Identifying the Spatial Scales of Neural Representations in Human Visual System using Functional MRI with Wavelet Transforms

by

Xueying Ren

Bachelor of Social Science in Economics, Shenyang Normal University, 2013

Master of Science in Financial Economics, City University of London, 2015

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Xueying Ren

It was defended on

July 1, 2021

and approved by

Melissa Libertus, Ph.D., Associate Professor, Department of Psychology, University of Pittsburgh

Timothy Verstynen, Ph.D., Associate Professor, Department of Psychology, Carnegie Mellon University

Committee Chair: Marc Coutanche, Ph.D., Assistant Professor, Department of Psychology, University of Pittsburgh

Identifying the Spatial Scales of Neural Representations in Human Visual System using Functional MRI with Wavelet Transforms

Xueying Ren, MS

University of Pittsburgh, 2021

Visual information is thought to be processed in a hierarchical fashion from simple features processing in early visual areas to more complex information processing in higher-order visual areas. However, much remains to be learned about how the human visual system represents information, and whether information is represented at fine or coarse spatial scales. Traditional methods, such as spatial smoothing, which have been used to investigate spatial scales, cannot confine individual spatial scales stringently. In this study, we applied a novel method– the wavelet transforms – that could overcome these limitations in quantifying spatial scales. The goals of our study were two-fold: we will examine 1) if information is represented at fine or coarse spatial scales in various sub-regions in human visual system, and 2) if there is any correlation between the spatial-scale dependent information representations and receptive field size across different sub-regions in human visual system. To answer those questions, we applied a dual-tree complex wavelet transform (dt-CWT) to fMRI data volumes, which were collected as eight participants viewed four visual categories. Five orthogonal spatial scales were generated using the wavelets, and a set of new features were defined that took scale and directionality information into account for information representations. Those new features generated by the wavelets were then submitted to a multi-class classification analysis with a XGBoost machine learning algorithm. We adopted a Bayesian approach to evaluate the statistical significance of the classification results.

Contradicting with earlier findings, we did not find evidence to support multi-scale dependent neural representations in human visual system. However, our results should be interpreted with caution as we looked at neural representations from a different perspective. Further studies are needed to gain more insights into the multi-scale dependent neural representations.

Table of Contents

List of Figures

1.0 Introduction

1.1 Visual hierarchical organization

Much remains to be learned about how human visual system represent visual information. It is a well-established fact that from individual neurons to larger functional brain regions, human brain is organized hierarchically on different spatial scales (Hilgetag & Goulas, 2020). Visual information is also thought to be processed in a hierarchical fashion from simple features processing in early visual areas to more complex information processing in higher-order visual areas (Grill-Spector & Malach, 2004). There is a large interest in whether the information in human visual system is represented on fine or large spatial scales during visual processing. For instance, earlier studies have suggested a fine-scale simple visual feature (e.g., orientation) representation in early visual areas such as primary visual cortex (Boynton, 2005; Haynes & Rees, 2005; Kamitani & Tong, 2005). However, others have proposed that visual features decoding reflects a large-scale organization in primary visual cortex (de Beeck, 2010). Same question also remains to be answered in higher-order visual areas along the ventral object visual pathway, where human brain can form more complex representations for a large number of objects.

This visual hierarchical organization also drew support from the increasing receptive field size (RF size; the area of the visual field to which a neuron responds) as one proceeds from posterior to anterior regions along the ventral visual pathway (Rousselet et al., 2004). Specifically, neurons in early visual areas respond to visual input in a small patch of the visual field, so they have small RF to represent precise position formation of the stimulus in the visual field. In contrast, neurons in higher-order visual areas receive input from lower-order visual areas and integrate

information from a larger area of the visual field, thus they have larger RF. However, one question that if information represented on fine or coarse spatial scales is related to the RF size of that region has yet to be fully discovered. Therefore, the aims of our study were two-fold: we will examine 1) if information is represented on fine or coarse spatial scales in various sub-regions in human visual system to gain more insight into neural representations; 2) If there is any correlation between the spatial-scale dependent information representations and RF size across different sub-regions in human visual system. To this end, we applied a novel method – wavelet transforms, which defines spatial scales rigorously comparing to conventional smoothing method, to investigate spatial-scale dependent information representations in human visual system.

1.2 Limitations of traditional multivariate methods

There is an increasing interests in multivariate pattern analysis for distributed brain activity patterns decoding (Haxby et al., 2001; Haynes & Rees, 2005; Kamitani & Tong, 2005; Kriegeskorte et al., 2008). The nature of the distributed 'multi-voxel' representations grants the possibility that information in different voxels is represented on different spatial scales, which makes one wonder at which spatial scale most discriminating information is represented that can be decoded. Conventional methods that have been used to probe this question is called spatial smoothing. For instance, Op de Beeck (2010) investigated if there is a fine-scale or large-scale spatial organization in primary visual cortex using a spatial smoothing method. Specifically, he used different Gaussian smoothing kernels (e.g., 2mm, 4mm, and 8mm) to examine how smoothing influences the decoding performance of grating orientation from the brain patterns in primary visual cortex. He found that smoothing did not decrease the decoding performance, which

supported a large-scale spatial organization in primary visual cortex. However, one limitation of this conventional smoothing method is that different spatial scales are non-orthogonal to each other such that information represented on different spatial scales might be overlapping (Penny, 2002). This limitation can be overcome using the wavelet transforms, which allows for orthogonal information representations on different spatial scales.

