PAPER • OPEN ACCESS

Real-time detection and discrimination of visual perception using electrocorticographic signals

To cite this article: C Kapeller *et al* 2018 *J. Neural Eng.* **15** 036001

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Real-time detection and discrimination of visual perception using electrocorticographic signals

C Kapeller[1,2], **H Ogawa**[3], **G Schalk**[6,8,9], **N Kunii**[4], **W G Coon**[5,6,7],
J Scharinger[2], **C Guger**[1,5] **and K Kamada** [3]

[1] Guger Technologies OG, Graz, Austria
[2] Department of Computational Perception, Johannes Kepler University, Linz, Austria
[3] Department of Neurosurgery, Asahikawa Medical University, Asahikawa, Japan
[4] Department of Neurosurgery, The University of Tokyo Hospital, Tokyo, Japan
[5] g.tec neurotechnology USA, Inc., Albany, NY, United States of America
[6] National Center for Adaptive Neurotechnologies, Wadsworth Center, New York State Department
of Health, Albany, NY, United States of America
[7] Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA,
United States of America
[8] Department of Biomedical Sciences, State University of New York at Albany, Albany, NY,
United States of America
[9] Department of Neurology, Albany Medical College, Albany, NY, United States of America

E-mail: kapeller@gtec.at

CrossMark

Abstract

Objective. Several neuroimaging studies have demonstrated that the ventral temporal cortex contains specialized regions that process visual stimuli. This study investigated the spatial and temporal dynamics of electrocorticographic (ECoG) responses to different types and colors of visual stimulation that were presented to four human participants, and demonstrated a real-time decoder that detects and discriminates responses to untrained natural images. *Approach.* ECoG signals from the participants were recorded while they were shown colored and greyscale versions of seven types of visual stimuli (images of faces, objects, bodies, line drawings, digits, and kanji and hiragana characters), resulting in 14 classes for discrimination (experiment I). Additionally, a real-time system asynchronously classified ECoG responses to faces, kanji and black screens presented via a monitor (experiment II), or to natural scenes (i.e. the face of an experimenter, natural images of faces and kanji, and a mirror) (experiment III). Outcome measures in all experiments included the discrimination performance across types based on broadband γ activity. *Main results.* Experiment I demonstrated an offline classification accuracy of 72.9% when discriminating among the seven types (without color separation). Further discrimination of grey versus colored images reached an accuracy of 67.1%. Discriminating all colors and types (14 classes) yielded an accuracy of 52.1%. In experiment II and III, the real-time decoder correctly detected 73.7% responses to face, kanji and black computer stimuli and 74.8% responses to presented natural scenes. *Significance.* Seven different types and their color information (either grey or color) could be detected and discriminated using broadband γ activity. Discrimination performance maximized for

combined spatial-temporal information. The discrimination of stimulus color information provided the first ECoG-based evidence for color-related population-level cortical broadband γ responses in humans. Stimulus categories can be detected by their ECoG responses in real time within 500 ms with respect to stimulus onset.

(Some figures may appear in colour only in the online journal)

1. Introduction

Real-time detection and discrimination of visual perception could lead to improved human-computer interfaces, and may also provide the foundations for new communication tools for people with serious neurological disorders such as amyotrophic lateral sclerosis (ALS).

Substantial research based primarily on functional magnetic resonance imaging (fMRI) has shown that categorization of visual perception is implemented by the brain across different regions on the ventral temporal cortex. Most notably, areas on or around the fusiform gyrus are well known to process face stimuli (Kanwisher *et al* 1997, Halgren *et al* 1999, Kadosh and Johnson 2007, Collins and Olson 2014), and can be used to discriminate visual stimuli of different categories (Grill-Spector and Weiner 2014). The left fusiform gyrus is known to process visually presented words (Cohen *et al* 2000, McCandliss *et al* 2003) and the inferior temporal gyrus has been shown to play an important role in recognition of numerals (Shum *et al* 2013).

The neural basis of face perception has also been investigated with electrocorticographic (ECoG) recordings. Initial work in this area investigated ECoG evoked responses to faces versus scrambled faces (Allison *et al* 1994), faces versus non-faces (Allison *et al* 1999), and more diverse stimuli including faces versus parts of faces versus scaled and rotated faces (McCarthy *et al* 1999) and faces versus bodies (Engell and McCarthy 2014a). In addition to investigating traditional evoked potentials, whose physiological origin is complex and unresolved (Makeig *et al* 2002, Mazaheri and Jensen 2006, 2008, Kam *et al* 2018), other studies have suggested that ECoG activity in the broadband γ (70–170 Hz) range is a general indicator of cortical population-level activity during auditory (Crone *et al* 2001, Edwards *et al* 2005, Potes *et al* 2012, 2014), language (Edwards *et al* 2009, Edwards *et al* 2010, Chang *et al* 2011, Pei *et al* 2011a, 2011b, Leuthardt *et al* 2012, Kubanek *et al* 2013), sensorimotor (Crone *et al* 1998, Miller *et al* 2007, Schalk *et al* 2007, Kubanek *et al* 2009, Wang *et al* 2012), attention (Ray *et al* 2008, Gunduz *et al* 2011, 2012), and memory (Jensen *et al* 2007, Sederberg *et al* 2007, Tort *et al* 2008, van Vugt *et al* 2010, Maris *et al* 2011) tasks. Physiologically, broadband γ has been shown to be a direct reflection of the average firing rate of neurons directly underneath the electrode (Miller *et al* 2009, Manning *et al* 2009, Whittingstall and Logothetis 2009, Ray and Maunsell 2011), and has been shown to drive the BOLD signal identified using fMRI (Logothetis *et al* 2001, Mukamel *et al* 2005, Niessing *et al* 2005, Engell *et al* 2012, Jacques *et al* 2016). Hence, more recent studies of visual perception investigated ECoG broadband responses to faces and other objects (Lachaux *et al* 2005, Tsuchiya *et al* 2008, Engell and McCarthy 2011, Engell and McCarthy 2014a, 2014b, Ghuman *et al* 2014), and used them to predict the N200 evoked response (Engell and McCarthy 2011), or to predict the onset and identity of visual stimuli (Miller *et al* 2016).

Different studies investigated the degree to which faces or other objects can be decoded from brain signals in individual trials. These offline studies reported detection performance of 85% for recognized faces on the ventral temporal cortex (VTC) (Tsuchiya *et al* 2008), 90.4% for faces and objects (Gerber *et al* 2017), 96% for faces and houses (Miller *et al* 2016), and about 60% for animals, chairs, faces, fruits and vehicles (Liu *et al* 2009). Another study reported 69% online accuracy in a target selection task of two overlaying images (Cerf *et al* 2010). ECoG's high signal-to-noise ratio even supports significant discrimination of two different faces or two different expressions of one face in single trials (Ghuman *et al* 2014). One study decoded twelve categories (excluding faces) during an object naming task with a mean rank accuracy of 76% (i.e. in a list of 100 objects, ranked by their probability to be selected by the classifier, the target object appears on position 24 on average) with a chance level of 50% (Rupp *et al* 2017). Another study decoded 24 different categories with an accuracy of 25% (chance level 4.2%) (Majima *et al* 2014).

The present study extends this large body of work via three experiments that decode type and color information in experiment I (offline), and (in real time) decode different computer-based stimuli in experiment II and natural image stimuli in experiment III. Specifically, ECoG signals were recorded in four patients while they were shown both color and greyscale versions of seven different types of visual stimuli (photos of faces, objects and bodies, images of line drawings and digits, and kanji and hiragana characters), thus creating a total of 14 stimulus classes. Experiment I investigated the spatial and temporal activity reflecting responses to visual stimulation in terms of discrimination performance at individual instants and sites, and classified across all types and colors in single trials. In addition, a real-time system was implemented to identify presented faces or kanji characters on a computer screen (experiment II), natural scenes with real faces (i.e. the faces of two experimenters and a mirror) and printed faces and kanji characters (experiment III).

Table 1. Clinical profiles of the four subjects. The IQ values in the table represent combined verbal and performance IQ scores. Handedness (H) was obtained and language lateralization (LL) was defined using WADA tests. '# of elec'. refers to the total number of electrode contacts in each patient. Subjects participated in up to three experiments (EXP).

Subject	Age	Sex	IQ	H	LL	Seizure focus	Covered hemisphere	# of elec.	EXP
A	26	M	96	R	L	Right temporal	Bilateral	188	1–3
B	20	F	74	R	L	Left temporal	Left	254	1
C	18	M	122	R	L	Occipital	Bilateral	254	1
D	29	F	85	R	L	Right	Right	144	2
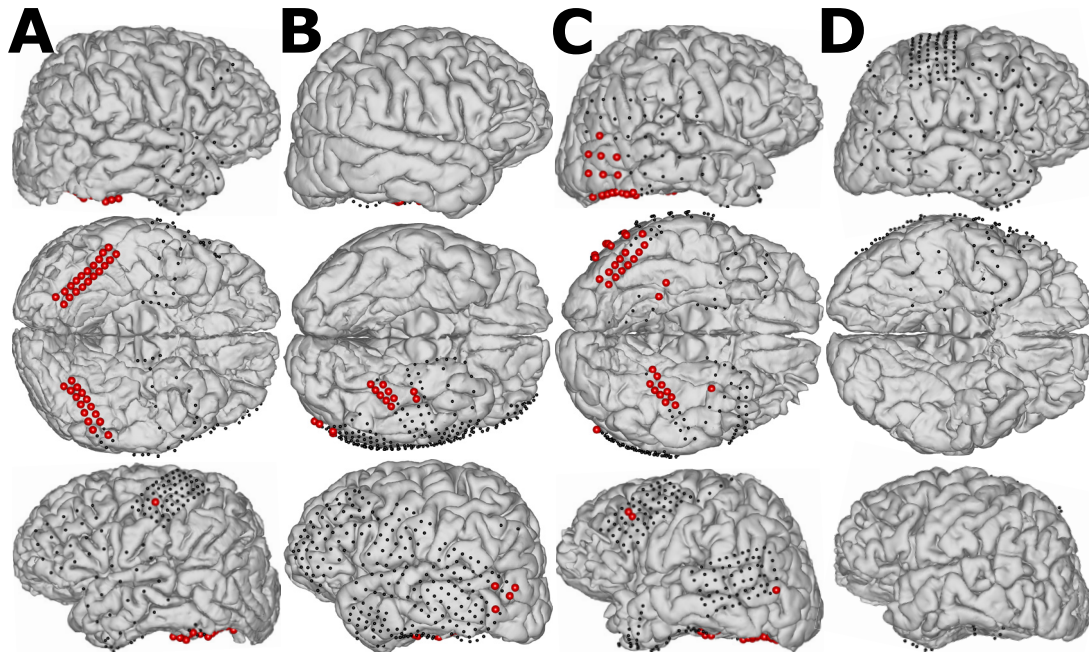
Figure 1. ECoG recording sites for subjects A–D. Black dots represent implanted subdural electrodes. Red balls highlight ECoG electrodes at locations that demonstrated significant broadband γ responses to visual stimuli for subjects who participated in experiment I (A–C). Subject D only participated in experiment II, in which all channels were spatially filtered.

