Functional Magnetic Resonance Imaging Data Clustering

by

Jinae Lee

(Under the direction of Cheolwoo Park)

Abstract

In this dissertation, we aim to evaluate brain activation using Functional Magnetic Resonance Imaging (fMRI) data and activation changes across time associated with practice-related cognitive control during eye movement tasks. FMR images are acquired from participants engaged in antisaccade (generating a glance away from a cue) performance at two time points: 1) at a pre-test before any exposure to the task, and 2) at a post-test, after one week of daily practice on antisaccades, prosaccades (glancing towards a target), or fixation (maintaining gaze on a target). The three practice groups are compared across the two time points. Since there are many problems inherent in fMRI data such as the huge data size, various sources of noise, ill-balanced groups, and temporal correlations, it is challenging to detect the activated regions of the brain. We propose a model-free clustering technique based on wavelet analysis to overcome the problems inherent in fMRI data. The proposed clustering technique is composed of several steps: detrending, data aggregation, wavelet transform and thresholding, the adaptive pivotal thresholding test, principal component analysis, and K-means clustering. The main clustering algorithm is built in the wavelet domain to account for temporal correlation. We apply the adaptive pivotal thresholding test based on wavelets to significantly reduce the high dimension of the data. We cluster the thresholded wavelet coefficients of the remaining voxels (the units of the images in the three dimensional space) using principal component analysis K-means clustering. Over the series of analyses, we find that the antisaccade practice group is the only group to show decreased activation from pre- to post-test in saccadic circuitry. In order to examine the proposed wavelet-based clustering approach, we perform a simulation study by adding artificial fMRI signals to the real resting-state data, and the results demonstrate its effectiveness in fMRI clustering analysis. We also

conduct Regions of interest (ROI) analysis to locate the regions in which attenuations occur. We apply bootstrap resampling and the mixed model with feature extraction approach to the eleven bilateral neural ROIs. We observe decreasing activation in the supplementary eye field, frontal eye field, superior parietal lobe, and cuneus for the antisaccade practice group.

INDEX WORDS: Bootstrap, Clustering, FMRI, Mixed model, Principal component analysis, Saccades, Wavelets

Functional Magnetic Resonance Imaging Data Clustering

by

Jinae Lee

M.S., The University of Georgia, 2006

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2013

Functional Magnetic Resonance Imaging Data Clustering

by

Jinae Lee

Approved:

Major Professor: Cheolwoo Park

Committee: Nicole A. Lazar
Jennifer E. McDowell
Paul Schliekelman
Lynn Seymour

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2013

Dedication

This dissertation is dedicated to my beloved husband, Dr. Hyungwook Park, who is the most precious gift God has ever given me. Everything was possible because of his endless love and sacrifice.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Magnetic Resonance Imaging (MRI) can be used to map changes in brain hemodynamics that correspond to mental processes, and it has extended traditional anatomical imaging to include maps of human brain function. Functional Magnetic Resonance Imaging (fMRI), which is a non-invasive neuroimaging technique conducted using an MRI system, makes it possible to observe the structures that participate in specific functions. FMRI can be considered as a technique for measuring brain activity in the sense that it can be used to produce activation maps showing which parts of the brain are involved in a particular mental operation or respond to a given stimulus. This new ability to directly observe brain function opens an array of new opportunities to advance our understanding of brain organization, as well as a potential new standard for assessing neurological status and neurosurgical risk. This technique can bring improved measurement of subtle brain activity that is associated with perceptual, motor, and cognitive processes. Such processes are usually evidenced by increased blood volume and blood flow levels in task-related neural circuits. Therefore, fMRI data typically are based on the "Blood Oxygen Level Dependent" (BOLD) signal, which is the local hemodynamic response relative to a particular task or stimulus.

In the past few decades, there has been a rapidly emerging body of literature in neuroimaging studies performed using fMRI to localize specific functions of the human brain. The field of fMRI is intrinsically interdisciplinary in that image acquisition and image analysis are propelled by researchers in neuroscience, psychology, physics, and statistics, among others. There are several common objectives in the analysis of fMRI data. These include localizing regions of the brain activated by a task, determining distributed networks that correspond to brain function, and making predictions about psychological or disease states. Each of these objectives can be achieved through the application of suitable statistical methods. Therefore statisticians play an important role in understanding the nature of the data and obtaining relevant results. Neuroscientists can use and interpret those results (Lindquist, 2008).