1.3 Wavelet transforms

A wavelet is a small wave-like oscillation. Wavelet transforms can be seen as a way of decomposing signal into a hierarchically organized scales, and wavelet coefficient at a particular location represents the response of the signal to the wavelet applied. Wavelet coefficients at coarser scales represent low-frequency information, whereas coefficients at finer scales represent higher frequency information (Bullmore et al., 2004). Wavelets have advantages of being naturally adaptable to local or nonstationary signal properties in scale and space comparing to traditional Fourier method (Bullmore et al., 2004; see Daubechies (1993), Graps (1995), or Mallat (1999) for a review). Wavelets are becoming a powerful mathematical tool for complex data analysis, however, application of wavelets in neuroimaging studies to better understand human brain has yet to be fully explored. Wavelets have been most often applied to visual 2-D images such as images decomposition (Puckett et al., 2020), images compression (Abu-Rezq et al., 1999; Angelidis, 1994) or images denoising (Alexander et al., 2000; Zaroubi & Goelman, 2000). Only few studies have used wavelets in human brain volume space to understand its spatial scale organization (Hackmack et al., 2012; Sajda et al., 2002). One such study was conducted by Hackmack et al. (2012). They applied wavelets to structural MRI to extract information on

different spatial scales for disease classification. They found that decoding accuracies changed greatly among different scales, and interestingly larger scales containing low frequency information showed superior classification performance than finer scales.

In this present study, we applied a wavelet analysis to investigate the spatial-scale dependent information representations in human visual system, including various of sub-regions that are involved in low-level and high-level visual processing. Specifically, we hypothesized that the amount of information is represented at different spatial scales across sub-regions in visual system. We also hypothesized that the change of the spatial-scale dependent information representations is related to the receptive field size change across sub-regions in human visual system.

2.0 Materials and Methods

2.1 Participants

In our study, we used a dataset that were originally collected and reported in a study by Coutanche and Thompson-Schill (2012). Specifically, data were obtained from ten participants (females = 9, age range = 19–27 years, age mean = 23 years). All participants were right-handed with normal or corrected-to-normal vision, and no neurological issues. Informed content was obtained from all participants prior to experiment, and they were compensated for their time monetarily. All research procedures were approved by the human subject review board of University of Pennsylvania approved. Two participants were removed from further analyses due to excessive motion on functional runs, leaving eight participants in the analyses.

2.2 Stimuli and experimental design

During the scanning section, participants first completed an anatomical scan followed by functional runs. Functional runs were obtained as they viewed blocks of four different categories: faces, places, man-made objects, and fruits. Data were originally collected in long and short runs, but only the data from the short runs were analyzed in our study (see Coutanche & Thompson-Schill, 2012 for expriment detail). There were 16 short runs of 36 TRs. Each run consists of one block of each category with a total of 16 blocks of each stimulus type in the short runs, and the blocks of categories were randomized within each run. Each block contained ten images with 1-

back task, where participants were required to respond on a button box when they saw an image repeat. Each block lasted 9 s with images presented for 400 ms followed by 500 ms inter-stimulus interval, and blocks were separated with 12 s of rest. Participants saw 144 unique photographic images of each stimulus type in a random order.

2.3 Image acquisition and preprocessing

Data were acquired using a 3T Siemens Trio system with an eight-channel head coil and foam padding for head stabilization. T1-weighted anatomical scans were obtained at the start of each session for each participant (TR = 1620 ms, TE = 3 ms, TI = 950 ms, voxel size = 0.977 mm× 0.977 mm×1.000 mm), which was followed by functional images with interleaved gradient-echo EPI (TR = 3000 ms, TE = 30 ms, field of view = 19.2cm×19.2cm, voxel size = 3.0mm×3.0mm×3.0mm) for twenty runs includind16 short runs and four long runs.

Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996) was used to preprocess the imaging data, and the first four TRs of each functional run were removed allowing for steady signal acquisition. All functional images were slice-time corrected, and a motion correction algorithm was used to register all volumes to a mean functional volume. A high-pass filter with threshold of 0.0159 was used to remove low-frequency trends from all runs. Voxel activation was scaled with a mean of 100 and a maximum limit of 200.

2.4 Regions of interest (ROIs)

2.4.1 ROIs for anatomical adjacency analysis

We adapted a brain atlas originally from Wang and colleagues' study (2015), which covers 17 topographically defined regions (V1v, V1d, V2v, V2d, V3v, V3d, V3A, V3B, hV4, VO1, VO2, PHC1, PHC2, LO1, LO2, TO1, TO2) in human visual system (see Wang et al., 2015 for detailed description). The brain atlas was in a standard surface space (i.e., *fsaverage* space), so we mapped the atlas to each individual subject's native volume space. For the purpose of our analyses, we grouped those regions to create five larger ROIs due to the minimum number of voxels required for the wavelet transforms (see Fig. 1). They were primary visual cortex (V1; V1v and V1d), dorsal-occipital (V2d, V3d, and V3A), ventral-occipital (V2v, V3v, and hV4), ventral occipitotemporal cortex (VOT; VO1, VO2, PHC1, and PHC2), and lateral occipitotemporal cortex (LOT; V3B, LO1, LO2, TO1, and TO2). The size of those five ROIs ranges from 150 to 350 voxels. We also included ventral temporal cortex (VT; reported in Coutanche et al., 2012), which was constructed from segmented gray matter of the parahippocampal and inferior temporal gyri as well as the fusiform and lingual gyri, consisting of around 1600 voxels.

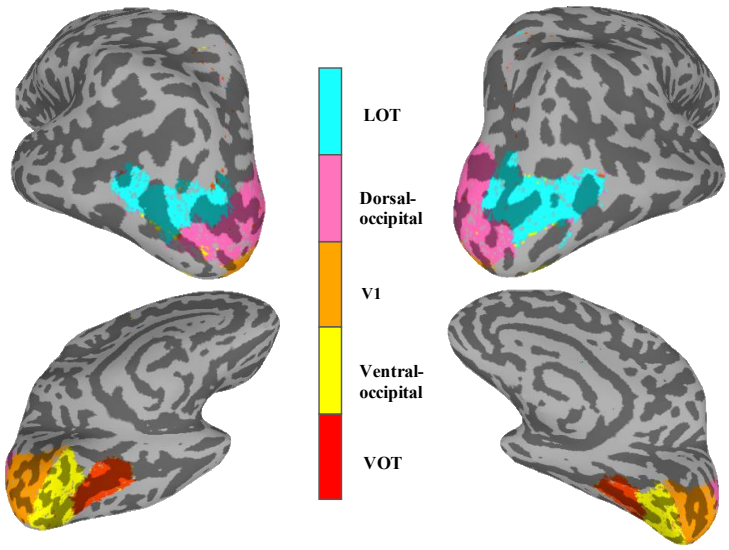

Figure 1 Five ROIs were created for the anatomical adjacency analysis. LOT: lateral occipitotemporal cortex; VOT: ventral occipitotemporal cortex.

