NeuroImage 137 (2016) 132–139

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Decoding of task-relevant and task-irrelevant intracranial EEG representations

M.E. van de Nieuwenhuijzen [a,*], N. Axmacher [b,c,d], J. Fell [d], C.R. Oehrn [d], O. Jensen [a], M.A.J. van Gerven [a]

[a] Radboud University, Donders Institute for Brain, Cognition and Behaviour, 6500 HE, Nijmegen, The Netherlands
[b] Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, D-44801 Bochum, Germany
[c] German Center for Neurodegenerative Diseases, D-53175 Bonn, Germany
[d] Department of Epileptology, University of Bonn, D-53105 Bonn, Germany

ARTICLE INFO

Article history:
Received 22 February 2016
Revised 31 March 2016
Accepted 2 May 2016
Available online 3 May 2016

Keywords:
Intracranial EEG
Auditory perception
Multivariate analysis
Task-relevance

ABSTRACT

Natural stimuli consist of multiple properties. However, not all of these properties are equally relevant in a given situation. In this study, we applied multivariate classification algorithms to intracranial electroencephalography data of human epilepsy patients performing an auditory Stroop task. This allowed us to identify neuronal representations of task-relevant and irrelevant pitch and semantic information of spoken words in a subset of patients. When properties were relevant, representations could be detected after about 350 ms after stimulus onset. When irrelevant, the association with gamma power differed for these properties. Patients with more reliable representations of irrelevant pitch showed increased gamma band activity (35–64 Hz), suggesting that attentional resources allow an increase in gamma power in some but not all patients. This effect was not observed for irrelevant semantics, possibly because the more automatic processing of this property allowed for less variation in free resources. Processing of different properties of the same stimulus seems therefore to be dependent on the characteristics of the property.

Introduction

In daily life, we encounter many stimuli that are a conjugation of simple and complex properties. For example, single-word utterances contain information about, among others, pitch, loudness, semantics, and speaker identity. However, not all stimulus properties are equally relevant in each situation. Therefore, it could be argued that the relevant properties for the task at hand have to be attended to, whereas irrelevant properties do not require as much attention.

Selectively attending to a stimulus or a specific stimulus property has been shown to modulate its neuronal processing. For example, the amplitude of event-related potentials (ERPs) in electroencephalography data decreases when an auditory stream is unattended (Hillyard et al., 1973; Näätänen et al., 1992; Woldorff and Hillyard, 1991), and brain activity measured with functional magnetic resonance imaging (fMRI) increases in regions corresponding to the attended property (Degerman et al., 2006; Downar et al., 2001; Johnson and Zatorre, 2006; Paltoglou et al., 2009). In the visual domain, this biases neuronal population activity as measured by fMRI such that specifically the attended property can be decoded from an ambiguous stimulus (Jehee et al., 2011; Kamitani and Tong, 2005, 2006; Niazi et al., 2014). Finally, neuronal firing rates of monkey single-unit recordings increase with selective spatial attention (Benson and Hienz, 1978).

Selective attention to one particular stimulus feature in the visual domain is associated with pronounced modulation of the neuronal response by task-relevance (e.g. Davidesco et al., 2013; Harel et al., 2014). However, it remains unclear whether the unattended property is still represented throughout the entire task interval between stimulus onset and task-related response. Therefore, in this study we explored the neuronal representations of the stimulus properties 'pitch' and 'semantics' of single spoken words throughout an auditory Stroop task, when these properties were either relevant (attended to) or irrelevant (unattended) (Haupt et al., 2009; Oehrn et al., 2014). We examined to what extent the identity of these properties are still represented when this identity was not relevant for the task at hand. Studies on feature-based attention (e.g. Krumbholz et al., 2007; O'Craven et al., 1999) suggest that unattended stimulus properties do not necessarily remain unprocessed. Thus, we hypothesise that although the specific identity of a stimulus property may not be relevant for the task, this identity is still processed and represented in the brain to a certain extent.

Since such neuronal representations of irrelevant stimulus properties are likely relatively weak, a method with high signal-to-noise ratio is required. We thus re-analysed existing intracranial electroencephalography (iEEG) data of a large group of patients (n = 21). These data have a superior signal-to-noise ratio compared to conventional scalp recordings, as well as a high temporal and spatial resolution,

* Corresponding author at: Radboud University, Donders Institute for Brain, Cognition and Behaviour, P.O. box 9104, 6500 HE, Nijmegen, The Netherlands.
E-mail address: m.vandenieuwenhuijzen@donders.ru.nl (M.E. van de Nieuwenhuijzen).

http://dx.doi.org/10.1016/j.neuroimage.2016.05.008
1053-8119/

because they are measured close to the brain without interference from the skull.

Conventionally, Stroop data, in which stimuli are presented with properties that are conflicting in meaning (such as the word 'high' spoken in a low pitch), are analysed based on this conflict between these stimulus properties. In this study, however, we focused on the effect of task-relevance on the representation of these properties. The amount of competition between the relevant (attended) and the irrelevant (unattended) property caused by the auditory Stroop paradigm depends on the automaticity with which the irrelevant property is processed. Here, the property 'semantics' is thought to be processed more automatically than the property 'pitch' (Haupt et al., 2009; Oehrn et al., 2014). It has been suggested, that when two conflicting properties are presented, activity related to the task-irrelevant property is suppressed (e.g. Iguchi et al., 2005; Liu et al., 2016; Mansouri et al., 2009; Polk et al., 2008). Because of its previously described role in attention, we hypothesise power in the alpha band to be related to this top-down suppression (8–12 Hz; e.g. Jensen et al., 2012; Jensen and Mazaheri, 2010; Klimesch et al., 2011; Klimesch, 2012). Furthermore, as gamma power (35–64 Hz) has also been associated with attentional processes, (Brovelli et al., 2005; Fries et al., 2008; Tallon-Baudry et al., 2005) and has been suggested to be increased in some cases related to an unattended stimulus (Martinovic et al., 2009), we also expect a role of this frequency band in the way different levels of task-relevance are represented.

We found that pitch and semantics of a spoken word were represented both when these features were relevant and when they were irrelevant. Furthermore, patients with a higher decodability of unattended pitch representations tended to have larger increases in gamma power. This effect was not observed for unattended semantics. Therefore, the specific relationship the representation of an irrelevant property has with frequency power seems to differ depending on the property itself.

