



## Identifying task-relevant spectral signatures of perceptual categorization in the human cortex

Ilya Kuzovkin, Juan R. Vidal, Marcela Perrone-Bertolotti, Philippe Kahane,  
Sylvain Rheims, Jaan Aru, Jean-Philippe Lachaux, Raúl Vicente

### ► To cite this version:

Ilya Kuzovkin, Juan R. Vidal, Marcela Perrone-Bertolotti, Philippe Kahane, Sylvain Rheims, et al.. Identifying task-relevant spectral signatures of perceptual categorization in the human cortex. *Scientific Reports*, 2020, 10 (1), 10.1038/s41598-020-64243-6 . hal-02860258

HAL Id: hal-02860258

<https://hal.science/hal-02860258>

Submitted on 16 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**OPEN**

# Identifying task-relevant spectral signatures of perceptual categorization in the human cortex

Ilya Kuzovkin<sup>1</sup>✉, Juan R. Vidal<sup>2,6,10</sup>✉, Marcela Perrone-Bertolotti<sup>3</sup>, Philippe Kahane<sup>4,5</sup>, Sylvain Rheims<sup>6,7</sup>, Jaan Aru<sup>1,8</sup>, Jean-Philippe Lachaux<sup>6,9</sup> & Raul Vicente<sup>1,10</sup>✉

Human brain has developed mechanisms to efficiently decode sensory information according to perceptual categories of high prevalence in the environment, such as faces, symbols, objects. Neural activity produced within localized brain networks has been associated with the process that integrates both sensory bottom-up and cognitive top-down information processing. Yet, how specifically the different types and components of neural responses reflect the local networks' selectivity for categorical information processing is still unknown. In this work we train Random Forest classification models to decode eight perceptual categories from broad spectrum of human intracranial signals (4–150 Hz, 100 subjects) obtained during a visual perception task. We then analyze which of the spectral features the algorithm deemed relevant to the perceptual decoding and gain the insights into which parts of the recorded activity are actually characteristic of the visual categorization process in the human brain. We show that network selectivity for a single or multiple categories in sensory and non-sensory cortices is related to specific patterns of power increases and decreases in both low (4–50 Hz) and high (50–150 Hz) frequency bands. By focusing on task-relevant neural activity and separating it into dissociated anatomical and spectrotemporal groups we uncover spectral signatures that characterize neural mechanisms of visual category perception in human brain that have not yet been reported in the literature.

Our capacity to categorize sensory information allows us to quickly process and recognize complex elements in our environment. Early studies revealed strong relations between the brain activity within certain localized networks and the neural representations of certain stimulus categories, as for example faces, bodies, houses, cars, objects and words<sup>1–7</sup>. These early assessments also revealed brain networks' capability to rapidly extract categorical information from short exposure to natural scenes<sup>8–10</sup> based on models of parallel processing across neural networks<sup>3,11</sup>. In both animal and human studies, visual cortices and particularly inferior temporal cortex (ITC) appears as a key region to integrate information at the object-level<sup>12–14</sup>. In humans, a great deal of observations of cortical response selectivity have been achieved using fMRI, but measuring direct neuronal activity<sup>15,16</sup> also revealed similar patterns. To further understand how stimulus features and perceptual experience is processed in neural networks, brain activity, especially in sensory cortices, has been decoded using a variety of methods and signals<sup>17–19</sup>. This decoding often relies on machine learning to avoid a priori selection of partial aspects of the data by the human observer, and unless additional analysis is performed on the model itself it does not emphasize the mechanisms of neuronal communication within and between neural networks involved in this processing.

A pervasive feature of electrophysiological neural activity are its spectral fingerprints. Neural oscillations have been proposed to reflect functional communication processes between neural networks<sup>20–23</sup>. Certain frequency bands are selectively associated with the operating of different cognitive processes in the human and animal

<sup>1</sup>Computational Neuroscience Lab, Institute of Computer Science, University of Tartu, Tartu, Estonia. <sup>2</sup>UMRS 449, Université Catholique de Lyon/Ecole Pratique des Hautes Etudes, 10 Place des Archives, 69002, Lyon, France.

<sup>3</sup>University Grenoble Alpes, University Savoie Mont Blanc, CNRS, LPNC, 38000, Grenoble, France. <sup>4</sup>Inserm, U1216, F-38000, Grenoble, France. <sup>5</sup>Neurology Department, CHU de Grenoble, Hôpital Michallon, F-38000, Grenoble, France.

<sup>6</sup>INSERM U1028, CNRS UMR5292, Lyon Neuroscience Research Center, Lyon, France. <sup>7</sup>Department of Functional Neurology and Epileptology, Hospices Civils de Lyon and Université Lyon, Lyon, France. <sup>8</sup>Institute of Biology, Humboldt University Berlin, Berlin, Germany. <sup>9</sup>Université Claude Bernard, Lyon, France. <sup>10</sup>These authors contributed equally: Juan R. Vidal and Raul Vicente. ✉e-mail: [ilya.kuzovkin@gmail.com](mailto:ilya.kuzovkin@gmail.com); [juanrvidal@gmail.com](mailto:juanrvidal@gmail.com); [raulvicente@gmail.com](mailto:raulvicente@gmail.com)

brain<sup>24–29</sup>, and lately, direct recordings from the human cortex have revealed the remarkable representation selectivity of broadband high-gamma activity (50–150 Hz)<sup>30–32</sup>. Human intracranial recordings have previously shown evidence of functional processing of neural networks related to perceptual category representation<sup>33</sup> and lately the prominence of broadband high-gamma activity in selective category responses in visual areas<sup>34–38</sup>. Yet, very little is known about the specific relation between the different components of the full power-spectrum, including high-gamma activity, and their level of selectivity in processing perceptual categories. Previous works have shown where and when perceptual category information can be decoded from the human brain, the approach introduced in this work adds to that line of research by allowing to identify spectrotemporal patterns that contribute to category decoding without the need to formulate a priori hypothesis on which spectrotemporal regions of interest are worth investigating.

In this work we capitalize on an extensive dataset of deep intracranial electrical recordings on 100 human subjects to decode neural activity produced by 8 different stimulus categories. We analyzed the decoding models built by a random forest classifier to disentangle the most informative components of the time-frequency spectrum related to the simultaneous classification of 8 different perceptual categories. Via *feature importance* analysis we quantified the contribution of each TF component into the decoding decision, which allowed us to identify the activity patterns that were either characteristic of the processing of a specific visual category or were shared by several categories. In addition to feature importance we analyzed the predictive power of each activity pattern and identified how informative was their spectral signature for the classification of visual categories. We tested the predictive power of broadband high-gamma activity in comparison to lower frequency activity as they reflect different communication mechanisms elicited by networks seemingly involved in distinct temporal windows of functional neuronal processing. Through the analysis of feature importance we show the specific neuronal spectral fingerprints from highly distributed human cortical networks that were elicited during automatic perceptual categorization. The uncovered spectral signatures provide insight into neural mechanisms of visual category perception in human brain.

## Results

**Feature importance allows to separate out the neural signals that are predictive of perceptual categorization from the mixture of stimulus-induced responses.** To identify spectrotemporal features that are characteristic of automatic perceptual categorization of a particular category we relied on time-frequency (TF) maps of the neural responses of intracranially implanted electrodes. Out of the total set of 11321 probes 11094 (98%) were responsive (see the Methods section on processing of neural data for details) to the stimuli from at least one of the categories. On one hand this provides us with abundance of data, on the other raises the question whether all of that activity was relevant to the processes that encode and process visual input.

Training a decoding model (see the Methods section on Random Forest as decoding model and the illustration on Fig. 1) for each of the probes allowed us to dissociate the *predictive probes* that exhibited activity that was useful for decoding from the rest of the *responsive probes* that did not carry such activity.

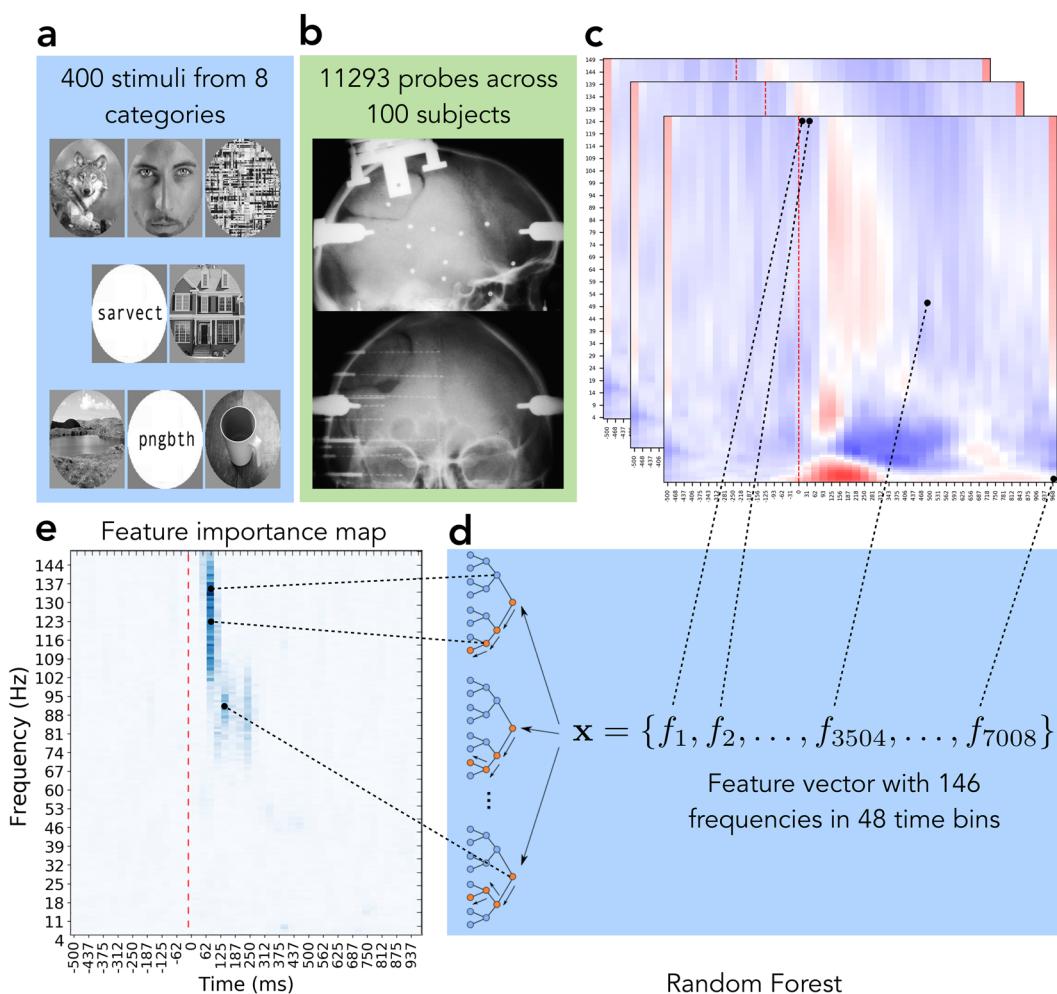
Green markers on Fig. 2a show the set of probes that are responsive to the *house* category, while the blue markers are the probes that are predictive of that category (4.8%, 535 probes). Figure showing the distribution of responsive and predictive probes for other categories are available in Supplementary Materials repository. Decoding models built on the neural responses of the predictive probes were successful at classifying at least one perceptual category ( $F_1 > 0.39$  for one or more classes), focusing on them in our further analysis allowed to work only with the locations that carry information relevant to the task of perceptual categorization.