2. Methods

2.1. Subjects

Four patients with epilepsy at Asahikawa Medical University (A and D) and The University of Tokyo Hospital (B and C) participated in this study. Each patient was temporarily implanted with subdural electrode grids to localize seizure foci and underwent neuro-monitoring prior to resective brain surgery. The grids consisted of platinum electrodes with an exposed diameter of 1.5–3.0 mm and an inter-electrode distance of 5–10 mm. After grid placement, each subject had postoperative computed tomography (CT) imaging to identify electrode locations in conjunction with preoperative magnetic resonance imaging (MRI). Table 1 provides an overview of the subjects and their clinical profiles. The study was approved by the institutional review boards of Asahikawa Medical University and The University of Tokyo Hospital. All subjects gave informed consent prior to the experiment.

Figure 1 shows the subjects' reconstructed brain models and indicates implanted electrode locations (dots). Each subject's brain model was reconstructed in FreeSurfer (Martinos Center for Biomedical Imaging, Cambridge, USA) using preoperative T1-weighted MRI data (Dale *et al* 1999). Then preoperative MRI data were co-registered to post-operative CT scans using SPM (Wellcome Trust Centre for Neuroimaging, London, UK) to localize electrode positions on the cortex (Penny *et al* 2007). Finally, the resulting 3D cortical models and electrode locations were visualized in NeuralAct (Kubanek and Schalk 2015).

2.2. Data acquisition

ECoG signals were recorded at the bed-side with a DC-coupled g.HIamp biosignal amplifier (g.tec medical engineering, Austria) after neuro-monitoring was completed—prior to resective surgery. Data were digitized with 24-bit resolution at 2400 Hz for offline assessment and 1200 Hz for real-time processing, synchronized with stimulus presentation using a photo diode, and stored using the g.HIsys real-time processing library (g.tec medical engineering GmbH, Austria). Ground (GND) and reference (REF) were located in dorsal parietal cortex (i.e. distant from task-related electrodes in the temporal lobe).
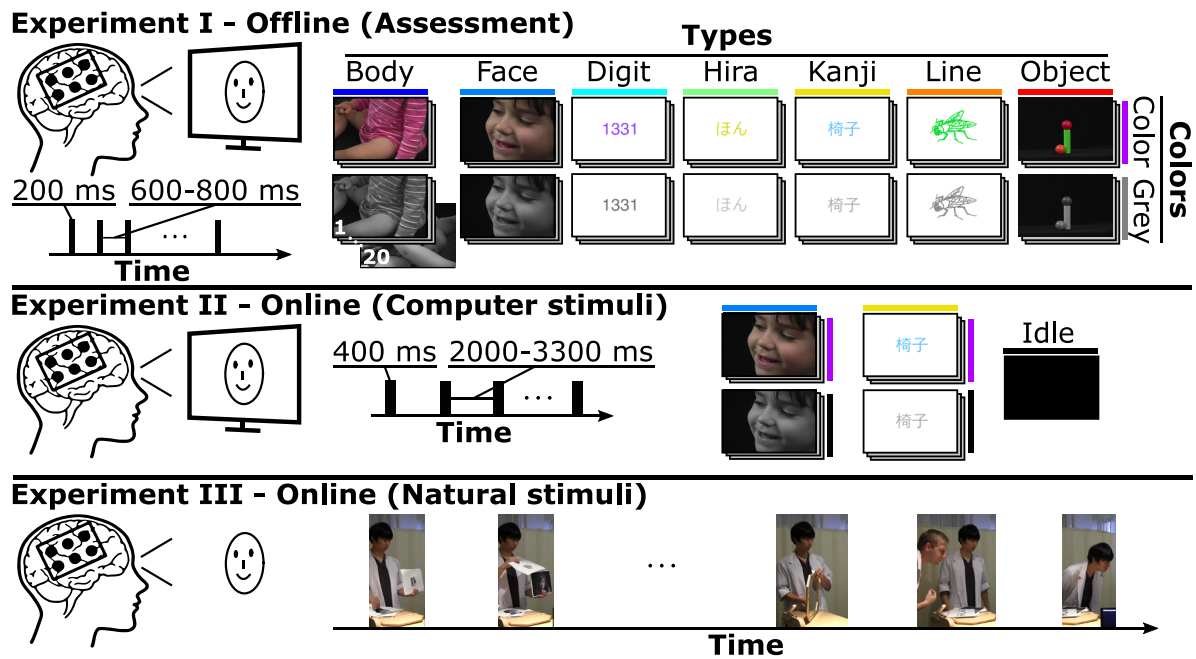
Figure 2. Presented stimulus types and experimental procedures for the three experiments. Experiment I, the assessment, collected ECoG responses from visual stimuli (200 ms presentation time) of seven types (*Body*, *Face*, *Digit*, *Hira*, *Kanji*, *Line* and *Object*). Examples show one out of 20 stimuli for each type in colored (*Color*) and greyscale (*Grey*) versions. Each stimulus occurred twice within the experiment (i.e. 40 stimuli per type and color, 560 stimuli in total). Experiment II evaluated the real-time discrimination performance of ECoG responses to presented *Face* and *Kanji* computer stimuli (400 ms presentation time), and to idle stimuli (black screen). Subjects viewed 30–40 stimuli of each type (180–240 trials in total) to calibrate the decoder and repeated the experiment with real-time discrimination (without getting any feedback). Experiment III tested the real-time discrimination performance of ECoG responses to natural stimuli (i.e. printed faces and kanji, mirror, real face) presented by the experimenter, one face presented by a co-experimenter and intermediate idle periods where nothing was shown.

2.3. Experimental procedure

The three experiments in this study are illustrated in figure 2. Experiment I assessed neural responses to visual stimuli using offline analysis. We also obtained online accuracies during real-time visual perception tasks, where the subjects looked at monitor-based stimuli in experiment II and natural images in experiment III.

During the assessment (experiment I) subjects A–C observed stimuli that were presented on a computer screen, which was placed about 80 cm in front of the subject. The stimuli were about 20 cm in size, and consisted of seven types ((i) *Body*, (ii) *Face*, (iii) *Digit*, (iv) *Hira* (Hiragana), (v) *Kanji*, (vi) *Line* and (vii) *Object*), all seven of which were shown in color (*Color*) or greyscale (*Grey*). This led to a total of 14 different classes for discrimination, which were presented sequentially in random order. Kanji and hiragana characters are components of the Japanese writing system and corresponded to the subjects' native language. Experiment I in figure 2 illustrates examples from 20 different stimuli for each class and shows the timeline of four out of 560 trials in the visual stimulation paradigm. Each trial consisted of a 200 ms presentation period and a subsequent black screen for 600–800 ms.

Experiment II employed real-time decoding of stimuli shown on a monitor, including images of faces and kanji characters, and an additional black screen as a new type. Thus, the three possible classification outcomes were *Face*, *Kanji*, and

Idle (i.e. neither *Face* nor *Kanji*, see figure 2, experiment II). Two subjects (A and D) participated in this discrimination experiment and were asked to observe a sequence of 30 (subject D) or 40 (subject A) stimuli of each type in randomized order with a presentation time of 400 ms each. Inter-stimulus-intervals (ISI) showed a black screen for 2.0–3.3 s. Each subject performed two runs (a total of 3 classes ·30 trials ·2 runs = 180 trials for subject D and $3 \cdot 40 \cdot 2 = 240$ trials for subject A), one for calibration and another to validate the real-time decoding performance.

Subject A also participated in experiment III, a real-world scenario with natural stimuli (see figure 2, experiment III), in which one of the people attending the experiment presented the subject with kanji characters and faces printed on pieces of paper, a mirror and two experimenters' faces—who appeared in front of the subject. A computer classified *Face*, *Kanji* and *Idle* in real time, and provided visual feedback about that type via a monitor next to the subject, by displaying a face, a kanji character or a black screen. The monitor output was not visible to the subject, but was recorded by a video camera that taped the experiment at a rate of 30 FPS for later synchronization of stimulus onset with ECoG data and for quantification of the decoder's performance. Frames of the video were synchronized with ECoG data based on the decoder output (i.e. the first video frame showing a kanji character on the monitor corresponded to the sample time at which the decoder classified a *Kanji* stimulus).
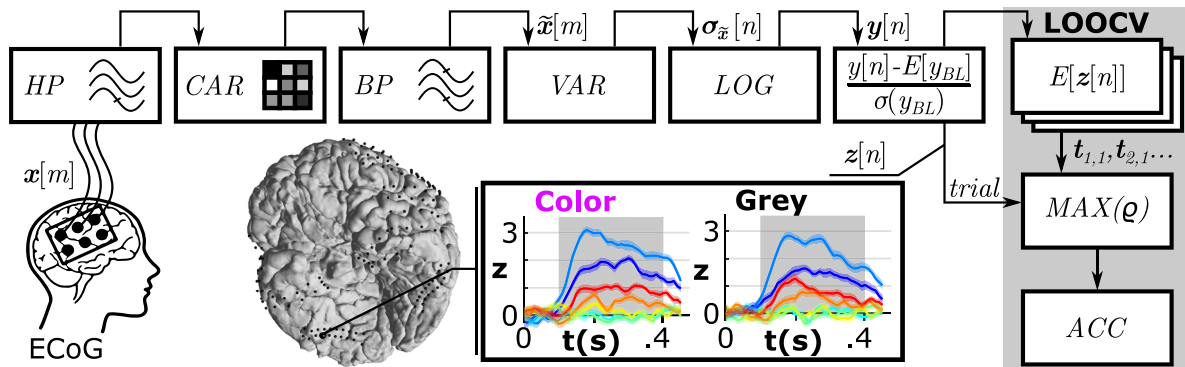
Figure 3. Feature extraction pipeline and classification method for the assessment (experiment I). Signal processing steps for the multi-channel ECoG signals $\mathbf{x}[m]$ included drift removal by a high-pass filter (HP), spatial filtering (CAR), temporal band-pass filtering (BP), variance estimation (VAR), log-transformation (LOG) and standardization (z-score). Colored time series show the mean z-scores ($\mathbf{z}[n]$) with standard errors for all stimulus types (color codes are based on figure 1) from ECoG electrode location 182 of subject A. Areas shaded in grey represent the active period used for discrimination. One active period (trial) of $\mathbf{z}[n]$ was correlated with templates ($t_{1,1}, t_{2,1}...$) based on the remaining trials of each stimulus type. The template that correlated most strongly (MAXρ) assigned the trial class according to the template class. A leave-one-out cross validation (LOOCV) yielded the classification accuracy (ACC) of all trials and stimulus types.