Materials and methods

Patients and paradigm

Intracranial EEG data of 22 patients with pharmacologically intractable epilepsy (mean age 36.4, $sd = 13.6$; 14 males; 21 right-handed, one ambidextrous) were recorded, while these patients performed an auditory version of the Stroop task (Haupt et al., 2009; Oehrn et al., 2014). One patient was excluded from further analysis because all responses in one condition were incorrect, resulting in 21 patients (mean age 35.7, $sd = 13.6$; 13 males; 20 right-handed, one ambidextrous).

The paradigm, as well as the data of 13 patients, was the same as described by Oehrn et al. (2014). Patients were presented with the German words for 'high' and 'low', spoken in either a high- or low-pitched male voice. The word meaning and the corresponding pitch either matched (congruent trials: the word 'high' spoken in a high-pitched voice, or the word 'low' spoken in a low-pitched voice) or were reversed (incongruent trials: the word 'high' spoken in a low-pitched voice, or the word 'low' spoken in a high-pitched voice). In addition, in control trials the German word for 'good' was spoken with either a high or a low pitch.

In one of two blocks, patients had to indicate whether the pitch was high or low, regardless of the word's semantics (pitch task). In the other block, they had to indicate whether the word meaning was 'high' or 'low', regardless of the pitch (semantic task). Four patients started with the pitch task. The rest of the patients started with the semantic task. Each block consisted of 40 congruent, 40 incongruent, and 40 control trials, which were randomly presented throughout the block. In each of these conditions the pitch was high in half of the trials. Responses were given by left and right button presses with the dominant hand. Response mapping was counter-balanced between participants. For the control trials no response was required in the semantic task. Only trials in which the response was correct were included in the analyses.

In each trial the spoken word was presented for 0.5 s, during which the response task was shown on a screen. After the word was spoken, the task instructions remained on screen for an additional 2 s, during which patients were still allowed to respond. Trials were separated by a variable inter-trial interval of 1.5–3.3 s, while a fixation cross was presented in the centre of the screen. Stimuli were presented with Presentation software (Version 14.5, Neurobehavioral Systems Inc.). This design has been shown previously to evoke a typical Stroop effect both in healthy subjects (Haupt et al., 2009), and in epileptic patients with implanted electrodes (Oehrn et al., 2014).

The digitized sound files, all voiced by the same male experimenter, were transposed either to a low or high pitch, such that the interval between the low- and high-pitched words was a fifth on the musical interval scale. The words were aligned with the Entropic Timescale Modification function of the GoldWave audio editing software (http://www.goldwave.com/) to ensure an equal length of 0.5 s.

Intracranial recordings

Depending on clinical criteria, patients were either implanted with subdural or depth electrodes or both, for diagnosis of the focus of pharmacologically intractable epilepsy. Subdural electrodes were made of stainless steel and consisted of strips or grids with a contact diameter of 4 mm and a centre-to-centre spacing of 10 mm. Depth electrodes had a diameter of 1.3 mm and contained cylindrical platinum electrodes of 2.5 mm every 4 mm. Electrodes were located over the frontal and temporal lobe, including the medial temporal lobe and hippocampus, with some strips extending into parietal and occipital areas (Fig. 1A). The location of the electrodes was dependent on the suspected epileptic focus. The data were recorded at a sampling frequency of 1000 Hz, referenced to linked mastoids and band-pass filtered from 0.01 Hz to 300 Hz, using the digital EPAS system (Schwarzer, Munich, Germany) and Harmonie EEG software (Stellate, Montreal, Canada). Measurements were performed in the Klinik für Epileptologie in Bonn, Germany. The study was approved by the local ethics committee. All patients gave written informed consent before participating in the study.

Artefact rejection

Electrodes located over the epileptic focus were excluded. The remaining data were further inspected visually with BrainVision Analyser 2 (Brain Products). Electrodes that showed more than occasional epileptic activity, such as spikes and high frequency high amplitude bursts, were excluded as well. The electrodes that were included for analysis are shown in Fig. 1A. In total, 432 electrodes were included (mean per patient = 20.67, $sd = 8.91$). Finally, trials in which artefacts were visually detected in the remaining electrodes were rejected. On average, this resulted in 194.19 ($sd = 23.76$) trials per patient that remained for analysis. Data were then exported and further analysed using MATLAB version 8.1.0.604, R2013a (The Mathworks Inc.) and FieldTrip, an open source Matlab toolbox for the analysis of neuroimaging data (Oostenveld et al., 2011).

Preprocessing

The data were low-pass filtered at 100 Hz with a Butterworth IIR filter, and a 50 Hz notch filter was applied to remove the line noise. After this, baseline correction was performed relative to the period between stimulus onset and 200 ms before. Subsequently, the data were downsampled to 300 Hz to reduce memory and CPU load. These time-domain data were used for further classification analyses.

For additional correlation analyses between classification accuracy and power in various frequency bands, time-frequency representations were calculated for the alpha (8–12 Hz) and gamma (35–64 Hz)