Functional connectivity in the brain is an important research topic in fMRI studies. Connectivity refers to networks that model or explain relationships between brain regions. It

involves finding regions or a network of regions in the brain that are related to a particular cognitive task (Lazar, 2008). Functional connectivity analysis aims to detect and characterize the coherent patterns of activity as a means of identifying brain systems by looking at temporal correlations (Lazar, 2008; Marco, 2009).

Among many approaches, fMRI clustering provides an effective way of discovering functional connectivity in the brain (Venkataraman et al., 2009). There are two dominant methods in fMRI clustering analysis: clustering based on the voxelwise hypothesis test and clustering based on grouping time series. The first method tests each voxel to determine whether it is activated in response to a given stimulus or not and constructs a statistical parametric map (SPM) of the brain using the test results. A typical method of constructing a SPM is to calculate a test statistic at each voxel and plot its testing result in a map (Lazar, 2008). Regarding the test statistic, one can use a simple t or F statistic to compare the mean levels of activation, or apply more sophisticated general linear models (Friston et al., 1994). These hypothesis-based approaches, however, may suffer from low signal-to-noise ratio (SNR) and multiple testing problems since analysis is performed at the voxel level (Worsley, 2003). The second clustering method is a data-driven approach and partitions a brain based on the patterns of voxel time courses without specifying a parametric model. It assumes that the temporal patterns of activation are organized in a spatially coherent fashion such that clustering extracts the main temporal patterns and partitions the brain by grouping similarly behaved functions together (Goutte et al., 1999).

FMRI analysis has given statisticians various challenges and opportunities. Besides different types of noises such as thermal, systematic, physiological, and motion-related noises in the data, two other main challenges are present in the analysis of fMRI data: correlation, both spatial and temporal, among observations, and a massive amount of data with hundreds of thousands of voxels over hundreds of time points (Lazar, 2008). Hence, fMRI clustering methods that circumvent these challenges have been proposed (Balslev et al., 2002; Fadili et al., 2000; Heller et al., 2006; Ye et al., 2009).

Cognitive control mediates the process through which tasks transition from new or unfamiliar to learned or skilled. With practice, a task becomes less effortful, resulting in improved performance and a reduced need for cognitive control (e.g, Chein and Schneider, 2005; Jansma et al., 2001; Schneider and Chein, 2003; Schneider and Shiffrin, 1977). Modifications in the neural circuitry supporting task performance also occur following practice (e.g, see Chein and Schneider, 2005; Kelly and Garavan, 2005, for reviews). These changes in brain

activation likely include changes in the specific neural circuitry supporting the response of interest as well as general changes reflecting a decreased need for cognitive control.

In this dissertation, changes in brain activation across time are examined before and after daily exposure to a set of eye movement tasks. Among the various types of eye movements, saccades rapidly redirect one's gaze to a location of interest. Prosaccades indicate more reflexive eye movements in the sense that a person simply redirects his or her gaze to a newly appearing visual cue. Antisaccades, however, indicate more complex and volitional eye movements in the sense that a person redirects his or her gaze to the mirror image location of a newly appearing peripheral visual cue. As such, a correct antisaccade response requires the inhibition of a glance toward the cue and generation of a voluntary saccade to an unmarked location in the opposite visual field.

Saccadic performance is supported by a network of subcortical and cortical regions, as identified via neuroimaging and other techniques (Camchong et al., 2008; Dyckman et al., 2007; Ford et al., 2005; Keedy et al., 2006; McDowell et al., 2008; Muri et al., 1998; O'Driscoll et al., 1995; Paus, 1996; Raemaekers et al., 2002; Sweeney et al., 2007). Although the basic circuitry is the same for prosaccades and antisaccades, the increased complexity of antisaccades is supported by increased activation of existing circuitry and/or the recruitment of additional neural regions into the circuitry. As such, the saccade circuitry provides a specific, well-studied system for understanding changes in brain activation associated with cognitive control and practice.