2.4. Signal processing for assessment

Figure 3 illustrates the feature extraction and classification method for assessment in experiment I. Recorded ECoG signals were denoted as $\mathbf{x}[m]$ (digitized multi-channel data at time m) and underwent a 2 Hz Butterworth high-pass (HP) filter (4th order) to remove DC drifts. Visual inspection of filtered data left 182, 247 and 246 channels (after exclusion of artifactual signals like epileptic activity, etc) for subsequent processing for subjects A–C, respectively. A common average reference (CAR) montage re-referenced the signals (Liu *et al* 2015) and a 110–140 Hz band-pass (BP) filter (Butterworth low and high pass filter, each of 4th order) extracted broadband γ activity. Given the time-frequency maps in figure 4, this band turned out to be most discriminant for individual classes. Next, the signals were temporally stabilized by computing the variance $\sigma_{\tilde{\mathbf{x}}}[n]$ based on 20 ms (50% overlap) epochs of $\tilde{\mathbf{x}}[m]$, and further normalized by log-transformation. This provided the output metric $\mathbf{y}[n]$, where n was an instant of the down sampled signals ($f_s = 100$ Hz). Data from each channel were further z-scored based on all samples of the baseline periods of all trials (-300 to 0 ms pre-stimulus interval), generating standardized data $\mathbf{z}[n]$. Information in $\mathbf{z}[n]$ was used to identify reactive ECoG locations and to discriminate ECoG responses to the visual stimulus types.

Channels were considered for classification only if the standardized data $\mathbf{z}[n]$ of any class was significantly higher for the task period compared to the baseline period. Thus, a Wilcoxon rank-sum test compared the average z-scores of a stimulus type's baseline periods (-300 to 0 ms pre-stimulus interval) with the average z-scores of the corresponding active periods (100–400 ms post-stimulus interval). This test was performed for each stimulus type and if a significant response was found ($p < 0.01$, Bonferroni corrected for the number of channels and tested stimulus types) the channels were considered for further analysis.

Standardized responses $\mathbf{z}[n]$ in selected channels (highlighted with red balls in figure 1) were discriminated by a pattern recognition approach. To do this, the assessment data were separated into $N_T = 40$ trials of each class i ($i \in \{1, 2...14\}$).

Each trial $z_{i,l}$ ($l \in \{1, 2...N_T\}$) consisted of a 100–400 ms post-stimulus interval of $\mathbf{z}[n]$. For pattern recognition templates were computed from 39 trials of each class, which were derived as follows:

$$\mathbf{t}_{i,k} = \frac{\sum_{l=1}^{N_T} \mathbf{z}_{i,l}}{N_T - 1}, l \neq k. \tag{1}$$

Each template vector $\mathbf{t}_{i,k}$ was calculated from training data, and was subsequently compared to the kth trial of each class (leave-one-out cross validation (LOOCV) approach, $k \in \{1, 2...40\}$). Thus, the remaining trial vector $\mathbf{z}_{i,k}$ was correlated with the template $\mathbf{t}_{i,k}$ leading to $r_{i,k}$, the correlation coefficient for class i and trial k.

$$r_{i,k} = \frac{\sigma_{\mathbf{t}_{i,k}, \mathbf{z}_{i,k}}}{\sigma_{\mathbf{t}_{i,k}} \sigma_{\mathbf{z}_{i,k}}}. \tag{2}$$

The correlation followed the definition of the Pearson's correlation coefficient with $\sigma_{\mathbf{t}_{i,k}, \mathbf{z}_{i,k}}$ as the covariance of $\mathbf{t}_{i,k}$ and $\mathbf{z}_{i,k}$, and $\sigma_{\mathbf{t}_{i,k}}$ and $\sigma_{\mathbf{z}_{i,k}}$ as the variance of $\mathbf{t}_{i,k}$ and $\mathbf{z}_{i,k}$, respectively. Correlation coefficients were computed for all 14 templates for each of the 14 test trials. Hence, for a given tested feature vector, the classifier determined the type and color that produced the highest correlation (MAX(ρ)). Results from 40 repetitions ($14 \cdot 40 = 560$ classifications in total) with new sets of templates yielded class specific positive rates (TPR) and an overall accuracy (ACC).

The same assessment approach was applied to paired conditions of colored and greyscale types to investigate any color or type specific bias that affected the discrimination performance. For paired conditions a test trial was correlated with the template vectors of the two selected classes and assigned to the class that correlated most.

Additionally, the assessment led to classification accuracies using temporal and spatial features only. Specifically, the temporal features contained concatenated $\mathbf{z}[n]$ of the selected channels for a dedicated 20 ms epoch and were classified by the pattern discrimination in 10 ms steps (from -300 to 450 ms relative to stimulus onset). A similar strategy for the
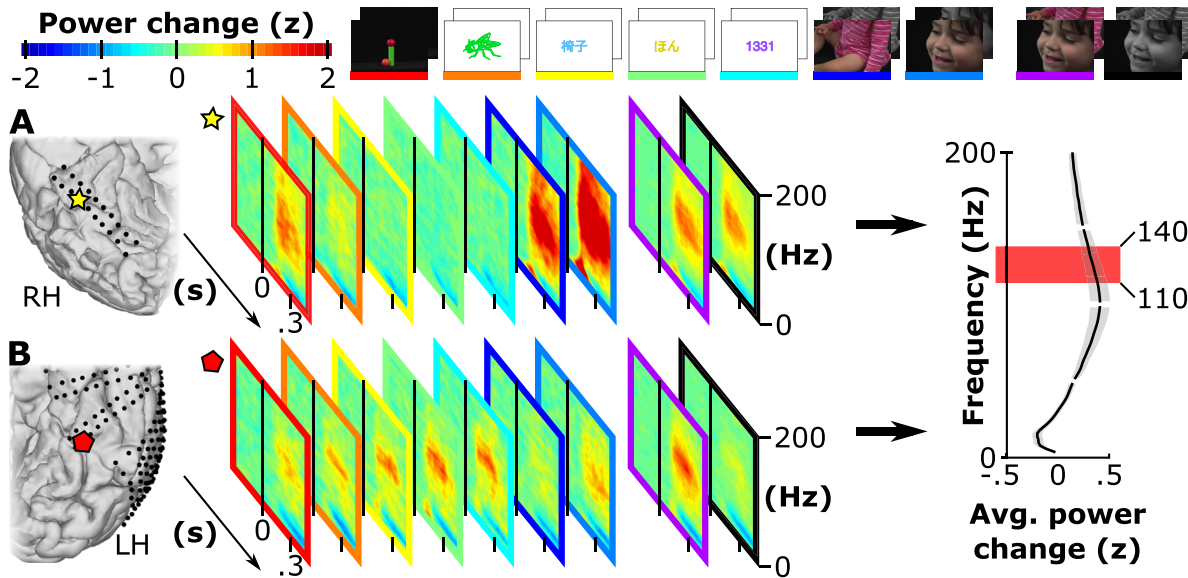
Figure 4. Time-frequency maps of two locations (yellow star and red diamond) of subjects A and B. Each map contains the standardized power change in z-scores for a given type (*Line*, *Digit*...) and color (*Grey* and *Color*) with respect to a reference window −300 to 0 ms relative to the visual stimulation onset. The yellow star location in A turned out to be *Face* selective, showing a strong broadband γ power increase for *Face* stimuli, whereas other most other locations remained silent or show little activation (like *Body*). A *Color* selective region was located in B (indicated by the red diamond). The average power change in z-scores of 120 ECoG locations on the VTC in subjects A–C, averaged over a period of 100–400 ms after stimulus onset, revealed the strongest change for the selected 110–140 Hz band (highlighted by the red bar).

spatial assessment included the pattern discrimination of concatenated $\mathbf{z}[n]$ over time (100–400 ms post-stimulus interval), tested for each selected channel.

For each assessment, an additional permutation test generated a random distribution of accuracies based on trial labels that were shuffled 1000 times. Hence, the rank of the assessment output in the random distribution gave the probability p for random classification. This probability was transformed into an activation index (AI) as follows (Schalk *et al* 2007, Kubanek *et al* 2009, Gunduz *et al* 2011, 2012, Gunduz *et al* 2016, Wang *et al* 2012, Liu *et al* 2015, Lotte *et al* 2015):

$$\mathrm{AI} = -\log(p). \qquad (3)$$

The AIs were used to highlight reliable discrimination for the temporal and spatial assessment, whereas the p values were used to indicate results that were significantly better than chance ($p < 0.05$).

### 2.5. Signal processing for online discrimination

Real-time processing of multichannel ECoG signals requires time efficient feature extraction methods that guarantee a certain processing time, independent from the number of recorded channels. At the same time, asynchronous detection of visual perception requires robust features that are stable over time to enable detection and discrimination of visual stimuli based on sliding windows. Hence, the signal processing pipeline used for the assessment had to be modified to fulfill the aforementioned requirements.

Before classifying ECoG data in real-time, it was necessary to first process calibration data. Figure 5 shows the required signal processing steps. First, a 4th order Butterworth

high-pass (HP) filter removed the DC drift of the recorded ECoG signals $\mathbf{x}[m]$ for visual inspection. If a channel contained power line interferences or epileptic waveforms, it was manually excluded. This led to 182 and 140 remaining channels for subjects A and D, respectively. Then, a 110–140 Hz band-pass (BP) filter extracted broadband γ activity $\mathbf{x}_{\mathrm{filt}}[m]$. Common spatial patterns (CSP) were computed from filtered signals to improve the signal-to-noise ratio (SNR) and reduce the feature dimensionality (Müller-Gerking *et al* 1999, Ramoser *et al* 2000, Guger *et al* 2000). Since CSPs maximize the signals' variance for one condition and minimize it for another condition, a set of spatial filters for three 'one-versus-all' conditions generated distinctive features for *Face*, *Kanji* and *Idle*. First, *Face* against combined *Idle* and *Kanji* stimuli, second *Kanji* against combined *Idle* and *Face* stimuli and finally, *Idle* against *Face* and *Kanji* stimuli. Hence, each combination resulted in a set of spatial filters sorted by their impact on the conditions' variance. The CSP filters were calculated from ECoG data from 100-600 ms post-stimulus. For further processing only the four most relevant filters of each paired condition were used (i.e. the filters that corresponded to the two highest and the two lowest eigenvalues (Blankertz *et al* 2008)), resulting in twelve feature channels in total. Specifically, spatial filters were applied as channel weights ($\mathbf{w}_{\mathrm{CSP},j}, j \in \{1, 2...12\}$) for all electrodes:

$$\tilde{x}_{\mathrm{CSP},j}[m] = \mathbf{w}_{\mathrm{CSP},j}^{T}\mathbf{x}_{\mathrm{filt}}[m]. \qquad (4)$$

