Color maps show the strength of the median difference by feature importance for both the low-frequency and gamma clusters in each participant. The black dots represent the electrodes that were not main contributors to the decoders (i.e., they were included as an input to the decoder models but were not selected by the objective threshold).



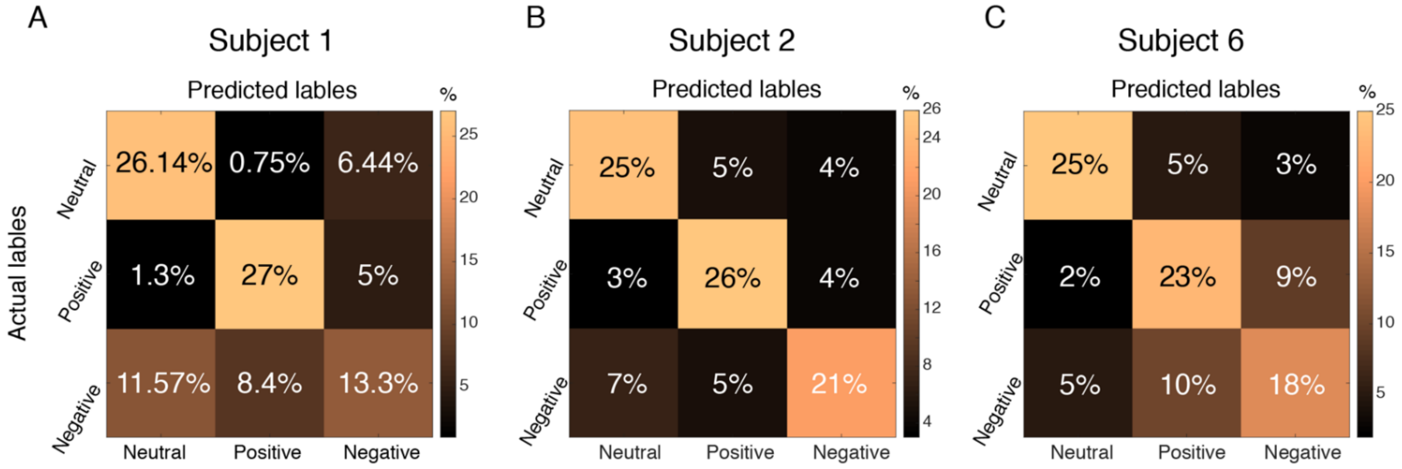
Supplementary Figure 10. *Personalized neural features from the negative decoders in 5 participants.* The values, that are illustrated on MNI brain template (Methods, section “Electrode localization”), are the median difference (negative affective behavior distribution minus neutral behavior distribution) scaled by the feature importance (a positive value) of the selected features that comprised the low-frequency (top row) and gamma (bottom row) clusters. Color maps show the strength of the median difference by feature importance for both the low-frequency and gamma clusters in each participant. The black dots represent the electrodes that were not main contributors to the decoders (i.e., they were included as an input to the decoder models but were not selected by the objective threshold).



