NeuroImage 186 (2019) 103–112

Contents lists available at ScienceDirect

NeuroImage

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

Coarse-to-fine information integration in human vision

Kirsten Petras [a,*], Sanne ten Oever [b], Christianne Jacobs [a], Valerie Goffaux [a,b,c]

[a] Research Institute for Psychological Science, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[b] Department of Cognitive Neuroscience, Maastricht University, Maastricht, the Netherlands
[c] Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

ARTICLE INFO

Keywords:
Spatial frequency
Coarse-to-fine
Electroencephalography
Multivariate decoding
Temporal generalization
Human face perception

ABSTRACT

Coarse-to-fine theories of vision propose that the coarse information carried by the low spatial frequencies (LSF) of visual input guides the integration of finer, high spatial frequency (HSF) detail. Whether and how LSF modulates HSF processing in naturalistic broad-band stimuli is still unclear. Here we used multivariate decoding of EEG signals to separate the respective contribution of LSF and HSF to the neural response evoked by broad-band images. Participants viewed images of human faces, monkey faces and phase-scrambled versions that were either broad-band or filtered to contain LSF or HSF. We trained classifiers on EEG scalp-patterns evoked by filtered scrambled stimuli and evaluated the derived models on broad-band scrambled and intact trials. We found reduced HSF contribution when LSF was informative towards image content, indicating that coarse information does guide the processing of fine detail, in line with coarse-to-fine theories. We discuss the potential cortical mechanisms underlying such coarse-to-fine feedback.

1. Introduction

Naturalistic visual input contains a wide range of spatial frequencies (**SF**) that are closely correlated in space and that are processed in an integrative fashion to support visual perception. Several theories have proposed that the visual system integrates visual input in a coarse-to-fine (**CtF**) manner. In this framework, coarse, low spatial frequency (**LSF**) information is processed first and quickly projects from primary visual cortex to higher level visual areas, which generate a feedback signal that subsequently guides the processing of finer-grained high spatial frequency (**HSF**) information (Marr, 1982; Watt, 1987; Schyns and Oliva, 1994; Bullier, 2001; Bar, 2003, 2004; Bar et al., 2006; Hegdé, 2008). For example, when you look at a dog, its general shape, carried by LSF, helps with the integration of finer details carried by HSF, such as the individual hairs of the dog's fur. Using LSF information to guide HSF processing is feasible, because both ranges are strongly correlated in natural images (i.e. the shape of a dog restricts where the fur can be). CtF strategies have been shown to be highly beneficial in terms of both speed and efficiency in many computer vision applications (e.g. Burt, 1988; Gavrila and Philomin, 1999; Zhou et al., 2013; Zhang et al., 2014), as well as in computational models of higher level vision (Kay and Yeatman, 2017). But despite an abundance of evidence for coarse over fine **precedence** in response to simple (see e.g. Jones and Keck, 1978; Parker, 1980; Musselwhite and Jeffreys, 1985; Parker and Dutch, 1987; Watt, 1987; Hughes et al., 1996; Mihaylova et al., 1999; Mazer et al., 2002 for dots, lines and gratings), as well as to more complex stimuli like faces and natural scenes (Parker et al., 1992, 1997; Schyns and Oliva, 1994; Halit et al., 2006; Peyrin et al., 2006; Vlamings et al., 2009; Goffaux et al., 2010; Lu et al., 2018), direct evidence for coarse-to-fine **integration** is still scarce. Coarse to fine integration requires not only that coarse image information is processed quicker than finer details, but also that this initial processing influences later stages of HSF information integration. To be beneficial, CtF integration should reduce either the required processing time and/or the metabolic cost of visual processing by alleviating resources from processing redundant detail information. In order to experimentally investigate CtF integration, several fundamental characteristics of natural visual input need to be mimicked. A first characteristic is that natural images are broad-band, i.e. they contain a combination of both low and high spatial frequencies. Second, the different ranges of SF composing an image are highly correlated in space: without such correlation (i.e., if LSF offers no information on the content of HSF), CtF integration is not computationally advantageous. Finally, natural images follow a roughly $1/SF^{\alpha}$ amplitude spectrum: the LSF components of a natural image generally have much higher contrast than its HSF components.

In spite of the broad-band nature of visual input, previous studies

* Corresponding author.
E-mail address: kirsten.petras@uclouvain.be (K. Petras).

https://doi.org/10.1016/j.neuroimage.2018.10.086
Received 29 April 2018; Received in revised form 17 October 2018; Accepted 31 October 2018
Available online 4 November 2018

addressed CtF processing mostly by comparing the neural responses to narrow-band spatial frequency filtered stimuli (e.g. Goffaux et al., 2010; De Cesarei and Codispoti, 2013; Musel et al., 2014). While they provide a solid foundation for coarse over fine *precedence*, they cannot address CtF *integration*, because SF ranges were presented separately in space, time, or both. Other studies have used so-called hybrid stimuli, combining LSF and HSF from different images, for example the LSF components of a female face with the HSF components of a male face (e.g. Schyns and Oliva, 1999). While hybrid stimuli present an observer with high and low frequencies simultaneously, the uncorrelated nature of LSF and HSF is in violation of the second characteristic of natural images, namely spatially correlated LSF and HSF signals. Additionally, the LSF and HSF stimuli used in past studies had different contrasts, which causes a considerable confound, because visual sensitivity and stimulus contrast are closely linked (Robson, 1966). Other studies equalized stimuli for contrast across SF conditions (Goffaux and Rossion, 2006; Peyrin et al., 2006; Vlamings et al., 2009; Mu and Li, 2013) which is in violation of the $1/SF^{\alpha}$ amplitude spectrum of natural images and might bias visual analysis strategies (Kauffmann et al., 2015). Notably, one approach to avoid the problems that arise through filtering and contrast matching is to mask out information in a given spatial frequency range by replacing the information with its phase-scrambled version. This strategy renders the masked frequency range uninformative without interfering with the natural amplitude spectrum (Näsänen, 1999; Ojanpää and Näsänen, 2003). However, while the resulting images meet the requirements of being broad-band and following a natural amplitude spectrum, the masking disrupts the correlation between LSF and HSF necessary for CtF integration to occur. In summary, by attempting to disentangle LSF and HSF at the level of the stimulus, all of the above research employed stimuli that departed considerably from the fundamental characteristics of naturalistic visual input which are essential to successful CtF integration.

In the present study, we employ a novel paradigm using multivariate pattern classification to tease apart LSF and HSF contributions during the processing of broad-band (BB) stimuli. Separating SF at the level of the cortical response rather than at the level of the stimulus allows us to preserve the fundamental properties of naturalistic broad-band vision, thus allowing to investigate CtF in more ecological settings than achieved so far. With Electroencephalography (EEG) we measure LSF- and HSF-related neural responses expressed as scalp topographies. Then, we train a set of linear classifiers to differentiate between these responses and test their generalization to broad-band visual stimulation over time. Namely, the classifier labels the broad-band response as either LSF or HSF, depending on which one is more prominent in the broad-band topographies; therefore, we interpret classifier prediction as **spatial frequency dominance**. Comparing SF dominance patterns over time allows us to directly address the essential prediction of CtF theories: LSF processing influences subsequent HSF processing in broad-band visual input. We include four different stimulus types: human faces, monkey faces, and both their phase-scrambled versions. Human faces were chosen as stimuli because they elicit a strong and well characterized EEG response (Eimer, 2011). Monkey faces share many of the same features without being equally ubiquitous in the normal human visual diet. While in both categories of intact images LSF is informative with respect to HSF, this is not the case for the phase-scrambled ones. Further, while both human and monkey faces show similar low level image properties and rely on largely similar resources for their processing, human face processing has been shown to be exceptionally efficient compared to other visual categories (Farah et al., 1998; Haxby et al., 2000). This effect is thought to strongly rely on LSF information (Sergent, 1982, 1986; Fiorentini et al., 1983; Goffaux and Rossion, 2006; Goffaux et al., 2010; Richler et al., 2011). Thus, responses to human and monkey faces should be similar if basic image properties drive coarse-to-fine integration but differ if higher level content information, which presumably is more readily available for human faces, is used instead. Consequently, we expected that when LSF is informative towards HSF content (i.e., in intact stimulation conditions), early LSF dominance would coincide with a reduced need to process the correlated HSF information, signified by reduced HSF dominance later in processing time (i.e. CtF integration) and that this effect would be stronger for human compared to monkey faces. To avoid any influence of semantic image content, we trained all classifiers exclusively on the phase-scrambled images.