Then, from each resultant time series $\tilde{x}_{\mathrm{CSP},j}[m]$ the variance $\sigma_{\tilde{x}_{\mathrm{CSP},j}}[n]$ was calculated from 500 ms epochs with a 97% overlap. These signals were log-transformed to normalize the data and to get $\mathbf{y}_{\mathrm{CSP}}[n]$, the normalized broadband γ power. Finally, three linear discriminant analyses (LDA) were trained
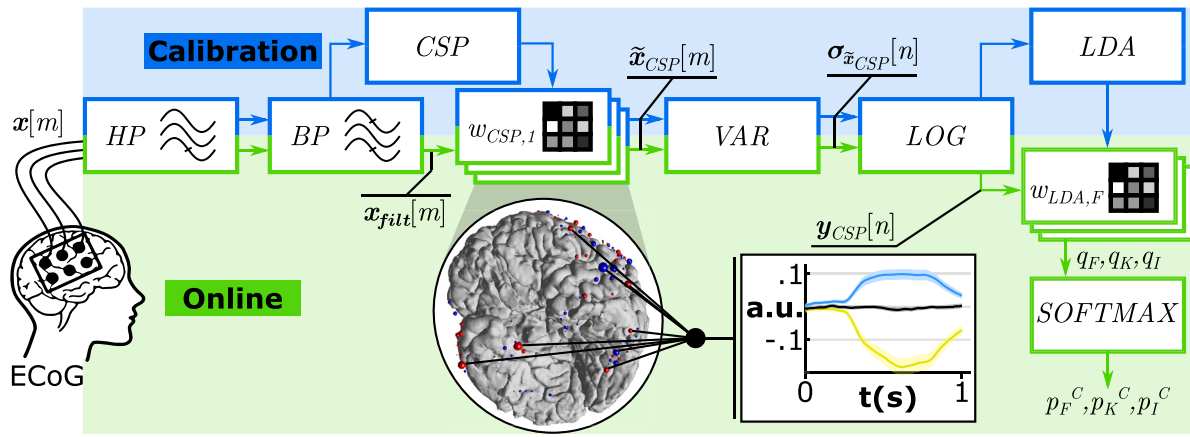
Figure 5. Feature extraction pipeline and classification method for real-time detection and discrimination in experiment II and III. Calibration: Recorded multi-channel ECoG signals $\mathbf{x}[m]$ were HP and BP filtered and submitted to a common spatial pattern (CSP) analysis that computed a set of spatial filters ($\mathbf{w}_{CSP}$). Spatially filtered signals $\tilde{\mathbf{x}}_{CSP}[m]$ underwent variance estimation (VAR) and log-transformation (LOG) and resulted in normalized $\mathbf{y}_{CSP}[n]$. A linear discriminant analysis (LDA) generated class specific weights ($\mathbf{w}_{LDA}$) for real-time processing. Online: Real-time processing steps included the HP and BP filtering and the spatial ($\mathbf{w}_{CSP}$) filtering, followed by the variance estimation (VAR) and log-transformation (LOG). The linear classifier ($\mathbf{w}_{LDA}$) weighted the features in $\mathbf{y}_{CSP}[n]$ and generated LDA outputs (q_F, q_K, q_I) for *Face*, *Kanji* and *Idle*. Finally, a Softmax function transformed the LDA output in complementary probabilities (p_{FC}, p_{KC}, p_{IC}). The diagram $\mathbf{y}_{CSP}[n]$ shows the real-time processing output for *Face* (blue), *Kanji* (yellow) and *Idle* (black) based on 182 combined ECoG locations in subject A.

to discriminate the twelve features of each class (30–40 trials per class of Face, Kanji and Idle), from data of the remaining classes. Each of the three combinations (denoted with $i \in \{F, K, I\}$) gave class specific weights $\mathbf{w}_{LDA,i}$.

After the calibration phase, the subsequent processing occurred in real time. Therefore, the ECoG data were sampled with 1200 Hz and read into the real-time processing in frames of 16 samples, which resulted in a processing rate of 75 Hz. In each processing step, data were HP and BP filtered, yielding $\mathbf{x}_{filt}[m]$ (as shown in figure 5). The twelve spatial filters were applied on $\mathbf{x}_{filt}[m]$ to get twelve time series $\tilde{\mathbf{x}}_{CSP}[m]$ for variance estimation and log-transformation, which yielded $\mathbf{y}_{CSP}[n]$. Subsequently, the three weight vectors $\mathbf{w}_{LDA,i}$ were applied to $\mathbf{y}_{CSP}[n]$ to get the three LDA outputs q_i for *Face*, *Kanji* and *Idle*:

$$q_i = \mathbf{w}_{LDA,i}^T \mathbf{y}_{CSP}[n]. \qquad (5)$$

Each LDA output was translated into a probability using a Softmax function (Sutton and Barto 1998):

$$p_{iC} = 1 - \frac{e^{q_i}}{\sum\limits_{j=1}^{3} e^{q_j}}. \qquad (6)$$

Here, p_{iC} was the complement probability that features represent class i. Hence, the classifier selected the class that corresponded to the lowest p_{iC}. In the case that no p_{iC} reached the confidence threshold of $p_{iC} < 0.05$, the output was automatically set to *Idle*. Finally, the activation index (AI) was calculated from the complementary probability p_{iC} according to the following equation:

$$\text{AI} = -\log(p_{iC}). \qquad (7)$$

In the real-time processing mode, when the AI crossed the significance threshold (p_{iC} or AI > 3), an image of a face, a kanji character or a black screen appeared on a feedback monitor.

This feedback was only visible to the experimenter, not the subjects.

3. Results

3.1. Assessment (experiment I)

3.1.1. Pairwise discrimination of colored and greyscale types. Figure 6 shows the TPR for each type of stimulation versus all each other type. TPR values were obtained by assigning the test trials to one of two template classes. Every matrix contains the TPR for each possible combination of classes. Subjects A and C reached very high TPRs (>90%) for *Face* stimulation versus all other types, except for *Color Face* versus *Grey Face*. Of course, the TPR reached its maximum if a type was compared with itself as illustrated by the diagonals. Figure 6 further depicts that the TPR minimized for comparisons of *Color* and *Grey* stimuli of the same type. For example, subject A correctly classified only 50% of *Grey Line* versus *Color Line* stimuli.

Table 2 contains the classification accuracies (50% chance) for each subject and each type. Accuracies correspond to the average TPR obtained from all pairwise discrimination tests of a certain type with any other type (i.e. the mean of each type's row and column TPR in figure 6). The highest classification accuracy of 97.9% was reached by subject C for *Color Face* stimulation. Subject A reached the second and third highest accuracies of 97.8% for *Color* and *Grey Face* stimulation. Subject B yielded the lowest classification accuracies (76.7%, 78.7% and 80.7%) for *Grey Digit* and *Body*, and for *Color Body* stimulation. *Face* stimulation achieved 92.3%, the highest average accuracy across all subjects, followed by *Body* (90.1%) and *Object* (90.0%) stimulation. The weakest average performance was found for Digit (88.0%) stimulation. Across all stimulus types and colors, the average
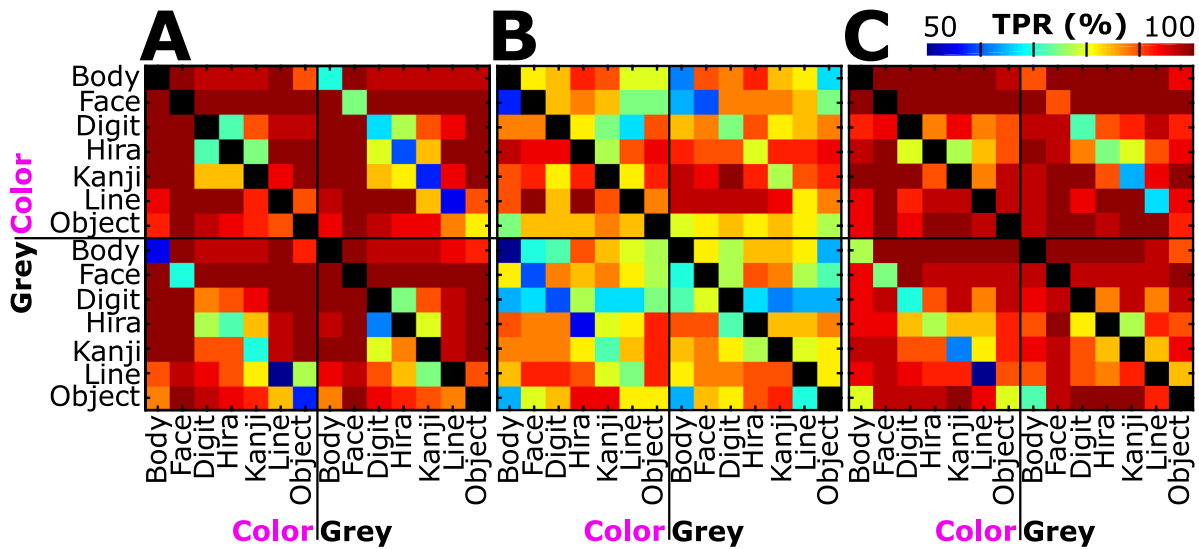
Figure 6. Results for pair-wise classification of colored (Color) and greyscale (Grey) stimulus types for subjects A–C. Colored squares indicate the true positive rate (TPR) for each type (rows) and color (columns) against every other type and color. A blue box indicates random classification, while perfect classification is highlighted in red (see color bar; 50% chance for paired classification). Diagonals are shown in black (i.e. no TPR available), as the same class templates used for discrimination were the same. The diagonals in the bottom left and top right quarter of each subject contain the TPR of colored stimuli against greyscale stimuli of the same type.

Table 2. Classification accuracies for each type (*Body*, *Face*...) for colored and greyscale stimuli (Colors) for subjects (S) A–C. The best three accuracies are bold, whereas the worst three accuracies are in italic.

S	Colors	Accuracy (%)							Average	
		Body	*Face*	*Digit*	*Hira*	*Kanji*	*Line*	*Object*		
A	*Color*	94.7	**97.8**	91.5	90.9	91.0	93.2	93.0	93.2	92.8
	Grey	94.8	**97.8**	91.5	89.0	91.3	89.4	92.2	92.3	
B	*Color*	*80.7*	80.8	83.6	88.4	86.5	86.3	84.1	84.3	83.1
	Grey	*78.7*	82.8	*76.7*	84.9	84.5	84.1	81.1	81.8	
C	*Color*	96.1	**97.9**	92.0	91.4	93.4	91.8	96.1	94.1	93.6
	Grey	95.6	97.0	92.8	90.7	90.9	90.6	93.8	93.0	
	Average	90.1	92.3	88.0	89.2	89.6	89.2	90.0	89.8	89.8