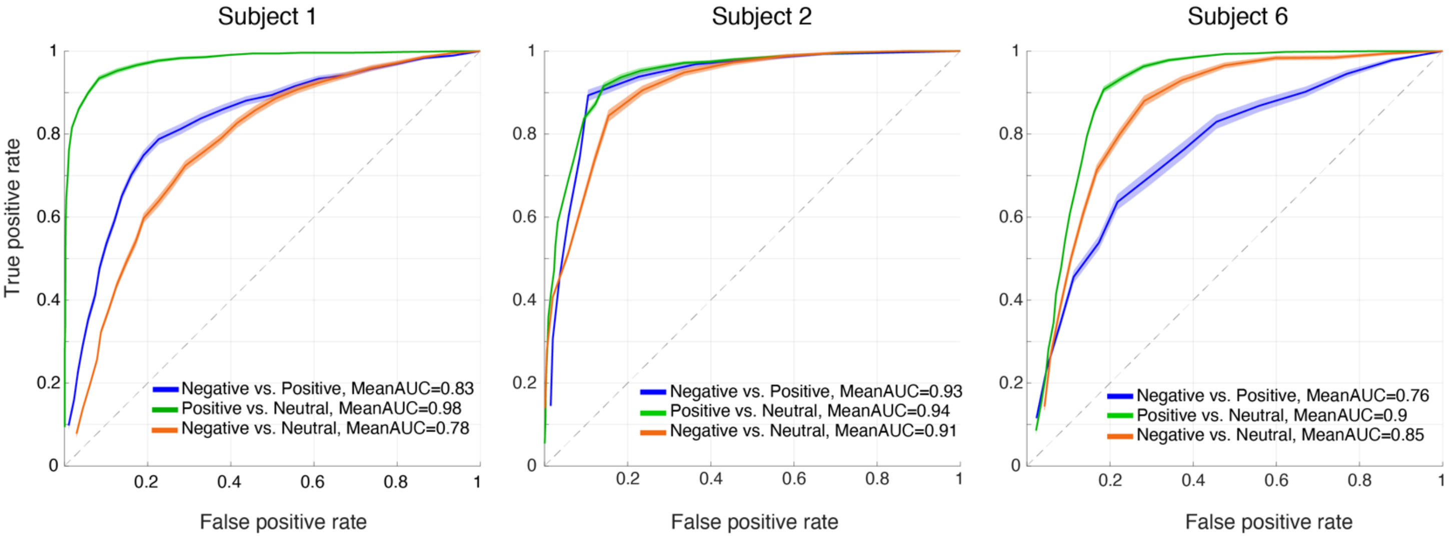
Supplementary Figure 11. *Clustering and normalization pipeline.*


within each subject for each decoder type
Feature ranking and selection from the random forest model
hierarchical clustering on selected features
low frequency cluster
gamma cluster
mostly features such as θ, α , β
mostly features such as Lγ , Hγ
Visualizing the feature distributions in regard to each class
3
2
1
0
-1
-2
ZScore
extracts median distribution for each selected feature and zscore ((value -mean)/std) across all selected features within each affective class
assigning the clusters of each features to the zscored medians and difference metric = zscore (median) of affect - zscore(-median) of neutral
pool across all subjects
Figure 3C and 3D
Figure 3E and 3F and Figure 4B/4D