We evaluate the changes across time in neural system activation after participants practice either specific or nonspecific eye movement tasks using fMRI. Participants concentrate on antisaccade performance during fMR image capture at two test sessions. The pre-test session is performed before any exposure to the task; the post-test session is performed after one week of daily practice on eye movement tasks. Participants are randomly divided into three different task groups: (a) antisaccades, (b) prosaccades, or (c) fixation (maintaining gaze on a target). Each participant practices daily a single type of eye movement only from the day after the pre-test to the day before the post-test. It is hypothesized that (1) the participants in the antisaccade practice group, which have task-consistent practice, would show changes in neural circuitry (as measured by the BOLD signal) and that (2) the participants in the prosaccade and fixation groups, which practice eye movement control tasks generally, but not antisaccades specifically, would not show changes in neural circuitry. It is further hypothesized that (i) for saccade tasks, saccade-related circuitry would show a decreased

BOLD signal over time as a result of the circuitry becoming more efficient and (ii) in the antisaccade practice group only, prefrontal cortex (PFC), frontal eye fields (FEFs), and striatum would show decreased BOLD signal due to decreased need for higher level cognitive control (Camchong et al., 2008; Munoz and Everling, 2004; Raemaekers et al., 2002).

The comparison of practice groups across test sessions is conducted using a combination of statistical techniques that are selected specifically to circumvent commonly acknowledged methodological problems. It is admitted that those methodological problems are associated with the structure of the data generated via fMRI and/or the model-based general linear model (GLM) analyses as they are typically applied. Various difficulties in fMRI data analysis have been addressed: the accurate modeling of the BOLD response, massive-sized data that are ill-balanced, and the influence of spatial and temporal correlations in the data.

In this dissertation, we propose a wavelet-based clustering method that can aggregate fMRI time series into a small number of clusters. Our clustering procedure reduces the temporal correlation in fMRI time series using the wavelet transform, significantly reduces the high dimension of the data using the wavelet-based adaptive pivotal thresholding test, and constructs data-driven clustered maps based on a principal component analysis (PCA) K-means clustering approach. Using the proposed clustering procedure, we successfully find brain activation changes across the two fMRI sessions associated with practice-related cognitive control during eye movement tasks. Furthermore, to assess the modifications in the saccade circuitry constrained to the superior cortex, we perform an analysis of variance (ANOVA) to evaluate the main effects of the practice group and the scan day (pre- and post-tests) as well as their interaction.

We attempt to empirically evaluate the performance of the proposed clustering procedure using a simulation study. We generate simulated examples by adding artificial fMRI signals to real resting-state data, and apply both K-means and PCA K-means clustering methods to the raw simulated data and the wavelet-transformed data separately to compare our clustering results. The simulation study shows that the wavelet-based clustering approach provides more accurate clustering results than the use the raw BOLD values, while PCA K-means clustering performs slightly better than K-means for the wavelet transformed data.

In order to locate the neural regions that show attenuation at the post-test, we conduct an analysis using 11 bilateral neural regions of interest (ROIs) previously identified using a similar task (Dyckman et al., 2007, and references therein). To achieve this goal, we apply two approaches: a bootstrap resampling-based test and a mixed model with feature

extraction. Both analyses reveal decreasing activation in the supplementary eye field, frontal eye field, superior parietal lobe, and cuneus for the antisaccade practice group.

The dissertation is organized as follows. In Chapter 2, we briefly review fMRI data and their properties, and existing fMRI clustering methods. We also describe the relevant statistical methods applied to the proposed clustering analysis. In Chapter 3, we illustrate the data used in our analysis, introduce the proposed clustering approach, and then present the analysis results. In Chapter 4, we empirically evaluate the performance of our wavelet-based clustering approach using simulated data. In Chapter 5, we identify the specific ROIs that show attenuations at the post-test for the antisaccade group. Finally, we conclude with future work in Chapter 6.

Chapter 2

Literature Review

In Section 2.1, we present a review of fMRI studies. First, we introduce fMRI data and their characteristics, describing how brain function can be measured. Second, we review the statistical analysis and issues in fMRI data: the most commonly used experimental designs and the problems inherent in the data that make the statistical analysis challenging. Third, we discuss two main streams in the fMRI clustering approach. In Section 2.2, we present reviews of statistical methods used in this dissertation. First, we introduce the principle of the wavelet transform. Second, we describe the adaptive pivotal thresholding test proposed by Park et al. (2011). Third, we explain the K-means clustering method. Finally, we introduce principal component analysis.