2.4.2 ROIs for RF size relativeness analysis

We were also interested if the amount of information represented on different spatial scales was related to the RF size of that region, so regions with similar RF size were grouped together. We referenced RF size from a Human Connectome Project (HCP), which consists of high-field (7T) fMRI scanning of 181 subjects in six 5-min population receptive field mapping runs. Stimuli consisted colorful object textures presented within a slowly moving aperture (see Benson et al., 2018 for detailed experimental description).We created four ROIs for RF size relativeness analysis (see Fig.2), and the size of those regions ranges from 200 to 400 voxels.

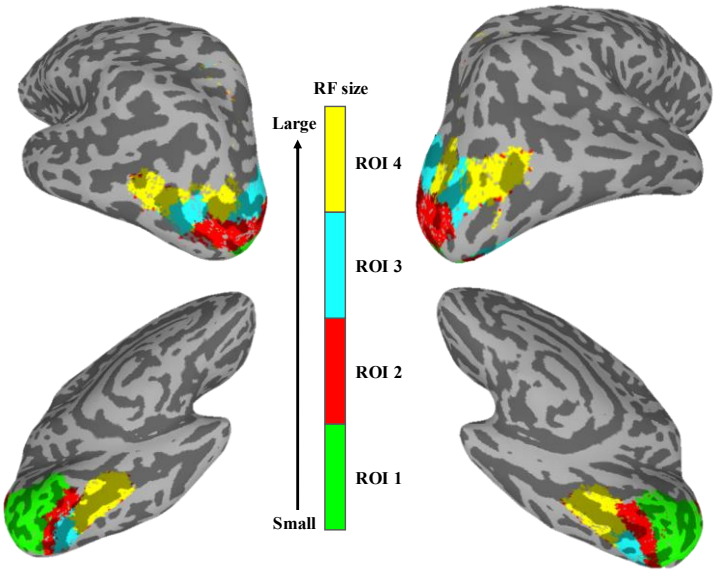

Figure 2 Four regions created for RF size relativeness analysis: ROI 1 (V1v, V1d, and V2v; RF size ranges from 0° to 0.5°), ROI 2 (V2d, V3v, and V3d; RF size ranges from 0.5° to 1°), ROI 3(V3A, hV4, and LO1; RF size ranges from 1° to 2°), and ROI 4 (V3B, LO2, TO1, TO2, VO1, VO2, PHC1, and PHC2; RF size ranges from 2° to 5°).

2.5 Wavelet decomposition

We applied a dual-tree complex wavelet transform (dt-CWT) to our preprocessed fMRI data volumes using *dtcwt* package in Python (Wareham et al., 2014). dt-CWT employs two real discrete wavelet transforms (DWT), which use different sets of filters. dt-CWT has the advantages of being nearly shift-invariant and directionally selective in two or higher dimensions (see Selesnick et al., 2005 for a detailed explanation). Being shift-invariant allows wavelet coefficients to not get perturbed largely by a small signal shift in the time or space domain. Being direction selective means that 28 orientated sub-bands can be isolated at each spatial scale for a 3-

dimensional fMRI data volume. Therefore, dt-CWT is highly suitable for handling large neuroimaging data (Selesnick et al., 2005). It is worth noting that, in our study, orientated sub-bands and scales were applied in the voxel volume space instead of 2-D image space.

For each subject and each TR, we applied dt-CWT and obtained five unique spatial scales from fine to coarse. Specifically, we covered scales ranging approximately from $[6\ mm]^3$ at scale I, $[12\ mm]^3$ at scale II, $[23\ mm]^3$ at scale III, $[45\ mm]^3$ at scale IV, to $[90\ mm]^3$ at scale V. One each scale, 28 oriented sub-bands were also generated, and each oriented sub-band contained complex wavelet coefficients. Wavelet coefficients represent the response of a signal to the wavelet applied at a specific location within a given sub-band at a particular scale. The magnitude of the coefficients depends on its scale and directionality information within a sub-band, and number of coefficients on one sub-band depends on its spatial scale. Specifically, in our study, scale I (fine scale) has coefficients of 32 x 32 x 21 on each sub-band, scale II has coefficients of 16 x 16 x 11 on each sub-band, and scale III has coefficients of 8 x 8 x 6 on each sub-band. There were coefficients of 4 x 4 x 3 at scale IV and 2 x 2 x 2 at scale V (large scale) on each sub-band (see Fig.3 for illustration). The variance of the coefficients (log-scaled) at each sub-band and each scale was then calculated to represent the "amount of information" (AOI hereafter), and those 28 AOI values (i.e., 28 new features for each category) were used for further classification analysis at each spatial scale.

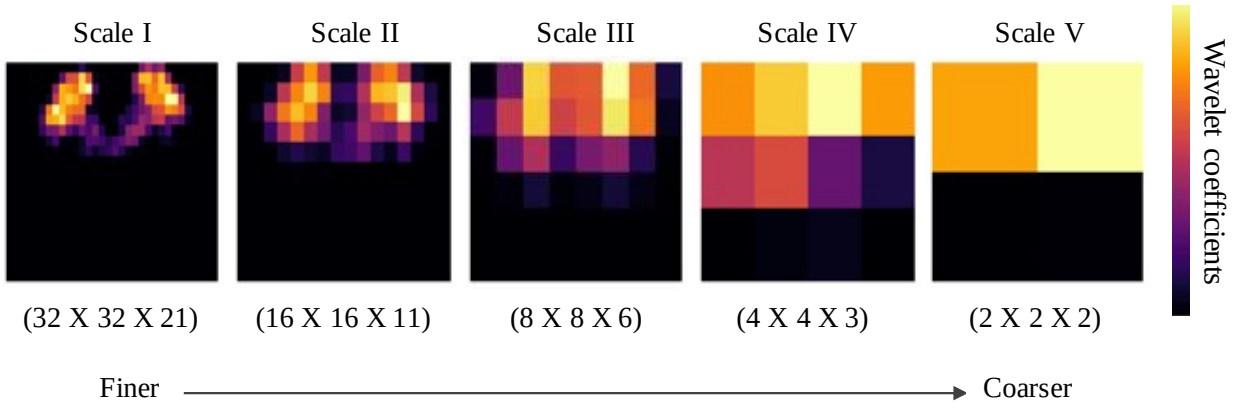

Figure 3 Wavelet transform illustration (condition = faces, one subject) with wavelet coefficients projected onto 2D surface for five different scales in VT. Scale I indicates finer spatial scale, whereas scale V indicates coarser spatial scale. Values in the pathesis indicate the number of coefficients transfromed using the wavelets on each scale. Note that color scale was adjusted for better visualization, not indicating actual values.