Predictive probes had a heterogeneous distribution in the brain, yet remained mostly concentrated in visual cortices and inferior temporal regions (76%), from BA17 to BA20, including early visual areas (BA 18, 19), fusiform gyrus (BA 37) and inferior temporal cortex (BA 20). A majority of the predictive probes were in fusiform cortex (average of 52% over all categories, Fig. 2b), followed by BA 19 (27%), across all category networks. The detailed numbers of the distribution of predictive and responsive probes across categories and Brodmann areas can be found in Table 1 of the Supplementary Materials Section.

Within the primary visual cortex, BA 17 and 18, the *scrambled* was the stimulus that elicited most predictive probes (28) amongst all stimulus categories (Fig. 2c), followed by *pseudowords* (13). Probes predictive of *faces* were mostly concentrated in BA19, BA37 and BA20 (72%, 108 out of 150). The low number of predictive probes in area 17 is explained by the fact that less than 1% of the implantation sites in the original dataset were located in primary visual cortex.

Previous studies have shown that perceptual category-selective networks are located in occipito-temporal cortex<sup>4,6,12</sup>. To test whether predictive power of the Random Forest model trained to decode activity of probes is coherent with known functional processing by cortical networks we evaluated the selectivity of the predictive power in three known functional networks: Fusiform Face Area (FFA)<sup>1</sup>, Visual Word Form Area (VWFA)<sup>7</sup> and Parahippocampal Place Area (PPA)<sup>39</sup>. We checked whether the probes located in each of these areas and the Random Forest model trained on these probe's activity to discriminate between 8 categories produces the highest predictive power for the category for which this area is known to be selective. Probes in FFA are associated with facial recognition and encoding facial information<sup>40–44</sup> and thus we expect their activity to be predictive of the *face* category, probes in VWFA should be predictive of *characters* and *pseudowords* categories<sup>42,45,46</sup> and probes in PPA should be responsive to *scenes* and *houses*<sup>39,47–49</sup>.

There were 12 probes in the FFA that were significantly (permutation test  $p < 1e - 4$ ) predictive (classification score  $F_1 > 0.39$ ) of a category: 5 were predictive of *faces*, 4 of *animals* (which mostly have faces on the image), 2 of *pseudowords* and 1 of *scrambled* images. Most probes that were in FFA and were predictive, carried information of the categories containing facial features.



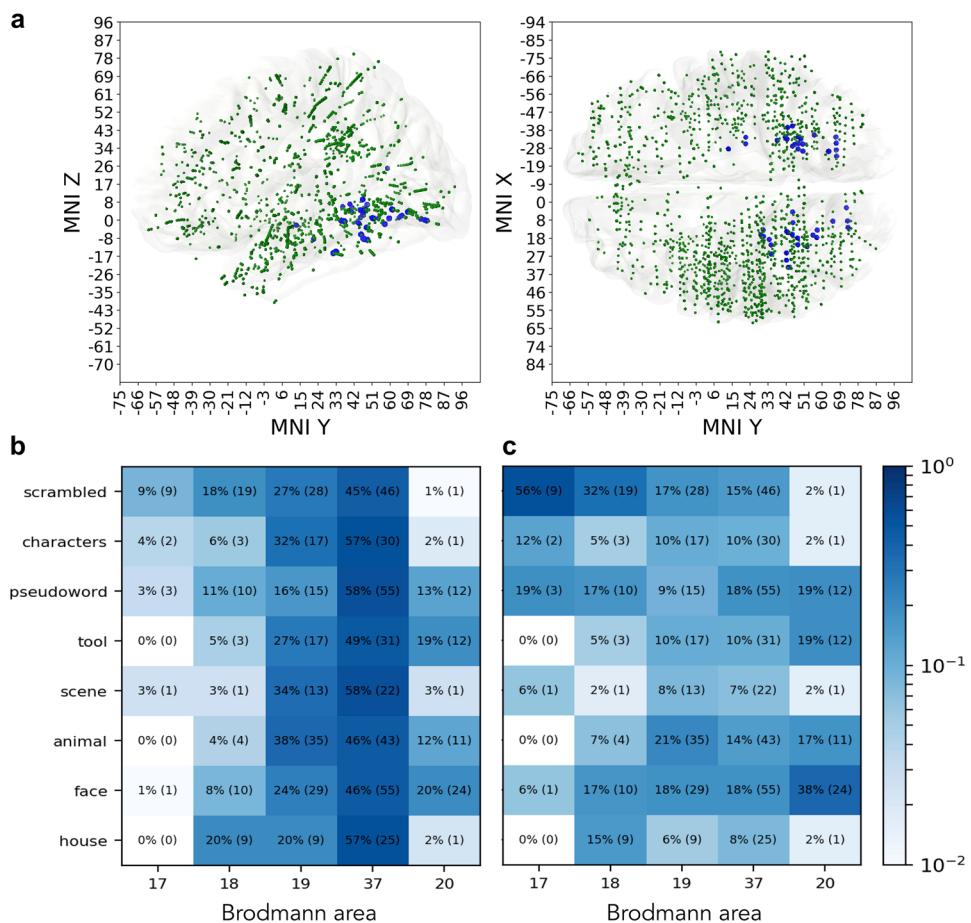
**Figure 1.** Major steps of the data processing pipeline. (a) Image stimuli from 8 categories were presented to test subjects. (b) Human brain responses to images were recorded with deep intracranial electrodes. (c) LFP signals were preprocessed and transformed into time-frequency domain. (d) Random Forest models were trained to decode image category from each electrode’s activity. (e) Feature importances of each model were calculated to identify the region on each electrode’s activity map that was relevant to visual object recognition. Notice how the final results on panel (e) tell us that high gamma activity in 90–120 ms window and the subsequent activity in the low gamma range in 120–250 ms window are the only bands and time windows in that particular electrode’s activity that are relevant for the classification task, while the spectrogram on panel c also shows that there was activity in early theta, beta and low gamma bands. Our analysis revealed that not all activity was relevant (or useful) for the classification of an object and showed which parts of the activity are actually playing the role in the process.

There were 8 probes in the VWFA that were predictive of a category: 5 were predictive of pseudowords, 2 of characters and 1 of faces. This points to the fact that the predictive probes in VWFA are predictive of the stimuli with written characters on them. These results confirm that predictive power of a Random Forest model trained on probes activity in VWFA reflects the functional role known to be carried by this area.

For probes in the PPA results were less selective. There were 23 probes inside that area that were predictive of a category: 5 were predictive of houses, 4 of scenes, 5 of characters, 5 of scrambled images, 2 of tools and 2 of pseudowords. The probes from PPA predicted not only houses and scenes, but also other categories. However, houses and scenes were among the categories that the probes from PPA were able to identify successfully in highest proportion as compared to the other categories.

These confirmatory findings give credibility to the methodology by which the probes that are identified as predictive of a certain category are involved in the processing of the stimuli that belong to that category.

Training per-probe decoding models not only allowed us to identify the predictive locations, but also to apply feature importance analysis to decoding models trained on local activity. Computing the feature importance across the time-frequency map (4–150 Hz and –500 to 1000 ms) allowed us to see which parts of neural activity are crucial for the decoding. Overlaying the importance over time-frequency map showed at which frequencies and at what times the activity that was important for the algorithm has occurred. This can be applied both on aggregated level, where the importance map is averaged over probes, and on individual probe level. “Feature



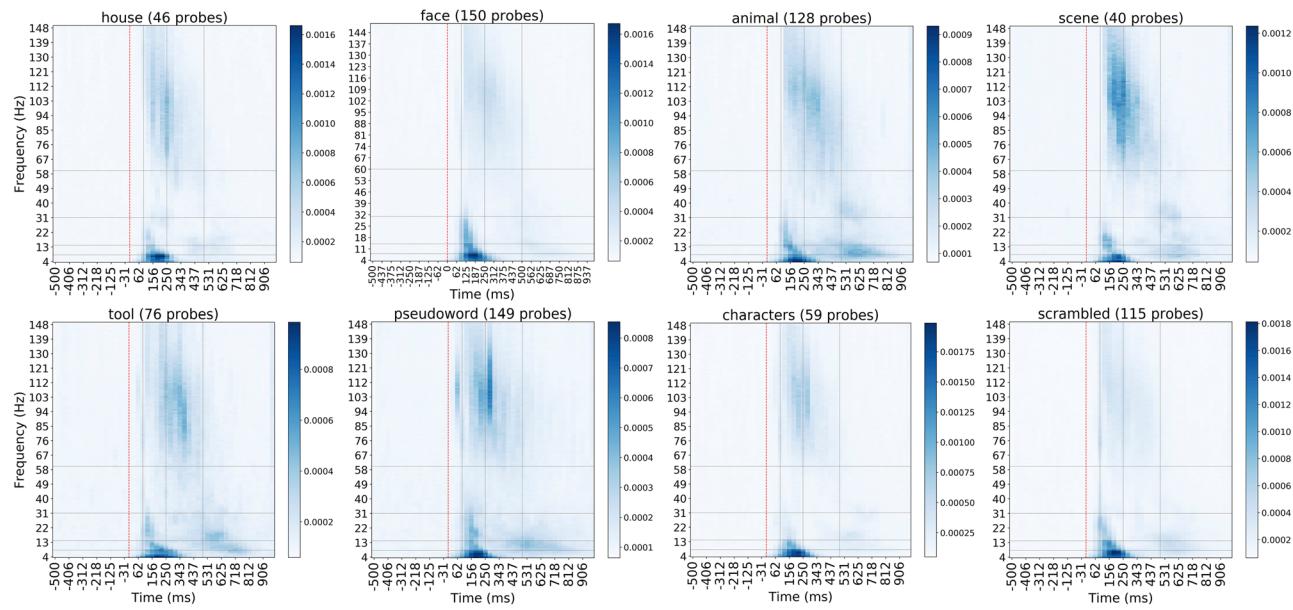
**Figure 2.** Distribution of predictive probes. (a) Green markers indicate all of the probes that were responsive to stimuli from the house category. Blue markers indicate only the predictive probes that carry information that is relevant to decoding the neural response as reaction to house stimulus. The house category was selected for this figure as an illustrative example, similar effect can be seen across all 8 categories, please see the Supplementary Materials repository for the visualization of the distributions of responsive and predictive probes in all eight categories. (b) Distribution of predictive probes over areas within each category (each row sums up to 100%). Color shows the percentage across categories. (c) Distribution of predictive probes over a category within each area (each column sums up to 100%). Color shows the percentage across areas.

importance of the analysis of task-relevant neural activity” part of the Methods section explains the application of probe importance map to filter irrelevant activity and obtain spectrotemporal signature of a particular category on a particular probe. Now we can use the feature importance map as a mask and perform the analysis of the activity itself, focusing only on the relevant parts of it. When applicable, this methodology helps to filter out irrelevant activity and allows to focus on the activity that is important to the scientific question under investigation.