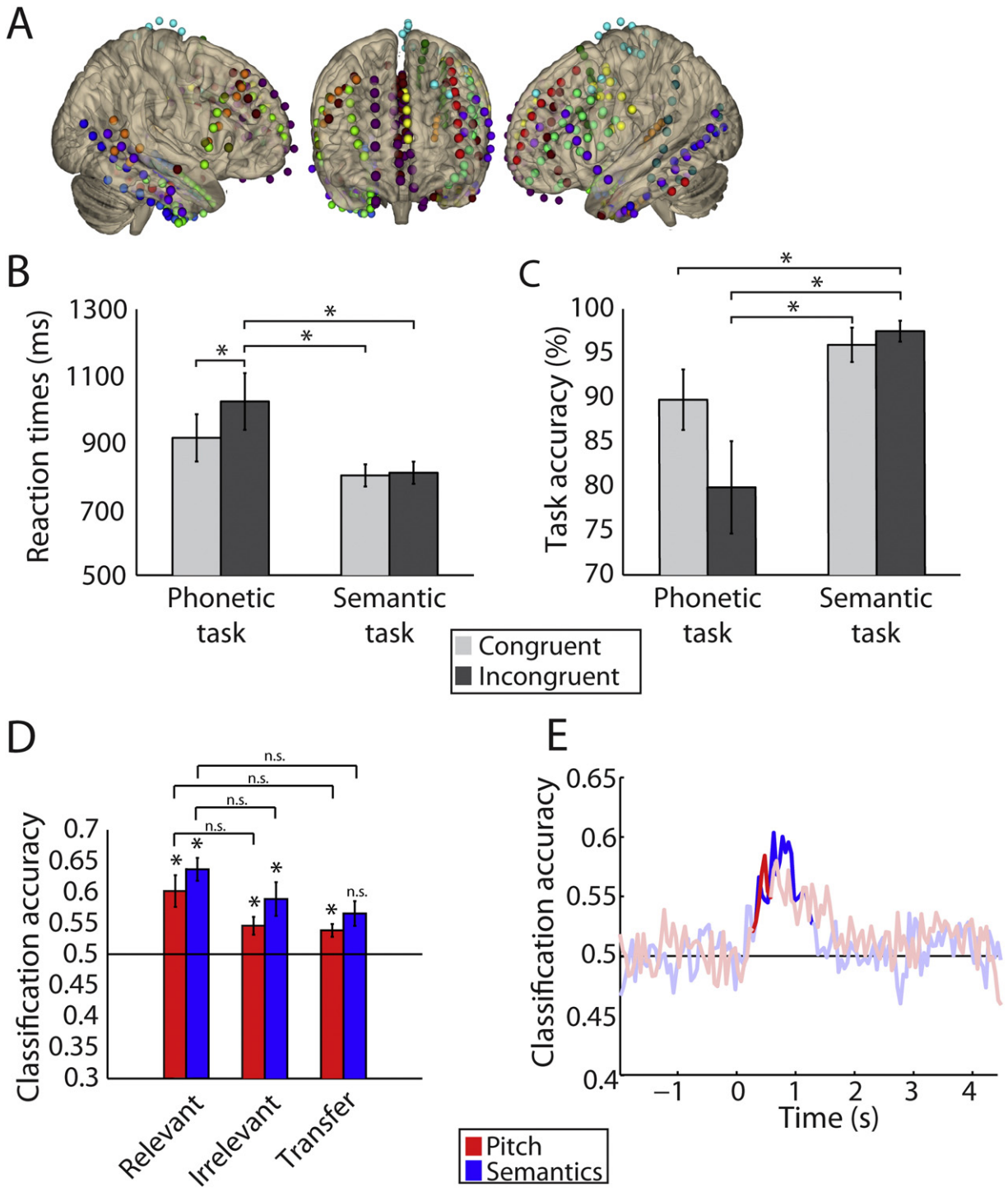

Fig. 1. Electrode coverage, behavioural results and classification accuracy. A) Location of all electrodes that passed artefact rejection and were used in further analyses. Different colours indicate electrodes of different patients. B & C) Reaction times (B) and task accuracy (C) for the different conditions (congruent and incongruent) and tasks (pitch task and semantic task). Asterisks indicate contrasts that were significant at $p < 0.05$ (Bonferroni corrected). D) Group averaged accuracies for pitch (red) and semantics (blue) when these features were task-relevant, irrelevant, and when the classifier was trained on the relevant property and tested on the irrelevant property (transfer). The black horizontal line indicates chance level. An asterisk indicates classification accuracies that are significant above chance level (Bonferroni corrected). Contrasts indicated with n.s. are not significant when correcting for multiple comparisons. E) Averaged classification accuracies over time. The red line indicates the classification accuracy trace for pitch, the blue line represents the trace for semantics. Stimulus onset was at zero seconds. The black horizontal line indicates chance level. The bright part of the blue line indicates the time points at which the classification accuracy trace for semantics significantly exceeded the baseline traces (before stimulus onset).

frequency band. Frequency bands had a resolution of 1 Hz, and were calculated using a Fourier analysis applied to sliding time-windows with a step size of 50 ms and an adaptive length, such that each window contained four cycles of the frequency of interest. These windows covered data ranging from 2 s before stimulus onset to 4.5 s after stimulus onset. Hanning tapers were applied to the data before Fourier analysis to smooth the data. To account for individual power differences due to, for example, electrode positioning, a relative baseline correction of the power spectrum was then performed against a baseline period of 500 to 200 ms before stimulus onset, as this period was not already

used for baseline correction of the raw signal. Power values of each patient were then averaged over trials, electrodes and time period between 0 and 1500 ms after stimulus onset to match the time-domain data to which the classifier was applied. This resulting power value was then correlated to classification accuracy (see below).

Furthermore, to assess whether these correlation effects are selective to the alpha and gamma band or whether they extend into other frequencies as well, successive frequency bins of 4 Hz each were computed following the same method as described above. This resulted in one power value averaged over time, electrodes and trials per patient for each frequency bin. The centre frequencies of these bins ranged from 3 to 88 Hz, and were shifted with a step size of 1 Hz. Power in each bin was correlated to classification accuracy (see below).

Classification

A linear support vector machine (SVM) algorithm was used for classification. The soft margin parameter *C*, which acts as a regularizer, was set to a default value of

$$C = 0.1 \cdot \left(\frac{1}{N} \sum_{i=1}^{N} k_{ii} - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} \right)$$

where *K* is a linear kernel such that k_{ij} is the inner product between trials *i* and *j*, and *N* is the total number of trials that constitute the training data. In short, this algorithm trains a classification model to optimally discern between data of two classes (for example between trials with a high pitch and trials with a low pitch). It then uses a subset of the data that was not used during the training phase to test how well the model generalizes, which prevents circularity ('double dipping'; Kriegeskorte et al., 2009).

Classifier performance was quantified in terms of accuracy (proportion of correctly classified trials). Because this is a binary classification problem, and because each class is equally likely to occur, chance level is 0.5. Classification accuracies at chance level indicate that the classifier is unable to differentiate between the two classes based on the classification model obtained during training. However, if the classifier performs above chance level this model is informative about the neuronal representations of the stimulus properties that are being discerned.

The number of trials included during classifier training and testing was always balanced, such that it was equal for both classes. Training and testing of the classifier was done on time-domain data. The feature space consisted of all selected electrodes per patient and all time points between stimulus onset and 1500 ms thereafter. Note that we did not apply pattern classification analysis to time-frequency-domain data, i.e. we did not assess whether a property was represented in the frequency domain or in a specific frequency band.

This algorithm was first applied to the relevant properties (i.e. discerning trials with a low pitch from trials with a high pitch in the pitch task, and discerning trials in which the word 'low' was spoken from trials in which the word 'high' was spoken the semantic task). Pitch classification was performed on all trials of the pitch task. Semantics classification was performed on all trials of the semantic task apart from the control trials. In addition, we applied this algorithm to the respective *irrelevant* properties (i.e. pitch in the semantic task, and semantics in the pitch task). Here, pitch classification was performed on all trials of the semantic task block (including control trials). Classification of semantics was performed on all trials in the pitch task, apart from the control trials.