2.1 FMRI and fMRI clustering

2.1.1 Introduction to fMRI data

FMRI is a specialized MRI scan for brain imaging which is used to study brain function in the form of a time series and/or longitudinal data. Scans are acquired through an MRI machine; therefore, we give descriptions of the MRI principle in simple terms. An MRI scanner is a large static magnetic field. The strength of the magnet in an MRI scanner is measured using a unit of measure known as a Tesla; scanners used on humans range from 1.5 Tesla (or 1.5 T) and up. By comparison, the magnetic field of the Earth measures only 0.5 Gauss (1 Tesla = 10,000 Gauss). Bodies of human beings are composed of atoms, which are the smallest pieces of an element that retains all the properties of that element, and an atom is made up of three different tiny particles: protons, electrons, and neurons. These three particles determine the atom's spin; only elementary particles with nonzero spin can be studied in an MR system. The usage of hydrogen isotope having a spin of $\frac{1}{2}$ is common in studies using an MR system because hydrogen is found in abundance in the tissues that are the target of MR imaging (for more detail, refer to Lazar, 2008).

The connection from an MR system to fMRI can be understood by correlation with neuronal activity. When particular sites in the brain display increased activity in response to a specific task or stimulus, more blood is supplied to the particular brain areas related to the task or affected by the stimulus, which means that the increased neuronal activity causes an increase in blood flow in those regions. An MR system tracks the changes in blood flow. This is possible because hemoglobin has different properties in a magnetic field. In 1936, the Nobel laureate Pauling and his student Coryell discovered that hemoglobin behaves differently in a magnetic field, depending on its binding to oxygen. That is, when hemoglobin is attached to oxygen, it weakly repels a magnetic field. Thus, oxygenated hemoglobin (Hb) is said to have a diamagnetic property in a magnetic field. Conversely, when hemoglobin is not attached to oxygen, the application of an external magnetic field causes it to become slightly magnetic. Thus, deoxygenated hemoglobin (dHb) is said to have a paramagnetic property in a magnetic field.

Both hemoglobin's diamagnetic and paramagnetic properties are subtle, but measurable. Thulborn and his colleagues in the early 1980s verified that MR pulse sequences show more MR signal for highly oxygenated blood and less MR signal for highly deoxygenated blood, and the magnitude of this effect increases with the square of the strength of the static magnetic field (i.e., strong static magnetic fields are essential for MR study) (Huettel et al., 2009). In the late 1980s, Ogawa and his research team showed that the magnetic susceptibility of paramagnetic dHb in blood vessels causes signal loss, resulting in the distortion of MR images; in contrast, diamagnetic Hb in blood vessels does not impair MR signals in the surrounding magnetic field. Magnetic susceptibility is defined as the intensity of magnetization of a substance when placed within a magnetic field. Most fMRI studies are based on this blood-oxygenation-level dependent (BOLD) contrast; therefore, the BOLD contrast becomes the basis for fMRI studies to identify the brain regions related to specific functions, and exploring the human brain using fMRI has flourished since the early 1990s.

The key role in identifying brain regions is not played by Hb, but by dHb. When neuronal activity increases in a specific brain area, blood flow increases in that area to supply glucose for refueling the neurons. The delivery of glucose is carried out with that of oxygen at the same time. This oxygen supply to the area is much greater than that consumed by the neurons. Hb is detached from oxygen in the blood vessels at a regular rate under usual conditions, so the contrast between Hb and dHb is constant. When a neuron is activated, however, the excessively supplied Hbs drive out the dHbs. Therefore, the BOLD contrast is

changed. The decreased amount of dHb decreases the signal loss, resulting in a brighter MR image. In conclusion, the BOLD contrast caused by neuronal activation occurs not because the Hb increases the MR signal but because it displaces the dHb that had been suppressing the MR signal intensity (Huettel et al., 2009).