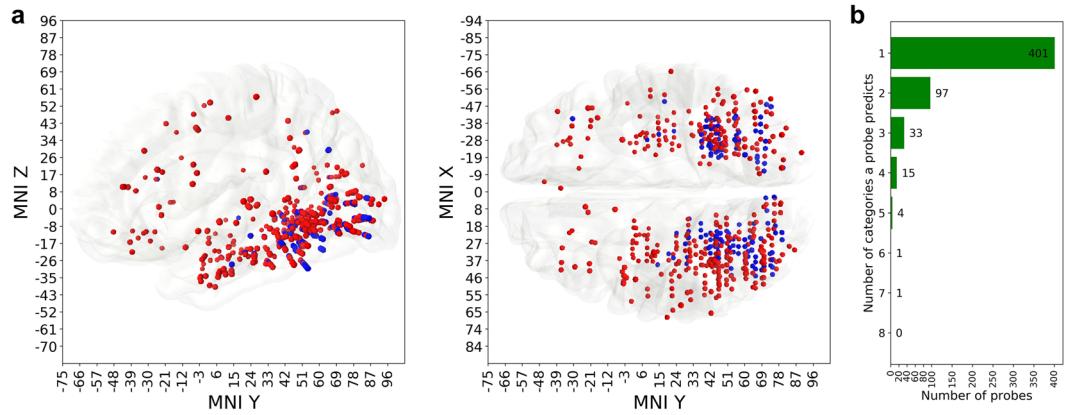
We took an average over importance maps of all individual probes within each category to obtain the global picture of where the category-specific activity lies in time and frequency space. Figure 3 visualizes average importance maps of each of eight categories and allows to single out the spectrotemporal signatures that are unique to specific categories and those that are less selective. From these importance maps we notice that certain TF components are distinctly present per category, as for example high (significantly higher than 81 out of 112 regions of interest, Mann-Whitney U  $p < 8.9e - 7$ , corrected) importance of the transient theta activity (theta burst) in all categories, or the almost absence of importance of broadband gamma (significantly lower than 10 out of 12 other regions of interest, Mann-Whitney U  $p < 8.3e - 5$ , corrected) in the control scrambled condition.

In the following sections we expand our analysis to the comparison of the feature maps and analyzing the activity under the regions that the method identifies as important.

**Polypredictive and monopredictive probes.** The analysis revealed two types of neural locations: *polypredictive* probes are predictive of multiple visual categories, while *monopredictive* are useful for decoding only one out of 8 different types of stimuli revealing a high degree of specialization (Fig. 4b). We considered a probe to be predictive of a category if cross-validation F<sub>1</sub> score for that category was higher than 0.39 (see the Methods for details on the threshold selection), which is a stricter condition than above-chance criterion (F<sub>1</sub> > 0.125). We use the term ‘predictive’ to highlight the difference in meaning with the classical terms ‘selective’ and ‘responsive’.



**Figure 3.** Average importance map of each of eight categories over probes predictive of that category. The color shows the relative importance of each spectrotemporal feature, indicating how informative that particular feature was for the task of decoding.

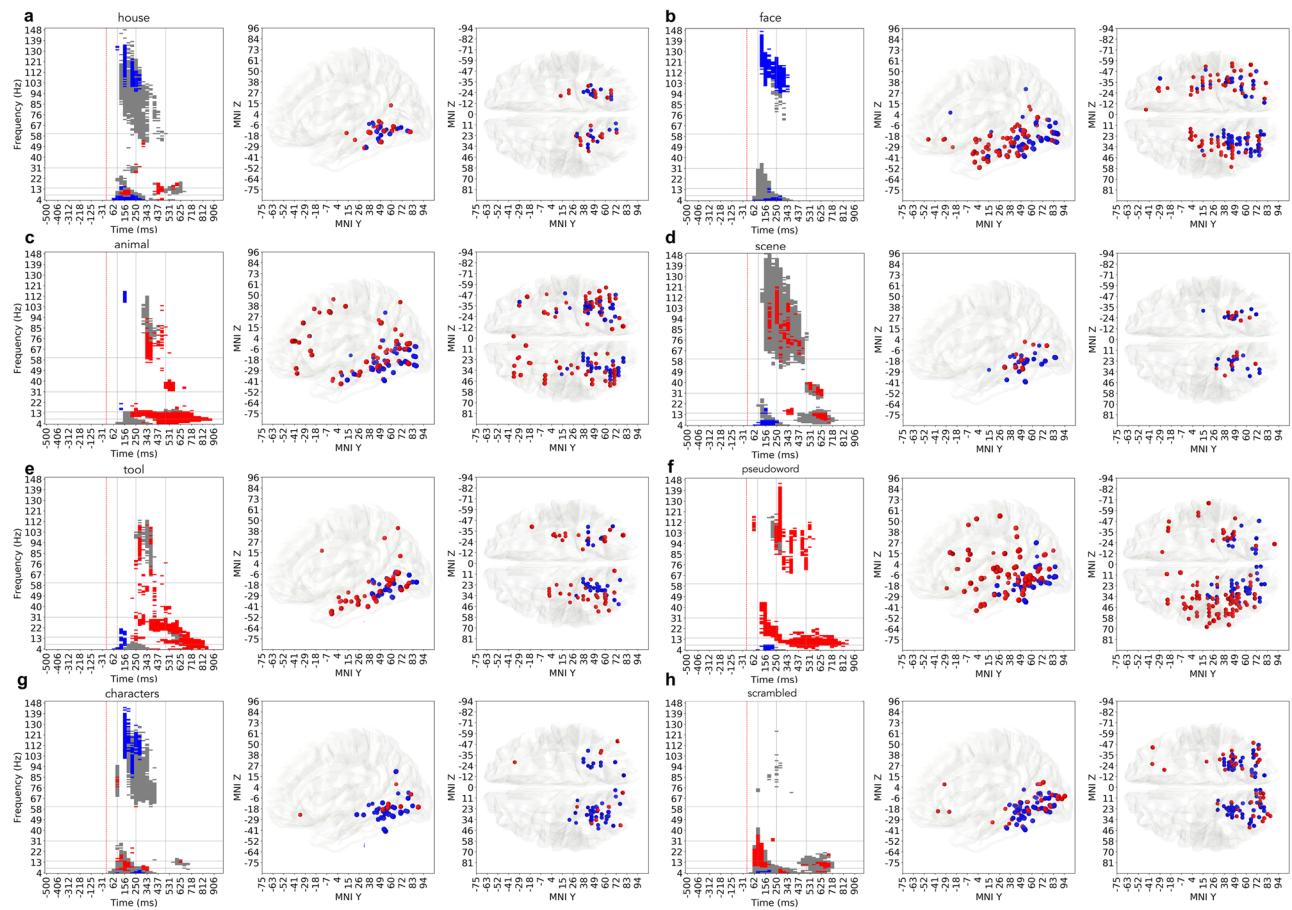


**Figure 4.** Anatomical distribution of mono- and polypredictive locations. (a) Red markers are the locations of monopredictive probes, blue markers are the locations of polypredictive ones. Polypredictive probes (145 unique locations) are mostly confined to visual areas and temporal lobe (both parts of the ventral stream), while monopredictive (specialized, 401 unique locations) probes are, in addition to visual areas, also found in frontal and parietal cortical structures. (b) The histogram shows how many categories are predictable by how many probes.

Both these terms describe the post-stimulus activity of an electrode in terms of the relationship of the post-stimulus response to the pre-stimulus baseline. The relevance of the electrode's activity to the classification task, however, is measured in this work using classifier's prediction score.

Figure 4a shows that polypredictive probes reside mainly (94%, 136 out of 145) in posterior occipital and posterior temporal, while the monopredictive probes extend, in addition to occupying similar posterior occipital and temporal locations, to frontal cortex (92%, 45 out of 49 probes in this area are monopredictive) and anterior temporal cortex (88%, 51 out of 58 probes). Both mono- and polypredictive probes are also observed in parietal cortex. Monopredictive probes that extend beyond ventral stream and temporal cortex pertain to the following perceptual categories: faces (orbitofrontal cortex), animals and pseudowords (dorsofrontal cortex, inferior frontolateral cortex, premotor cortex), and, to a smaller extent, scrambled images (prefrontal cortex).

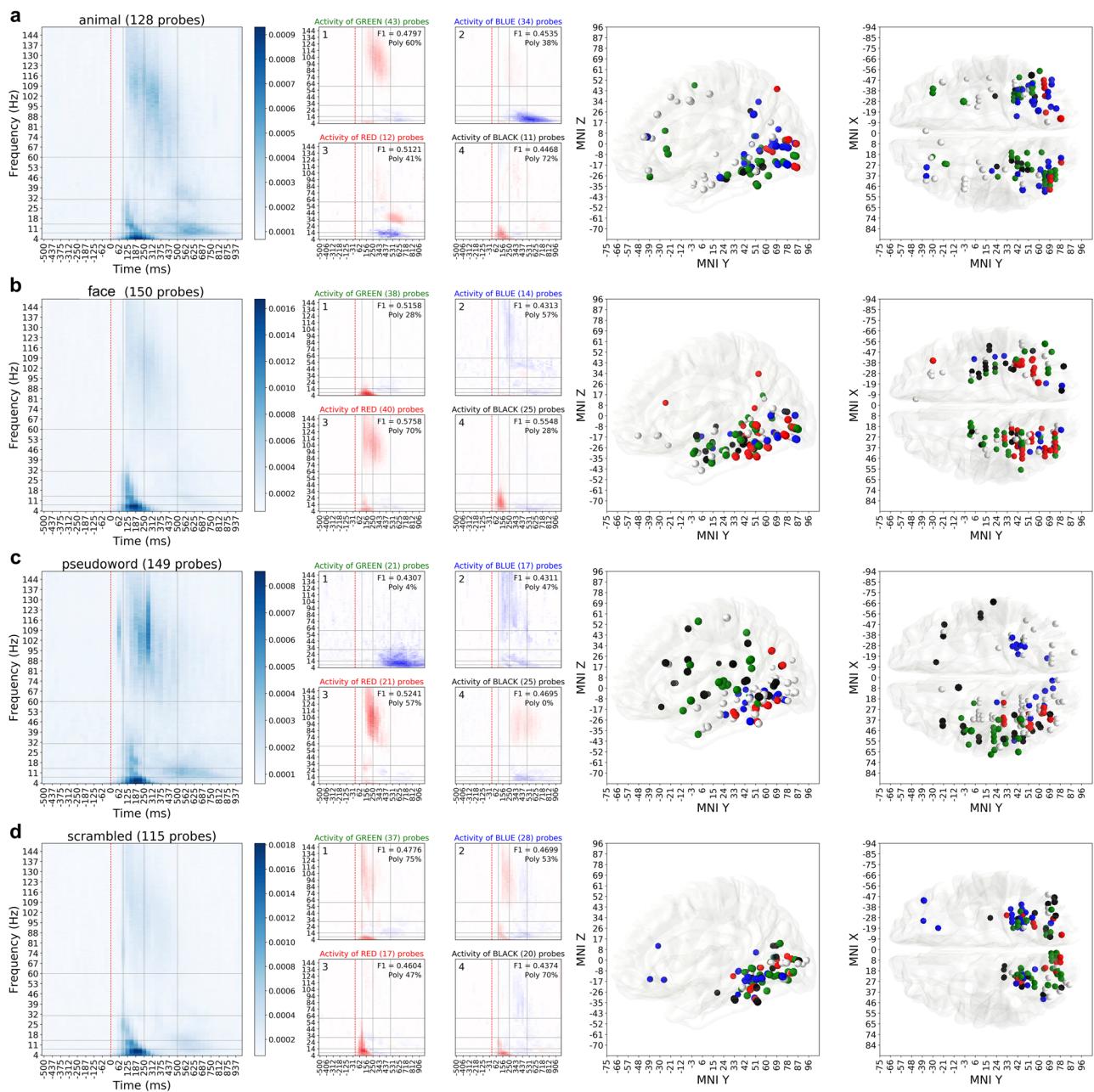
The unique association of specific TF feature importance components with either polypredictive and monopredictive probes was category specific, as shown in Fig. 5a–h. While all of the data presented on these figures shows statistically significant differences between monopredictive and polypredictive neural locations, we will focus only on a few that were supported by the strongest signal in the data. For face stimuli, most of the feature importance in the early broadband gamma response was significantly ( $4\sigma$ ) higher in polypredictive probes as