2. Methods

2.1. Participants

21 healthy volunteers (mean age 26.7, range 21–34, 10 female) with normal or corrected to normal visual acuity participated in the EEG experiment. Data were collected as part of a technical development project aimed at improving EEG source reconstruction. Therefore, a subset of the participants also took part in a separate fMRI experiment not further described here and all EEG recordings were performed while participants were lying on their back in a mock-scanner. All participants gave written informed consent and were reimbursed for their participation in the form of gift vouchers (7.5/hour). The EEG experiment lasted 55 min with an additional set-up time of 15–35 min. The subjects who also participated in the MRI experiment additionally underwent a 10-min anatomical scan and two functional localizer runs of 12 min each. All procedures have been approved by the ethical committee of the Université Catholique de Louvain (approval number: 2016/13SEP/393).

2.2. Stimuli

Stimuli were created from seven original frontal views each of female human and monkey faces. Images were scaled and aligned so that the distance between eyes and mouth was equal for all images, cropped to the same oval occlusion and pasted onto a uniformly gray background. They were then converted to grayscale and luminance and root-mean-square (RMS) contrast was equalized using custom made functions in Matlab (Mathworks, 2016). Images of faces usually follow a roughly $1/SF^{\alpha}$ amplitude spectrum, were α is close to 2, giving considerably higher spectral energy to low than to high spatial frequencies. Since visual processing is strongly influenced by the contrast of the stimulus (De Valois and De Valois, 1980; Boynton et al., 1996), it is common practice to equalize RMS contrast between stimuli by enhancing contrast in HSF images after filtering (Vlamings et al., 2009; Goffaux et al., 2010; Mu and Li, 2013; Kauffmann et al., 2015). However, since such a manipulation would result in non-natural amplitude spectra we developed the following normalization to avoid contrast differences between LSF and HSF stimuli while maintaining a natural spectrum for the composite broad-band (BB) images (see Fig. 1). We first constructed 5th order Butterworth frequency filters with cut-offs chosen such that, averaged over all images, the summed Fourier-space amplitudes of LSF and HSF images were equal. Since most energy is contained in extremely low SF, our cutoff manipulation would lead to LSF images containing only very coarse visual information if applied to full spectrum images. To prevent such large imbalances in information content, we excluded the first 5 cycles per face (cpf) prior to determining the filter cut-offs. The Fourier amplitude of each image was then multiplied with the resulting filters and transformed back into image space. To create BB images, the Fourier-amplitudes of the original images were multiplied with the sum of the low (5–8 cpf) and the high (>8cpf) SF filters and then also transformed back into image space. Importantly, there was no post-hoc contrast enhancement of HSF images. Instead, the design of the Butterworth-filters ensured that mean contrast was 'naturally' similar between HSF and LSF images, namely 0.0185 and 0.018 respectively (see Fig. 1).

To derive meaningless noise images with power spectra similar to those of the original images, we created phase-scrambled versions of full spectrum images by shuffling the phase information of the Fourier-Transform. Operations in Fourier space require input images to be rectangular and cannot be restricted to the region occupied by the face.

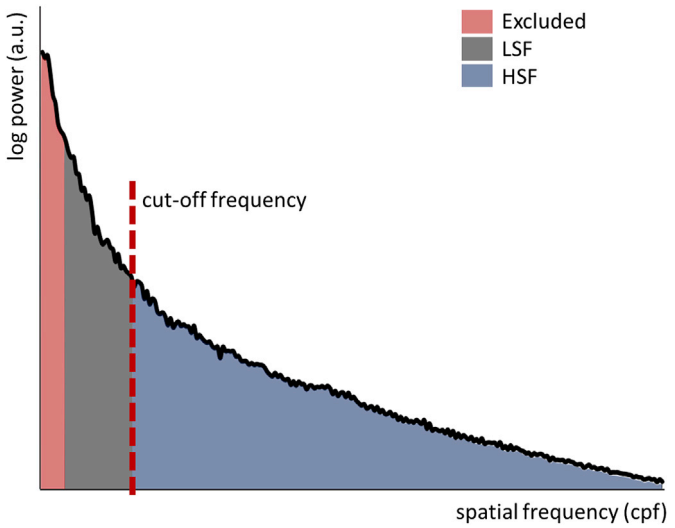

Fig. 1. The cut-off frequency for the Butterworth filters was determined as the frequency for which on average over all images, the area under the curve of the Fourier amplitude spectrum was equal to either side of the cut-off. To avoid very small bandwidths in the LSF condition, we excluded the first 5 cycles per image before determining the cut-off. Note that this is a conceptual illustration and not to scale.

Therefore, the phase-scrambling may artificially increase LSF energy due to the low frequency uniform background leaking into the oval face area and vice versa. To minimize the effects of such leakage, we employed the following iterative approach to phase scrambling: Images were cropped to the square that most closely matched the occlusion and phase-scrambled in Fourier space before being transformed back into image space. The original, non-scrambled face pixels were pasted onto the resulting phase-scrambled images. This procedure was repeated 100 times so that face areas and backgrounds had similar power spectra before moving on to the next step. The scrambled face areas were then pasted back onto the original gray backgrounds. In effect, by making the spectral properties of face and background pixels more similar, this procedure minimized contaminations of the scrambled image by the uniform background (see Fig. 2). The images were then subjected to the same Butterworth filter procedure as the phase-intact images to derive LSF, HSF and BB versions. Images were displayed on a standard LCD monitor (1900 × 1200px) with an absolute size of the face area of 5.8 cm (max. height of the oval shape) x 4.7 cm (max. width of the oval shape) at a viewing distance of approx. 60 cm thus covering ~5.5° x ~4.5° of visual angle.

2.3. Procedure

During the experiment, participants were passively viewing intact and scrambled images of human faces and monkey faces in the 3 filter