The hemodynamic response indicates the measure of the changes in the ratio of oxygenated to deoxygenated blood, and the estimation of hemodynamic response is of interest in statistical research in fMRI (Lazar, 2008). The shape of the hemodynamic response depends on the properties of stimuli and the related brain functions. Figure 2.1 shows an outline of a typical hemodynamic response function (HRF) of a voxel to a single stimulus. After the stimulus is given, there is a time lag until the change in the hemodynamic response is observed (i.e., delay). We also notice this in our data, and take the lags into account in our analysis in Section 5.3. As seen in the figure, after a few seconds of delay, the hemodynamic response rapidly increases, and then attains its maximum value, known as the peak of the hemodynamic response. As the neuronal activation stops, the hemodynamic response decreases below the starting baseline and remains at this level. This phenomenon is known as the post-stimulus undershoot (PSU). The cause of PSU is still under debate (van Zijl et al., 2012). Finally, the BOLD signal returns to the baseline, which is the condition before the neuron is activated.

The difference between MRI and fMRI is that the former depicts anatomical structure of the brain and other body areas to identify tissue types, whereas the latter depicts the changes in the brain activity over time relative to the specific tasks or stimuli given in the intended experiments. The developments in MR technology finally allow us to obtain a functional image of the brain by measuring the BOLD signal relative to a particular task or stimulus. MRI has eminent advantages compared with other medical imaging technologies. Unlike computed tomography (CT), which uses X-rays, it does not require radiation exposure. It is also different from positron emission tomography (PET), which requires the injection of radioactive materials into the participant's body. However, a CT scan is relatively lower in cost than an MRI scan, and a PET scan can provide direct information about metabolic processes, which the MRI scan can not. In actuality, these alternatives for MRI scans are often used in clinics and hospitals. The last two decades have seen the domination of fMRI in the field of neuroimaging due to its low invasiveness (non-surgery), lack of radiation exposure, and relatively wide availability.

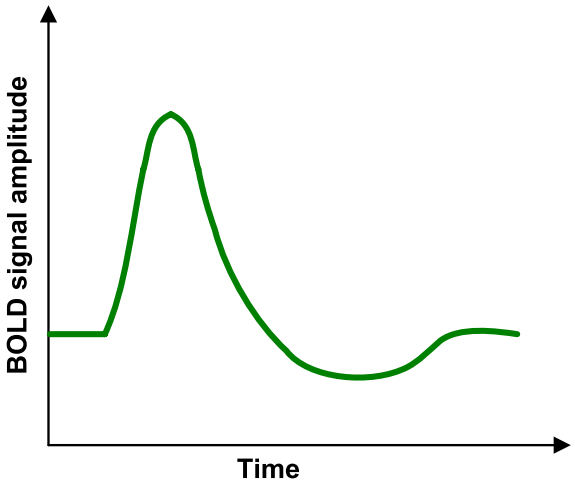


Figure 2.1: Typical shape of the hemodynamic response for a single stimulus. The y- and x-axes indicate BOLD signal and scan time point in seconds, respectively. The hemodynamic response lags behind neuronal activation, and then hits its peak. As the neuronal activation stops, the hemodynamic response decreases below initial baseline levels and remains at this level, and then the BOLD signal returns to initial baseline levels. The shape of the hemodynamic response varies with the properties of stimuli and the activated brain regions.

2.1.2 Statistical issues and prominent topics in fMRI data

FMRI data are acquired via experiments in which the human brain is the target. Participants (subjects) are required to visit the MRI scanner center once in most studies, whereas more visits may be needed in some studies (by days or by weeks); each visit is called a *session* and these last from 30 minutes to a couple of hours. In each session, participants lie on a table that slides into the doughnut-shaped static magnetic field, and then the MR system produces one or more *runs* of functional images along with anatomical images. An image acquired in two-dimensional space (single cross-section) is called a *slice*, and the number of slices varies depending on the purpose of the study. The number of acquired slices can range from one to four for a specific region, up to more than 30 for the whole brain. The slice can lie on a plane of data in the axial, coronal, or sagittal direction. A three-dimensional image depicting the real brain shape, called a *volume*, can be constructed by stacking the slices.