**Figure 5.** Statistically significant differences between the importance of monopredictive and polypredictive probes' activity. Gray regions indicate the areas of TF spectrum where both monopredictive and polypredictive probes exhibit important ( $2\sigma$  from the mean) activity. On top of it, if one of the groups is statistically significantly ( $4\sigma$  difference) more important than another, the region is colored with blue (polypredictive) or red (monopredictive) to show which of two kinds of neural specialization dominates this TF region in terms of importance. For example decoding of scenes (d) involves early theta activity of polypredictive (blue) probes, followed broadband gamma activity that is significantly important (gray), and is slightly dominated by monopredictive (red) probes, then followed by late alpha activity produced predominantly by monopredictive neural locations. (a) house, (b) face, (c) animal, (d) scene, (e) tool, (f) pseudoword, (g) characters, (f) scrambled.

compared to monopredictive probes, indicating that the most useful information for distinguishing faces from other visual categories is coded in that region of time-frequency space and is carried by polypredictive probes (Fig. 5b). Decoding of animals and tools relied on the activity patterns produced by monopredictive neural locations in late broadband gamma range ( $>300$  ms) and in even later (350–600 ms) alpha/beta range, with very little involvement of the activity of polypredictive probes. Scenes and houses also show strong feature importance in late alpha and beta band responses of monopredictive probes ( $4\sigma$  higher). Interestingly, for characters (Fig. 5g), feature importance in the early broadband gamma range was dominant for polypredictive probes ( $4\sigma$  higher than monopredictive), while the opposite was true for the pseudowords (Fig. 5f) – the late broadband gamma revealed to be dominant for monopredictive probes, also note the difference in the anatomical locations that were the most useful for the decoding of the pseudowords compared to the locations that were useful for decoding characters. Pseudowords also elicited a significantly stronger TF feature importance in monopredictive probes in late (350–750 ms) low-frequency (4–12 Hz) range, similar to animal and tool stimulus categories. Finally, an interesting observation was that animals and faces share most of their polypredictive probes (51%) indicating a large overlap of categorization networks of these two categories.

**Further decomposition of important activity reveals clusters of distinct time-frequency patterns.** We ran clustering analysis of the probes predictive of a category based on their activity to see which probes in the category-network behave in a similar way. Left column of Fig. 6 shows an averaged feature importance map for a given category. We look into the regions of the time-frequency map that are indicated as important by the feature importance map, extract baseline-normalized activity in those regions and cluster the probes according to that activity using hierarchical complete linkage clustering with cosine distance (see the Methods



**Figure 6.** Detailed analysis of spectral activity of (a) animals, (b) faces, (c) pseudowords and (d) scrambled images. For this figure we have selected the four categories that contained the most interesting observations, same visualization of activity importance, cluster assignments and the anatomical locations of those clusters for other four categories are available in the Supplementary Materials repository. Leftmost column contains the importance maps extracted from Random Forest models and shows where in time and frequency the important activity is. Second column visualizes the four largest (by the number of recording sites) clusters of activity patterns inside those spectrotemporal regions that are deemed important. The numbers in the top right corner of each cluster's activity pattern show the average predictive power ( $F_1$  score) of the probes in that cluster and proportion of polypredictive locations that exhibited this particular pattern of activity. Note how every cluster has a designated color: green, blue, red or black. This color of the cluster matches the color of MNI location markers in the last two columns, that show sagittal and dorsal views of the brain. White markers show the probes that have predictive power, but their activity pattern does not belong to any of the four major clusters.

section on hierarchical clustering for details). The second column of Fig. 6 shows the activity of four most populated clusters for each category. Each cluster represents the activity pattern exhibited by the probes in that cluster. Only the probes whose activity had predictive power ( $F_1 > 0.39$ ) are included in this analysis. As the final step we identified the anatomical locations of the probes from each cluster to see whether difference in the activity patterns could be attributed to the functional regions of the brain. The visualization of this step in the last two columns of Fig. 6.

This analysis allowed us make a number of *global* and *category-specific* observations. The set of visual categories presented in our data is diverse enough to consider category-specific findings to be general and emerge under any comparable set of visual stimuli.

The first global observation was that it is not only broadband gamma activity that is useful for the decoder's (Random Forest) performance, but low-frequency activity also contributed significantly (41% of predictive probes exhibited only low-frequency activity in the regions of importance), sometimes overshadowing the activity of higher frequency bands altogether (for `face` and `scrambled` stimuli low frequency activity was significantly more important than broadband gamma activity, Mann–Whitney U test  $p < 1e - 7$ , corrected). Most clusters were composed of a combination of low and high-frequency components (Fig. 6, second column) and were mostly (87%) located in occipito-temporal cortices, though some electrodes in parietal and frontal cortex (7%) also appeared to contribute with predictive responses in the decoding process (Fig. 6, two right columns), especially for such stimulus categories as `animal` and `pseudoword`.

The second observation spanning across all categories was that the classifier used not only the increases in power to perform the classification, but also relied on power decreases in different brain networks (7 out of 32 dominant activity clusters consisted solely from the activity patterns characterized by power decrease). The most prominent examples are the clusters `faces-2` (Fig. 6b), `animals-2` (Fig. 6a), `tools-2`, `pseudowords-1`, `pseudowords-2` (Fig. 6c), `scrambled-1` and `scrambled-2` (Fig. 6d). For example, to decode `face` or `pseudowords` from the activity of the blue cluster network, the RF classifier used broadband gamma power decreases located in posterior inferior temporal cortex and inferior occipital gyrus. None of the probes for which the decrease in activity was identified as important for decoding were located in classically defined Default Mode Network<sup>50,51</sup>.

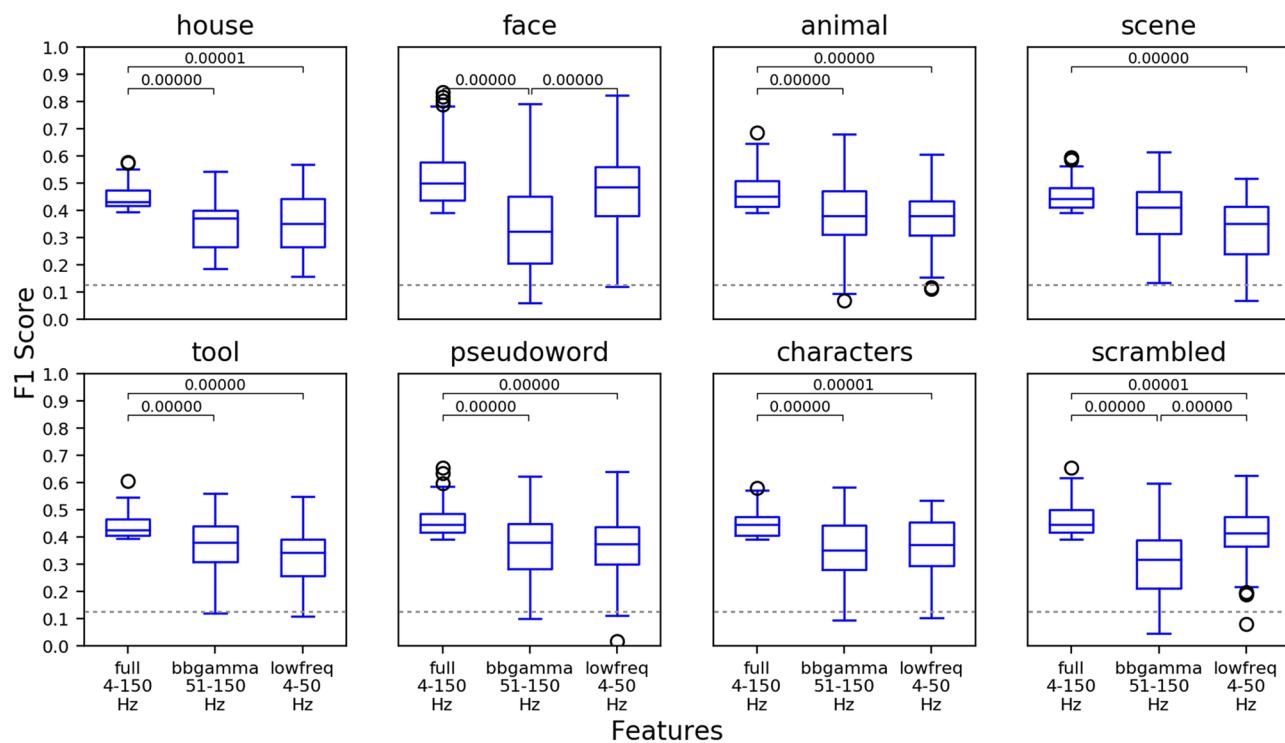
Across all categories, the earliest component that often appeared in clusters was the brief power increase (mean non-zero power increase was 2.8 times the baseline in the region of interest) in the low-frequency interval (4–25 Hz), which for one group of probes can be associated to an almost instantaneous broadband gamma power increase (6b, cluster 3, mean broadband gamma increase of 1.9 times the baseline), but remains the only source of important activity for another group of probes (6b, cluster 1).