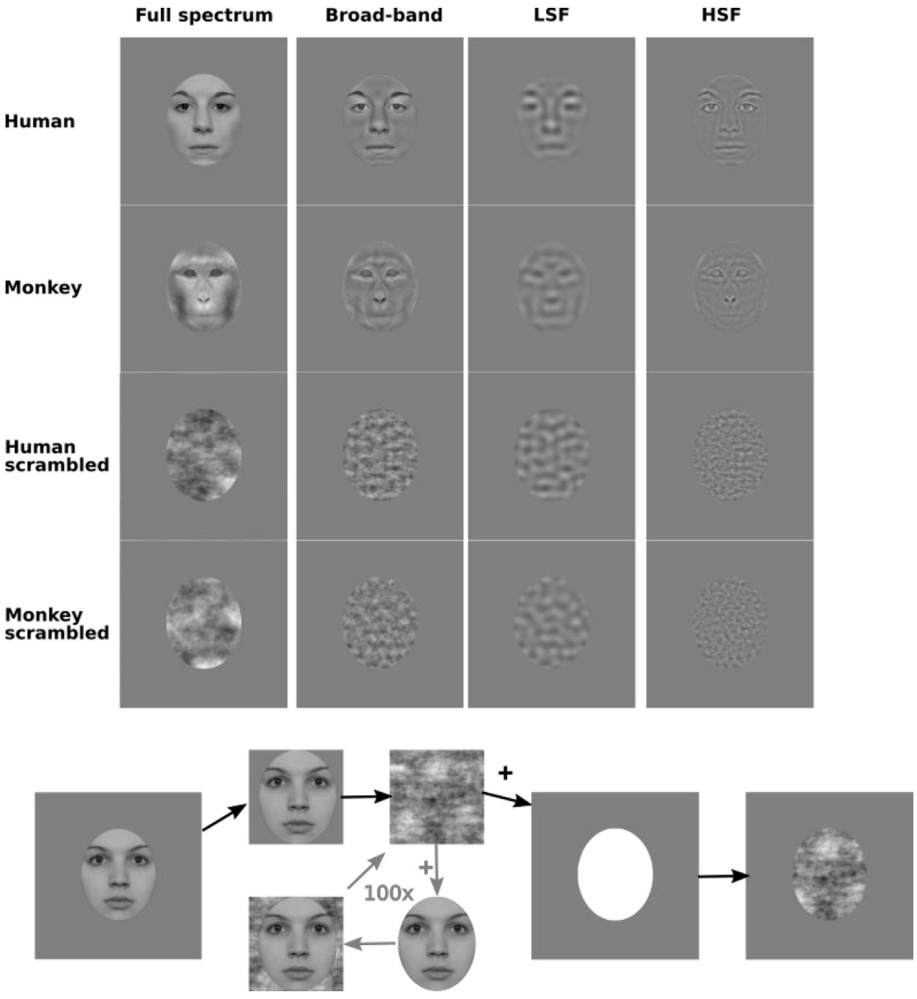

Fig. 2. Upper panel: Example stimuli. Full spectrum grayscale pictures of human and monkey faces (left-most column, note that the original images were never shown to the participant) were filtered to contain low and high, only low, or only high spatial frequencies. All faces were aligned at the horizontal connecting the eyes and scaled such that the vertical distance between eyes and mouth was equal across images. Lower panel: Images with similar amplitude and orientation spectra as the original face images, but without any semantic content were created. Face images on gray background were cropped to the closest square surrounding the oval face outline and scrambled in Fourier-phase. To reduce the loss of spatial frequency information due to the uniform background leaking into the oval image area, the intact face was pasted back onto the scrambled image and the whole process was repeated 100 times before adding the original gray background.

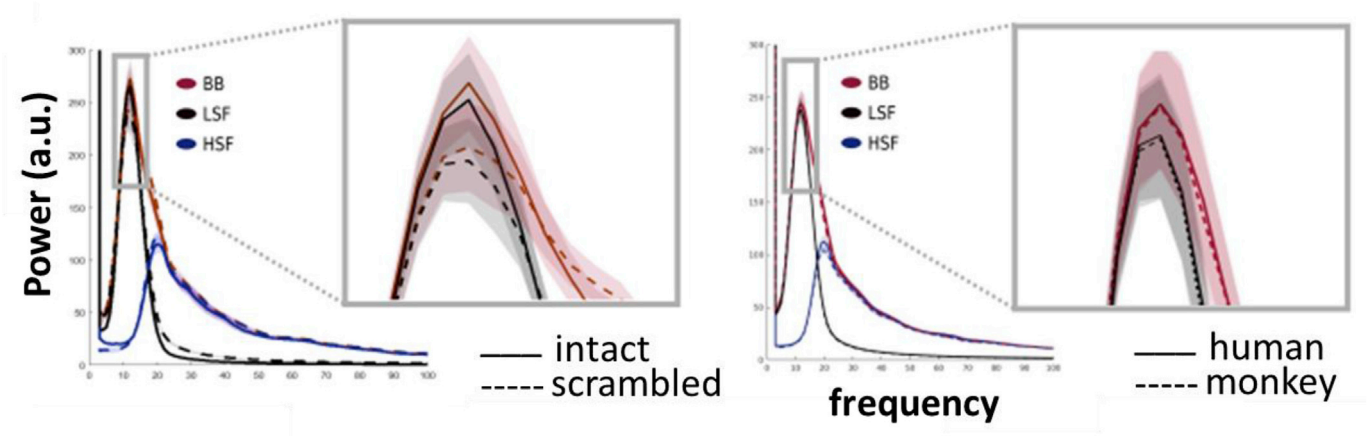

Fig. 3. Left: Power spectrum comparing intact and scrambled images. Spectra were averaged over human and monkey images. Right: Power spectra comparing human and monkey images. Note that differences between categories are minimal and well within standard error margins.

conditions (LSF, HSF and BB). Intact images and scrambled images were presented in blocks. Each stimulus was on screen for 300 ms with a dark grey fixation cross superimposed in the center of the image. Block order was counterbalanced between participants. Stimulus order within each block as well as inter-trial-interval (1.5–2.5 s) were counterbalanced via an m-sequence (Buračas and Boynton, 2002) to avoid any potential order effects. Participants underwent a total of 1440 trials subdivided into 6 blocks. Each block contained 206 experimental trials and 34 catch trials in which participants had to indicate whether a black shape appearing on the current image was a circle or a square by pressing one of two buttons on a button box. Catch trials were included to ensure participants paid attention to the stimuli (Performance: mean 97.8% correct, range 89.2%–99%). They, as well as the trials immediately following them, were excluded from further EEG analysis.

2.4. EEG recording and analysis

An EGI geodesic sensor net (Electrical Geodesics Inc., Eugene, OR, USA) was soaked in a potassium chloride solution for 10 min to facilitate conductivity of the sponges placed within plastic cups between scalp and electrodes. The net was fitted on the participant's head such that electrode Cz, which served as a reference for the remaining 255 electrodes, was located on the intersection of the midlines between nasion and inion and the preauricular points. We then carefully removed hair trapped between electrodes and scalp and re-moistened sponges that did not meet our scalp-electrode impedance requirements (<40 kΩ at the beginning of the recording). EEG signals were amplified with a high input-impedance of ~200 MΩ using a Net-Amps dense array amplifier (Electrical Geodesics Inc., Eugene, OR, USA) and sampled at a rate of 1000 Hz. To prevent sponges from drying out during the long recording session, we wrapped the participants head in cling film held together with tape which also served to keep electrodes in place and close to the scalp once participants laid on their backs. We could thus maintain scalp-electrode impedances <50 kΩ throughout the recordings for almost all electrodes. Participants were placed in a mock-MRI-scanner (no magnetic field present) and viewed a screen at a distance of 60 cm via a mirror mounted to the (not connected) head coil.

EEG data were preprocessed using the Fieldtrip toolbox (Oostenveld et al., 2011) under Matlab. Data were re-referenced to the average of all electrodes and band pass filtered to 0.5–40 Hz. Trials were segmented into epochs ranging from −500 ms to 1500 ms around stimulus onset and down-sampled to 256 Hz. We then manually removed trials with excessive muscle artifacts (but not blinks) and subjected the remainder of the data to an ICA algorithm (Runica, implemented in Fieldtrip) and subsequently removed components containing ocular artifacts. Finally, to reduce the number of features and as a consequence the computational demands for classification, only a region of interest consisting of the 47 electrodes covering occipital and temporal regions were a-priori selected (see Fig. 4 for locations).

2.5. Classification

The aim of the study was to investigate the contribution of low and high SF over the time course of broad-band image-processing. To avoid any influence of semantic content on our classifiers, we only used phase-scrambled image trials for training. Classifiers were trained to discriminate between LSF and HSF scrambled trials. We then used this "low vs high scrambled" model and evaluated its prediction in response to broad-band scrambled and intact trials. This provided us with information about the dominance of either low or high spatial frequency in the EEG signal during broad-band viewing conditions.

We used the support vector machine algorithm provided by libsvm (Chang and Lin, 2011) running under Matlab for all classifications. We employed a linear kernel and kept the regularization parameter c equal to 1 throughout classifications. Epochs included the −100 to 600 ms time window around stimulus onset (rounded to fit sampling rate). Then, we created time bins by shifting a 12 ms window in 8 ms steps (rounded to fit sampling rate) resulting in 90 bins. For each time bin we averaged amplitudes over the 4 samples per channel contained in it and, separately for all conditions, transformed feature-vectors into unit vectors such that:

$$\widehat{u} = \frac{u}{|u|}$$

where |u| is the norm of the original vector u given by

$$|u| = \sqrt{\sum_{n=1}^{n} x_n^2}$$

To validate our LSF vs. HSF model we pseudo-randomly split the training data for each subject into training and validation sets retaining 20% of the original data for model validation. This procedure was repeated 150 times to ensure the greatest variability in fold composition that was computationally feasible. Statistical significance of above chance classification was assessed for each data point via two-sided paired samples t-tests against the empirical chance level of each participant. Data across all participants was corrected for multiple comparisons using cluster based permutation testing (Oostenveld et al., 2011) at a cluster alpha of 0.05 using the maximum sum of all t values within a cluster as a test statistic and testing against the Monte-Carlo estimates of the critical values from the permutation distribution (10 000 permutations). The empirical chance distribution was derived by label permutations. At each randomization, two models were generated and assessed:

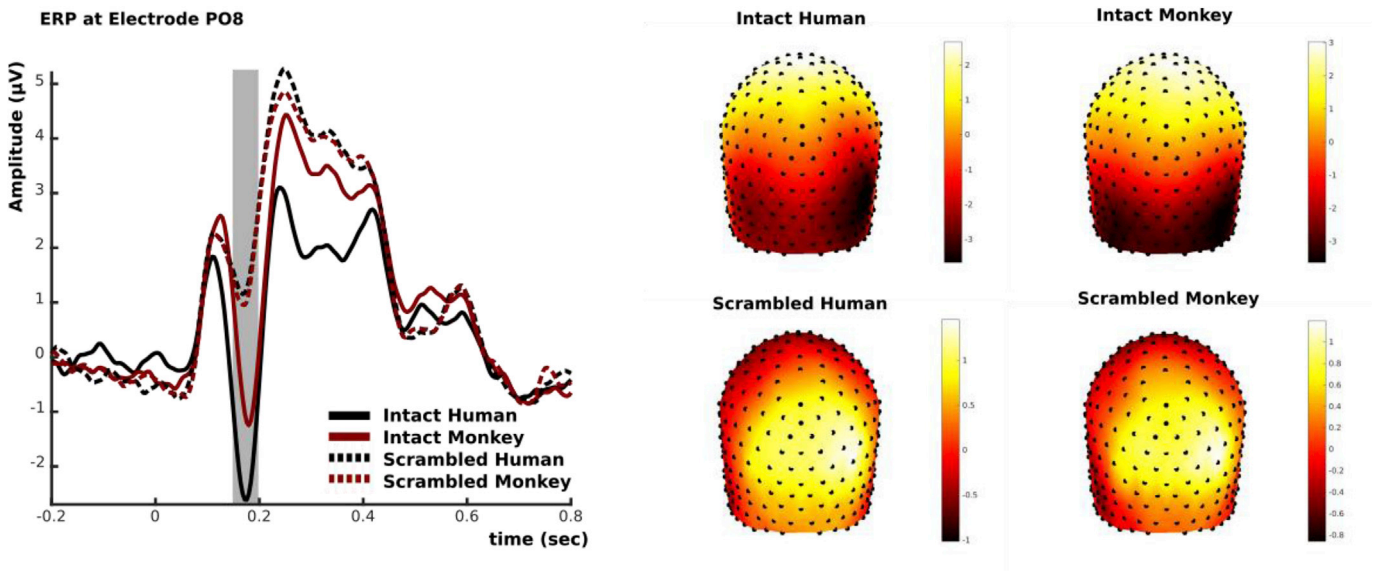

Fig. 4. Left: Event related potential at electrode PO8. Note the typical N170 response (grey shaded area) that is most prominent for human faces and least prominent for scrambled images. Right: Averaged EEG scalp topographies at the latency of the N170 component (160–180 ms). Note that both human faces and monkey faces evoke a right lateralized negativity over occipito-temporal areas.

one with the original class labels and one with trial labels randomly permuted before classification performance was assessed. All reported accuracies for within condition classification are the average over all randomizations and participants. Please note that, because we kept class frequencies balanced in all training and validation sets, the commonly used area under the curve (AUC) and plain accuracy are the same. We therefore report accuracy throughout the results section.

The generalization of the LSF/HSF scrambled classifier was assessed on all available trials for each BB condition (human, monkey, intact and scrambled) for each of the 150 models per time bin. To assess classifier generalization performance all BB trials were arbitrarily labeled as HSF. Off-chance generalization was interpreted as LSF or HSF dominance dependent on bias direction (<50% as LSF and >50% as HSF dominance). The statistical procedure (paired samples t-tests against empirical chance and cluster correction for multiple comparisons on the group level) was identical to the model validation procedure.

For between condition contrasts (Intact vs. Scrambled and Human vs. Monkey) we contrasted the individual condition cross-time matrices with each other and tested for statistical significance using paired sample t-tests. Results were again cluster corrected at a cluster alpha of 0.05.

3. Results

3.1. Successful spatial frequency decoding in scrambled images

The SVM classifiers were able to separate LSF and HSF trials significantly better than empirical chance starting from around 90 ms after stimulus onset, [$p < 0.001$, cluster-based permutation tests for multiple comparisons using the maximum sum of cluster t-values against empirical chance (see methods for details), cluster statistic = 5153.4] until about 560 ms after stimulus onset (Fig. 5). Although class frequencies were balanced in all training and testing folds, on average over all iterations, trials and subjects, classifiers predicted topographies as belonging to LSF-trials 51,2% and HSF-trials 48.8% of the time. When training time and testing time match (no generalization case) classification shows no bias (see suppl. figure 3). To assess the classification stability over time, we tested generalization of each classifier to every time-sample of the cross-validation trials (for procedure see King and Dehaene (2014) and methods). If the evoked scalp patterns used for classifier training and prediction remain similar across those time samples a classifier trained at a given time-sample t can still successfully predict SF class at a later (or earlier) time-sample t'. As expected, the highest classification accuracy was found around the diagonal of the time-generalization matrix; i.e. when training and testing times were identical (no generalization). While the performance of classifiers trained at early time windows (<200 ms post stimulus onset) rapidly decayed with temporal distance between training and testing samples, classifiers trained on time windows later than 200 ms post stimulus onset remained efficient in predicting SF condition for up to 200 ms. This difference in temporal generalizability could indicate either a stabilization of the cortical representation over time, or it could simply be due to a larger variance in the later evoked responses. All models (i.e. trained classifiers) resulting from the cross-validation procedure were stored to use for the subsequent analyses.

Notably, SF dominance is not only determined by the evolution of testing time (pattern changes in x direction of the temporal generalization matrix), but also by the training time, i.e. the model used to perform the prediction (patterns change along the y-direction, see Fig. 5 C & D and Fig. 6). This model evolution indicates that the classifier used different criteria to label a trial as LSF or HSF dependent on the progress of visual processing. Based on visual inspection we differentiated at least two different main time-courses of model validation. The first generalization pattern depends on models trained at times ranging from ~110 ms to ~220 ms after stimulus onset and the second on models trained at times ranging from ~230 ms to ~440 ms after stimulus onset. These latencies coincide with the P1—N1 and the P2—P3 ERP complex, respectively. We did not foresee this shift of critical information used by the classifier but speculate that over the course of visual processing different features, potentially originating from different cortical sites, are indicative of the spatial frequency content of the processed stimulus (see discussion). We however remain cautious to draw any strong conclusions based on post-hoc explanations.

3.2. Coarse-to-fine pattern of spatial frequency dominance in broad-band image processing

We applied the models trained to differentiate between LSF and HSF phase-scrambled stimuli to broad-band images of intact and phase-scrambled human and monkey faces (where LSF and HSF contributed equally to the overall image contrast) and assessed their predictions.