Furthermore, we trained the classifier on trials in which a property was relevant, and tested the model on trials in which that property was *not* relevant (e.g. training on low versus high pitch discrimination in the pitch task and testing on low versus high pitch discrimination in the semantic task). This transfer learning can be interpreted as a test of similarity of the neuronal signals: if the neuronal signal of task-relevant pitch is similar to the neuronal signal of irrelevant pitch, the test data will fit the classification model that was generated based on the training data. In this case, classification accuracy will be high. However, if the signals of the train and test data differ too much for the testing data to fit the training model, the resulting classification accuracy will be at chance level.

In the previously mentioned analyses, the classification algorithm was applied to time-domain data from stimulus onset to 1500 ms afterwards as a whole, i.e. resulting in one classification accuracy per patient for this entire time window. To obtain a more fine-grained temporal representation of classification accuracy, data from consecutive windows of 50 ms were averaged over time and each window was used as an input to the classification algorithm, resulting in a feature space for each window of one averaged time-domain amplitude per electrode. These 50 ms windows spanned the interval of 2 s before stimulus onset to 4.5 s after stimulus onset with the outer 0.5 s potentially overlapping with the previous or next trial due to the variable inter-trial interval. This resulted in 130 classification accuracy values for each patient (130 windows of 50 ms spanning a total of 6500 ms), forming a trace of accuracies over time which indicated when class information was detectable from the iEEG signal.

To assess the relationship between power in different frequency bands and the ability of the classifier to detect the representations of the irrelevant stimulus properties in the time-domain, classification accuracies obtained for the irrelevant properties were correlated with power in the different frequency bands mentioned above (see Preprocessing). Power was averaged over all electrodes and over time between stimulus onset and 1500 ms thereafter to match the data used to obtain the classification accuracies. Also, only the trials that were used in the calculation of these classification accuracies were included in the average power calculation. Spearman correlation (ρ) was used to correlate the classification accuracies with power in the different frequency bands.

Using a classification method allows for a dissociation between the representations of relevant pitch and of irrelevant semantics and vice versa, even though these representations exist in the same trial. After all, in each trial of the pitch task, pitch is relevant and semantics are irrelevant. However, because both incongruent and congruent trials were included (pitch congruent mean per patient = 15.9, $sd = 3.15$; pitch incongruent mean per patient = 14.52, $sd = 4.58$; semantic congruent mean per patient = 17.26, $sd = 2.55$; semantic incongruent mean per patient = 17.4, $sd = 2.07$), the class of the relevant property (e.g. high or low pitch), was not consistently related to the class of the irrelevant property (e.g. 'high' or 'low' semantics). Therefore, the classification contrast made on one property (high versus low relevant pitch), is unrelated to the contrast of the other property ('high' or 'low' unattended semantics), disentangling the representations of relevant pitch from irrelevant semantics and vice versa. In addition, this renders it unlikely that semantics are driving pitch classification or that pitch drives classification on semantics.

Statistical analysis

As the data were not normally distributed, we used non-parametric testing. One-sample and pairwise tests were performed using the Wilcoxon signed-rank test. Multiple comparisons were in these cases corrected for using Bonferroni correction. For correlations over multiple frequency bands, multiple comparisons were corrected for with the FDR.

For classification over time, the classification accuracies after stimulus onset were compared to the baseline accuracies before stimulus onset and corrected for multiple comparisons using cluster-based permutation testing implemented in FieldTrip (Maris and Oostenveld, 2007). In short, this method tests the largest sum of neighbouring *t*-values whose corresponding *p*-value exceeded a threshold of 0.05

against the maximum sum obtained when class labels were reshuffled randomly for 500 permutations.

Results

Behavioural results

First, we assessed whether the task indeed induced a Stroop-like effect, as has been observed before by Oehrn et al. (2014) and Haupt et al. (2009). Pair-wise Wilcoxon signed-rank tests revealed a slowing in reaction times for incongruent pitch trials compared to congruent pitch trials ($Z = 2.83$, $p = 0.005$), as well as slower reaction times on incongruent pitch trials compared to congruent ($Z = 3.18$, $p = 0.001$) and incongruent semantic trials ($Z = 3.04$, $p = 0.002$; Fig. 1B). The other three contrasts were not significant (all $Z < 2.45$, all $p > 0.01$, Bonferroni corrected α for six contrasts is 0.008). Similarly, task accuracy was lower on incongruent pitch trials compared to congruent ($Z = 3.10$, $p = 0.002$) and incongruent semantic trials ($Z = 3.18$, $p = 0.001$; Fig. 1C). Furthermore, task accuracy was lower in congruent phonetic trials than in semantic incongruent trials ($Z = 2.76$, $p = 0.006$). The other three contrasts were not significant (all $Z < 2.58$, all $p > 0.01$, Bonferroni corrected α for six contrasts is 0.008). This larger Stroop effect for pitch discrimination suggests that semantics is the more automatic process in terms of the auditory Stroop task, as has been suggested before with this paradigm (Haupt et al., 2009; Oehrn et al., 2014).

Pitch and semantic information can be decoded from single-trial neuronal representations

We applied the classification algorithm to trials in which pitch was relevant (i.e. the pitch task) to distinguish between high and low pitch based on time-domain data from stimulus onset to 1500 ms afterwards. Additionally, we applied this same method to trials in which semantic content was relevant (i.e. during the semantic task) to distinguish between the word meanings 'high' and 'low'. As can be seen in Fig. 1D, classification accuracies rose above chance level for both pitch (mean classification accuracy: 0.64 ± 0.08; mean classification accuracy $\pm$ *sd*: 0.60 ± 0.12; $Z = 3.02$, $p = 0.003$) and semantics ($Z = 3.92$, $p < 0.0001$). Classification accuracy was not related to the number of features used for classification (pitch: $\rho = 0.37$, $p = 0.10$; semantics: $\rho = 0.38$, $p = 0.09$).

Temporal distribution of relevant property representations

Classification over consecutive time bins showed classification accuracies above chance level, as quantified by the baseline before stimulus onset, from 325 to 525 ms after stimulus onset for pitch (cluster with largest summed *t*-values $t(18) = 16.66$, $p = 0.004$). For semantics, classification accuracies were found to be above chance level from 375 to 1225 ms after stimulus onset (cluster with largest summed *t*-values $t(18) = 67.10$, $p = 0.002$). Average peak classification accuracies (pitch: mean = 0.58, *sd* = 0.07; semantics: mean = 0.60, *sd* = .08) were reached at 475 ms and 625 s, respectively (see Fig. 1E). No differences were detected between these time courses (cluster with largest summed *t*-values: $t(18) = 2.95$, $p = 0.40$).