Studying the anatomical locations of the probes belonging to different clusters of activity revealed interesting observations. Figure 6c, `pseudowords`, clusters 1 and 3 show a clear example how clustering by activity patterns leads to assigning the probes into functionally different anatomical areas. The gamma-band increase signature captured by cluster 3 occurs only in the left hemisphere (red markers on Fig. 6c), the late theta-alpha power decrease captured by cluster 1 also occurs only in the left hemisphere (green markers) and is spatially clearly distinct from probes in cluster 3. Because it is known that `pseudoword` stimuli elicit top-down language-related (orthographic, phonological and semantic) analysis, which elicits highly left-lateralized networks identifiable in iEEG recordings<sup>52,53</sup>, we know that this observation reflects a functional brain process. This dissociation in both the spectrotemporal and anatomical domains provides us with valuable data on the locations and associated activity patterns emerging during automatic perceptual categorization and highlights the benefit of disentangling the activity into functionally and anatomically disassociated clusters.

Finally, the relevance of the different components in the TF domain for the Random Forest classification process was assessed. Specifically, we tested whether the activity in the broadband gamma range, commonly present on most clusters across categories, is in general the most valuable neural signature for category networks as compared to the low-frequency parts of the spectrum. To test whether broadband gamma was solely the most informative frequency interval we statistically compared predictive power of three intervals: broadband gamma (50–150 Hz), low-frequency (4–50 Hz) and full spectrum (4–150 Hz). Overall, across 7 perceptual categories out of 8 (except for `scenes`), using the full spectrum was more informative than using the broadband gamma interval or the low-frequency interval alone (Mann–Whitney U test,  $p < 0.001563$ , corrected to the number of clusters compared, see Fig. 7), which is in line with the results reported by<sup>54</sup>. Importantly, for `scrambled` images and `faces` the broadband gamma carried less (Mann–Whitney U  $p < 1e - 7$ , corrected) decoding-relevant information than the lower frequencies.

## Discussion

In the present work we explored the bottom-up approach to the analysis of human intracerebral neural activity. Due to a rich dataset and powerful methodology we were able to uncover facts about neural processing of automatic visual categorization that we would not necessarily address in a hypothesis-driven study. We trained a machine learning model to decode 8 different perceptual categories from human intracerebral neural spectral activity and analyzed the resulting statistical model to expose what information does the model rely on in order to make decoding decisions. We then drew the parallels between the patterns of neural activity that the machine learning algorithm deemed important and the functional role of that activity for the task of decoding perceptual categories in human brain. This allowed us to distinguish between the spectral activity that is relevant for the task from the activity that is not. The study was conducted using a large dataset consisting of 4528400 local field potential recordings, from a cohort of 100 human patients whose combined electrode implantations spanned across all major brain cortical areas. The decoding process confronted neural responses elicited by 8 different perceptual categories (from 8 different stimulus sets), which allowed obtaining a high resolution in the dissociation between neural response patterns across categories. All classifications were operated on a broad frequency spectrum ranging from 4 to 150 Hz and allowed to distinguish degrees of selectivity of neural responses and which spectral components most strongly enable this selectivity. Previous works have shown where and when perceptual category information can be decoded from the human brain, our study adds to that line of research by identifying spectrotemporal patterns that contribute to category decoding without the need to formulate a priori hypothesis on which spectral components and at which times are worth investigating.



**Figure 7.** Comparison of predictive power of the electrodes from three different sets of features: full spectrum (4–150 Hz), broadband gamma alone (50–150 Hz) and lower frequencies alone (4–50 Hz) across categories. The bracket with the p-value indicates a significant difference according to Mann–Whitney U test. The dashed line indicates change classification level for an 8-class problem.

The classifier model first allowed us to globally identify two types of neural responses: those that were predictive of a certain category and those that did not predict any category despite eliciting strong amplitude modulation across multiple frequency bands. Surprisingly, when comparing the level of predictability of probe responses we found that only 4.8% of the responsive probes were predictive of a category. This very low percentage highlights an important fact regarding the level of “selectivity” of a neural responses. In this decoding approach, the level of single-probe neural response selectivity depends on the diversity and overall quantity of the comparison/reference group to which it is compared to. Stimulus-induced neural signal selectivity is thus a graded quality that can be assessed through multiple comparisons with a broad variety of stimulation conditions. This result also implies that although any stimulus can elicit a local neural response throughout the cerebral cortex, in the light of our results, there is a high probability of it being non-predictive of any of the categories or being polypredictive of several categories at once.

In line with a vast literature on the localization of category related networks<sup>1–7,12–14</sup> predictive probes concentrated mostly in the inferior temporal cortex, namely the fusiform gyrus (BA 37), yet surprisingly for some categories, probes in primary visual cortex were also predictive of these categories. This effect is probably related to the specifics of the physical content of certain images that uniquely characterize certain categories amongst all others, as for example the content in high-contrast edge information in **scrambled** and written text stimuli. The existence of polypredictive neural locations could be an important finding in the context of the debate concerning the functional organization of the inferior temporal cortex. Specifically, this finding contradicts a modular organisation in which each region is involved in processing a single category<sup>55</sup>, and instead favours a more distributed, overlapping representation of categories across the IT<sup>56</sup>.

Predictive probes were subsequently classified according to their level of selectivity towards a single or multiple visual categories. Polypredictive probes (36%) clustered in visual cortices and inferior temporal cortex and were associated with early spectral components (<300 ms) such as broadband gamma power increases and a transient theta burst shortly after stimulus presentation. Monopredictive probes (64%) were abundant in these same regions, but extending uniquely in frontal, parietal, superior temporal and anterior limbic cortex. Their activity was strongly associated with the later (>300 ms) time and with power suppression of spectral importance features, versus baseline, in the theta (4–7 Hz), alpha (8–15 Hz) and beta bands (16–40 Hz). In a subgroup of probes the associated power suppression of the feature importances extended into the broad gamma band (50–150 Hz).

Importantly, the capacity to ascribe category selectivity to predictive probes (mono vs polypredictive probes) arises from the fact that the decoding model was trained to discriminate between all 8 categories simultaneously. The separation between mono and polypredictive probes revealed specific effects in terms of network localization and time-frequency components. The high concentration of polypredictive probes (and local networks) in early visual cortices, from primary visual cortex up to inferior temporal cortex is coherent with the idea that networks

in the ventral visual stream progressively integrate more complex features into object representations, thus becoming progressively more selective, and converge within median temporal lobe to more stimulus-invariant representations<sup>15</sup>. This progressive information integration by spectral features of neuronal responses across the visual hierarchy has been recently connected with the computations carried out by deep convolutional neural networks trained to solve the task of visual recognition<sup>57</sup>.

Globally, the random forest data classification provided results that are coherent with current knowledge on 1) the implication of networks located in visual cortex and inferior temporal cortex in processing visual categories, 2) the timing of object categorization in the human brain and 3) the role of broadband gamma responses in processing category-selective information within these networks. Previous studies have shown that certain stimulus categories elicit clustered cortical responses of highly localized networks in the occipito-temporal ventral stream such as the fusiform-face-area (FFA) and the visual-word-form area (VWFA)<sup>17</sup>. Yet, other studies have broadened this scope by showing that certain categories, as for example faces, rely on the involvement of a larger brain-wide distributed network<sup>34,58</sup>. Our classification analysis shows that the spatial extent of this network distribution is category specific, certain stimuli eliciting larger network responses, such as for faces, animals and pseudowords, as compared to scenes, houses and scrambled images which concentrate in the fusiform cortex, the parahippocampal cortex and primary visual cortex respectively.

Our results largely agree with previous works trying to decode visual object categories over time with MEG<sup>59,60</sup> or intracranial recordings<sup>61</sup>. All these studies converge on the result that perceptual categories can be decoded from human brain signals as early as 100 ms. Our current work goes a step beyond these previous investigations by demonstrating which spectral components underlie this fast decoding. Previous intracranial studies have also shown that broadband gamma is modulated by information about object categories<sup>34,37,38</sup>. Moreover, broadband gamma has been suggested as a proxy to population spiking output activity<sup>30,62–64</sup>. It has since then been considered as a hallmark of local population processing<sup>31</sup>. Our classification results however show that broadband gamma is not the sole selectivity marker of functional neural processing, and that higher decoding accuracy can be achieved by including low-frequency components of the spectrum. For certain stimulus categories, as scrambled images, the broadband gamma range is even outperformed by the predictive power of the low-frequency range.

To understand which spectral components play a specific role in stimulus categorization we analyzed the decision process that drives the decoding model and identified the combined spectrotemporal regions that are informative for the output of the random forest classification procedure. This allowed us 1) to identify the category-selective spectral components of high importance for the automatic visual categorization process, and 2) identify the correlates functional involvement of positive as well as negative power modulations (increases and decreases versus baseline) in early and late time windows of neural processing involved in visual categorization.

While the distinctive activity of polypredictive neural locations is mostly reflected by early TF components (i.e. broadband gamma and theta burst in faces), the sustained decrease in power in the alpha/beta band was extended in space and time. This process is probably dependent on the degree of difficulty for the networks in reaching a perceptual decision and which appeals to the involvement of top-down processing required to resolve perceptual ambiguity elicited by the different stimulus categories. For example, animal and tool stimuli are highly diverse in their physical image structure, as compared to face stimuli. This affects the efficiency of bottom-up process in extracting category information, often associated with increase in gamma activity, and probably in parallel triggers top-down processes through selective activity modulation in low-frequency channels<sup>65</sup>. In our data, this latter phenomenon could be mirrored by a decrease of predictive power in the low-frequency range. Studies have shown that power modulations reflect changes in network connectivity<sup>66</sup> and that top-down processes, eliciting a decrease in power in the alpha-beta band, are accompanied by an increase in distant network connectivity<sup>67</sup>.

Finally, we also show that certain probes elicit decreased broadband gamma responses (versus baseline) while representing a significant feature importance for the classification model. It has been shown that neural activity in the Default Mode Network can be negatively modulated by attending sensory stimulation<sup>50</sup>, and intracranial studies have found that this was reflected by decreases (versus baseline) in the broad gamma range<sup>68–70</sup>. Here we found no evidence of such power decreases in probes located in the DMN<sup>50</sup>. However, the random forest classifier singled-out broad spectral patterns of power decreases at probes located in visual regions and beyond for categories faces, pseudowords and characters. This is the first time, to our knowledge, that power decreases in the broadband gamma range outside the DMN have been associated with highly functional neural signal classification of perceptual categories. Their functional significance should be studied in the future as they could reflect an important phenomenon of communication regulation between networks during perceptual decision making of visual categories.

In this work we studied the information that allows local cortical populations of neurons to make predictions of the category of a visual stimulus. Expanding on this work by including more subject data in the future might allow us to make a transition from the observations of local activity and the analysis of its role to being able to detect signatures of global decision-making processes. It is possible that these signatures would be reflected in specific spectral fingerprints as many classic theories would suggest<sup>22,71–73</sup>. The methodology proposed in this study can facilitate the search of those fingerprints without the need to formulate a priori hypothesis about which spectrotemporal components are worth investigating.