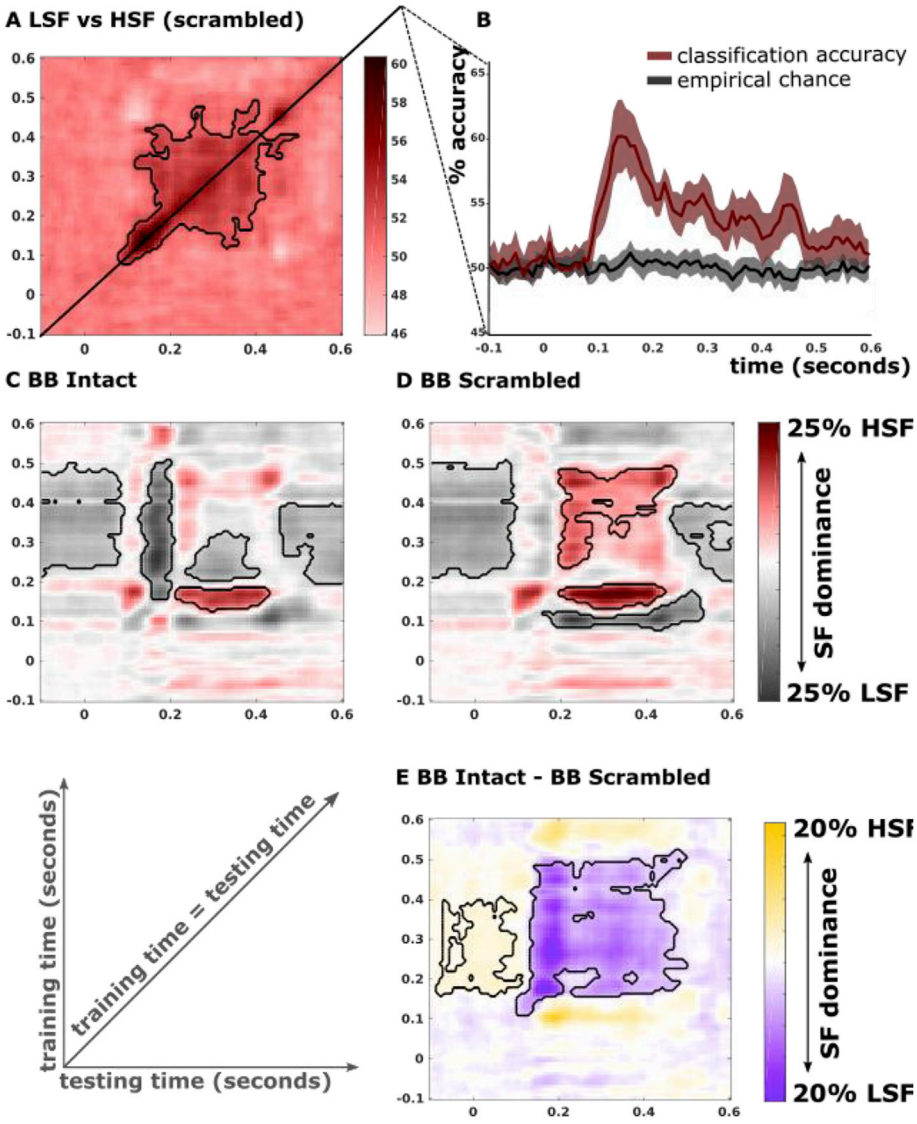

Fig. 5. (A) Time generalized spatial frequency decoding between LSF and HSF scrambled images. The black contour indicates decoding performance significantly above empirical chance. (B) Time-course of spatial frequency decoding against empirical chance. Note that this is identical to the diagonal of the matrix in (A). Margins indicate standard error of the mean. (C) Spatial frequency dominance in intact broad-band images. Black contour indicates significant low or high spatial frequency dominance. (D) Spatial frequency dominance in scrambled broad-band images. (E) Difference between C and D. Black contour indicates significantly stronger low spatial frequency dominance in intact images.

Using the same temporal generalization approach as above, we applied each of the LSF vs. HSF classifiers to each of the BB trial time windows. Because LSF and HSF components contribute equally to our BB images, classifier prediction now indicates the relative resemblance of the current scalp pattern to those associated with LSF and HSF trials, hereafter referred to as **SF dominance**, rather than SF class membership.

For all BB conditions, classifier prediction deviated significantly from chance only when 'training-time' fell into the window of successful classification between LSF and HSF trials, that is starting from 90 ms after stimulus onset on (see Fig. 5C and D). In other words, when a model was successfully trained to classify LSF vs. HSF scrambled images, this model generalized to predict BB trials as either LSF or HSF in a non-random fashion. These findings support the validity of our generalization results.

SF dominance patterns evoked by intact BB images (pooled over human and monkey faces) contain three post stimulus clusters of significant LSF dominance [128 ms-207 ms testing time, cluster statistic = 1206.6, $p < 0.01$], [238 ms-395 ms testing time, cluster statistic = 728.8, $p < 0.05$] and [451 ms-600 ms testing time, cluster statistic = 1147.3, $p < 0.01$] and one cluster of significant HSF dominance [215 ms-427 ms testing time, cluster statistic = 589.1, $p < 0.05$]. In contrast, the SF dominance matrices evoked by scrambled BB images contain only two post-stimulus clusters of significant LSF dominance [151 ms-530 ms testing time, cluster statistic = 831.7, $p < 0.01$] and [458 ms-600 ms testing time, cluster statistic = 738.6, $p < 0.01$] as well as two separate clusters of significant HSF dominance [200 ms-466 ms testing time, cluster statistic = 1588.6, $p < 0.01$] and [199 ms-466 ms testing time, cluster statistic = 762, $p < 0.05$]. All latency ranges indicated here refer to testing time (x-axes of the temporal-generalization plots). Interestingly, we also find significant LSF dominance in the pre-stimulus period for both intact and scrambled BB conditions [Intact: -100 ms - 90 ms testing time, cluster statistic = 2523, $p < 0.001$; Scrambled: -100 ms–100 ms testing time, cluster statistic = 3188, $p < 0.001$]. This pre-stimulus LSF dominance did not differ significantly between conditions (see Fig. 6 E). Because the stimulus is removed from the screen during the baseline period, what remains is the very low SF of the uniformly gray screen (0 cpi). Therefore, the pre-stimulus (empty) screen is sensibly labeled as LSF by the classifier (also see suppl. figure 3).

3.3. Coarse-to-fine integration for intact, but not scrambled images

Low and high spatial frequencies are spatially aligned in intact images, satisfying the prerequisite for CtF integration. This is not the case for scrambled images. To statistically evaluate the differences in SF dominance between intact and scrambled BB conditions we contrasted predictions for BB-intact and BB-scrambled conditions. Results show significantly larger LSF dominance in intact compared to scrambled

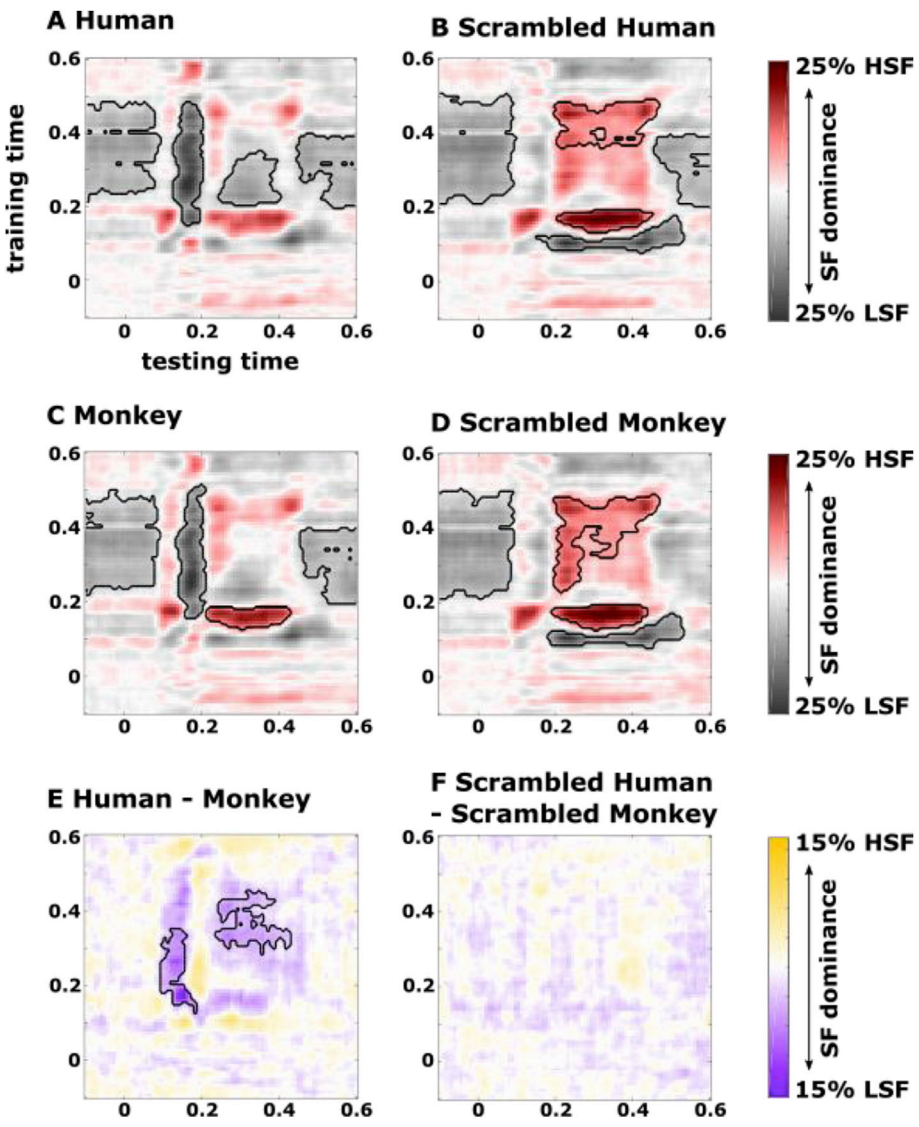

Fig. 6. (A) and (B): Spatial frequency dominance in intact and scrambled human face images. Black contour indicates significant low or high spatial frequency dominance. (C) and (D): Spatial frequency dominance in intact and scrambled monkey face images. (E) Difference between (A) and (C). Black contour indicates significantly stronger low spatial frequency dominance in human, compared to monkey images. (F) Difference between (B) and (D). No differences were found in spatial frequency dominance between scrambled human and scrambled monkey faces.