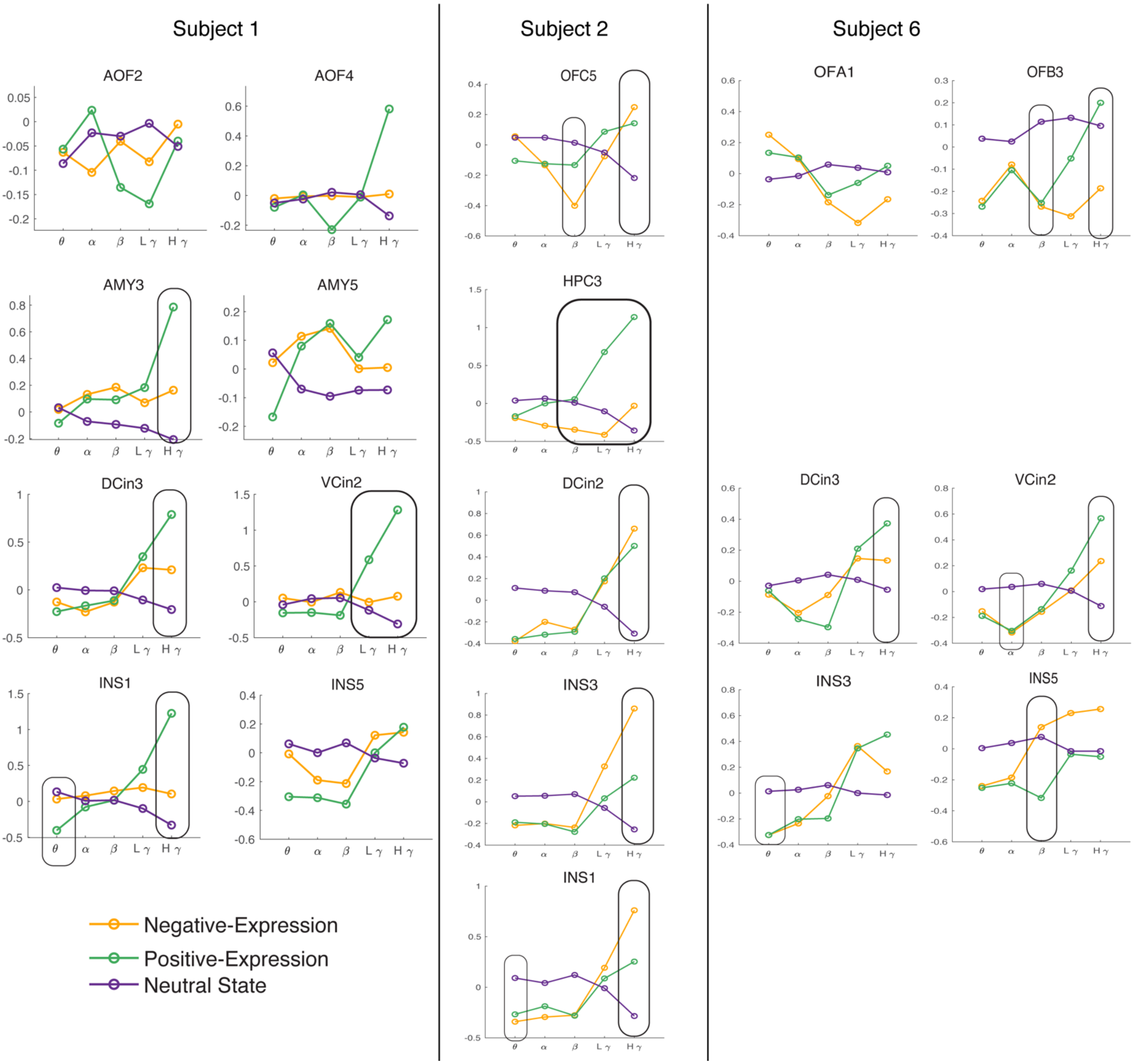
Supplementary Figure 12. *Confusion matrices from the multiclass decoders.* Percentages represent the number of labels of each class over the total number of labels within each fold and dataset, which are then averaged across all 100 runs of the RF models for each participant. Note the ideal separation = 33%. Color bars and percentages show the mean of the confustion matrix values across all 100 runs**.**



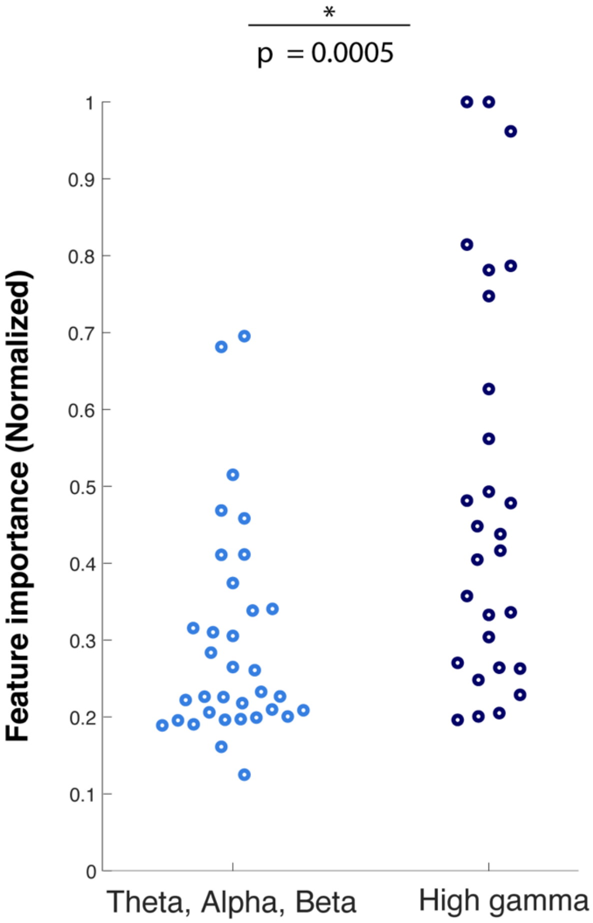
Supplementary Figure 13. *Comparison of the decoding performance (ROC curves) for the three participants in whom the multiclass decoder was trained.* The green lines represent results from positive vs. neutral decoders, the orange lines are negative vs. neutral decoders and the blue lines are the results from positive vs. negative models. Shadings represent standard error of mean across 100 runs of decoders.