## Methods

**Patients and recordings.** 100 patients of either gender (49 females, 51 males) of the mean age of  $33.18 \pm 10.13$  years with drug-resistant partial epilepsy and candidates for surgery were considered in this study and recruited from Neurological Hospitals in Grenoble and Lyon (France). All patients were stereotactically implanted with multi-lead EEG depth electrodes (DIXI Medical, Besançon, France). All participants provided

written informed consent, and the experimental procedures were approved by local ethical committee of Grenoble hospital (CPP Sud-Est V 09-CHU-12). All methods were performed in accordance with the guidelines and regulations of the ethical committee of Grenoble hospital. Recording sites were selected solely according to clinical indications, with no reference to the current experiment. All patients had normal or corrected to normal vision.

**Electrode implantation.** 11 to 15 semi-rigid electrodes were implanted per patient. Each electrode had a diameter of 0.8 mm and was comprised of 10 or 15 contacts of 2 mm length, depending on the target region, 1.5 mm apart. The coordinates of each electrode contact with their stereotactic scheme were used to anatomically localize the contacts using the proportional atlas of Talairach and Tournoux<sup>74</sup>, after a linear scale adjustment to correct size differences between the patient's brain and the Talairach model. These locations were further confirmed by overlaying a post-implantation MRI scan (showing contact sites) with a pre-implantation structural MRI with VOXIM® (IVS Solutions, Chemnitz, Germany), allowing direct visualization of contact sites relative to brain anatomy.

All patients voluntarily participated in a series of short experiments to identify local functional responses at the recorded sites<sup>34</sup>. The results presented here were obtained from a test exploring visual recognition. All data were recorded using approximately 120 implanted depth electrode contacts per patient using SD LTM Express, Micromed system for signal acquisition with a sampling rate of 512 Hz, high-pass filter 0.15 Hz, low-pass filter 500 Hz. Data were obtained from a total of 11321 recording sites.

**Stimuli and task.** The visual recognition task lasted for about 15 minutes. Patients were instructed to press a button each time a picture of a fruit appeared on screen (visual oddball paradigm). Only non-target stimuli were used in this work, the target *fruit* category was excluded from the analysis. Non-target stimuli consisted of pictures of objects of eight possible categories: houses, faces, animals, scenes, tools, pseudo words, consonant strings, and scrambled images. Natural images for the appropriate categories were obtained from the public domain on the Internet. Pseudoword and character images were generated by inserting a desired string onto a white background. All the included stimuli had the same average luminance. All categories were presented within an oval aperture of  $2 \times 3$  in visual angle (illustrated on Fig. 1a) at a distance of 70–90 cm using NeuroBehavioral Systems (NBS) Presentation® software. Stimuli were presented for a duration of 200 ms every 1000–1200 ms in series of 5 pictures interleaved by 3 second pause periods during which patients could freely blink. Patients reported the detection of a target through a right-hand button press and were given feedback of their performance after each report. A 2 second delay was placed after each button press before presenting the follow-up stimulus in order to avoid mixing signals related to motor action with signals from stimulus presentation. Altogether, responses to 400 unique natural images were measured per subject, 50 from each category.

**Processing of neural data.** The analyzed dataset consisted of 4528400 local field potential (LFP) recordings – responses from 11321 recording sites to 400 stimuli. To remove the artifacts the signals were linearly detrended and the recordings that contained values  $\geq 10\sigma_{images}$ , where  $\sigma_{images}$  is the standard deviation of voltage values (in the time window from -500 ms to 1000 ms) of that particular probe over all stimuli, were excluded from data. All electrodes were re-referenced to a bipolar reference and the reference electrodes were excluded from the analysis. The signal was segmented in the range from -500 ms to 1000 ms, where 0 marks the moment when the stimulus was shown. The -500 to -100 ms time window served as a baseline.

To quantify the power modulation of the signals across time and frequency we used standard time-frequency (TF) wavelet decomposition<sup>75</sup>. The signal  $s(t)$  was convoluted with a complex Morlet wavelet  $w(t, f_0)$ , which has Gaussian shape in time ( $\sigma_t$ ) and frequency ( $\sigma_f$ ) around a central frequency  $f_0$  and defined by  $\sigma_f = 1/2\pi\sigma_t$  and a normalization factor. To achieve good time and frequency resolution over all frequencies we slowly increased the number of wavelet cycles with frequency,  $\frac{f_0}{\sigma_f}$  was set to: 6 for high (61–150 Hz) and low (31–60 Hz) gamma, 5 for beta (15–30 Hz), 4 for alpha (9–14 Hz) and 3 for theta (4–8 Hz) frequency ranges. This method allowed to obtain better frequency resolution than applying a constant cycle length<sup>76</sup>. The square norm of the convolution results in a time-varying representation of spectral power, given by:  $P(t, f_0) = |w(t, f_0) \cdot s(t)|^2$ . Baseline normalization was performed by dividing the average power after stimulus onset (0 to 1000 ms) in each frequency by the average power of that frequency in the baseline window (-500 to -100 ms). Each LFP recording was transformed from 768 data points (1.5 seconds of voltage readings at 512 Hz sampling rate) into a matrix of size  $146 \times 48$  where each row represents a 1 Hz frequency band from 4 Hz to 150 Hz and columns represent 31.25 ms time bins. Value in each cell of that matrix is the power of that specific frequency averaged over 16 time points.

Further analysis was done only on the electrodes that were responsive to the visual task. In each frequency band we compared each electrode's average post-stimulus band power to the average baseline power with a Wilcoxon signed-rank test for matched-pairs. Only the probes that showed a post-stimulus response that is statistically significantly ( $p$ -value  $\leq 0.005$ , corrected for multiple comparisons with the false discovery rate (FDR) procedure<sup>77</sup>) different from the baseline response in at least two frequencies were preserved for future analysis. Please note that eliciting a significant response in at least 2 out of 146 frequencies is a relaxed requirement. The use of such a relaxed criterion allowed us to include into analysis not only the areas that had a strong response in the visual areas, but also the responses from other brain areas that might reflect downstream processes related to automatic perceptual categorization. This was possible due to the fact that the proposed method, given sufficiently large dataset, will not be hindered by the additional volume of irrelevant data and is able to detect narrow phenomena even in the large corpus of data.

To anatomically localize the source of each signal in subject's brain each electrode's MNI coordinates were mapped to a corresponding Brodmann brain area<sup>78</sup> using Brodmann area atlas from MRICron<sup>79</sup> software.

To confirm that probe's predictiveness of a certain category implies that the probe belongs to the network selective of that category we ran a set of experiments on three well-known functional areas: Fusiform Face Area (FFA)<sup>1</sup>, Visual Word Form Area (VWFA)<sup>7</sup> and Parahippocampal Place Area (PPA). Following Montreal Neurological Institute (MNI) coordinates of FFA reported in<sup>80</sup> and<sup>81</sup> we defined FFA bounding box as  $x \in [-44, -38], y \in [-61, -50], z \in [-24, -15]$  in the left hemisphere and  $x \in [36, 43], y \in [-55, -49], z \in [-25, -13]$  in the right hemisphere. Based on the Table 1 from<sup>82</sup> we defined VWFA area as MNI bounding box  $x \in [-50, -38], y \in [-61, -50], z \in [-30, -16]$  in the left hemisphere. From MNI coordinates reported in<sup>83</sup> and<sup>84,85</sup> we defined PPA bounding box to be  $x \in [-31, -22], y \in [-55, -49], z \in [-12, -6]$  in the left hemisphere and  $x \in [24, 32], y \in [-54, -45], z \in [-12, -6]$  in the right hemisphere.

**Random Forest as a decoding model.** A Random Forest<sup>86</sup> is a collection of decision trees, where each tree gets to operate on a subset of features. Each tree is assigned a random set of features and it has to find the decision boundaries on those features that lead to best classification performance. At each branching point the algorithm must decide using which feature will be most efficient in terms of reducing the entropy of class assignments to the data points under current branch of the decision tree. To achieve that, the feature that is most useful will be selected first and will be responsible for largest information gain. For example, if the activity of a probe at 52 Hz at 340 ms is high when a subject is presented with a face and low for all other categories, decision tree will use that fact and rely on the “52 Hz at 340 ms” feature, thus assigning it some importance. How high the importance of a feature is will depend on how well does this feature distinguish faces from all other categories. As Random Forest is a collection of trees and the same feature will end up being included into several different trees, being important in many trees contributes to the overall importance of a feature (for the exact computation see the section on feature importance below).

We treated each electrode's responses as a separate dataset consisting of 400 data points (one per stimulus image), and 7008 features – time-frequency transformation of LFP response into 146 frequencies and 48 time bins. For each electrode we trained a Random Forest with 3000 trees and used 5-fold cross-validation to measure the predictive power of the neural activity recorded by each of the electrodes. Per-class  $F_1$  score, a harmonic mean of precision and recall of a statistical model, provides us with a metric of success of the classification. The parameters were selected by performing informal parameter search. Random Forest was the algorithm of choice in our analysis due to its previous application to spectrotemporal features<sup>87</sup> and, more importantly, due to the interpretability of the resulting models. Highly interpretable structure of the resulting decision trees allowed us to follow the decision-making process that led each particular model to a decoding decision, and to estimate each feature's contribution to this process, allowing to identify and characterize category-specific spectral signatures. We used scikit-learn<sup>88</sup> implementation of the above-mentioned methods with default parameters unless indicated otherwise.