Above-chance classification accuracies for irrelevant properties

Next, to test to what extent a property remains represented in the brain when it is not relevant to the task, we applied the classification algorithm to the irrelevant property, i.e. classifying pitch during the semantic task and vice versa. Average classification accuracies for irrelevant pitch classification remained above chance level ($Z = 2.94$, $p = 0.003$) and were not decreased compared to classification accuracies for relevant pitch (0.55 ± 0.07; $Z = 2.57$, $p = 0.01$; not significant when Bonferroni corrected for 10 multiple comparisons: six tests comparing accuracies to chance level and four tests comparing between group level accuracies).

Similarly, average classification accuracies of irrelevant semantics remained above chance level ($Z = 2.84$, $p = 0.0045$) and were not reduced compared to classification accuracies obtained for relevant semantics (0.59 ± 0.12; $Z = 1.36$, $p = 0.18$; see Fig. 1D). Again, classification accuracy was not related to the number of features used for classification (pitch: $\rho = 0.23$, $p = 0.32$; semantics: $\rho = 0.46$, $p = 0.04$; not significant when Bonferroni corrected for four multiple comparisons).

Representations of irrelevant properties are similar to representations of relevant properties

Next, we compared the representations of relevant and irrelevant properties. We trained classifiers on trials in which a stimulus property was relevant, and tested on trials in which that same property was irrelevant. For pitch information, transfer learning accuracies were above chance level (0.54 ± 0.05; $Z = 2.84$, $p = 0.004$), indicating a commonality in representations of relevant and irrelevant pitch. Furthermore, compared to the classification accuracies obtained for relevant pitch, transfer accuracies did not decrease ($Z = 1.93$, $p = 0.05$; see Fig. 1D). This suggests that there is no difference when testing on relevant or irrelevant pitch with a model trained on relevant pitch, supporting the notion that the representation of relevant and irrelevant pitch does not differ.

For semantic information, transfer learning accuracies only just failed to rise above chance level (0.57 ± 0.09; $Z = 2.80$. $p = 0.0051$; not significant when Bonferroni corrected for 10 multiple comparisons, see above). However, this accuracy was not reduced compared to classification accuracies obtained for relevant semantics ($Z = 2.52$, $p = 0.01$; not significant when Bonferroni corrected for 10 multiple comparisons, see above). This suggests that there may be a small difference in the representation of relevant and irrelevant semantics. However, we did not have sufficient power to test where these differences would originate if they would indeed exist.

It is unlikely that classification was driven by the motor response. After all, the motor response was the same regardless of which property was relevant (pitch or semantics), but there was no relationship between classification accuracies when the pitch or semantics was relevant ($\rho(19) = 0.04$, $p = 0.86$). Moreover, of the six patients that had individual classification accuracies above chance level for pitch and of the four patients that had above-chance classification accuracies for semantics, only one patient had a classification accuracy above chance for both pitch and semantics. Classification accuracies seem therefore quite different for pitch and semantics, which is not what would be expected if the common motor response would play a role. Finally, as the motor response was only related to the relevant property and not to the irrelevant property, a motor effect would imply that classification on the irrelevant property would not be possible or at least would do worse than classification on the relevant property. However, as shown above, this was not the case, rendering an effect of the motor response on classification unlikely.

Classification accuracy of irrelevant properties is differentially related to gamma power

Finally, we assessed the relationship between classification accuracies of irrelevant properties and power in the alpha and gamma frequency band. Power in the gamma band seemed to be behaviourally relevant, as patients that showed increased power had faster reaction times ($\rho(19) = -0.55$, $p = 0.01$; Fig. 2A). This was not the case for alpha (alpha: $\rho(19) = -0.44$, $p = 0.05$). Note that for average power and reaction times no clear distinction could be made between relevant pitch and irrelevant semantics in the pitch task and vice versa in the semantic task. After all, unlike for classification, no contrast was made

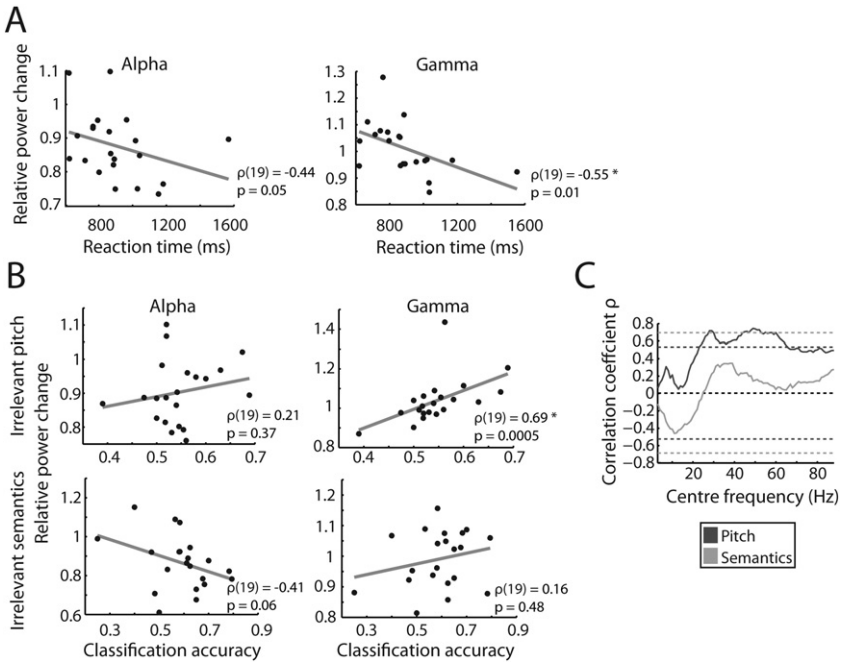

Fig. 2. Relation between irrelevant property representations and power in the alpha and gamma band. A) Correlations over patients of power change in the alpha (left plot) and gamma band (right plot), with average reaction times over all trials. An asterisk at the correlation coefficient ρ indicates a Bonferroni corrected significant correlation. B) Correlations over patients of classification accuracies of irrelevant pitch (upper plots) and irrelevant semantics (lower plots), with power changes in the alpha (left plots) and gamma frequency band (right plots). An asterisk at the correlation coefficient ρ indicates a Bonferroni corrected significant correlation. C) Correlations over patients of classification accuracy of irrelevant pitch (dark grey line) and semantics (light grey line), with power in consecutive frequency bands with a width of 4 Hz and centre frequencies between 3 and 88 Hz. The dotted black line indicates a correlation of zero. The dotted dark grey (pitch) lines indicate the threshold for FDR-corrected significant correlations between unattended pitch accuracies and the various frequency bands. The light grey dotted lines (semantics) indicate the estimated threshold for FDR-corrected significant correlations between unattended semantics accuracies and the various frequency bands. Because none of the correlations with classification accuracies for semantics were significant, the correlation corresponding to the FDR-corrected alpha value for semantics was estimated based on the pitch correlation with a *p*-value closest to this FDR-corrected alpha value. Note that this means that for the correlation between frequencies and classification accuracies for unattended semantics accuracies none of the correlations were significant, whereas the correlations based on unattended pitch were significant between 23 and 63 Hz.