2.6 Multi-class classification using XGBoost algorithm

For each spatial scale, we submitted those AOI values for a multi-class classification analysis using the Extreme Gradient Boosting algorithm (XGBoost) (Chen & Guestrin, 2016) in Python. XGBoost is computationally efficient and has advantages of fast parallel processing and high flexibility (Torlay et al., 2017), which is highly applicable to neuroimaging studies with superior performance at classification problems (Tahmassebi et al., 2018; Yu et al., 2018).

A leave-one-run-out cross-validation scheme (LORO-CV) was applied to validate the generalizability of classifier performance. Data were partitioned based on the 16 runs (16-fold CV). By leaving one run out, classifier was trained on 15 runs of the data to learn the AOI vectors

of four classes (faces, man-made objects, fruits, and places) and tested on the unseen remaining run of the data. This procedure was repeated 16 times so that each run was used as the test set once.

2.7 Bayesian statistical analysis

Bayesian statistical analysis is a trending method that has been used increasingly in scientific research. In this current study, we adopted Bayesian statistics to investigate if there was rigorous evidence for spatial scale-dependent information representation in human visual system. In this study, we conducted a likelihood ratio test (LRT) to test the effects of scale-dependent information representation. LRT is calculated as a likelihood ratio of two hypotheses given a dataset. Specifically, LRT is the probability of the data (D) given the alternative hypothesis (H_1) divided by the probability of the data given the null hypothesis (H_0), which is often denoted as:

$$LRT = \frac{likelihood\ of\ data\ given\ H_1}{likelihood\ of\ data\ given\ H_0} = \frac{P(D|H_1)}{P(D|H_0)}$$

In our study, the null hypothesis H_0 is that there is no reliable association between the data (i.e., AOI for different categories) and the labels that can be learned by the classifier, and H_1 is the opposite.

For the null hypothesis that there is no reliable relationship between the data and labels that can be learned by the classifier, we conducted a permutation test (Golland & Fischl, 2003), which involves assigning new labels to observations to test the null hypothesis (Etzel, 2017; Stelzer et al., 2013). Specifically, we used balanced block permutation test (Schreiber & Krekelberg, 2013) by shuffling entire blocks of labels rather than single TR labels in order to maintain the temporal correlations within the block as in the non-shuffled dataset. Blocks of each class were balanced to

ensure equal occurrence on all classes for each permutation. The blocks of labels were shuffled for each of the 16 CV folds. For each permutation run, we estimated the likelihood of the prediction using the predicted class probabilities (from the null model) given by the XGBoost classifier.

To test the alternative hypothesis, we adopted a bootstrap approach, which is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement (Efron & Tibshirani, 1994). Specifically, we performed block-wise resampling by using 15 runs of data (60 blocks) with replacement as the training dataset, then tested on the remaining run. Similar to the permutation test, for each bootstrap run, we calculated the likelihood of the prediction using the predicted class probabilities (from the alternative model) given by the XGBoost classifier.

In order to combine the permutation and bootstrap tests to obtain the distribution for the likelihood ratio effects, we first estimated the likelihood of the null model with a single permutation run. Next, we estimated the likelihood of the alternative model with a single bootstrap run. We then calculated the likelihood ratio test as the ratio of the likelihood obtained from the null and alternative models (based on the aforementioned LRT formula). This procedure was repeated for 200 integrations for each CV run (overall 3200 in total). The advantage of this method is that it provides both a "significance test" and an evidentiary test. The probability of the LRT < 1 indicates an approximated p-value that allows for null hypothesis judgement, and the mean of LRT suggests the degree of evidence for the alternative model.

3.0 Results

3.1 Anatomical adjacency analysis

For each ROI in the anatomical adjacency analysis, we conducted multi-class classification based on the differences in AOI for different categories at each spatial scale. Bayesian statistical method was used, and LRT were calculated by estimating a likelihood function with a permutation test and a bootstrap approach. The probability of LRT < 1 was calculated (p-values) to obtain an approximated p-value to estimate the probability that the observed data being consistent with the null model (e.g., the smaller p-values, the greater chance to reject null model). The mean of LRT was also calculated which can be interpreted as the degree of evidence for the alternative model. Based on the results for p-values (*ps* > 0.3), there were no significant effects observed for the alternative model at any scale and in any region. Similarly, for the degree of evidence, there were no subjects that showed evidence to support the alternative model at any scale and in any region (ranges from 0.7 to 1.7). Full results for p-values and degree of evidence were shown in Figure 4 and 5.