images [105 ms - 482 ms testing time cluster statistic = 6224.8, $p < 0.001$ (see Fig. 5 E). Because we use binary classifiers, larger LSF dominance can equally be interpreted as reduced HSF dominance. However, the early LSF component is present only in the intact, but not in the scrambled trials whereas the late HSF component is present only in the scrambled, but not in the intact images. We therefore interpret the contrast as indicating an early LSF dominance and a late HSF reduction for intact, compared to scrambled images in line with CtF integration when low spatial frequencies are predictive of high spatial frequency information.

3.4. High level category dependence of spatial frequency dominance

The use of intact and phase-scrambled trials allowed us to compare between conditions in which low spatial frequency is or is not informative about high spatial frequency. Comparing human trials to monkey trials taps into a finer functional distinction. Indeed human and monkey faces are two visually similar classes of stimuli in which low spatial frequency is similarly informative towards high spatial frequencies in the image space. Images of human and monkey faces also share similar amplitude spectra (see Fig. 3) and cortical resources for their processing. However, the human visual system has been shown to be particularly sensitive to conspecific faces and highly efficient in their processing indicating further specialization for this stimulus class (Bruce and Young, 1986). Generalizing SF classifiers to BB images of human and monkey faces separately, we found EEG responses to human faces to be significantly LSF dominated in the post stimulus period [128 ms-200 ms testing time, cluster statistic = 1221.6, $p < 0.01$], [246 ms-400 ms testing time, cluster statistic = 909.7, $p < 0.05$] and [460 ms-600 ms testing time, cluster statistic = 1044.7, $p < 0.01$]. While responses to monkey faces also showed early [144 ms-207 ms testing time, cluster statistic = 1146.8, $p < 0.01$] and later [458 ms–600 ms testing time, cluster statistic = 1133.4, $p < 0.01$] LSF dominance, they additionally demonstrated an intermediate period of HSF dominance [215 ms-419 ms testing time, cluster statistic = 638.7, $p < 0.05$].

Directly comparing SF dominance evoked by intact human faces to SF dominance evoked by intact monkey faces, we find significantly higher LSF dominance for the human face trials [96 ms-175 ms testing time, cluster statistic = 520.3, $p = 0.01$; 246 ms-420 ms testing time, cluster statistic = 667.6, $p < 0.01$]. Again, our use of binary classifiers limits the interpretation of directionality of the effect. Only in combination with the generalization patterns against empirical chance does it become apparent, that the stronger LSF dominance later in trial-time for human face trials is in fact driven by the absence of the HSF dominance period present in monkey trials. Importantly, this difference in SF dominance patterns is not driven by purely low-level image properties: results show

strikingly similar patterns for both categories when classification is based on their scrambled versions (see Fig. 6 F). Both human and monkey phase-scrambled images evoked LSF as well as HSF dominant patterns in the post stimulus period [Human scrambled LSF: 159 ms-529 ms testing time, cluster statistic = 946.6, $p < 0.05$; 458 ms–600 ms testing time, cluster statistic = 867.9, $p < 0.05$]; [Human scrambled HSF: 199 ms–458 ms testing time, cluster statistic = 3134.2, $p < 0.001$; 207 ms–443 ms testing time, cluster statistic = 946.6, $p < 0.05$] and [Monkey scrambled LSF: 183 ms–521 ms testing time, cluster statistic = 760.5, $p < 0.05$]; [Monkey scrambled HSF: 199 ms–466 ms testing time, cluster statistic = 1486.3, $p < 0.01$; 199 ms–435 ms testing time, cluster statistic = 812.3, $p < 0.05$]. We found no significant differences between the two categories of phase-scrambled stimuli ($P > 0.05$).

We therefore conclude that intact, more than scrambled, and human, more than monkey, face images evoke early LSF dominant responses. Furthermore, the late HSF dominance is significantly reduced for intact and particularly for human face trials. These findings are in line with CtF theories that predict a LSF precedence along with a reduction of HSF processing load when LSF is predictive of HSF content.

4. Discussion

Coarse-to-fine theories are a highly influential alternative to feedforward models of visual processing. So far, the main empirical support for CtF processing came from studies presenting LSF or HSF in isolation and showing coarse over fine temporal precedence. Here we directly investigated the CtF integration underlying the processing of more naturalistic broad-band input. In intact images, i.e. when image LSF were aligned to, and therefore informative about, HSF content, we found a reduction in the late contribution of HSF to the neural responses to BB, suggesting that informative LSF does modulate the processing of HSF information. Further, we found this pattern to be more prominent in response to human faces, for which the human visual system has developed robust perceptual templates and therefore strong predictions compared to monkey faces. These results are consistent with coarse-to-fine theories of visual processing where coarse LSF information guides the integration of fine HSF details.

While it is generally assumed that coarse-to-fine integration relies on feedback projections from higher level inferior-temporal or parietal visual cortices, little is known about the nature of such feedback. One possibility is that feedback to early visual cortex is retinotopically organized in a similar fashion as feedforward input. Because LSF and HSF components in natural images are closely correlated in visual space, a retinotopic organization would allow for LSF evoked feedback to be projected onto the same (albeit larger because of larger receptive field sizes in higher level visual cortex) retinotopic locations, which are then reached by the delayed HSF-related feedforward signal. Evidence from animal electrophysiology suggests that LSF feedback and HSF feedforward input to primary visual cortex temporally coincide (Bullier, 2001). Further support for the temporal feasibility of CtF perception comes from a recent study that found response latencies of SF selective clusters in monkey visual cortex (V1 — V4) to reflect CtF organization (Lu et al., 2018). Therefore, retinotopic feedback to early visual cortex presents a feasible mechanism for LSF information to directly influence the processing of the HSFs in close spatial proximity. This is in line with our results for intact versus scrambled images where we observed reduced HSF dominance in response to intact images, but not in response to scrambled images in which LSF and HSF content are not correlated in space. However, purely retinotopic feedback cannot explain the differences in EEG generalization between human and monkey faces. Human faces evoked a stronger LSF dominance early in the trial and a subsequently reduced HSF dominance compared to monkey faces, although in both stimulus categories LSF are equally indexing HSF, and the categories do not vary much in terms of local contrast changes (similar prominence of e.g. eyes and mouth across exemplars). A recent fMRI study (Revina et al., 2017) showed classification of natural scenes that generalized across spatial frequency, but not across feedforward and feedback conditions. Because fMRI measures metabolic demands within small voxels, retinotopic feedback should evoke similar patterns to feedforward input from the same stimulus and should thus allow generalization between feedforward and feedback information. The absence of retinotopic generalization in Revina et al. (2017) is inconsistent with simple retinotopic mapping of feedback although this null result has to be considered with caution.