based on high and low pitch and 'high' and 'low' semantics to tease these cases apart.

High classification accuracies of irrelevant properties suggest that these properties are still detectable from the iEEG signals, whereas classification accuracies at chance level indicate that the representation cannot be detected. A high classification accuracy, hence a clear representation of the irrelevant property, may have resulted from a failure in the suppression of this representation or a faulty recruitment of attentional resources towards it. To test this hypothesis, we correlated the patient-specific classification accuracies of irrelevant properties with the individual power in the alpha and gamma band (Fig. 2B). Classification accuracies for irrelevant pitch were significantly correlated with activity in the gamma ($\rho(19) = 0.69$, $p = 0.0005$) but not alpha frequency range ($\rho(19) = 0.21$, $p = 0.37$). By contrast, for irrelevant semantics, classification were neither correlated with alpha ($\rho(19) = -0.41$, $p = 0.06$) nor with gamma power ($\rho(19) = 0.16$, $p = 0.48$). These results show that in patients in which irrelevant pitch was better detectable, gamma power was higher.

As shown in Fig. 2C, the correlations in the gamma band did not carry over to lower frequencies, although it did seem to include part of the higher beta band. This analysis revealed that the effect of the correlation with gamma power extended between centre frequencies of 23 Hz ($\rho(19) = 0.53$, $p = 0.01$) and 63 Hz ($\rho(19) = 0.54$, $p = 0.001$), with the highest correlation at 49 Hz ($\rho(19) = 0.74$, $p = 0.0001$; all *p*-values FDR-corrected). This may seem odd given the notch filter at 50 Hz. However, it should be kept in mind that frequency bins with a width of 4 Hz were used, meaning that for this peak correlation frequencies below 50 Hz (47–51 Hz) were also included. Furthermore, as the plateau of high correlations extended widely beyond 50 Hz, it is unlikely that these effects were only due to any residual effect of line noise.

Assessing correlations between classification accuracy and frequency power in either frontal or temporal electrodes yielded no further distinction in correlation significance for different brain areas (all absolute $\rho(19) < 0.64$, all $p > 0.02$; not significant when Bonferroni corrected for four multiple comparisons). Moreover, no significant correlations were observed between classification accuracies of relevant properties and either alpha or gamma band activity (all absolute $\rho(19) < 0.20$, all $p > 0.39$), suggesting that the correlation between gamma power and classification accuracies of irrelevant pitch is specific to the irrelevant representation. Finally, we found no significant correlations between reaction time and classification accuracy of either property, both relevant and irrelevant (all absolute $\rho(19) < 0.28$, all $p > 0.23$).

Discussion

In this study we investigated the effect of task-relevance on the representation of pitch and semantics of spoken words. We found that pitch and semantics were represented in the brain not only when the property was relevant, but also when this was irrelevant. Furthermore, we observed that an individual's gamma power was related to the stability of irrelevant pitch representations. This suggests a differential effect of task-relevance on the representation of stimulus properties, which we hypothesise could be operationalized by attention, depending on the automaticity of the processing of these properties.

Neuronal representations of relevant properties

We found that relevant property representations were accurate enough to allow for stable single-trial distinctions between the two classes (high versus low pitch, and 'high' versus 'low' semantics; Fig. 1D). Furthermore, the temporal representation of these relevant properties started with a sharp increase in accuracy, followed by a peak around 475–625 ms before returning to baseline (Fig. 1E). This peak seems to occur just before the time of conflict resolution of the Stroop task

(about 725 ms after stimulus onset; Oehrn et al., 2014). It may well be that the signals before accuracies rise above chance level are a mixture of both the relevant and irrelevant property, which could be difficult to disentangle. After 300–500 ms however, the properties may start to be segregated into a relevant and irrelevant information stream, enabling above-chance classification. After this segregation it is then possible to resolve the conflict.

This relatively late onset of decodability could in addition be explained by electrode placement. As no electrodes were located over the primary auditory cortex (Brodmann area 41), and only 11 electrodes (3.2%) were located over the remainder of the auditory cortex (Brodmann area 22; no electrodes were located over Brodmann area 42), it is unlikely that consistent differences in early neuronal representations could be detected in this study. Therefore, the seemingly late response we observed may be related to predominantly higher-order task-specific processing in temporal and frontal cortex. It should be noted that although it was not possible with this electrode placement to look at early low-level representations of pitch, this placement did allow us to assess the representation of pitch and semantics throughout the remainder of the task period in terms of the task and task-relevance.

Finally, it could well be that information was present during a longer time period than we detected. As the presence of information was defined by whether the classification accuracies were higher than those during baseline, this definition is largely dependent on the signal-to-noise ratio of the underlying accuracy traces. We observed relatively low classification accuracies, which could be explained by the heterogeneity of electrode localization. After all, classification accuracies from patients with electrodes only over non-informative brain areas are unlikely to rise above chance level. This would result in a relatively large variance between patients, as well as decreased average classification accuracies. In turn, this could result in accuracies not being judged to be above chance level, although this may actually be the case for some patients.

Differential effects of irrelevant property representations

Although classification of a distracter stimulus based on fMRI data has yielded average classification accuracies at chance level for the visual domain (Woolgar et al., 2015), we showed that representations of irrelevant properties are detectable, comparable to the representations of the relevant properties (Fig. 1D). Irrelevant pitch and semantics are therefore likely still being processed. This is in line with what we hypothesised based on studies on feature-based attention showing that when one property of a stimulus is attended to, another irrelevant or unattended property is still processed to a certain extent (Krumbholz et al., 2007; O'Craven et al., 1999).