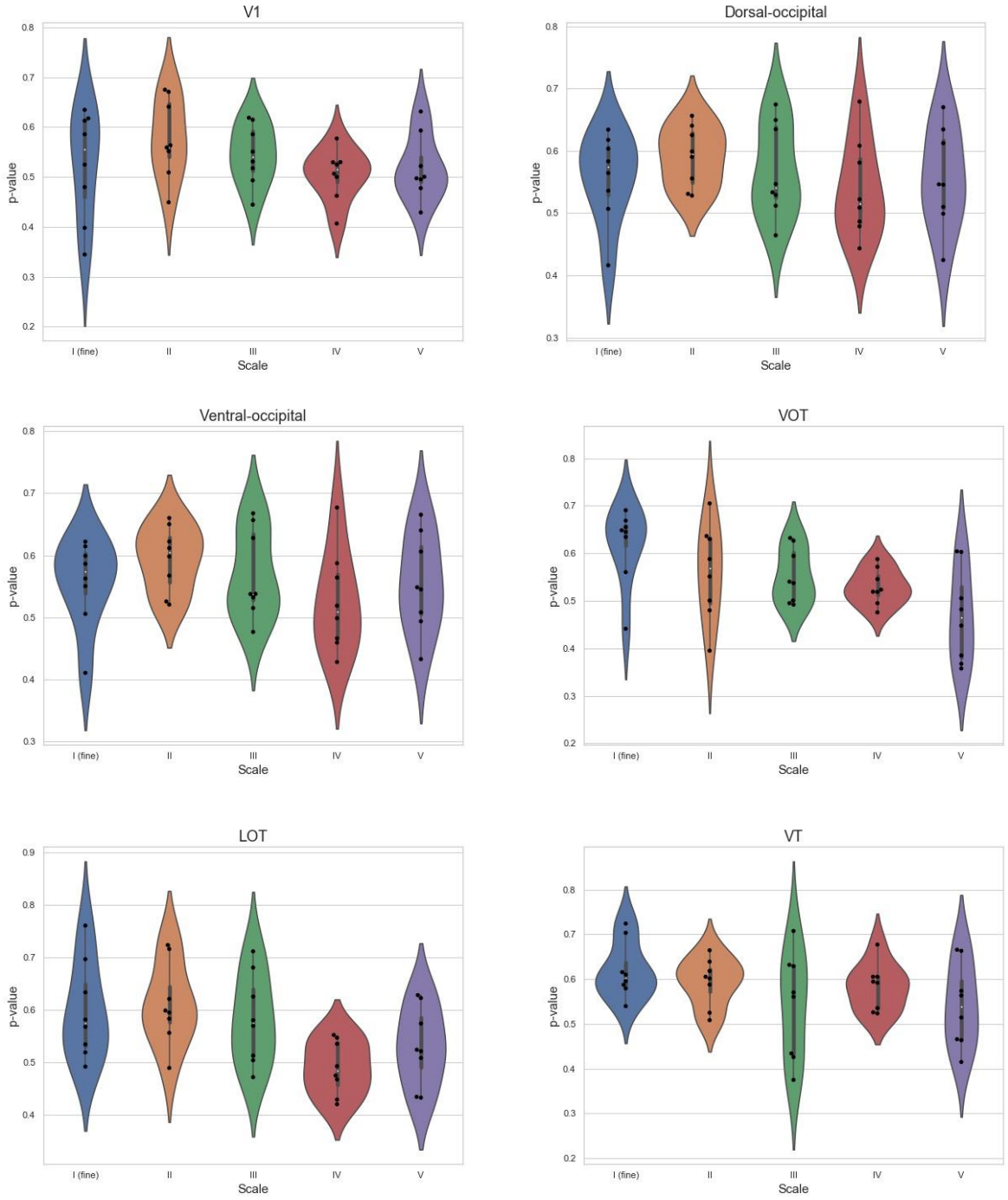

Figure 4 Results of p-values at each scale in six adjacency analysis ROIs. Black dots indicate the p-values for eight subjects.