The unit of image in a three-dimensional space is called a *voxel*, which comprises millions of neurons; a voxel corresponds to the unit of image in a two-dimensional space, a *pixel*. A single slice typically consists of a 64×64 or 128×128 matrix of voxels, resulting in 4096 or 16384 voxels. Most of the statistical analyses of fMRI data are performed using voxels as the basic unit. Thus, fMRI data are indicated by a four-dimensional coordinate, such as the form of "$x \times y \times z$ at time point t."

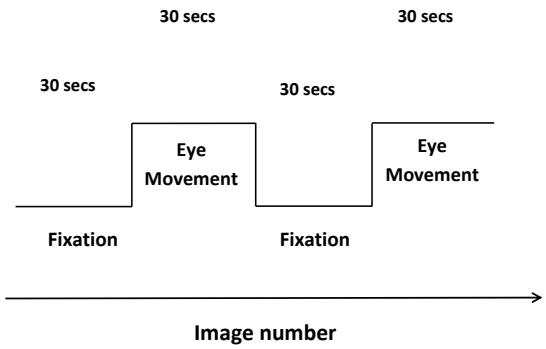


Figure 2.2: Stimulus path of a simple block design.

There are currently two primary approaches to the design of fMRI experiments: block design and event-related design. Figure 2.2 shows a simple example of a block design experiment. In block design experiments, the stimulus is given in blocks of a predetermined length (typically 10 to 30 s in duration). One stimulus type is presented per block. For example, in an antisaccade task experiment, subjects are requested to glance toward the opposite side of the target dot during the *eye movement* blocks (i.e., the task block) and to stare at the center during the *fixation* blocks (i.e., the control block), in which no visual stimulus is given. Blocks of the experimental task alternate with blocks of rest, or control. Because of the "up-down" pattern shown in the figure, block design is also referred to as "boxcar design." Of interest is the comparison between levels of activation during the task and during the control. In complex experiments, more than two types of stimuli are given. When the goal of an experiment is to find the activated brain regions or to determine whether or not specific areas of the brain are activated in response to a stimulus or a task, the block design is useful and powerful.

However, when the goal of an experiment is to estimate the shape of the hemodynamic response function (HRF) associated with a single stimulus, task or event, the event-related design is useful. In other words, this indicates when we are interested in the pattern of the transient changes in brain regions across time. In the event-related design, one or more events appear individually at random, distinguished by an interstimulus interval (ISI), which implies that the event does not occur like clockwork. ISI is referred to as the temporal interval of off-stimulus until another stimulus is given. It is not necessary that those event types are identical, but there should be sufficient time between the presentation of stimuli to avoid having the effects overlap with one another. Since event-related studies observe the transient changes of brain activity related to individual stimuli given in a short time, high temporal resolution should be assured for precise results. Temporal resolution indicates the minimal time length of signal change accurately separated out by fMRI; therefore, it is important to describe the ability to distinguish changes in a signal through time.

Indeed, block design studies, which are comparatively less sensitive to temporal resolution, were dominant in early fMRI experiments when low temporal resolution was typical. As the technology has improved, studies using event-related design have become more common, and make the wide adaptability possible in fMRI experiments. Estimation of the HRF can be improved when events are presented in relative isolation or when sophisticated strategies are applied to separate the responses to closely spaced events (Huettel et al., 2009). In summary, while block design is optimal for detecting activity in the brain, event-related design is optimal for estimating the parameters of the hemodynamic response associated with individual events (Buračas and Boynton, 2002). More recently, hybrid, or mixed designs, which combine aspects of block and event-related designs, have been used (Lazar, 2008).

The statistical challenges in the analysis of fMRI data are multiple because of the characteristics of the fMRI data and the procedures for data collection. First, fMRI data have natural temporal and spatial correlation; however, a definite solution for these two correlations has not yet emerged. It is believed that the temporal correlation is embedded since fMRI data are acquired across time while stimuli are presented in a discrete sequence over time. The response to a stimulus at the current time point can be affected by, or react to the stimulus at the previous time point. Spatial correlation arises from the fact that voxels are neighboring one another. When specific voxels are activated by a stimulus of particular interest, it is natural to suspect that the voxels surrounding those voxels also are affected. When the relationships between the measurements at neighboring locations is

decided relying on physical distances, such as Euclidean distance, spatial correlation makes the problem more complicated because those distances disregard the fact that voxels in different locations could be activated simultaneously by a single stimulus (i.e., coactivation). If there are indeed spatio-temporal correlations present in the data, ignoring them introduces bias in the assumed significance levels, resulting in tests that may be conservative or liberal in direction, depending on the experimental design (Lazar, 2008). Since the spatial distribution of the activated regions is expected to be in the vicinity of the neural population concerned with the accomplished task, Fadili et al. (2000) indicate that voxels may be selected based on anatomical criteria, use the cross-correlation function with fuzzy clustering analysis, and compare task and resting data sets if possible.