Although the classification accuracies for relevant and irrelevant properties were not found to differ, this does not necessarily mean that they are represented similarly. Therefore, we tested for a possible similarity in representations by training on the relevant property and testing on the irrelevant property. Although we did not detect the classification accuracies for this transfer learning to be decreased compared to the accuracies obtained for the relevant property, the transfer learning accuracy for semantics did not rise significantly above chance level. This could be a result of the relatively low signal-to-noise ratio in this study, but this may just as well mean that the representations of semantics differ when they are relevant as compared to irrelevant. Future research should assess whether there is indeed a difference in these representations, and if so, activation patterns in which brain regions drive these differences.

In addition, we observed a correlation between individual gamma power and classification accuracies for irrelevant semantics, but not irrelevant pitch (Fig. 2B). Gamma power has been suggested to play a role in attention (Brovelli et al., 2005; reviewed by Fell et al., 2003; Tallon-Baudry et al., 2005). Furthermore, especially low gamma has been thought to decrease for unattended stimuli (Pitts et al., 2014; Sokolov et al., 2004), and is thought to play a role in active suppression of this irrelevant stimulus (Sokolov et al., 2004). If this decrease in gamma would not occur, or to a lesser extent, then this suppression may be lifted and the property would be detectable by classification again. This is indeed what we observe in our correlation results. However, what would underlie this differentiation in gamma power? Martinovic et al. (2009) found that in some cases gamma power related to an unattended stimulus did increase. This was mainly the case when that stimulus was familiar during a task with a low load. It may well be that in the current study some subjects had resources to spare for the familiar pitch property during the easier semantic task, while others did not. During the more difficult pitch task, however, there were fewer resources to spare, hence gamma could not increase as easily in response to the irrelevant semantic property.

The correlation with classification accuracies of irrelevant pitch is first observed at 23 Hz. Although low gamma has sometimes been defined to start as early as 20 Hz, and hence could be explained in the same framework as above, this could also be regarded as the high beta band. An increase in this frequency has been hypothesised to signify a decrease in mental flexibility in extreme cases (Engel and Fries, 2010). Potentially this is the case in some of the patients in this study, resulting in an inability to suppress irrelevant pitch, even though it should be relatively easy to do so. This effect may be specific to irrelevant pitch, as semantics are thought to be the more automatic process in this paradigm, rendering it difficult to suppress in all subjects and thus reducing variability based on beta power.

Alternatively, it may be that the correlation with gamma power is related to gamma activity as a correlate of pitch, as previous studies have shown activity in the high gamma range (80–120 Hz) during pitch perception (e.g. Kumar and Schönwiesner, 2012; Sedley et al., 2012), as well as around 40 Hz (e.g. Crone et al., 2001; Ross et al., 2005). Activity in the gamma band would then signify the processing of pitch, and hence influence to what extent pitch could be decoded. After all, the better irrelevant pitch is processed, the easier the classification algorithm can distinguish the two classes that make up the pitch category. However, the correlation we observed between classification accuracies for irrelevant pitch and gamma power was strongest for frequencies below 70 Hz, as opposed to the higher frequencies that have been related to pitch processing (Fig. 2C). Furthermore, if gamma band activity was indeed an indication of the processing of pitch, we would also observe a correlation between gamma and classification accuracies for *attended* pitch. This, however, was not the case, suggesting that the correlation with gamma band activity was not related to pitch processing per se.

As semantics has been thought to be the more automatically processed property, one could argue that in order to successfully perform the auditory Stroop task, this automatic processing has to be suppressed actively. This is in line with the theory that when two conflicting properties are presented, activity of the irrelevant property is suppressed (Mansouri et al., 2009). Alpha oscillations are thought to serve as a mechanism for active suppression of task-irrelevant processes (e.g. Jensen et al., 2012; Jensen and Mazaheri, 2010; Klimesch et al., 2011; Klimesch, 2012). However, we only observed a non-significant trend that patients with lower levels of alpha power, hence less suppression, showed a larger detectability of the representations of irrelevant semantics by the classifier (Fig. 2B). This absence of a significant correlation could be due to the high inter-patient variability and small number of data points, possibly destabilizing a minor correlation. This hypothesis is supported by the overall strength of the correlations for frequencies surrounding the alpha band instead of a strong correlation only for the alpha band itself. Further research should validate these trends to determine whether alpha oscillations could regulate the amount of automatic processing of semantics.

Differences in task difficulty may affect the attentional modulation of neuronal activity (Altmann et al., 2008), and indeed, in this study the semantic task was easier than the pitch task, in line with the observed

Stroop effect. However, it can be argued that task difficulty is in fact related to the automaticity with which a property is processed. After all, the more automatically a property is processed, the easier it will be extracted for processing, and the harder it is to suppress that property when it is irrelevant, rendering the task more difficult. Along these lines, the pitch task would be more difficult simply because the irrelevant semantic property is processed more automatically. In this respect, automaticity could be the construct through which task difficulty modulates the different representations.

In summary, pitch and semantics seem to be represented in the brain both when these properties are relevant and when they are irrelevant. Furthermore, whereas the detectability of irrelevant pitch representations was related to an increase in gamma power, this was not the case for semantics. We suggest an attentional role for gamma power, dependent on the extent to which the property is processed automatically. Processing of different properties of the same stimulus is therefore not trivial, but seems to be highly dependent on the characteristics of the property.

Acknowledgements

MvdN was funded by The Netherlands Organisation for Scientific Research (NWO-MaGW) under grant number 404.10.500. NA and CO were supported by an Emmy Noether grant of the DFG (AX82/2-1). NA received additional funding by the German Research Foundation via SFB 874 and project AX82/3, and together with JF via SFB 1089. OJ gratefully acknowledges funding from The Netherlands Organization for Scientific Research (NWO): a VICI grant (453-09-002). We would like to thank Amirhossein Jahanbekam for help with plotting of the electrode localizations.