Table 3. Overall offline classification accuracies (ACC) for colors (Color versus Grey), types without color separation (7-Types) and types with color separation (T&C) for subjects (S) A–C. Random accuracies (RAND) and p values derived from a randomization test with scrambled trial labels.
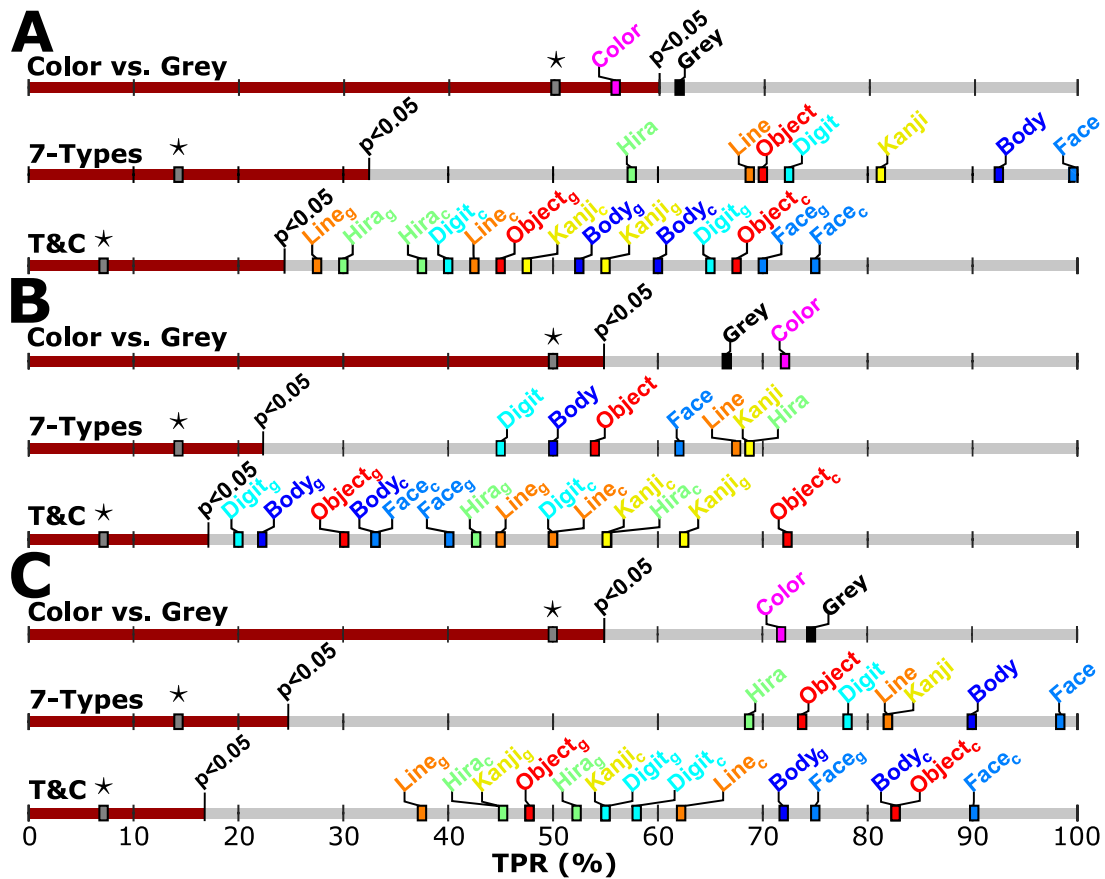
S	Color versus Grey			7-Types			T&C		
	ACC (%)	RAND (%)	*p*	ACC (%)	RAND (%)	*p*	ACC (%)	RAND (%)	*p*
A	58.8	50.0	0.0865	77.5	14.3	<0.0002	51.1	7.1	<0.0004
B	69.5	50.0	<0.0005	59.5	14.3	<0.0002	43.6	7.1	<0.0001
C	73.0	50.0	<0.0005	82.1	14.3	<0.0002	61.6	7.1	<0.0001
All	67.1			73.0			52.1		

accuracies were 92.8%, 83.1% and 93.6% for subjects A–C, respectively. Accuracies above 65.0% (for A) and 62.5% (for B and C) were statistically better than chance ($p < 0.05$).

3.1.2. Overall discrimination of types and colors. Table 3 contains the classification accuracy after discrimination of *Color* and *Grey* images ('Color versus Grey'), whereby subject C achieved the highest accuracy of 73.0% (50% chance). It contains also the classification accuracy after discrimination of stimulus types without color separation ('7-Types') when the seven types were classified against each other. Again,

subject C reached the highest accuracy of 82.1% (14.3% chance). 'T&C' in table 3 contains the accuracy after classification of all 14 colored and greyscale stimulus types against each other. Here subject C reached 61.6% (7.1% chance). Interestingly, '7-Types' performed better than 'Colors versus Grey'. Although 'T&C' performed worst, because of the 14 different classes, all subjects achieved highly significant accuracies ($p < 0.0004$).

Figure 7 illustrates the TPRs for 'Color versus Grey', '7-Types' and 'T&C'. In the 'Color versus Grey' assessment, subject C achieved a TPR of 74.3% for *Grey* and 71.8% for

Figure 7. TPR for colors (Color versus Grey), types (7-Types) and both (T&C) after a leave-one-out cross validation test for three subjects (A–C). Stars indicate the expected random accuracy for each test. The red bar ends at the significance border ($p < 0.05$) of an empirically derived random distribution based on scrambled trial labels. As an example, subject A had a 50% chance level for 'Color versus Grey' with a threshold TPR of 60.0% ($p < 0.05$), a chance level of 14.3% for '7-Types' with a threshold TPR of 32.5% ($p < 0.05$), and a chance level of 7.1% for 'T&C' with a threshold TPR of 25.0% ($p < 0.05$).

Table 4. Asynchronous classification accuracy (Accuracy) of the computer stimuli in experiment II and natural stimuli in experiment III for subjects A and D. Accuracy with sub-sampling (w ss) includes classification results (ACC) from data sub-sampled randomly 50 times, with equally balanced conditions (i.e. *Face*, *Kanji* and *Idle*). Accuracy without sub-sampling (w/ss) contains the classification results (ACC) including all *Idle* data. Random (RAND) accuracies for the three classes reflect the mean accuracy after a permutation test with 1000 times shuffled trial labels. The duration (DUR) shows the length of each experiment. The latency (LAT) shows the time delay between the stimulus onset and the actual real-time decoder output with best match.

S	Stimulus	DUR (s)	LAT (ms)	Accuracy (w ss)			Accuracy (w/ss)		
				ACC (%)	RAND (%)	p	ACC (%)	RAND (%)	p
D	Computer	275	440	66.7	33.3	<0.001	91.1	83.7	<0.001
A	Computer	275	468	80.8	33.3	<0.001	90.7	75.6	<0.001
A	Natural	39	468	74.8	33.3	<0.001	82.8	69.7	<0.001

Color, whereas subject B correctly classified *Color* (72.1%) more often than *Grey* (66.8%). Subject A identified 61.8% of the *Grey* stimuli, but did not reach significant TPR for *Color*. In the '7-Types' classification *Face* performed best for subjects A and C. Indeed, all *Face* stimuli were identified in subject A. In subject B the TPR maximized for *Hira* and reached 68.8%. In the 'T&C' mode, the classification worked best for *Color Face* in subjects A (75.0%) and C (90.0%), and *Color Object* in subject B (72.5%). *Grey Line* in subjects A (27.5%) and C (37.5%) and *Grey Digit* in subject B (20.0%) performed worst.

3.1.3. Temporal and spatial characteristics. The temporal pattern of the TPR is plotted in figure 8 for each type (*Body*, *Face*...), and *Grey* and *Color*. Each star shows the time point with the highest average TPR (classification accuracy) of all types. The classification accuracy reached its peak at 240, 210 and 250 ms after stimulus onset for subject A–C, respectively. The classification accuracies (7.1% chance) at those times were 41.8%, 24.6% and 41.8%, respectively, and thus led to an average peak accuracy of 36.1% at about 233 ms across subjects. Responses to *Face* and *Body* stimuli gave a high TPR after 200 ms and remained stable until about 400 ms for
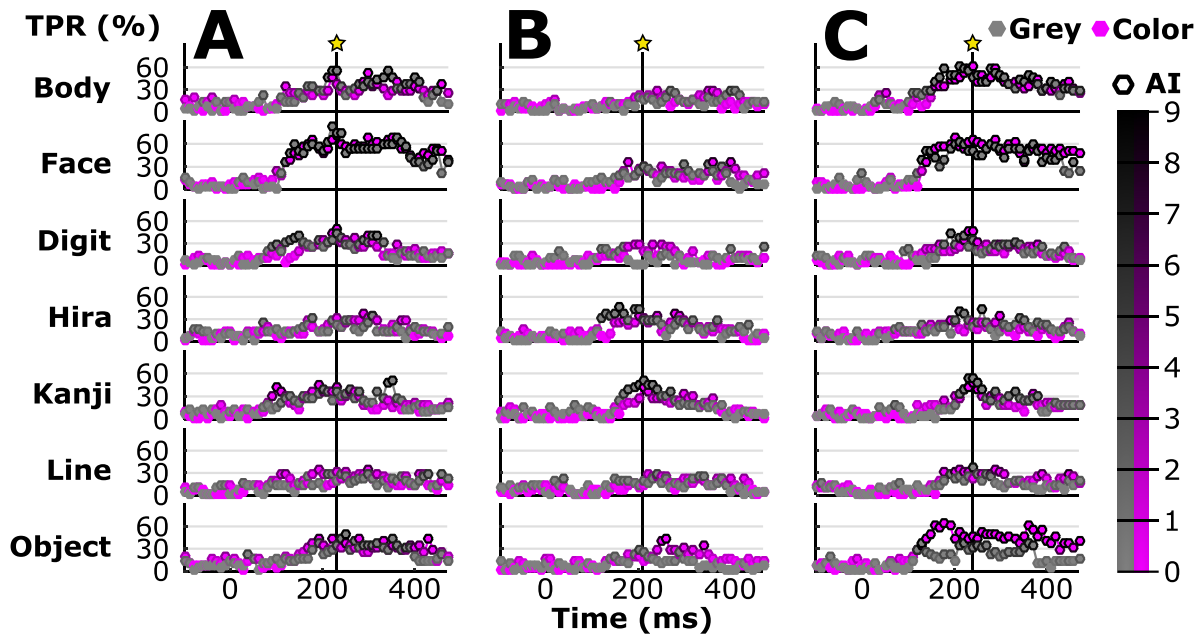
Figure 8. TPR and activation index (AI) over time for types and colors. Each time segment (20 ms epochs with 50% overlap) led to a feature vector and resulted in an independent classification output. Thus, the curve represents the TPR for individual segments and the edge color of each bullet shows the AI (black edges indicate reliable activation), which was derived from a randomization test with scrambled trial labels. Stars with vertical lines represent the times for which average TPR and AI maximized for types and colors.
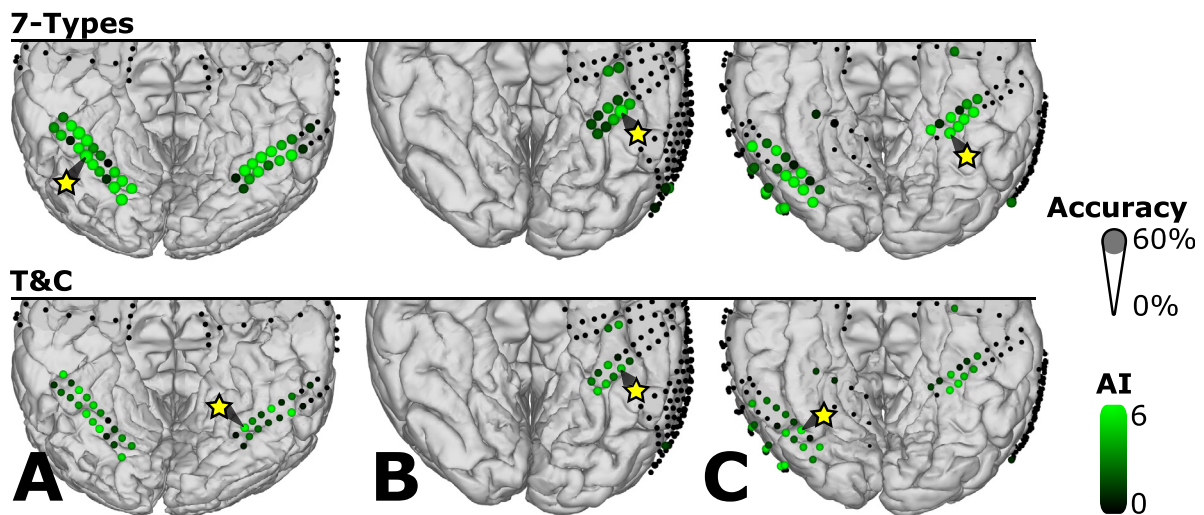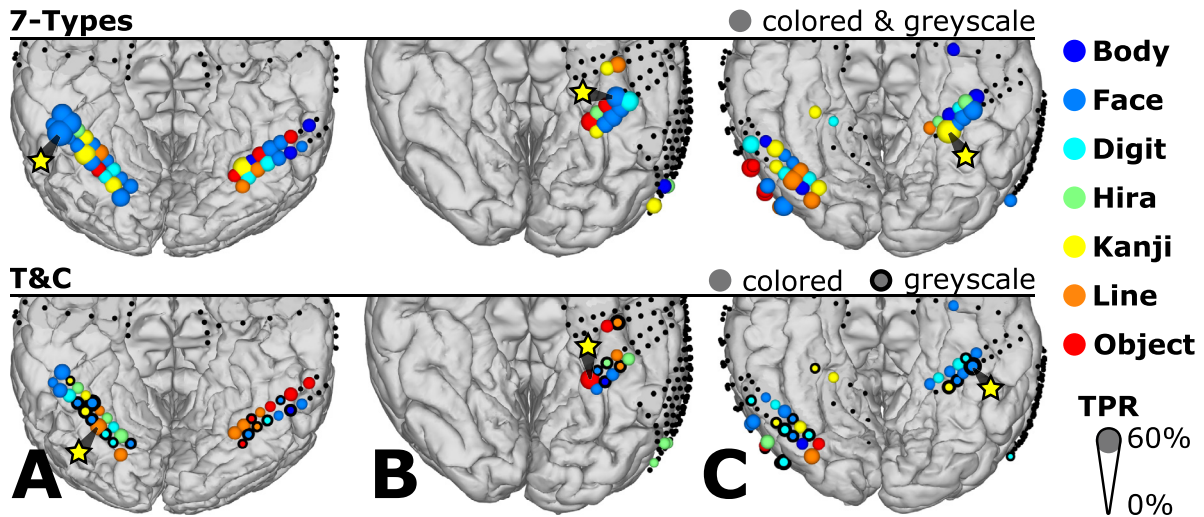