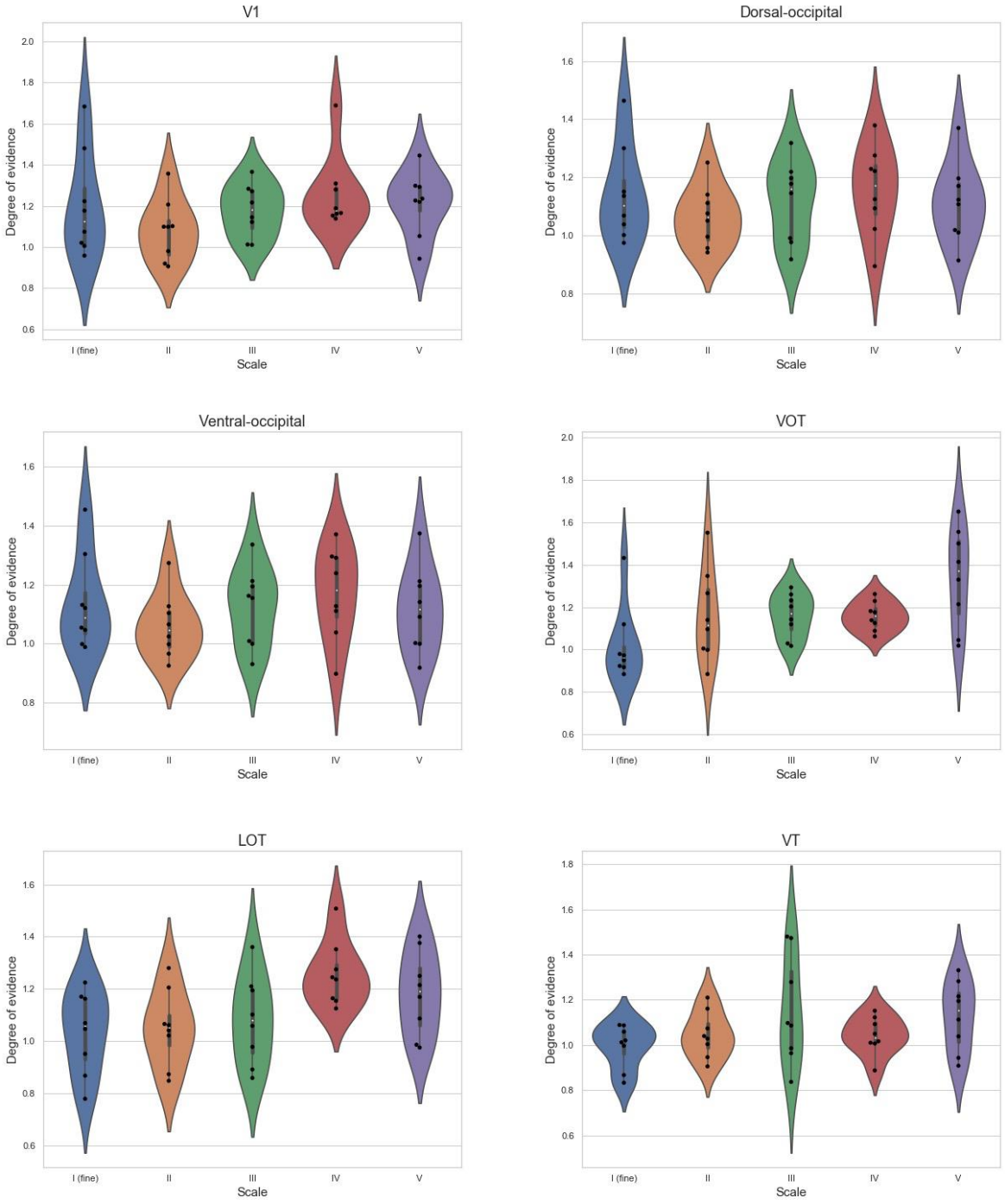

Figure 5 Results of degree of evidence at each scale in six adjacency analysis ROIs. Black dots indicate the degree of evidence for eight subjects.

3.2 RF size relativeness analysis

We were also interested if there is any correlation between the spatial-scale dependent information representation and receptive field size across different sub-regions in human visual system. To answer this question, we grouped regions with similar RF size to created four larger ROIs with increasing RF size. Based on the results for p-values ($ps > 0.3$), there were no significant effects observed for the alternative model at any scale and in any region. Additionally, there were no subjects that showed evidence to support the alternative model at any scale and in any region (ranges from 0.8 to 1.9). Full results for p-values and degree of evidence were shown in Figure 6 and 7.

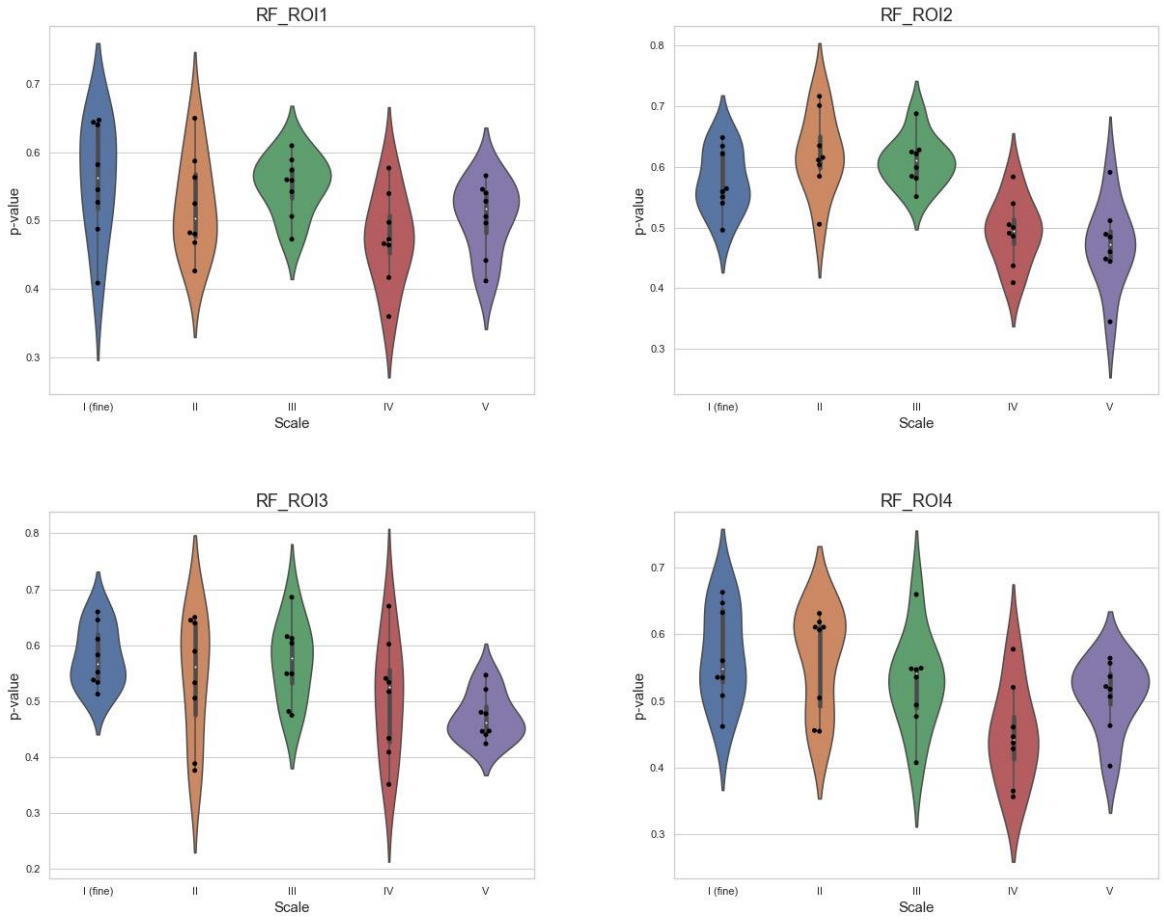

Figure 6 Results of p-values at each scale in four RF analysis ROIs. Black dots indicate the p-values for eight subjects.