An alternative feedback mechanism might rely on non-retinotopic, higher level information about the stimulus. This notion is supported by a recent study on the macaque face processing hierarchy that shows feedback to carry features of higher level areas (Schwiedrzik and Freiwald, 2017). The authors manipulated stimulus expectation and found that when the expected stimulus sequence was violated, the prediction error measured in low level face selective areas reflected identity specificity and view invariance, properties generally represented in higher level face selective patches. These results indicate that feedback from higher level areas preserves the tuning properties of its origin and makes those properties available as predictions to regions earlier in the visual hierarchy. Such recurrent predictive feedback could explain our results for both intact and scrambled image categories. While the LSF components in scrambled images do not allow for any coherent predictions of HSF content, the intact images do. Further, it has been shown that the processing of human face images strongly relies on LSF information (Goffaux et al., 2003; Goffaux and Rossion, 2006; Goffaux, 2009; Quek et al., 2018). This highly efficient face perception at a glance has been attributed to the strong expertise human observers have developed early in life in processing faces. For monkey faces, the same expertise is usually not achieved (Pascalis et al., 2005). Therefore, the differences in SF dominance patterns between human and monkey trials could indicate the need to integrate more HSF information in the latter case to reduce uncertainty about the image when LSF is not sufficient to form an adequate prediction. In future studies, progressively obscuring content information from experimental stimuli while leaving spatial correlations between LSF and HSF intact could provide a way to more finely investigate to which extent retinotopic proximity of LSF and HSF information on the one hand and higher level content information on the other hand are driving CtF integration. Also note that in our experiment the first 5 cycles (which according to Quek et al. (2018) are sufficient for face detection) were excluded to avoid contrast dissimilarities between SF conditions. LSF information was limited to 5–8 cycles whereas our HSF range (>8 cycles) included what is more commonly referred to as mid-range frequencies. This choice of filter cut-offs might have reduced effect sizes in our study and potentially obscured additional differences in SF dominance over time. Future experiments should test different ranges of SF to allow for a more precise mapping of SF integration.

In summary, our findings indicate that LSF information modulates the processing of HSF details, in line with CtF theories. The reduced HSF dominance in intact human face-, compared to intact monkey face images, along with the evidence about the high-level content of feedback information provided by Revina et al. and Schwiedrzik and Freiwald favors a high-level interpretation of feedback. Although we expect CtF to be the general mode of spatial frequency integration, our experiment only contrasted highly relevant (face-) stimuli with entirely meaningless noise. We therefore cannot conclude from our data that the described effects will indeed generalize to other image categories. Future research including a larger range of stimuli is needed to assess whether reduced HSF dominance in CtF integration could qualify as a general mechanism or whether our results are limited to face processing.

Where exactly the initial LSF processing takes place and hence where the feedback comes from remains unresolved. In the current study, we used EEG-topography based decoding for the assessment of SF dominance, exploiting EEG's high temporal resolution at the cost of spatial specificity. All classifiers were trained on EEG topographies formed by 47 occipital-temporal electrodes, which cover the full range of the visual hierarchy, from early visual cortex to high level category selective

regions in the inferior temporal cortex. As the visual signal progresses along the processing hierarchy, different regions become responsible for the amplitude distribution patterns observed on the scalp surface. To reduce potential confounds of stimulus category on SF decoding, we trained all classifiers exclusively on data from scrambled image trials. However, because scrambled images are likely not processed in the same higher level visual regions, which is reflected in the scalp patterns (see suppl. Figure 1), generalization performance between scrambled and intact image trials might suffer. Supplementary figure 2 illustrates the reduction in generalization performance over trial time as a function of such progressive diversification of cortical sources contributing to the observed scalp pattern. Importantly though, generalization performance remains significantly above chance until about 300 ms into testing time. The low spatial resolution of EEG recordings is a limiting factor in our experiment. Conclusions about the potential origin of LSF feedback require an extension of our paradigm to include parcellations of the cortical sources. A promising candidate for the source of feedback in our example are the high-level face-selective visual regions in the occipito-temporal cortex (i.e., fusiform face area or FFA, and occipital face area, or OFA). Both regions have repeatedly been linked to EEG scalp topographies in response to faces at a latency consistent with the N170 ERP component (Yovel et al., 2008; Nguyen and Cunnington, 2014). FFA has also been shown to represent the LSF structure of the face before its finer details (Goffaux et al., 2010). In our own data, the earliest latencies at which we find differences in SF dominance between stimulus categories are compatible with typical N170 latencies, thus potentially reflecting sources in FFA/OFA. Given the recent advancements in EEG and MEG source reconstruction methods (Henson et al., 2010; Colclough et al., 2015; Farahibozorg et al., 2018), our paradigm can easily be extended to analyze distinct regions of interest as well as the relationships among them.

In our experiment, SF dominance also varied as a function of the temporal window that the classification model was based on. This indicates that the pattern of information about SF which the classifier extracts changes with time. Although SF classification is robust at a latency of ~90 ms–~560 ms post stimulus onset, for models trained at the later portion of this time window the generalizability across time increases. There are several explanations for this. First, visual evoked potentials display high temporal variability across trials, with larger uncertainty later in processing time (Ouyang et al., 2016). Due to this temporal smearing, markers of a given neuronal representation appear to be more widely spread across trials, although the representation might be transient in each individual trial. Further, due to EEG's poor spatial resolution, changes in contributions between close-by sources might be missed.

We believe that our approach based on pattern classification and generalization from single processes to integrated processing conditions provides an ideal tool to characterize complex components of human cognition without the need to interfere with the process under study. The principle of recognizing the building blocks of any complex task and subsequently delineating their neural footprints from the integrated whole can be applied to various problems in neuroscience research in which interactions between processes or the integration of components are the target of investigation. This integrative approach may be transferred to other neuroimaging techniques, such as MEG and fMRI.

In conclusion, we found evidence for category-dependent spatial frequency dominance patterns in the processing of broad-band images. Intact images, especially of human faces, evoked patterns consistent with a coarse-to-fine parsing of the visual stimulus. Our results demonstrate the potential of multivariate decoding techniques to separate stimulus features at the level of the neural response, thus facilitating the use of more ecologically valid stimuli.

Acknowledgements

KP, CJ and VG are supported by the Belgian National Fund for Scientific Research. StO is supported by the Netherlands Organisation for Scientific Research, Grant number 453-15-008. We thank Scannexus Maastricht for providing access to the EEG recording equipment and Miriam Heynckes for her help with participant recruitment and recording. We also thank two anonymous reviewers for providing valuable feedback on the initial manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2018.10.086.