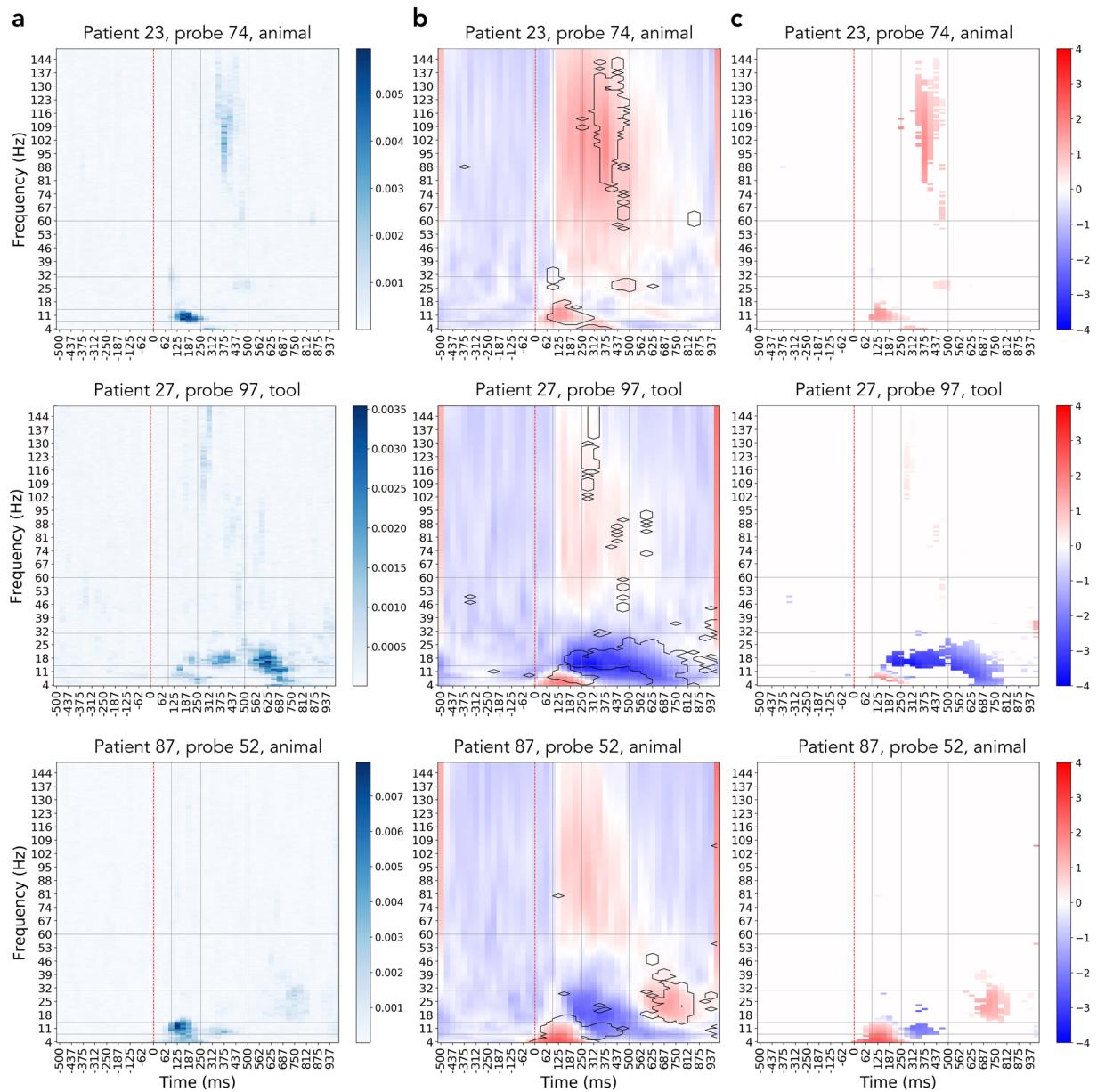
As the first step of the decoding analysis we estimated which of 11321 electrodes have predictive power. For that we split each electrode's 400-sample dataset into 320 samples for training and 80 for prediction estimation. Repeating this procedure 5 times provided us with 400 predictions that we could compare to the true categories. By running a permutation test 100000 times on electrodes with randomly permuted class labels we estimated that 99.999th percentile (equivalent to significance threshold of  $p \leq 0.00001$ ) of  $F_1$  score is 0.390278.  $F_1$  score is an aggregated metric of the performance of a classifier that combines both the *precision* (the ratio of the data samples that truly belong to a category among the ones that were assigned to that category by the model) and *recall* (the ratio of data samples that were correctly identified to belong to a category to the total number of samples of that category in the dataset) into one number:  $F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ . In total 787 electrodes had a predictive power of  $F_1 > 0.390278$  in at least one of the categories. Although none of the results of this work depend on direct comparison of  $F_1$  scores, we have estimated the noise level of this metric by computing standard deviation over 5 folds of each electrode and calculated the average standard deviation of  $F_1$  metric to be  $\sigma = 0.1008$  over all predictive electrodes. For each of those electrodes a Random Forest model was retrained once more on whole data (400 samples instead of 320) and that model was used for calculating feature importances and, ultimately, for understanding which parts of the recorded activity were relevant for visual object recognition in human brain.

**Feature importance for the analysis of task-relevant neural activity.** During the process of constructing the decision trees, Random Forest relies on some features more than on the others. We chose *Gini impurity*<sup>89</sup> as a measure of which features should be used to make the branching decisions in the nodes of a tree. This score, along with the number of times each particular feature was used across trees, informed us on the relative importance of each particular feature with respect to other features. Gini impurity  $G$  is calculated as

$$G = \sum_{i=1}^{n_c} p_i(1 - p_i), \quad (1)$$

where  $n_c$  is the number of categories and  $p_i$  is the proportion of class  $i$  in a node. To pick a feature for a parent node, Gini impurity of both child nodes of that parent are calculated and used to estimate the *reduction in impurity* that will be achieved by picking that particular features as the branching factor for the node. The feature that decreases impurity the most is selected to be the branching factor of that parent node. The reduction in impurity is calculated as

$$I = G_{\text{parent}} - G_{\text{left child}} - G_{\text{right child}} \quad (2)$$



**Figure 8.** Using the importance map to filter out irrelevant activity. The three rows show three different examples of how filtering of the activity by importance is beneficial: in patient 23, probe 74 we see that only later portion of the broadband gamma activity increase was useful for identifying this activity as a response to the animal stimulus; patient 27, probe 97 shows that although there is an increase in broadband activity, the actually useful information contained in decrease in the lower frequency bands; patient 87, probe 52 demonstrates that for decoding this particular probe's activity one must focus on the activity in lower frequencies at specific time and, despite prominent presence, ignore the increase in broadband gamma. **(a)** Probe's importance map, color codes the relative importance of each spectrotemporal feature within the map. **(b)** Full spectrotemporal activity of the probe, features with importances one standard deviation higher than the average (in contour) mark the regions of activity that were useful for the decoding model. Color shows the activity in the power spectrum normalized by the baseline. **(c)** Activity of the probes filtered by the importance mask, only the relevant activity is preserved. Color shows the activity in the power spectrum normalized by the baseline.

and is called *node importance*. Feature importance of a feature  $f$  is estimated by calculating the sum of Gini impurity reductions over all samples in the dataset that were achieved with the use of a particular feature  $f$  and normalizing it by the total number of samples. Figure 1e is a visual representation of relative feature importance, color intensity shows the importance of each of 7008 (146 frequencies  $\times$  48 time bins) spectrotemporal features from one probe. In total our analysis has produced 787  $\times$  8 such images – one for each probe-class pair.

The importance map computed as depicted on Fig. 1 is a global map for all 8 categories, such that the regions that are highlighted on the map are important for distinguishing between all 8 categories. There is, however, a

way to look at category-specific importances as well. The final set of nodes of a decision tree, called *leaves*, are the end-points of the classification process and each leaf is associated with a certain category. If we take one TF activity map (TF components are the features) and start traversing a decision tree following the rules set by the nodes of the tree, we will end up in a certain leaf. That leaf will be associated with a certain category, for example, with faces. The fact that we followed the rules and ended up in that leaf indicates that the TF map we used as the input to the tree probably comes from a trial where a face stimulus was shown to the subject. In order to get category-specific feature importance map we took all the leaves associated with a category, traversed the tree backwards and tracked all the features that were used on the path from the leaf to the root of the tree. This way we got a list of features (TF components) that were used to identify a neural response as belonging to a certain category. Random Forest feature importance allowed us to identify which sub-regions of neural activity (TF maps) are relevant for decoding. It showed that only a small portion of activity is actually crucial for identifying the categories.

To compare importance maps between each other we fit a normal distribution on the difference between two maps and considered statistically significant the differences that are bigger than  $\mu + 4\sigma$ . One spectrotemporal importance map consists of 7008 values. To filter out false positives we stipulated that only 1 false positive out of 7008 pixels can be tolerated and tuned the threshold accordingly. That requirement resulted in the p-value of 0.0001427 and confidence level of 99.99%, corresponding to  $3.89\sigma$ , which we rounded up to  $\sigma = 4.0$ .

**Hierarchical clustering to reveal types of activity patterns.** To further analyze the spectrotemporal signatures elicited by different visual categories in different parts of human brain we clustered filtered activity patterns and identified the most prominent groups. The result of this analysis is shown in the second column of Fig. 6. For each category, the four most populated (in terms of the number of probes) clusters of activity patterns elicited by this category are shown.

To do the clustering we first took each probe's category-specific activity separately by averaging probe's responses to 50 images of each particular category in time-frequency domain. We then masked the activity with the category importance map (as shown on Fig. 8), leaving only those features out of  $146 \times 48$  that have importance score larger than  $\mu + \sigma$ , where  $\mu$  is the average importance score for that category and  $\sigma$  is one standard deviation of the score distribution.

Masked activity patterns were hierarchically clustered using Eq. 3 to calculate the distance between a pair of clusters  $U$  and  $V$  as the maximal cosine distance between all of the clusters' member observations (complete linkage clustering):

$$d(U, V) = \max \left( \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \right) \quad \forall \mathbf{u} \in U, \forall \mathbf{v} \in V \quad (3)$$

SciPy<sup>90</sup> implementation of the hierarchical clustering methods was used in this work. Resulting clustering assignments were visually inspected and corrected.

## Data and code availability

The final pre-processed spectrotemporal data are available for download under Academic Free License 3.0 from <https://web.gin.g-node.org/ilyakuzovkin/Spectral-Signatures-of-Perceptual-Categorization-in-Human-Cortex>. The code that was used to produce this data from the raw recordings and to perform all of the subsequent analysis steps is available at <https://github.com/kuz/Spectral-signatures-of-perceptual-categorization-in-human-cortex>. All raw human brain recordings that support the findings of this study are available from Lyon Neuroscience Research Center but restrictions apply to the availability of these data, which were used under license for the current study, and are not publicly available. Raw data are however available from the authors upon reasonable request and with permission of Lyon Neuroscience Research Center.

Received: 18 June 2019; Accepted: 11 March 2020;

Published online: 12 May 2020

## References

1. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
2. Epstein, R., Harris, A., Stanley, D. & Kanwisher, N. The parahippocampal place area: Recognition, navigation, or encoding? *Neuron* **23**, 115–125 (1999).
3. Peelen, M. V., Fei-Fei, L. & Kastner, S. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* **460**, 94 (2009).
4. Malach, R. *et al.* Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl Acad. Sci.* **92**, 8135–8139 (1995).
5. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
6. Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L. & Haxby, J. V. Distributed representation of objects in the human ventral visual pathway. *Proc. Natl Acad. Sci.* **96**, 9379–9384 (1999).
7. Cohen, L. *et al.* The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307 (2000).
8. Potter, M. C. & Faulconer, B. A. Time to understand pictures and words. *Nature* **253**, 437 (1975).
9. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *nature* **381**, 520 (1996).
10. Li, F. F., VanRullen, R., Koch, C. & Perona, P. Rapid natural scene categorization in the near absence of attention. *Proc. Natl Acad. Sci.* **99**, 9596–9601 (2002).