Supplementary Figure 14. *Median values of the spectral features extracted from example electrodes during positive, negative, and neutral behaviors in three participants.* Black rectangles highlight the selected features from the RF models.



Supplementary Figure 15. *Feature importance of the selected spectro-spatial features within the theta, alpha, beta, and high gamma bands pooled from the three participants in whom the multiclass decoders were trained. Low gamma was excluded because it did not reach statistical significance among the three behavioral classes.*



Supplementary Table 1. Demographic information, information about the seizure foci, sampled hemisphere, mesolimbic coverage, and available mesolimbic coverage after data cleaning from seizure activity.

Subject Identifier	Age	Gender	Seizure foci	Hemisphere	Mesolimbic coverage	Mesolimbic coverage after electrode cleaning (number of channels)
Subject 1	35	M	Mesial temporal	Right	INS, AMY, DCin, VCin, HPC, OFC,	OFC (4), AMY(3), DCin(2), VCin(2), INS(5)
Subject 2	21	F	hippocampus	Left	INS, AMY, DCin, VCin, HPC, OFC	OFC (3), HPC(1), DCin(2), INS(5)
Subject 3	33	F	Posterior superior frontal gyrus	Right	INS, AMY, DCin, VCin, HPC, OFC	AMY(2), DCin(2), VCin(2), HPC(4), OFC(6)
Subject 4	20	F	Right parietal calcified lesion	Bilateral	INS, HPC, CIN, OFC	RINS(3), RCin (2), ROFC(1), RHPC(2), LINS(3), LCin(1)
Subject 5	20	F	Mesial temporal	Left	INS, AMY, DCin, HPC, OFC	OFC(18), AMY(4), HPC(1), INS(1), DCin(2)
Subject 6	34	F	Temporal lobe	Right	INS, AMY, DCin, VCin, HPC, OFC	INS(3), DCin(2), VCin(3), OFC(4)
Subject 7	30	M	Mesial and lateral temporal lobe	Right	INS, DCin, VCin, OFC	INS(2), DCin(1), OFC(5)
Subject 8	36	M	Mesial Temporal	Left	INS, AMY, DCin, HPC, OFC	INS(5), AMY(5), DCin(2), HPC(3)
Subject 9	20	M	Hippocampus RNS (NA)	Right	INS, DCin, VCin, HPC, OFC	INS(2), DCin(2), HPC(6), OFC(3)
Subject 10	43	M	Left frontal	Left	INS, AMY, HPC, OFC, VCin	INS(1), AMY(2), HPC(1), OFC(4), VCin(2)
Subject 11	27	F	Anterior lateral temporal lobe	Right	AMY, HPC, ACin, PCin, OFC	OFC(6), HPC(2), ACin(1)

Supplementary Table 2. Annotation instructions used by the human raters to code the affective behaviors

Affective Behaviors	Definition
Smiling	The patient is smiling when showing their teeth with a large grin
Laughing	Patient is laughing, including chuckling.
Crying	Patient is crying
Positive verbalization	Patient says something in the context of conversation that indicates a positive state. For example, “I love coffee! This made my day!”
Negative verbalization	Patient says something in the context of conversation that indicates a negative state. For example, “I’m having the worst day of my life”
Discomfort	The patient verbally indicates (without being prompted by medical staff or family) that they are in pain. They could also be exhibiting physical symptoms such as holding their head for long periods of time, holding an icepack on their head, or moaning.