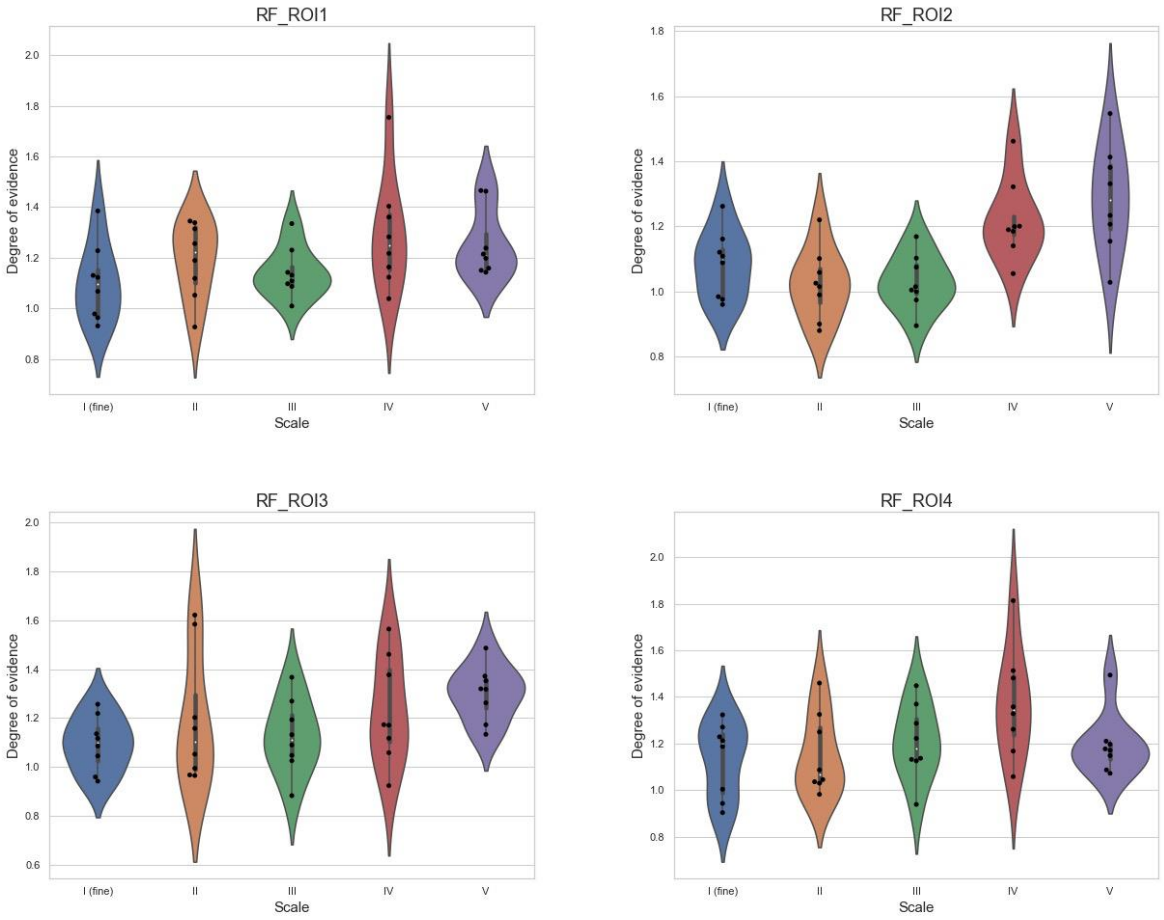

Figure 7 Results of degree of evidence at each scale in four RF analysis ROIs. Black dots indicate the degree of evidence for eight subjects.

4.0 Discussion

In this study, we implemented a novel method to analyze spatial-scale dependent information representations in human visual system. Specifically, we applied a dt-CWT to the fMRI data volumes and generated five orthogonal spatial scales with 28 oriented sub-bands at each spatial scale. We defined the "amount of information" (AOI) by calculating the variance of the coefficients at each orientated sub-band and each spatial scale, and those 28 values capturing the AOI were submitted for a classification analysis using XGBoost algorithm at each spatial scale. Bayes statistical method was adopted to test our hypothesis that 1) the amount of information is represented at different spatial scales across sub-regions in visual system, and 2) the change of the spatial-scale dependent information representations is related to the RF size change across sub-regions in human visual system. There were no significant effects found for spatial-scale dependent information representations in the brain regions considered in this study. We were also interested if the amount of information represented at different spatial scales was related to the RF size of that brain region. However, we did not find such association.

Firstly, our study did not show evidence that information is represented on multiple spatial scales in human visual system, which conflicts with previous findings. For instance, Brants and colleagues (2011) examined the scale organization for category and exemplar selectivity in ventral visual cortex using a spatial smoothing method, and confirmed the existence of multi-scale organization in ventral visual pathway. However, due to the limitations of spatial smoothing, information represented on different spatial scales might be overlapping or important information getting lost during smoothing, which rendered the multi-scale arguments susceptible to skepticism. In contrast, we used wavelet transforms to define spatial scales rigorously and examined spatial-

scale dependent information representations in several sub-regions in human visual systems, and did not find the evidence to support the multi-scale representations in human visual system. However, the discrepancy between our study and others could be due to the different methodologies being used.

It is a well-established fact that visual information processing changes from simple visual features in early visual areas (e.g., V1, V2) to more complex visual feature in higher-order visual areas (e.g., VT). For example, primary visual cortex has a columnar organization that can encode orientation information, which might prompt people to think that information in this area might represent at finer spatial scales. However, we did not find such evidence. In other words, there were no significant differences in AOI represented at any spatial scales for those categories. Our findings conflict with earlier work (de Beeck, 2010; Furmanski & Engel, 2000; Sasaki et al., 2006; Serences et al., 2009). For example, Op de Beeck (2010) also suggested a large-scale spatial organization in early visual area with the evidence showing that spatial smoothing did not decrease the decoding performance. However, when defining the spatial scales in a more rigorous way, we did not find such spatial-scale dependent information representations. Another possibility is that higher-level categorical information is not represented in this area, since this area is more designated for simple features representations.

Higher-order visual areas can form complex object representations by aggregating the information from earlier areas. Why higher-order visual areas can generate distinct representations for a large number of objects (Haxby et al., 2001)? Whether information in this brain area is represented at fine or coarse spatial scales? However, we did not find evidence to support distinct spatial-scale dependent information representations in ventral-occipito-temporal regions, area VT and lateral occipitotemporal region. One explanation would be that most of the information

represented in those areas are shared or overlapped across different categories. This is in line with the distributed theory (Haxby et al., 2001; Ishai et al., 1999; Tanaka, 1996, 1997) in which information for different categories is represented diffusely in higher-order visual areas. An alternative explanation would be that there are no spatial-scale dependent representations in those areas. However, the interpretation still remains open that worth further investigation.

Another goal of our study was to find out if the change of the spatial-scale dependent information representations is related to the RF size change across different sub-regions in human visual system. Specifically, we hypothesized that, if there were spatial-scale dependent representations in human visual system, in regions with a larger RF size, more information (discriminating amount of information) would represent at either finer or coarser spatial scale. We did not find such association. However, our findings do not rule out the possibility that such relationship do exist, and future studies should aim to investigate this potential correlation.