References

Altmann, C.F., Henning, M., Döring, M.K., Kaiser, J., 2008. Effects of feature-selective attention on auditory pattern and location processing. NeuroImage 41 (1), 69–79.
Benson, D., Hienz, R., 1978. Single-unit activity in the auditory cortex of monkeys selectively attending left vs. right ear stimuli. Brain Res. 159, 307–320.
Brovelli, A., Lachaux, J.-P., Kahane, P., Boussaoud, D., 2005. High gamma frequency oscillatory activity dissociates attention from intention in the human premotor cortex. NeuroImage 28 (1), 154–164.
Crone, N., Boatman, D., Gordon, B., Hao, L., 2001. Induced electrocorticographic gamma activity during auditory perception. Clin. Neurophysiol. 112, 565–582.
Davidesco, I., Harel, M., Ramot, M., Kramer, U., Kipervasser, S., Andelman, F., Neufeld, M.Y., Goelman, G., Fried, I., Malach, R., 2013. Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. J. Neurosci. 33 (3), 1228–1240.
Degerman, A., Rinne, T., Salmi, J., Salonen, O., Alho, K., 2006. Selective attention to sound location or pitch studied with fMRI. Brain Res. 1077, 123–134.
Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D., 2001. The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. NeuroImage 14 (6), 1256–1267.
Engel, A.K., Fries, P., 2010. Beta-band oscillations—signalling the status quo? Curr. Opin. Neurobiol. 20 (2), 156–165.
Fell, J., Fernández, G., Klaver, P., Elger, C.E., Fries, P., 2003. Is synchronized neuronal gamma activity relevant for selective attention? Brain Res. Rev. 42 (3), 265–272.
Fries, P., Womelsdorf, T., Oostenveld, R., Desimone, R., 2008. The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. J. Neurosci. 28 (18), 4823–4835.
Harel, A., Kravitz, D.J., Baker, C.I., 2014. Task context impacts visual object processing differentially across the cortex. Proc. Natl. Acad. Sci. U. S. A. 111 (10), E962–E971.
Haupt, S., Axmacher, N., Cohen, M.X., Elger, C.E., Fell, J., 2009. Activation of the caudal anterior cingulate cortex due to task-related interference in an auditory Stroop paradigm. Hum. Brain Mapp. 30 (9), 3043–3056.
Hillyard, S., Hink, R., Schwent, V., Picton, T., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–182.
Iguchi, Y., Hoshi, Y., Tanosaki, M., Taira, M., Hashimoto, I., 2005. Attention induces reciprocal activity in the human somatosensory cortex enhancing relevant- and suppressing irrelevant inputs from fingers. Clin. Neurophysiol. 116 (5), 1077–1087.
Jehee, J.F.M., Brady, D.K., Tong, F., 2011. Attention improves encoding of task-relevant features in the human visual cortex. J. Neurosci. 31 (22), 8210–8219.
Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front. Hum. Neurosci. 4 (186), 1–8.
Jensen, O., Bonnefond, M., VanRullen, R., 2012. An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn. Sci. 16 (4), 200–206.
Johnson, J.A., Zatorre, R.J., 2006. Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. NeuroImage 31, 1673–1681.
Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8 (5), 679–685.
Kamitani, Y., Tong, F., 2006. Decoding seen and attended motion directions from activity in the human visual cortex. Curr. Biol. 16 (11), 1096–1102.
Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 16 (12), 606–617.
Klimesch, W., Fellinger, R., Freunberger, R., 2011. Alpha oscillations and early stages of visual encoding. Front. Psychol. 2 (118), 1–11.
Kriegeskorte, N., Simmons, W., Bellgowan, P., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540.
Krumbholz, K., Eickhoff, S.B., Fink, G.R., 2007. Feature- and object-based attentional modulation in the human auditory "where" pathway. J. Cogn. Neurosci. 19 (10), 1721–1733.
Kumar, S., Schönwiesner, M., 2012. Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling. J. Neurosci. 32 (39), 13348–13351.
Liu, Y., Bengson, J., Huang, H., Mangun, G.R., Ding, M., 2016. Top-down modulation of neural activity in anticipatory visual attention: control mechanisms revealed by simultaneous EEG-fMRI. Cereb. Cortex 26, 517–529.
Mansouri, F.A., Tanaka, K., Buckley, M.J., 2009. Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. Nat. Rev. Neurosci. 10, 141–152.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190.
Martinovic, J., Gruber, T., Ohla, K., Müller, M., 2009. Induced gamma-band activity elicited by visual representation of unattended objects. J. Cogn. Neurosci. 21 (1), 42–57.
Näätänen, R., Teder, W., Alho, K., Lavikainen, J., 1992. Auditory attention and selective input modulation: a topographical ERP study. Neuroreport 3, 493–496.
Niazi, A.M., Van den Broek, P.L.C., Klanke, S., Barth, M., Poel, M., Desain, P., Van Gerven, M.A.J., 2014. Online decoding of object-based attention using real-time fMRI. Eur. J. Neurosci. 39 (2), 319–329.
O'Craven, K., Downing, P., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. Nature 401, 584–587.
Oehrn, C.R., Hanslmayr, S., Fell, J., Deuker, L., Kremers, N.A., Do Lam, A.T., Elger, C.E., Axmacher, N., 2014. Neural communication patterns underlying conflict detection, resolution, and adaptation. J. Neurosci. 34 (31), 10438–10452.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011 (156869), 1–9.
Paltoglou, A.E., Sumner, C.J., Hall, D.A., 2009. Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. Hear. Res. 257, 106–118.
Pitts, M.A., Padwal, J., Fennelly, D., Martínez, A., Hillyard, S.A., 2014. Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. NeuroImage 101, 337–350.
Polk, T.A., Drake, R.M., Jonides, J.J., Smith, M.R., Smith, E.E., 2008. Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: a functional magnetic resonance imaging study of the Stroop task. J. Neurosci. 28 (51), 13786–13792.
Ross, B., Herdman, A.T., Pantev, C., 2005. Right hemispheric laterality of human 40 Hz auditory steady-state responses. Cereb. Cortex 15 (12), 2029–2039.
Sedley, W., Teki, S., Kumar, S., Overath, T., Barnes, G.R., Griffiths, T.D., 2012. Gamma band pitch responses in human auditory cortex measured with magnetoencephalography. NeuroImage 59 (2), 1904–1911.
Sokolov, A., Pavlova, M., Lutzenberger, W., Birbaumer, N., 2004. Reciprocal modulation of neuromagnetic induced gamma activity by attention in the human visual and auditory cortex. NeuroImage 22, 521–529.
Tallon-Baudry, C., Bertrand, O., Hénaff, M.-A., Isnard, J., Fischer, C., 2005. Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. Cereb. Cortex 15 (5), 654–662.
Woldorff, M., Hillyard, S., 1991. Modulation of early auditory processing during selective listening to rapidly presented tones. Electroencephalogr. Clin. Neurophysiol. 79, 170–191.
Woolgar, A., Williams, M.A., Rich, A.N., 2015. Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. NeuroImage 109, 429–437.