Supplementary Table 3. Annotation instructions used by the human raters to code the neutral behaviors

Other Behaviors	Definition
MedON/MedOFF	Medical staff present/absent
FamON/FamOFF	Family members or friends present/absent
ResearchON/ ResearchOFF	Research staff present/absent
ConFam	The patient is engaged in a conversation, either talking or listening, (lasting more than 10 seconds) with family or friends
ConMed	The patient is engaged in a conversation (lasting more than 10 seconds) with medical staff
ConRes	The patient is engaged in a conversation (lasting more than 10 seconds) with research staff
Comp	Patient is actively using a computer device including iPads *this includes using computers/ipads during research testing
Drink	The patient is drinking – start annotation when the patient is putting the cup to their mouth and drinking. Then annotation is off when the patient removes the cup from their mouth to stop drinking. Say for example that the patient is holding a cup in their hand and talking, this is not drinking. Drinking is ONLY when cup is going to mouth, physically drinking, and then the moment cup is pulled away from their mouth, turn the annotation off.
Eat	The patient is eating. Eating is turned on when the patient brings a fork to their mouth, chews, and when they stop chewing, then turn the annotation off.
Headp	The patient is listening to something on their headphones. Turning Headp ON also applies to when a Patient is listening to music on their phone or a book on tape.
PersCare	Patient was personally caring for themselves which includes activities such as going to the restroom, brushing their hair, washing themselves etc.
Phone	The patient is verbally talking into a phone or a patient is texting/surfing the web etc. on their phone. *turn off phone when the patient is not actively engaged with it for 10 seconds or more.
Read	Patient is reading and must be actively engaged with it for 10 seconds or more.
Search	The patient is actively searching for an item in or around their hospital bed
Seizure	The patient is having a seizure
Sleep/Eye closure	The patient has their eyes closed and is not moving for more than 30 seconds. SleepON begins as soon as the patient closes their eyes.
TestMed	Medical staff are conducing medical tests on the patient such as taking blood pressure, changing IV, playing with any machine attached to the patient, etc.
TResearch	Research Staff are administering research tasks to the patient
TV	TV is on in the patient’s hospital room

Supplementary Table 4. Number of instances of positive and negative affective behavior for each participant after neural data cleaning. NA = not used in decoding.

Subject Identifier	Number of positive samples	Number of negative samples	Number of Affectless samples	Number of Rest samples	Percentage of affectless samples overlap with sleep	Percentage Of affect samples overlap with conversation	Hours	Number of channels (features)
Subject 1	164	133	499	53	26%	45%	14	17(85)
Subject 2	160	28	499	439	0%	77%	6	11 (55)
Subject 3	149	5 (NA)	499	25	46%	94%	11	16(80)
Subject 4	151	12(NA)	336	44	36%	95%	4	12(60)
Subject 5	161	0(NA)	277	146	0%	79%	4	26(130)
Subject 6	133	65	499	499	45%	83%	17	12(60)
Subject 7	51	46	499	499	11%	62%	17	8(40)
Subject 8	42	15(NA)	499	103	0%	91%	6	15(75)
Subject 9	55	5(NA)	499	17(NA)	9%	63%	8	13(65)
Subject 10	47	3(NA)	499	274	28%	94%	19	10(50)
Subject 11	4(NA)	34	499	499	19%	44%	10	9(45)

Supplementary Table 5. Median distributions of the selected features across participants from the positive (n = 10) and negative (n = 5) decoders.

Frequency band	Normalized median of spectro-spatial features from positive decoders For the positive class	Normalized median of spectro-spatial features from positive decoders For the neutral class	Normalized median of spectro-spatial features from negative decoders For the negative class	Normalized median of spectro-spatial features from negative decoders For the neutral class
High Gamma	-0.37 (n = 86)	0.8 (n=86)	0.45 (n=33)	-0.94 (n=33)
Low Gamma	0.0012 (n= 65)	0.44(n=65)	0.79 (n=17)	0.12 (n=17)
Beta	0.81 (n= 37)	-0.82 (n = 37)	-0.46 (n = 23)	-0.81 (n = 23)
Alpha	-0.34 (n= 30)	-0.62 (n= 30)	-0.61 (n= 17)	0.36 (n= 17)
Theta	0.32 (n= 55)	-0.85 (n=55)	-0.6 (n=17)	-0.07 (n=17)

Supplementary Table 6. Multi-comparison tests compared the top selected features from the full models for the positive decoders (see also Extended Data Figure 6).