11. Rousselet, G. A., Fabre-Thorpe, M. & Thorpe, S. J. Parallel processing in high-level categorization of natural images. *Nat. Neurosci.* **5**, 629 (2002).
12. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536 (2014).
13. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
14. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
15. Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102 (2005).
16. Kreiman, G., Koch, C. & Fried, I. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci.* **3**, 946 (2000).
17. Haynes, J.-D. & Rees, G. Neuroimaging: decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523 (2006).
18. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl Acad. Sci.* **103**, 3863–3868 (2006).
19. Kamitani, Y. & Tong, F. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* **16**, 1096–1102 (2006).
20. Fries, P. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
21. Buzsaki, G. *Rhythms of the Brain* (Oxford University Press, 2006).
22. Siegel, M., Donner, T. H. & Engel, A. K. Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* **13**, 121 (2012).
23. Michalareas, G. *et al.* Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* **89**, 384–397 (2016).
24. Vidal, J. R., Chaumon, M., O'Regan, J. K. & Tallon-Baudry, C. Visual grouping and the focusing of attention induce gamma-band oscillations at different frequencies in human magnetoencephalogram signals. *J. Cognit. Neurosci.* **18**, 1850–1862 (2006).
25. Wyart, V. & Tallon-Baudry, C. Neural dissociation between visual awareness and spatial attention. *J. Neurosci.* **28**, 2667–2679 (2008).
26. Jensen, O. & Mazaheri, A. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* **4**, 186 (2010).
27. VanRullen, R. Perceptual cycles. *Trends Cognit. Sci.* **20**, 723–735 (2016).
28. Engel, A. K. & Fries, P. Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* **20**, 156–165 (2010).
29. Dalal, S. S. *et al.* Spanning the rich spectrum of the human brain: slow waves to gamma and beyond. *Brain Structure Funct.* **216**, 77 (2011).
30. Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E. & Crone, N. E. High-frequency neural activity and human cognition: past, present and possible future of intracranial eeg research. *Prog. Neurobiol.* **98**, 279–301 (2012).
31. Parvizi, J. & Kastner, S. Promises and limitations of human intracranial electroencephalography. *Nature neuroscience* **1** (2018).
32. Fox, K. C., Foster, B. L., Kucyi, A., Daitch, A. L. & Parvizi, J. Intracranial electrophysiology of the human default network. *Trends in cognitive sciences* (2018).
33. McCarthy, G., Puce, A., Gore, J. C. & Allison, T. Face-specific processing in the human fusiform gyrus. *J. Cognit. Neurosci.* **9**, 605–610 (1997).
34. Vidal, J. R. *et al.* Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and erp response selectivity. *Front. Hum. Neurosci.* **4**, 195 (2010).
35. Davidesco, I. *et al.* Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. *J. Neurosci.* **33**, 1228–1240 (2013).
36. Hamamé, C. M. *et al.* Functional selectivity in the human occipitotemporal cortex during natural vision: Evidence from combined intracranial eeg and eye-tracking. *NeuroImage* **95**, 276–286 (2014).
37. Privman, E. *et al.* Enhanced category tuning revealed by intracranial electroencephalograms in high-order human visual areas. *J. Neurosci.* **27**, 6234–6242 (2007).
38. Fisch, L. *et al.* Neural “ignition”: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* **64**, 562–574 (2009).
39. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598 (1998).
40. Parvizi, J. *et al.* Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14915–14920 (2012).
41. Ghuman, A. S. *et al.* Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* **5**, 5672 (2014).
42. Kadipasaoglu, C. M., Conner, C. R., Whaley, M. L., Baboyan, V. G. & Tandon, N. Category-selectivity in human visual cortex follows cortical topology: a grouped iceeg study. *PLoS one* **11**, e0157109 (2016).
43. Jonas, J. *et al.* A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc. Natl Acad. Sci.* **113**, E4088–E4097 (2016).
44. Jonas, J. *et al.* Beyond the core face-processing network: Intracerebral stimulation of a face-selective area in the right anterior fusiform gyrus elicits transient prosopagnosia. *Cortex* **72**, 140–155 (2015).
45. Lochy, A., Van Reybroeck, M. & Rossion, B. Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proc. Natl Acad. Sci.* **113**, 8544–8549 (2016).
46. Hirshorn, E. A. *et al.* Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl Acad. Sci.* **113**, 8162–8167 (2016).
47. Aguirre, G. K., Zarahn, E. & D'Esposito, M. An area within human ventral cortex sensitive to “building” stimuli: evidence and implications. *Neuron* **21**, 373–383 (1998).
48. Mégevand, P. *et al.* Seeing scenes: topographic visual hallucinations evoked by direct electrical stimulation of the parahippocampal place area. *J. Neurosci.* **34**, 5399–5405 (2014).
49. Bastin, J. *et al.* Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages. *Hum. brain Mapp.* **34**, 1357–1370 (2013).
50. Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain's default network. *Ann. N. Y. Acad. Sci.* **1124**, 1–38 (2008).
51. Raichle, M. E. The brain's default mode network. *Annu. Rev. Neurosci.* **38**, 433–447 (2015).
52. Juphard, A. *et al.* Direct evidence for two different neural mechanisms for reading familiar and unfamiliar words: an intra-cerebral eeg study. *Front. Hum. Neurosci.* **5**, 101 (2011).
53. Mainy, N. *et al.* Cortical dynamics of word recognition. *Hum. brain Mapp.* **29**, 1215–1230 (2008).
54. Miller, K. J., Schalk, G., Hermes, D., Ojemann, J. G. & Rao, R. P. Spontaneous decoding of the timing and content of human object perception from cortical surface recordings reveals complementary information in the event-related potential and broadband spectral change. *PLoS computational Biol.* **12**, e1004660 (2016).
55. Kanwisher, N. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl Acad. Sci.* **107**, 11163–11170 (2010).
56. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
57. Kuzovkin, I. *et al.* Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Communications Biology* **1** (2018).

58. Ishai, A., Schmidt, C. F. & Boesiger, P. Face perception is mediated by a distributed cortical network. *Brain Res. Bull.* **67**, 87–93 (2005).
59. Carlson, T., Tovar, D. A., Alink, A. & Kriegeskorte, N. Representational dynamics of object vision: the first 1000 ms. *J. Vis.* **13**, 1–1 (2013).
60. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455 (2014).
61. Liu, H., Agam, Y., Madsen, J. R. & Kreiman, G. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* **62**, 281–290 (2009).
62. Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* **29**, 13613–13620 (2009).
63. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
64. Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J. & Hsiao, S. S. Neural correlates of high-gamma oscillations (60–200 Hz) in macaque local field potentials and their potential implications in electrocorticography. *J. Neurosci.* **28**, 11526–11536 (2008).
65. Bastos, A. M. *et al.* Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390–401 (2015).
66. Tewarie, P. *et al.* Relationships between neuronal oscillatory amplitude and dynamic functional connectivity. *Cerebral Cortex* (2018).
67. Gaillard, R. *et al.* Converging intracranial markers of conscious access. *PLoS Biol.* **7**, e1000061 (2009).
68. Ossandón, T. *et al.* Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance. *J. Neurosci.* **31**, 14521–14530 (2011).
69. Jerbi, K. *et al.* Exploring the electrophysiological correlates of the default-mode network with intracerebral eeg. *Front. Syst. Neurosci.* **4**, 27 (2010).
70. Dastjerdi, M. *et al.* Differential electrophysiological response during rest, self-referential, and non-self-referential tasks in human posteromedial cortex. *Proc. Natl Acad. Sci.* **108**, 3023–3028 (2011).
71. Rodriguez, E. *et al.* Perception's shadow: long-distance synchronization of human brain activity. *Nature* **397**, 430 (1999).
72. Varela, F., Lachaux, J.-P., Rodriguez, E. & Martinerie, J. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* **2**, 229 (2001).
73. Engel, A. K., Fries, P. & Singer, W. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* **2**, 704 (2001).
74. Talairach, J. & Tournoux, P. *Referentially oriented cerebral MRI anatomy: an atlas of stereotaxic anatomical correlations for gray and white matter* (Thieme, 1993).
75. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. theory* **36**, 961–1005 (1990).
76. Delorme, A. & Makeig, S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. Neurosci. methods* **134**, 9–21 (2004).
77. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).
78. Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde* (Barth, 1909).
79. Rorden, C. Micron [computer software] (2007).
80. Harris, R. J., Young, A. W. & Andrews, T. J. Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proc. Natl Acad. Sci.* **109**, 21164–21169 (2012).
81. Axelrod, V. & Yovel, G. Successful decoding of famous faces in the fusiform face area. *PLoS One* **10**, e0117126 (2015).
82. Price, C. J. & Devlin, J. T. The myth of the visual word form area. *Neuroimage* **19**, 473–481 (2003).
83. Bastin, J. *et al.* Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *J. Neurosci.* **33**, 10123–10131 (2013).
84. Park, S. & Chun, M. M. Different roles of the parahippocampal place area (ppa) and retrosplenial cortex (rsc) in panoramic scene perception. *Neuroimage* **47**, 1747–1756 (2009).
85. Hamamé, C. M. *et al.* Dejeire's reading area revisited with intracranial eeg selective responses to letter strings. *Neurology* **80**, 602–603 (2013).
86. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
87. Westner, B. U., Dalal, S. S., Hanslmayr, S. & Staudigl, T. Across-subjects classification of stimulus modality from human meg high frequency activity. *PLoS computational Biol.* **14**, e1005938 (2018).
88. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
89. Breiman, L. *Classification and regression trees* (Routledge, 2017).
90. Jones, E. *et al.* SciPy: Open source scientific tools for Python, [Online; accessed 2016-11-08] (2001).

## Acknowledgements

We would like to thank Benoît Chatard, Lorella Minotti and Mathilde Petton for their work on the intracranial dataset that is the foundation of this research project, and also for their help recording, organizing and curating the data. I.K. and R.V. thank the financial support from the Estonian Research Council through the personal research grants PUT438 and PUT1476. This work was supported by the Estonian Centre of Excellence in IT (EXCITE), funded by the European Regional Development Fund. J.A. was also supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 799411. J.P.L. and S.R. received funding from the “Programme Avenir Lyon Saint-Etienne” of Université de Lyon, within the program “Investissements d'Avenir”, ANR convention No. ANR-11-IDEX-0007; the LabEx CORTEX, ANR convention No. ANR-11-LABX-0042; and European Union Seventh Framework Programme (FP7/2007–2013) under Grant Agreement No. 604102 (Human Brain Project); iEEG-only recordings were made possible thanks to IHU CESAME, within the program “Investissements d'Avenir” (ANR-10-IBHU-0003). J.P.L. received funding from HBP Grant Agreements No. 785907 (SGA2) and No. 7202070 (SGA1). M.P.B. received funding from the Institut Universitaire de France.

## Author contributions

I.K., J.R.V. and R.V. conceived the idea of the research project and developed the methodology. J.-P.L. and S.R. curated the data. J.-P.L., S.R. and P.K. provided the study materials and other resources required to run the experiments. M.P.-B. and J.R.V. conducted the experiments and collected the data. I.K. performed the analysis of the data and implemented the methodology. I.K. and J.R.V. wrote the original draft of the manuscript. I.K. prepared the visualizations. I.K., J.R.V., J.A. and R.V. reviewed and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-64243-6>.

**Correspondence** and requests for materials should be addressed to I.K., J.R.V. or R.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020