Figure 9. Spatial distribution of the average classification accuracy for types without color separation (7-Types) and types with color separation (T&C) for subjects A–C. Diameters show the average classification accuracy. Only channels with significant activation in the channel selection test were considered for classification and are marked with different AI scale values in green. All other recording locations were excluded from EXP1 and are indicated with small black dots. Yellow stars mark sites that showed the best discrimination performance between types (with and without color separation).

subject A and C. The biggest TPR difference could be found for subject C for *Object*. Subject B attained the highest TPR of about 58% for *Kanji* at about 210 ms. Interestingly, *Body*, *Face* and *Object* produced a high TPR over a long period of about 150–400 ms, while all other stimulus types showed a much shorter peaks.

Figure 9 summarizes the classification accuracy for each selected electrode channel discriminating all '7-Types' and all seven types and two colors ('T&C'). Yellow stars label those ECoG electrode locations that provided the highest classification accuracy for each subject. For the '7-Types' comparison, the highest classification accuracies (14.3% chance) were

26.4%, 24.3% and 21.6% for subjects A–C, respectively. Note that each accuracy resulted from a single channel. The corresponding average peak accuracy across subjects was 24.1%. For the 'T&C' assessment the highest classification accuracies (7.1% chance) reached 12.5%, 10.4% and 11.6% for subjects A–C, respectively. Here, the average peak accuracy across subjects was 11.5%.

Figure 10 shows the highest TPR of all stimulus types and electrode locations for '7-Types' or 'T&C'. Subject A reached a high TPR for Face around the area indicated with the star. In the '7-Types' condition, the types with the highest TPR (14.3% chance) were *Face*, for subject A (62.5%) and B

7-Types

⬤ colored & greyscale

T&C

⬤ colored ◉ greyscale

● **Body**
● **Face**
● **Digit**
● **Hira**
● **Kanji**
● **Line**
● **Object**

TPR
60%
0%

A **B** **C**

Figure 10. Spatial distribution of the classes with the highest TPR for types without color separation (7-Types) and types with color separation (T&C) for subjects A–C. Diameters show the TPR and the colors indicate the types with the highest TPR. Only channels with significant activation in the channel selection test were considered for classification (locations with colored dots). All other recording locations were excluded from experiment I and are indicated with small black dots. Yellow stars highlight sites with the highest TPR.
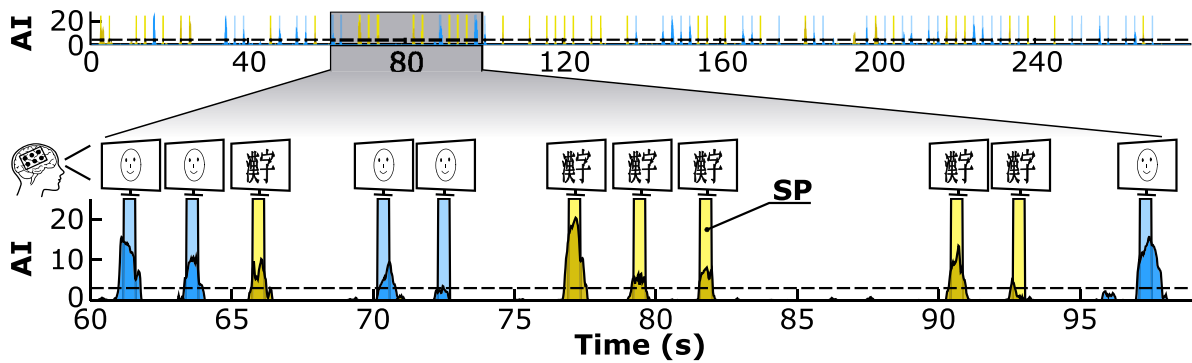
Figure 11. Real-time classification output (AI) over time for the computer stimuli in experiment II of subject A. Stimulus presentation (SP) times of *Face* (blue) or *Kanji* (yellow) computer stimuli are overlapped with the AI of Face (blue) and Kanji (yellow).
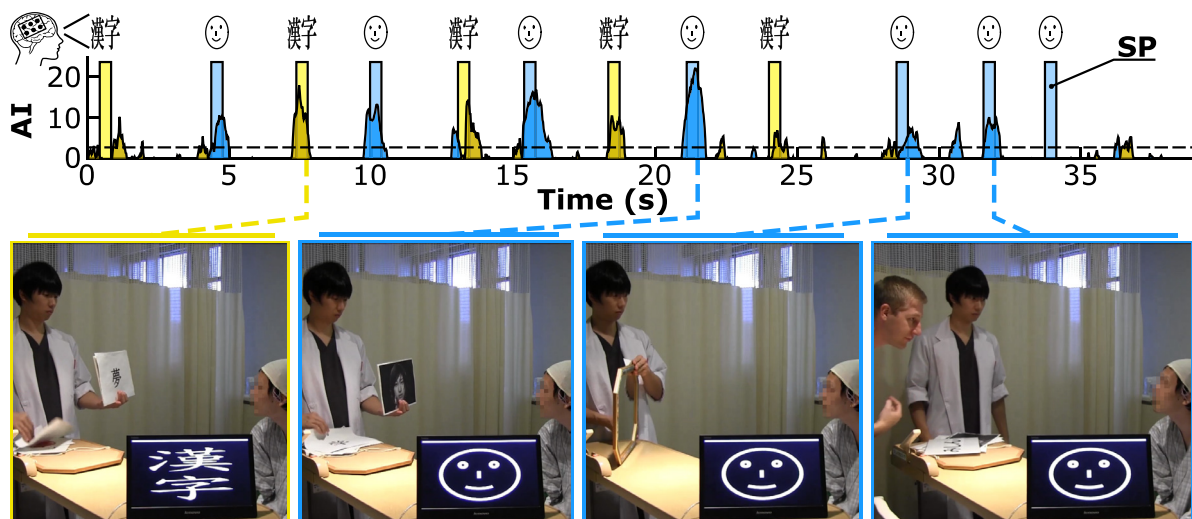
Figure 12. Real-time classification output (AI) over time for the natural stimuli in experiment III of subject A. Stimulus presentation (SP) times of natural *Face* (blue) or *Kanji* (yellow) stimuli are overlapped with the AI of Face (blue) and *Kanji* (yellow). These four photographs were taken from a video during the experiment and show the experimenter(s) (on the left) and the subject (on the right). From left to right, the pictures show: (1) the experimenter holding a printed kanji and the brain–computer interface (BCI) system successfully decoding *Kanji*; (2) the experimenter holding a printed face; (3) the experimenter holding a mirror; (4) a second experimenter. The video monitor on the bottom demonstrates that the brain signals were classified in real time.

(47.5%), and *Kanji* for subject C (61.3%). Adding the color information in 'T&C', the types with the highest TPR (7.1% chance) were *Color Line* (37.5%), *Color Object* (42.5%) and *Grey Face* (42.5%) for subjects A–C, respectively.

3.2. Online discrimination of computer and natural stimuli (experiment II and III)

Table 4 lists the total duration of data collection, the latency of the real-time classification output with respect to stimulus onset, the asynchronous classification accuracies and the corresponding random accuracies for subjects A and D. The actual stimulus and the decoder output matched best after shifting the decoder output 440–467 ms backwards in time, and thus showed the processing speed for real-time classification. In experiment II the real-time decoder correctly identified 73.7% of the computer stimuli for both subjects on average. The highest accuracy of 80.80% was achieved by subject A in the computer stimulus run. Even in the natural run subject A achieved an accuracy of 74.82% and performed better than subject D performed in the computer stimulus run. The latency of the decoder during the natural run was fixed to the 467 ms obtained in experiment II.

Figures 11 and 12 illustrate the AI over time for the computer stimuli in experiment II and the natural stimuli in experiment III. The decoder classified this output in real time into *Face* or *Kanji* when AI exceeded the dashed significance line ($AI > 3$, corresponding to $p < 0.05$), and *Idle* otherwise. The AI time series in both figures were corrected for the mean latency of the cortical responses (i.e. 440–468 ms) and thus shifted compared to the stimulus presentation bars.

4. Discussion

Many neuroimaging studies have demonstrated that ventral temporal cortex and inferior temporal gyrus are well known to contain specialized regions that process visual stimuli, and represent objects, words, numbers, faces and other categories. Some electrophysiological studies using electrocorticography (ECoG) have corroborated and extended these findings by identifying broadband ECoG responses to visual stimuli in the γ band. The present study provides the first human electrocorticographic evidence for color-related population-level cortical broadband γ responses, and demonstrated that neural categories established using stimuli presented on a video screen may generalize to the presentation of real-world visual scenes.

Results in this study were obtained by offline (experiment I) and online analyses (experiment II and III), which were fundamentally different in their decoding strategy. Specifically, the synchronous classification strategy during offline analysis revealed subject and location specific differences of individual categories, whereas the online decoder aimed to asynchronously detect and decode neural categories in real time.