Subject Identifier	Top 10 Features from Positive vs affectless decoders	Multi-comparison test among regions based on AUC
Subject 1	INS1 $H\gamma$, VCin3 $H\gamma$, DCin3 $H\gamma$, VCin2 $H\gamma$, AMY4 $H\gamma$, AMY2 $H\gamma$, AMY3 $H\gamma$, INS2 $H\gamma$, INS1- θ, INS3 $H\gamma$	Ins, AMY, VCin
Subject 2	HPC3 $H\gamma$,DCin2 $H\gamma$,INS1 θ, HPC3 $L\gamma$, OFC2 $H\gamma$, DCin2 θ, INS3 θ, INS3 β, INS2 $H\gamma$, INS5 $H\gamma$	**HPC, Ins**, DCin
Subject 3	HPC2 θ,HPC2 Lγ, AMY2 β , HPC2 $H\gamma$, AOFC2 $H\gamma$, AOFC2 $L\gamma$, HPC4 β, HPC4 $H\gamma$, AMY2 α, HPC1 β	**HPC, AMY,** VCin, OFC
Subject 4	L_DCin2 $H\gamma$, L-INS2 θ, L-INS3 $H\gamma$, RH3 θ, L_DCin2 $L\gamma$, RH3 Lγ , L-INS1 $H\gamma$, RH4 Hγ,L-DCin2 α, L-INS3 $L\gamma$	**L-Ins**, L-DCin, R-HPC
Subject 5	DCin3 Lγ, DCin4 Lγ , OFC8 Lγ, INS5- θ, HPC2- $H\gamma$, DCin4 β, OFC7 Lγ, OFC3 Lγ, AMY2 β, OFC3 Hγ	**DCin, OFC,** HPC, AMY
Subject 6	VCin3 $H\gamma$, VCin2 $H\gamma$, VCin4 $H\gamma$, DCin3 $H\gamma$, INS3 $H\gamma$, INS3 $L\gamma$, INS3 θ, INS5 β, VCin3 β, DCin2 $H\gamma$	DCin, Ins VCin, OFB
Subject 7	AOFC2 $H\gamma$, AOFC3 $H\gamma$, INS1- $H\gamma$, AOFC3 $L\gamma$, INS1- $L\gamma$, AOFC2$L\gamma$, POFC2 $H\gamma$, AOFC1 $H\gamma$, INS4- $H\gamma$, DCin2- $H\gamma$	No significance
Subject 8	AMY4 $H\gamma$, AMY4 θ , INS3 $H\gamma$, INS5 β, INS5 $H\gamma$, AMY4 α, AMY4 Lγ, DCin3 θ, AMY3 α, INS1- α	Ins, AMY
Subject 9	HPC5 Lγ, HPC6 α, HPC4 α, HPC2 Lγ , HPC3 Lγ , INS7 $H\gamma$, INS6 $H\gamma$, HPC4 Lγ, INS7 Lγ, HPC6 Lγ	HPC, Ins
Subject 10	OFC2 Lγ, OFC4 Lγ, OFC3 Lγ, OFC1 Lγ, OFC4 Hγ, OFC1 Hγ, OFC3 Hγ, HPC3 θ, OFC2 Hγ, INS1 θ	**OFC,** VCin, Ins, HD, AMY (no other significance between all 4 regions, they all have high AUC)

Supplementary Table 7. Multi-comparison tests compared the top selected features from the full models for the negative decoders (see also Extended Data Figure 7).

Subject Identifier	Top 10 Features	Multi-comparison test among regions based on AUC
Subject 1	INS1 $H\gamma$, DCin3 $H\gamma$, DCin3 Lγ, INS4 α, VCin2 $H\gamma$, AMY2 $H\gamma$, AMY4 $H\gamma$, AMY4 β, INS5 β , INS4 Lγ	Ins, VCin, DCin, Amy
Subject 2	INS3 $H\gamma$, DCin1 θ, INS4 $H\gamma$, $\mathrm{INS5}$ $H\gamma$, OFC5 β, INS1 $H\gamma$, HPC3 β, INS1 θ , INS1α, INS1 β	HPC, Ins
Subject 6	VCin4 $H\gamma$, INS5 θ, INS4 $L\gamma$, OFB1 $L\gamma$, OFB3 $L\gamma$, OFA1$L\gamma$, INS3 $L\gamma$, OFB3 β, OFA3 $L\gamma$, VCin2 $H\gamma$	Ins, DCin
Subject 7	AOFC1 θ, AOFC3 θ , INS4 $L\gamma$, POFC2 β, POFC4 β, AOFC2 $L\gamma$, POFC4 α, AOFC3 α, INS1 β, AOFC2 $H\gamma$	POFC, Insula, AOFC
Subject 11	HPC2- θ, POFC1 Hγ, DCIN4 β, DCIN4 $H\gamma$, AOFC1 θ, AOFC1 Lγ, POFC2 Lγ, DCIN4 θ, HPC4β, HPC4 Hγ, DCIN4 α	**HPC,** DCin