Second, fMRI data are contaminated by noise arising from various sources, which might make the signal related to brain activity relatively weak. The sources of noise may be divided into two main groups: mechanical noise and subject noise. The former is caused by the scanner or image processing. It has been studied in depth, and there are many methods to handle it. However, the latter comes from the unavoidable fact that the experimental subjects are human, so they will breathe and cannot remain perfectly still while in the scanner. When our vision is moving in a specific direction, our head easily follows in the same direction. This head motion can induce additional spatial correlation. Thus, it is routine to put foam padding around the head so that the subject cannot move suddenly. Due to this noise contamination of fMRI data, preprocessing steps must be performed on the acquired images to control the noise in the acquired data before the actual statistical analysis.

Third, fMRI data are ill-balanced, as there is a severe lack of balance between the different classes of voxels. The voxels responding to a stimulus are a small portion of the voxels in the brain. Under these circumstances, some statistical techniques such as clustering analysis generally do not perform well since the identified clusters tend to be dominated by one large class. Since the task-related voxels are the minority, clustering results can be dominated by the majority of non-task related voxels.

Fourth, fMRI studies conduct MRI scans for multiple subjects. A main interest in an fMRI study is not exploring the brain of a designated person, but exploring the brain area related to a particular ability of human beings or studying the brains of specific groups composed of subjects sharing common characteristics (e.g., schizophrenia group).

Many statistical approaches, ranging from simple t tests performed at each voxel (an analysis which ignores the spatial and the temporal correlations in the data) to complex

Bayesian hierarchical spatio-temporal models, have been applied at the level of individual subjects. Many underlying connections to statistics warrant further investigation in fMRI data analysis. See Lazar (2008) for a recent survey of statistical analysis techniques. Therefore, aggregation of the information from multiple subjects and further comparison between groups are important statistical issues.

2.1.3 FMRI CLUSTERING

Connectivity refers to networks that model or explain how different brain regions are connected to one another. In the current literature, there are two dominant methods of assessing connectivity in fMRI data: correlation-based methods and structural equation models. The former looks at the temporal correlations between spatially remote neurophysiological events to determine which areas of the brain coactivate, so this is the measure of functional connectivity. The latter looks at the influence that one neural system has over another by attaching some notion of causality to the relationships that are found among the components, so this is the measure of effective connectivity and it is used mostly in social science and econometrics. The objective of an effective connectivity analysis is to estimate parameters that represent influences among regions that may change with respect to experimental tasks, whereas functional connectivity analysis aims to detect and characterize coherent patterns of activity as a means of identifying brain systems (Lazar, 2008; Marco, 2009). Functional connectivity analysis is the mainstream in that much of the current fMRI research aims to identify brain regions activated in response to stimulus. Huettel et al. (2009, p386) define functional connectivity as "how the activity of one brain region influences activity in another brain region." Functional connectivity is defined in terms of statistical dependencies among neurophysiological measurements.

FMRI clustering analysis provides an effective way of finding functional connectivity in the brain. There are two dominant methods of clustering fMRI data: clustering based on the voxelwise hypothesis test and clustering based on the patterns of time series. The first method assumes that different voxels in the brain have different responses to stimulus, tests whether each voxel is activated or not, and constructs a *statistical parametric map* of the brain based on the test results. One classical method is to calculate the t statistics at each voxel to compare the mean levels of activation. Then usually the result from each voxel for a given subject can be plotted in a map of t values (Huettel et al., 2009). If there is more than one task condition, this method can be extended to generate an F statistic at each voxel to

compare levels of activity over all conditions, leading to an F map for each subject (Lazar, 2008). Under the assumption that the time series at each voxel is independent of the others, the general linear model can be applied to the time series at each voxel. For example, we can build a linear model for stimulus and treat data as a combination of factors (Friston et al., 1994). Most of these hypothesis-based approaches, however, assume that voxels are independent without considering their spatial correlations.