The assessment showed that all types Body, Face, Digit, Hira, Kanji, Line and Object could be classified with a grand average accuracy of 89.8%, and for each type, the accuracy was ≥88% (see table 2). The best performance was achieved for *Grey Face* and *Color Face* yielding 92.3% classification accuracy on average. Subject B achieved a lower accuracy than subjects A and C, which may result from missing coverage of the right fusiform gyrus, the location of fusiform face area (FFA). In contrast, subjects A and C had at least partly coverage of the left and right fusiform gyri and showed almost perfect *Face* classification in pairwise discrimination, which is consistent with the 86–96% correctly detected faces reported elsewhere (Tsuchiya *et al* 2008, Gerber *et al* 2017, Miller *et al* 2016).

Aside from the detection of face-related neural responses, it is noteworthy that the accuracy for *Color* and *Grey Digits* reached 92.8% and 92.0% for subject C. Interestingly, electrode sites in subject C covered the right inferior temporal gyrus, which has been identified as a number form area (Shum *et al* 2013).

Overall discrimination showed that the best classification accuracy of 72.9% was achieved when the '7-Types' were discriminated from each other based on class templates obtained from broadband γ responses. Other decoders have utilized event-related potentials (ERP), achieving a discrimination performance of about 60% for five stimulus categories (Liu *et al* 2009), or single neuron recordings, leading to 69% correctly assigned image labels in a two class selection task (Cerf *et al* 2010).

Furthermore, the discrimination of colored and greyscale stimuli yielded 67.1% correct classification. In the present study ECoG signals for color discrimination were obtained mainly from visual area VO1, which has been reported to be color and object selective (Brewer *et al* 2005), and further to be responsive to color changes (Brouwer and Heeger 2013). A previous reported decoder based on fMRI utilized signals obtained from visual area VO1 and discriminated responses to eight colors with an accuracy of 48% after more than 15 repetitions (Brouwer and Heeger 2009).

The discrimination of all 14 classes in 'T&C' gave the lowest accuracy of 52.1%, but contained of course 14 different classes. Subjects A and C performed better for types than for colors, but in contrast, subject B performed better for colors than for types. Therefore, the electrode location could play an important part for color or type separation. Several ECoG locations showed the highest classification accuracy for *Face* or *Kanji* stimuli. Those locations were spread across the cortex and support the model of alternating face and letterstring selective cortex regions around the middle fusiform sulcus (Matsuo *et al* 2015). Notably, ECoG locations that showed the highest TPR to *Face* stimuli were grouped into clusters of bigger regions than for *Kanji* locations. One cluster was located on the right FFA in subject A and turned out to be face selective and causally involved in face processing after systematic electrical cortical stimulation (Schalk *et al* 2017). Such a face selective cluster of ECoG locations has also been previously reported in another electrical stimulation study (Parvizi *et al* 2012). In the current study two of these clusters were found in subject A (one in each hemisphere), whereas only one cluster, located in the left hemisphere, was found in subjects B and C. This can be most likely explained by missing or only partial coverage of the right fusiform gyrus.

Another interesting finding is that features obtained from *Kanji* locations enabled the decoder to discriminate even between *Hiragana* and *Kanji* stimuli. Such a discrimination task has not been presented elsewhere and shows that the letterstring locations reported in Matsuo *et al* (2015) can be further subdivided into more specific regions.

The assessment further showed that even a single ECoG electrode location decoded specific stimulus types with an accuracy of 24.1% in the '7-Types' discrimination. Although this is already remarkable, combined information from multiple locations revealed the 72.9% accuracy of the '7-Types'. For real-time processing, it was important to efficiently consider multiple electrodes. This was realized with the CSPs that automatically weighted each electrode according to its importance for the classification task. Therefore, the most important electrodes were considered automatically, resulting in higher classification accuracy than single channel analysis.

The spatial distribution of type-specific information remained stable throughout experiments, whereas the onset of broadband γ activity varied from trial to trial and caused a different temporal pattern for each repetition. Hence, it is important to train the classifier on multiple trials and utilize moving variance windows for real-time classification. It is likely that the known relationship between modulatory activity in the α band and cortical population-level broadband γ activity (phase-amplitude coupling (Canolty *et al* 2006, Coon *et al* 2016, Coon and Schalk 2016)) resulted in variable broadband γ responses. Hence, in real-time mode the variance was calculated from 500 ms windows and induced, together with the response time of the subject, a delay of 400–500 ms with respect to the stimulus onsets. However, this latency did not affect the performance of the decoder in the present study, as the feedback was not presented to the subject. Still, the observed delay here is much shorter than the reported 1–10 s of the online decoder presented in Cerf *et al* (2010). This is mainly due to the different experimental design, which required the subject to voluntarily activate stimulus selective neurons. Shorter latencies reported in other studies were obtained offline (Miller *et al* 2016) and did not report asynchronous classification over time (Tsuchiya *et al* 2008, Liu *et al* 2009, Majima *et al* 2014).

The TPR over time in figure 8 revealed that Body, Face and Object generated distinctive broadband γ activity over a relatively long period from 150 to 400 ms. This was much wider than for *Digit*, *Hira*, *Kanji* and *Line*, which indicates that processing stimuli of types like *Body*, *Face* or *Object* requires more time and thus is a more complex cognitive task.

Face, *Kanji* and *Idle* phases could be separated in real time with accuracies between 66.7–80.8% after about 4 minutes of training in experiment II. These accuracies were achieved without giving feedback to the subject. With longer training periods, and with feedback to subjects, performance would probably increase further. The feedback may help the subject focus on the required tasks and maintain concentration, in addition to facilitating learning. The performance difference between subject A (80.80%) and D (66.67%) can be most likely explained by the dense electrode coverage of the ventral temporal cortex of subject A (66 recording sites versus 20 in subject D).

Spontaneous online detection of visual stimuli in the real-world scenario in experiment III demonstrated a surprisingly high accuracy of 74.82%. Notably, the real-world stimuli (e.g. the face of the experimenter) were not part of the stimuli used for training with the visual stimuli shown on the computer screen. Thus, the real-world scenario was not only based on new and independent data, but also on a different set of stimuli than the artificial stimuli shown on the computer in experiment II. In fact, natural stimuli included images of kanji and faces printed on a sheet of paper, but also real human faces of the experimenters and the subject through a mirror. This showed that, in subject A, the same cortical regions process information from natural stimuli and from trained faces and kanji characters shown on a computer monitor.

Another issue relevant to real-world applications is the additional cortical activity due to eye motion and moving visual targets, described as motion related augmentation of broadband γ activity on the lateral, inferior and polar occipital regions (Nagasawa *et al* 2011). Such activation patterns could interfere with the expected features from the training runs and therefore impair the classification performance. Furthermore, time-locking the onset of neural responses due to natural stimuli is much more challenging than time-locking the onset of responses resulting from stimuli presented via a computer.

4.1. Conclusion

Real-time detection and discrimination of visually perceived natural scenes is even possible when the system is trained on different data than was presented on a computer screen. This could lead to improved human-computer interfaces such as those proposed in the context of passive BCIs (van Erp *et al* 2012). Specifically, learning the identity of a perceived (or perhaps even covertly attended) visual object could be useful for constraining or otherwise informing the options of an interface. This ability may also prove useful for establishing new communication options for people that have lost the ability to communicate, such as people with amyotrophic lateral sclerosis (ALS).

ORCID iDs

C Kapeller ⬤ https://orcid.org/0000-0002-7330-895X

References

Allison T, Ginter H, McCarthy G, Nobre A C, Puce A, Luby M and Spencer D D 1994 Face recognition in human extrastriate cortex *J. Neurophysiol.* **71** 821–5

Allison T, Puce A, Spencer D D and McCarthy G 1999 Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli *Cerebral Cortex* **9** 415–30

Blankertz B, Tomioka R, Lemm S, Kawanabe M and Muller K R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56

Brewer A A, Liu J, Wade A R and Wandell B A 2005 Visual field maps and stimulus selectivity in human ventral occipital cortex *Nat. Neurosci.* **8** 1102–9

Brouwer G J and Heeger D J 2009 Decoding and reconstructing color from responses in human visual cortex *J. Neurosci.* **29** 13992–4003

Brouwer G J and Heeger D J 2013 Categorical clustering of the neural representation of color *J. Neurosci.* **33** 15454–65

Canolty R T, Edwards E, Dalal S S, Soltani M, Nagarajan S S, Kirsch H E, Berger M S, Barbaro N M and Knight R T 2006 High gamma power is phase-locked to theta oscillations in human neocortex *Science* **313** 1626–8

Cerf M, Thiruvengadam N, Mormann F, Kraskov A, Quiroga R Q, Koch C and Fried I 2010 On-line, voluntary control of human temporal lobe neurons *Nature* **467** 1104–8

Chang E F, Edwards E, Nagarajan S S, Fogelson N, Dalal S S, Canolty R T, Kirsch H E, Barbaro N M and Knight R T 2011 Cortical spatio-temporal dynamics underlying phonological target detection in humans *J. Cogn. Neurosci.* **23** 1437–46

Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff M A and Michel F 2000 The visual word form area *Brain* **123** 291–307

Collins J A and Olson I R 2014 Beyond the FFA: the role of the ventral anterior temporal lobes in face processing *Neuropsychologia* **61** 65–79

Coon W G and Schalk G 2016 A method to establish the spatiotemporal evolution of task-related cortical activity from electrocorticographic signals in single trials *J. Neurosci. Methods* **271** 76–85

Coon W, Gunduz A, Brunner P, Ritaccio A L, Pesaran B and Schalk G 2016 Oscillatory phase modulates the timing of neuronal activations and resulting behavior *NeuroImage* **133** 294–301

Crone N E, Boatman D, Gordon B and Hao L 2001 Induced electrocorticographic gamma activity during auditory perception *J. Clin. Neurophysiol.* **112** 565–82

Crone N E, Miglioretti D L, Gordon B and Lesser R P 1998 Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band *Brain* **121** 2301–15

Dale A, Fischl B and Sereno M 1999 Cortical surface-based analysis. I. Segmentation and surface reconstruction *NeuroImage* **9** 179

Edwards E, Nagarajan S S, Dalal S S, Canolty R T, Kirsch H E, Barbaro N M and Knight R T 2010 Spatiotemporal imaging of cortical activation during verb generation and picture naming *NeuroImage* **50** 291–301

Edwards E, Soltani M, Deouell L Y, Berger M S and Knight R T 2005 High gamma activity in response to deviant auditory stimuli recorded directly from human cortex *J. Neurophysiol.* **94** 4269–80

Edwards E, Soltani M, Kim W, Dalal S S, Nagarajan S S, Berger M S and Knight R T 2009 Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex *J. Neurophysiol.* **102** 377–86

Engell A D and McCarthy G 2011 The Relationship of gamma oscillations and face-specific ERPs recorded subdurally from occipitotemporal cortex *Cerebral Cortex* **21** 1213–21

Engell A D and McCarthy G 2014a Face, eye, and body selective responses in fusiform gyrus and adjacent cortex: an intracranial EEG study *Frontiers Human Neurosci.* **8** 642