Supplementary Table 8. Multiclass decoder performance (F1-Score) related to figure 6-A.

Subject	neutral class F1-Score, p-value from shuffled models	Positive class F1-Score, p-value from shuffled models	Negative class F1-Score, p-value from shuffled models	Accuracy All

S1	0.73 +- 0.01
	Median = 0.73
	p=$4 * 10^{-34}$
		0.77+- 0.013
		Median = 0.77
		p=$4 * 10^{-31}$
			0.44 +- 0.015
			Median = 0.47
			p=$8 * 10^{-12}$
				0.66+- 0.006
				Median = 0.66

S2	0.73+- 0.016
	Median = 0.75
	p=$1.5 * 10^{-27}$
		0.75 +- 0.017
		Median = 0.75
		p=$1.1 * 10^{-28}$
			0.70+- 0.02
			Median = 0.71
			p=$2.5 * 10^{-19}$
				0.72 +-0.0127
				Median = 0.73

S6	0.77 +-0.014
	Median = 0.79
	p=$7.65 * 10^{-30}$
		0.65 +- 0.02
		Median = 0.70
		p=$1.6 * 10^{-19}$
			0.59 +- 0.018
			Median = 0.6
			p=$7.3 * 10^{-17}$
				0.67 +- 0.014
				Median = 0.66

Average of all three subjects	0.74 ± 0.013	0.72 ± 0.037	0.57 ± 0.07	0.68 ± 0.016

Supplementary Table 9. Multiclass decoder performance related to figure 6-C.

Subject	Accuracy
Insula	0.62 ± 0.006
OFC	0.52 ± 0.006
Dorsal ACC	0.58 ± 0.008
Ventral ACC	0.58 ± 0.007

Supplementary Table 10. P-value regarding statistical test for panel F in figure 2. All values are obtained by two-sided non-parametric pairwise ranksum test across n=100 datasets.

Subject Identifier	p-values
Subject 1	$1.8 * 10^{-34}$
Subject 2	$4.8 * 10^{-34}$
Subject 3	$5.5 * 10^{-32}$
Subject 4	$3.18 * 10^{-34}$
Subject 5	$6.41 * 10^{-34}$
Subject 6	$4.8 * 10^{-34}$
Subject 7	$3 * 10^{-13}$
Subject 8	$2 * 10^{-22}$
Subject 9	$8 * 10^{-27}$
Subject 10	$2.63 * 10^{-31}$

Supplementary Table 11. P-value regarding statistical test for panel G in figure 2. All values are obtained by two-sided non-parametric pairwise ranksum test across n=100 datasets

Subject Identifier	p-values
Subject 1	$1 * 10^{-23}$
Subject 2	$5 * 10^{-25}$
Subject 6	$2.4 * 10^{-26}$
Subject 7	$5.3 * 10^{-8}$
Subject 11	$2.5 * 10^{-21}$

SI References

1. P. Krolak-Salmon, *et al.*, An attention modulated response to disgust in human ventral anterior insula: Disgust in Ventral Insula. *Ann. Neurol.* **53**, 446–453 (2003).

2. A. Touroutoglou, M. Hollenbeck, B. C. Dickerson, L. Feldman Barrett, Dissociable large-scale networks anchored in the right anterior insula subserve affective experience and attention. *NeuroImage* **60**, 1947–1958 (2012).

3. Y. Zhang, *et al.*, The Roles of Subdivisions of Human Insula in Emotion Perception and Auditory Processing. *Cereb. Cortex* **29**, 517–528 (2019).