References

Bar, M., 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. J. Cognit. Neurosci. 15 (4), 600–609. MIT Press.
Bar, M., 2004. 'Visual objects in context', *Nature reviews*. Neuroscience 5 (8), 617.
Bar, M., et al., 2006. Top-down facilitation of visual recognition. Proc. Natl. Acad. Sci. U. S. A. 103 (2), 449–454. National Acad Sciences.
Boynton, G.M., et al., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. 16 (13), 4207–4221.
Bruce, V., Young, A., 1986. Understanding face recognition. Br. J. Psychol. 77 (3), 305–327.
Bullier, J., 2001. Integrated model of visual processing. Brain Res. Rev. 36 (2), 96–107. Elsevier.
Buračas, G.T., Boynton, G.M., 2002. Efficient design of event-related fMRI experiments using M-sequences. Neuroimage 16 (3), 801–813.
Burt, P.J., 1988. Smart sensing within a pyramid vision machine. Proc. IEEE 76 (8), 1006–1015. IEEE.
De Cesarei, A., Codispoti, M., 2013. Spatial frequencies and emotional perception. Rev. Neurosci. 24 (1), 89–104. De Gruyter.
Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3), 27.
Colclough, G.L., et al., 2015. A symmetric multivariate leakage correction for MEG connectomes. Neuroimage 117, 439–448. Elsevier.
Eimer, M., 2011. 'The Face-sensitive N170 Component of the Event-related Brain Potential', the Oxford Handbook of Face Perception, vol. 28. Oxford University Press Oxford, UK, pp. 329–344.
Farah, M.J., et al., 1998. What is" special" about face perception? Psychol. Rev. 105 (3), 482. American Psychological Association.
Farahibozorg, S.-R., Henson, R.N., Hauk, O., 2018. Adaptive cortical parcellations for source reconstructed EEG/MEG connectomes. Neuroimage 169, 23–45. Elsevier.
Fiorentini, A., Maffei, L., Sandini, G., 1983. The role of high spatial frequencies in face perception. Perception 12 (2), 195–201. SAGE Publications Sage UK: London, England.
Gavrila, D.M., Philomin, V., 1999. 'Real-time object detection for" smart" vehicles', in *Computer Vision, 1999*. In: The Proceedings of the Seventh IEEE International Conference on. IEEE, pp. 87–93.
Goffaux, V., 2009. Spatial interactions in upright and inverted faces: Re-exploration of spatial scale influence. Vis. Res. 49 (7), 774–781. Elsevier.
Goffaux, V., et al., 2010. From coarse to fine? Spatial and temporal dynamics of cortical face processing. Cerebr. Cortex 21 (2), 467–476. Oxford University Press.
Goffaux, V., Gauthier, I., Rossion, B., 2003. Spatial scale contribution to early visual differences between face and object processing. Cognit. Brain Res. 16 (3), 416–424. Elsevier.
Goffaux, V., Rossion, B., 2006. Faces are" spatial"–holistic face perception is supported by low spatial frequencies. J. Exp. Psychol. Hum. Percept. Perform. 32 (4), 1023. American Psychological Association.
Halit, H., et al., 2006. Is high-spatial frequency information used in the early stages of face detection? Brain Res. 1117 (1), 154–161. https://doi.org/10.1016/j.brainres.2006.07.059. Elsevier.
Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends Cognit. Sci. 4 (6), 223–233. Elsevier.
Hegdé, J., 2008. Time course of visual perception: coarse-to-fine processing and beyond. Prog. Neurobiol. 84 (4), 405–439. Elsevier.
Henson, R.N., et al., 2010. A Parametric Empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Hum. Brain Mapp. 31 (10), 1512–1531. Wiley Online Library.
Hughes, H.C., Nozawa, G., Kitterle, F., 1996. Global precedence, spatial frequency channels, and the statistics of natural images. J. Cognit. Neurosci. 8 (3), 197–230. MIT Press.
Jones, R., Keck, M.J., 1978. Visual evoked response as a function of grating spatial frequency. Investig. Ophthalmol. Vis. Sci. 17 (7), 652–659. The Association for Research in Vision and Ophthalmology.
Kauffmann, L., et al., 2015. Rapid scene categorization: role of spatial frequency order, accumulation mode and luminance contrast. Vis. Res. 107, 49–57. Elsevier.
Kay, K.N., Yeatman, J.D., 2017. Bottom-up and top-down computations in word-and face-selective cortex. eLife 6, e22341 eLife Sciences Publications Limited.
King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cognit. Sci. 18 (4), 203–210. Elsevier.
Lu, Y., et al., 2018. Revealing detail along the visual hierarchy: neural clustering preserves acuity from V1 to V4. Neuron 98, 1–12. Elsevier.

Marr, D., 1982. Vision: a Computational Investigation into the Human Representation and Processing of Visual Information, vol. 1. Freeman and Company, WH San Francisco (2).
Mathworks, 2016. MATLAB R2016b. The MathWorks, Inc., Natick, Massachusetts, United States.
Mazer, J.A., et al., 2002. Spatial frequency and orientation tuning dynamics in area V1. Proc. Natl. Acad. Sci. Unit. States Am. 99 (3), 1645–1650. National Acad Sciences.
Mihaylova, M., Stomonyakov, V., Vassilev, A., 1999. Peripheral and central delay in processing high spatial frequencies: reaction time and VEP latency studies. Vis. Res. 39 (4), 699–705. Elsevier.
Mu, T., Li, S., 2013. The neural signature of spatial frequency-based information integration in scene perception. Exp. Brain Res. 227 (3), 367–377. https://doi.org/10.1007/s00221-013-3517-1. Springer-Verlag.
Musel, B., et al., 2014. Coarse-to-fine categorization of visual scenes in scene-selective cortex. J. Cognit. Neurosci. 26 (10), 2287–2297. MIT Press.
Musselwhite, M.J., Jeffreys, D.A., 1985. The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli. Vis. Res. 25 (11), 1545–1555. Elsevier.
Näsänen, R., 1999. Spatial frequency bandwidth used in the recognition of facial images. Vis. Res. 39 (23), 3824–3833. Elsevier.
Nguyen, V.T., Cunnington, R., 2014. The superior temporal sulcus and the N170 during face processing: single trial analysis of concurrent EEG–fMRI. Neuroimage 86, 492–502. Elsevier.
Ojanpää, H., Näsänen, R., 2003. Utilisation of spatial frequency information in face search. Vis. Res. 43 (24), 2505–2515. Elsevier.
Oostenveld, R., et al., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1.
Ouyang, G., Sommer, W., Zhou, C., 2016. Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: a solution based on a novel application of residue iteration decomposition. Int. J. Psychophysiol. 109, 9–20. Elsevier.
Parker, D.M., 1980. Simple reaction times to the onset, onset, and contrast reversal of sinusoidal grating stimuli. Atten. Percept. Psychophys. 28 (4), 365–368. Springer.
Parker, D.M., Dutch, S., 1987. Perceptual latency and spatial frequency. Vis. Res. 27 (8), 1279–1283. Elsevier.
Parker, D.M., Lishman, J.R., Hughes, J., 1992. Temporal integration of spatially filtered visual images. Perception 21 (2), 147–160. SAGE Publications Sage UK: London, England.
Parker, D.M., Lishman, J.R., Hughes, J., 1997. Evidence for the view that temporospatial integration in vision is temporally anisotropic. Perception 26 (9), 1169–1180. SAGE Publications Sage UK: London, England.
Pascalis, O., et al., 2005. Plasticity of face processing in infancy. Proc. Natl. Acad. Sci. Unit. States Am. 102 (14), 5297–5300. National Acad Sciences.
Peyrin, C., et al., 2006. Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. Brain Cognit. 62 (3), 214–220. Elsevier.
Quek, G.L., Liu-Shuang, J., Goffaux, V., Rossion, B., 2018. Ultra-coarse, single-glance human face detection in a dynamic visual stream. Neuroimage 176, 465–476.
Revina, Y., Petro, L.S., Muckli, L., 2018. Cortical feedback signals generalise across different spatial frequencies of feedforward inputs. Neuroimage 180 (Part A), 280–290.
Richler, J.J., Cheung, O.S., Gauthier, I., 2011. Holistic processing predicts face recognition. Psychol. Sci. 22 (4), 464–471. Sage Publications Sage CA: Los Angeles, CA.
Robson, J.G., 1966. Spatial and temporal contrast-sensitivity functions of the visual system. *Josa*. Optical Society of America 56 (8), 1141–1142.
Schwiedrzik, C.M., Freiwald, W.A., 2017. High-level prediction signals in a low-level area of the macaque face-processing hierarchy. Neuron 96 (1), 89–97. Elsevier.
Schyns, P.G., Oliva, A., 1994. From blobs to boundary edges: evidence for time-and spatial-scale-dependent scene recognition. Psychol. Sci. 5 (4), 195–200. SAGE Publications Sage CA: Los Angeles, CA.
Schyns, P.G., Oliva, A., 1999. Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. Cognition 69 (3), 243–265. Elsevier.
Sergent, J., 1982. About face: left-hemisphere involvement in processing physiognomies. J. Exp. Psychol. Hum. Percept. Perform. 8 (1), 1. American Psychological Association.
Sergent, J., 1986. Microgenesis of face perception. In: Aspects of Face Processing. Springer, pp. 17–33.
De Valois, R.L., De Valois, K.K., 1980. Spatial vision. Annu. Rev. Psychol. 31 (1), 309–341. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.
Vlamings, P.H.J.M., Goffaux, V., Kemner, C., 2009. Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? Journal of vision. The Association for Research in Vision and Ophthalmology 9 (5), 12.
Watt, R.J., 1987. Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *JOSA A*. Optical Society of America 4 (10), 2006–2021.
Yovel, G., et al., 2008. The face-selective ERP component (N170) is correlated with the face-selective areas in the fusiform gyrus (FFA) and the superior temporal sulcus (fSTS) but not the occipital face area (OFA): a simultaneous fMRI-EEG study. Journal of Vision. The Association for Research in Vision and Ophthalmology 8 (6), 401.
Zhang, J., et al., 2014. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. In: European Conference on Computer Vision. Springer, pp. 1–16.
Zhou, E., et al., 2013. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 386–391.