Engell A D and McCarthy G 2014b Repetition suppression of face-selective evoked and induced EEG recorded from human cortex *Human Brain Mapp.* **35** 4155–62

Engell A D, Huettel S and McCarthy G 2012 The fMRI BOLD signal tracks electrophysiological spectral perturbations, not event-related potentials *NeuroImage* **59** 2600–6

Gerber E M, Golan T, Knight R T and Deouell L Y 2017 Persistent neural activity encoding real-time presence of visual stimuli decays along the ventral stream *NeuroImage* **161** 67–79

Ghuman A S, Brunet N M, Li Y, Konecky R O, Pyles J A, Walls S A, Destefino V, Wang W and Richardson R M 2014 Dynamic encoding of face information in the human fusiform gyrus *Nat. Commun.* **5** 5672

Grill-Spector K and Weiner K S 2014 The functional architecture of the ventral temporal cortex and its role in categorization *Nat. Rev. Neurosci.* **15** 536–48

Guger C, Ramoser H and Pfurtscheller G 2000 Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI) *IEEE Trans. Rehabil. Eng.* **8** 447–56

Gunduz A, Brunner P, Daitch A, Leuthardt E C, Ritaccio A L, Pesaran B and Schalk G 2011 Neural correlates of visual–spatial attention in electrocorticographic signals in humans *Frontiers in Human Neurosci.* **5** 89

Gunduz A, Brunner P, Daitch A, Leuthardt E C, Ritaccio A L, Pesaran B and Schalk G 2012 Decoding covert spatial attention using electrocorticographic (ECoG) signals in humans *NeuroImage* **60** 2285–93

Gunduz A, Brunner P, Sharma M, Leuthardt E C, Ritaccio A L, Pesaran B and Schalk G 2016 Differential roles of high gamma and local motor potentials for movement preparation and execution *Brain-Comput. Interfaces* **3** 88–102

Halgren E, Dale A M, Sereno M I, Tootell R B, Marinkovic K and Rosen B R 1999 Location of human face-selective cortex with respect to retinotopic areas *Human Brain Mapp.* **7** 29–37

Jacques C, Witthoft N, Weiner K S, Foster B L, Rangarajan V, Hermes D, Miller K J, Parvizi J and Grill-Spector K 2016 Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex *Neuropsychologia* **83** 14–28

Jensen O, Kaiser J and Lachaux J P 2007 Human gamma-frequency oscillations associated with attention and memory *Trends Neurosci.* **30** 317–24

Kadosh K C and Johnson M H 2007 Developing a cortex specialized for face perception *Trends Cogn. Sci.* **11** 367–9

Kam J *et al* 2018 Differential sources for 2 neural signatures of target detection: an electrocorticography study *Cerebral Cortex* **28** 9–20

Kanwisher N, McDermott J and Chun M M 1997 The fusiform face area: a module in human extrastriate cortex specialized for face perception *J. Neurosci.* **17** 4302–11

Kubanek J and Schalk G 2015 NeuralAct: a tool to visualize electrocortical (ECoG) activity on a three-dimensional model of the cortex *Neuroinformatics* **13** 167–74

Kubanek J, Brunner P, Gunduz A, Poeppel D and Schalk G 2013 The tracking of speech envelope in the human cortex *PloS One* **8** e53398

Kubanek J, Miller K, Ojemann J, Wolpaw J and Schalk G 2009 Decoding flexion of individual fingers using electrocorticographic signals in humans *J. Neural Eng.* **6** 066001

Lachaux J P, George N, Tallon-Baudry C, Martinerie J, Hugueville L, Minotti L, Kahane P and Renault B 2005 The

many faces of the gamma band response to complex visual stimuli *NeuroImage* **25** 491–501

Leuthardt E, Pei X M, Breshears J, Gaona C, Sharma M, Freudenburg Z, Barbour D and Schalk G 2012 Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task *Frontiers Human Neurosci.* **6** 99

Liu H, Agam Y, Madsen J R and Kreiman G 2009 Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex *Neuron* **62** 281–90

Liu Y, Coon W, de Pesters A, Brunner P and Schalk G 2015 The effects of spatial filtering and artifacts on electrocorticographic signals *J. Neural Eng.* **12** 056008

Logothetis N K, Pauls J, Augath M, Trinath T and Oeltermann A 2001 Neurophysiological investigation of the basis of the fMRI signal *Nature* **412** 150–7

Lotte F, Brumberg J S, Brunner P, Gunduz A, Ritaccio A L, Guan C and Schalk G 2015 Electrocorticographic representations of segmental features in continuous speech *Frontiers Human Neurosci.* **9** 97

Majima K, Matsuo T, Kawasaki K, Kawai K, Saito N, Hasegawa I and Kamitani Y 2014 Decoding visual object categories from temporal correlations of ECoG signals *NeuroImage* **90** 74–83

Makeig S, Westerfield M, Jung T P, Enghoff S, Townsend J, Courchesne E and Sejnowski T J 2002 Dynamic brain sources of visual evoked responses *Science* **295** 690–4

Manning J R, Jacobs J, Fried I and Kahana M J 2009 Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans *J. Neurosci.* **29** 13613–20

Maris E, van Vugt M and Kahana M 2011 Spatially distributed patterns of oscillatory coupling between high-frequency amplitudes and low-frequency phases in human iEEG *NeuroImage* **54** 836–50

Matsuo T *et al* 2015 Alternating zones selective to faces and written words in the human ventral occipitotemporal cortex *Cerebral Cortex* **25** 1265–77

Mazaheri A and Jensen O 2006 Posterior α activity is not phase-reset by visual stimuli *PNAS USA* **103** 2948–52

Mazaheri A and Jensen O 2008 Asymmetric amplitude modulations of brain oscillations generate slow evoked responses *J. Neurosci.* **28** 7781–7

McCandliss B D, Cohen L and Dehaene S 2003 The visual word form area: expertise for reading in the fusiform gyrus *Trends Cogn. Sci.* **7** 293–9

McCarthy G, Puce A, Belger A and Allison T 1999 Electrophysiological studies of human face perception. II: response properties of face-specific potentials generated in occipitotemporal cortex *Cerebral Cortex* **9** 431–44

Miller K J, Leuthardt E C, Schalk G, Rao R P, Anderson N R, Moran D W, Miller J W and Ojemann J G 2007 Spectral changes in cortical surface potentials during motor movement *J. Neurosci.* **27** 2424–32

Miller K J, Schalk G, Hermes D, Ojemann J G and Rao R P 2016 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 Comput. Biol.* **12** e1004660

Miller K J, Sorensen L B, Ojemann J G and den Nijs M 2009 Power-law scaling in the brain surface electric potential *PLoS Comput. Biol.* **5** e1000609

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I and Malach R 2005 Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex *Science* **309** 951–4

Müller-Gerking J, Pfurtscheller G and Flyvbjerg H 1999 Designing optimal spatial filters for single-trial EEG classification in a movement task *Clin. Neurophysiol.* **110** 787–98

Nagasawa T, Matsuzaki N, Juhász C, Hanazawa A, Shah A, Mittal S, Sood S and Asano E 2011 Occipital gamma-oscillations modulated during eye movement tasks:

Simultaneous eye tracking and electrocorticography recording in epileptic patients *NeuroImage* **58** 1101–9

Niessing J, Ebisch B, Schmidt K E, Niessing M, Singer W and Galuske R A 2005 Hemodynamic signals correlate tightly with synchronized gamma oscillations *Science* **309** 948–51

Parvizi J, Jacques C, Foster B L, Withoft N, Rangarajan V, Weiner K S and Grill-Spector K 2012 Electrical stimulation of human fusiform face-selective regions distorts face perception *J. Neurosci.* **32** 14915–20

Pei X, Barbour D L, Leuthardt E C and Schalk G 2011a Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans *J. Neural Eng.* **8** 046028

Pei X, Leuthardt E C, Gaona C M, Brunner P, Wolpaw J R and Schalk G 2011b Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition *NeuroImage* **54** 2960–72

Penny W D, Friston K J, Ashburner J T, Kiebel S J and Nichols T E 2007 *Statistical Parametric Mapping: the Analysis of Functional Brain Images* 1st edn (New York: Academic)

Potes C, Brunner P, Gunduz A, Knight R T and Schalk G 2014 Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing *NeuroImage* **97** 188–95

Potes C, Gunduz A, Brunner P and Schalk G 2012 Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening *NeuroImage* **61** 841–8

Ramoser H, Muller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6

Ray S and Maunsell J 2011 Different origins of gamma rhythm and high-gamma activity in macaque visual cortex *PLoS Biol.* **9** e1000610

Ray S, Niebur E, Hsiao S S, Sinai A and Crone N E 2008 High-frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention *Clin. Neurophysiol.* **119** 116–33

Rupp K, Roos M, Milsap G, Caceres C, Ratto C, Chevillet M, Crone N E and Wolmetz M 2017 Semantic attributes are encoded in human electrocorticographic signals during visual object recognition *NeuroImage* **148** 318–29

Schalk G, Kapeller C, Guger C, Ogawa H, Hiroshima S, Lafer-Sousa R, Saygin Z M, Kamada K and Kanwisher N 2017 Facephenes and rainbows: causal evidence for functional and anatomical specificity of face and color processing in the human brain *Proc. Natl Acad. Sci.* **114** 12285–90

Schalk G, Kubanek J, Miller K, Anderson N, Leuthardt E, Ojemann J, Limbrick D, Moran D, Gerhardt L and Wolpaw J 2007 Decoding two-dimensional movement trajectories using electrocorticographic signals in humans *J. Neural Eng.* **4** 264

Sederberg P B, Schulze-Bonhage A, Madsen J R, Bromfield E B, Litt B, Brandt A and Kahana M J 2007 Gamma oscillations distinguish true from false memories *Psychol. Sci.* **18** 927–32

Shum J, Hermes D, Foster B L, Dastjerdi M, Rangarajan V, Winawer J, Miller K J and Parvizi J 2013 A brain area for visual numerals *J. Neurosci.* **33** 6709–15

Sutton R S and Barto A G 1998 *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT Press) (https://doi.org/10.1016/S1364-6613(99)01331-5)

Tort A B, Kramer M A, Thorn C, Gibson D J, Kubota Y, Graybiel A M and Kopell N J 2008 Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task *PNAS USA* **105** 20517–22

Tsuchiya N, Kawasaki H, Oya H, Howard M A III and Adolphs R 2008 Decoding face information in time, frequency and space from direct intracranial recordings of the human brain *PLoS One* **3** e3892

van Erp J, Lotte F and Tangermann M 2012 Brain-computer interfaces: beyond medical applications *Computer* **45** 26–34

van Vugt M K, Schulze-Bonhage A, Litt B, Brandt A and Kahana M J 2010 Hippocampal gamma oscillations increase with memory load *J. Neurosci.* **30** 2694–9

Wang Z, Gunduz A, Brunner P, Ritaccio A L, Ji Q and Schalk G 2012 Decoding onset and direction of movements using electrocorticographic (ECoG) signals in humans *Frontiers Neuroeng.* **5**

Whittingstall K and Logothetis N K 2009 Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex *Neuron* **64** 281–9