It is worth noting that one feature that sets our study apart from previous work is that we looked at information representations from a different perspective, i.e., wavelet transformed spatial-scale dependent representations, so our findings should be interpreted with more caution.

One limitation of our study is the relatively small sample (participants) size, which might cause the evidence to be less detectable. Future studies should aim to include more participants for evidence detection. Another limitation would be the restricted ROIs sizes. We were interested in how spatial scale-dependent information representations changes across different sub-regions in human visual systems, therefore, those sub-regions were relatively small. However, there is a trade-off between small region size and the amount of information represented in that region. In other words, small ROIs would restrict the amount of the information represented, which might

render less detectable differences across different categories. Wavelet transforms analysis would be most beneficial when it comes to larger dataset with more information imbedded within.

In sum, in this study, we adopted a novel method to explore the information representations in human visual system by using a new set of features that take scale and directionality information into account. Wavelet transforms not only can define fine versus coarse spatial scales more rigorously, which is not possible using conventional spatial smoothing method, but also can provide a new perspective to look at neural representations. Specifically, we covered scales ranging from $[6\ mm]^3$ at scale I to $[90\ mm]^3$ at scale V, and our results did not support the multi-scale neural representations in human visual system. It is possible that the limitations of our study restricted our findings, and our results do not rule out the possibility of multi-scale neural representations in the visual system. Future studies should continue to investigate this matter to gain more insights into both the new methodology and the multi-scale neural representations in human visual system.

Bibliography

Abu-Rezq, A., Tolba, A. S., Khuwaja, G. A., & Foda, S. (1999). Best parameters selection for wavelet packet-based compression of magnetic resonance images. *Computers and Biomedical Research*, *32*(5), 449–469.

Alexander, M., Baumgartner, R., Windischberger, C., Moser, E., & Somorjai, R. (2000). Wavelet domain de-noising of time-courses in MR image sequences. *Magnetic Resonance Imaging*, *18*(9), 1129–1134.

Angelidis, P. (1994). MR image compression using a wavelet transform coding algorithm. *Magnetic Resonance Imaging*, *12*(7), 1111–1120.

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., Van Essen, D. C., Yacoub, E., Ugurbil, K., Winawer, J., & Kay, K. (2018). The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *Journal of Vision*, *18*(13), 23. https://doi.org/10.1167/18.13.23

Boynton, G. M. (2005). Imaging orientation selectivity: Decoding conscious perception in V1. *Nature Neuroscience*, *8*(5), 541–542.

Brants, M., Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2011). Multiple scales of organization for object selectivity in ventral visual cortex. *NeuroImage*, *56*(3), 1372–1381. https://doi.org/10.1016/j.neuroimage.2011.02.079

Bullmore, E., Fadili, J., Maxim, V., Şendur, L., Whitcher, B., Suckling, J., Brammer, M., & Breakspear, M. (2004). Wavelets and functional magnetic resonance imaging of the human brain. *Neuroimage*, *23*, S234–S249.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Coutanche, M. N., & Thompson-Schill, S. L. (2012). The advantage of brief fMRI acquisition runs for multi-voxel pattern detection across runs. *NeuroImage*, *61*(4), 1113–1119.

Daubechies, I., & Bates, B. J. (1993). *Ten lectures on wavelets*. Acoustical Society of America.

de Beeck, H. P. O. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage*, *49*(3), 1943–1948.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Etzel, J. A. (2017). MVPA significance testing when just above chance, and related properties of permutation tests. *2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 1–4.

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience*, *3*(6), 535–536.

Golland, P., & Fischl, B. (2003). Permutation tests for classification: Towards statistical significance in image-based studies. *Biennial International Conference on Information Processing in Medical Imaging*, 330–341.

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*, *2*(2), 50–61.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.*, *27*, 649–677.

Hackmack, K., Paul, F., Weygandt, M., Allefeld, C., Haynes, J.-D., Initiative, A. D. N., & others. (2012). Multi-scale classification of disease using structural MRI and wavelet transform. *Neuroimage*, *62*(1), 48–58.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Penny, W. (2002). Wavelet smoothing of fMRI activation images. In *Technical Report*. Wellcome Department of Imaging Neuroscience UCL, UK.

Puckett, A. M., Schira, M. M., Isherwood, Z. J., Victor, J. D., Roberts, J. A., & Breakspear, M. (2020). Manipulating the structure of natural scenes using wavelets to study the functional architecture of perceptual hierarchies in the brain. *NeuroImage*, 117173.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*(8), 363–370.

Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., & Tootell, R. B. (2006). The radial bias: A different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron*, *51*(5), 661–670.

Schreiber, K., & Krekelberg, B. (2013). The statistical analysis of multi-voxel patterns in functional imaging. *PLoS One*, *8*(7), e69328.

Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. C. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, *22*(6), 123–151.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *Neuroimage*, *65*, 69–82.

Tahmassebi, A., Gandomi, A. H., McCann, I., Schulte, M. H., Goudriaan, A. E., & Meyer-Baese, A. (2018). Deep learning in medical imaging: FMRI big data analysis via convolutional neural networks. In *Proceedings of the Practice and Experience on Advanced Research Computing* (pp. 1–4).

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*(1), 109–139.

Torlay, L., Perrone-Bertolotti, M., Thomas, E., & Baciu, M. (2017). Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Informatics*, *4*(3), 159–169.

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Wareham, R., Forshaw, S., & Roberts, T. (2014). *dtcwt: A Python dual tree complex wavelet transform library*. DOI.

Yu, W., Na, Z., Fengxia, Y., & Yanping, G. (2018). Magnetic resonance imaging study of gray matter in schizophrenia based on XGBoost. *Journal of Integrative Neuroscience*, *17*(4), 331–336.

Zaroubi, S., & Goelman, G. (2000). Complex denoising of MR data via wavelet analysis: Application for functional MRI. *Magnetic Resonance Imaging*, *18*(1), 59–68.