The second clustering method is to partition the brain via the time courses of voxels without assuming a specific model. This approach leads us to a way to find functional connectivity between brain regions with the assumptions that the pattern of activation has a structure and can be divided into a few types of similar activations (Goutte et al., 1999). One can extract the main patterns in time and partition the data set by grouping similarly shaped waveforms together. The most popular clustering tools are K-means (Balslev et al., 2002), fuzzy clustering (Baumgartner et al., 1998; Fadili et al., 2000), Principal Component Analysis (Baumgartner, 2000), Independent Component Analysis (McKeown and Sejnowski, 1998; McKeown et al., 2003), minimal spanning trees (Hartigan, 1975), hierarchical methods (Stanberry et al., 2003), mixture models (Filzmoser et al., 1999), and dynamic clustering (Baune et al., 1999). These procedures also provide the cluster indices, labels for each voxel denoting cluster identification.

There are several statistical issues in clustering fMRI time courses. As pointed out in Section 2.1.2, spatio-temporal correlation, fMRI noise, ill-balanced data problems, and massive size of data from multiple subjects are main concerns. In order to reduce the brain volume prior to clustering, initial screening should be performed to exclude voxels that are likely to be inactive. However, this is not a straightforward way to distinguish active from inactive voxels, as we have already seen (e.g., Gibbons et al., 2004). Some possible strategies are to use a loose F-test (Balslev et al., 2002; Goutte et al., 2001), a self-similarity test for the autocorrelation function (Fadili et al., 2000), or to remove white noise areas. White noise areas belong in an inactive area, but it indicates the areas in which voxels do not have autocorrelation in their time courses. Also, if one clusters the fMRI data while ignoring the temporal correlation, then the results are unsatisfactory and clusters do not form according to the similarity of their pattern of response to the stimulus.

Several metric distances, such as the Euclidean and Mahalanobis distance, can be considered in clustering analysis. Golay et al. (1997) consider a metric based on the correlation between stimulus and time courses and suggest clustering voxels on the basis of the

cross-correlation function in order to reduce high noise level and improve performance. The appropriate number of clusters is another nontrivial challenge. It is known that too small a number of clusters makes it difficult to distinguish inactive from active voxels because the active voxels get clustered together with inactive voxels. And too large a number of clusters causes a difficulty of interpretation because the active voxels may be partitioned into several clusters. There are several methods used for the selection of the cluster number in fMRI studies: the silhouette value (Ye, 2003), the likelihood approach after removing correlation (e.g., cross-validation (Balslev et al., 2002), AIC, BIC), the hierarchical approach, (e.g., a tree-like structure which starts with K clusters and combines similar clusters one by one (Stanberry et al., 2003)), and unsupervised fuzzy c-means analysis (Fadili et al., 2000).

2.2 Review of statistical methods

2.2.1 Wavelets

Wavelets can be explained by the mathematical analogue of a wave whose amplitude starts out at zero, increases, and then decreases back to zero. It is a mathematical function used to divide a given function or continuous-time signal into different scale components. Wavelets are particularly useful for signal and image analysis since wavelets are localized in both frequency and time. Another good aspect of wavelets is their fast convergence rates of wavelet estimators on computational complexity. We briefly introduce wavelets in this section and recommend Vidakovic (1999) for more details.

Let $\{\phi, \psi\}$ be a pair of compactly supported scaling and wavelet functions. Let $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ be their dilated and translated versions where j and k are the scale and location parameters, respectively. The collection of dilations and translations of ϕ and ψ, $\{\phi_{J_0,k}, k = 0, \ldots, 2^{J_0} - 1;\ \psi_{j,k}, j \geq J_0 \geq 0, k = 0, \ldots, 2^j - 1\}$, generates an orthonormal basis of $L_2(\mathbb{R})$ space. Therefore, any function $f(x)$ in $L_2(\mathbb{R})$ can be expressed as

$$f(x) = \sum_{k=0}^{2^{J_0}-1} a_{J_0,k}\phi_{J_0,k}(x) + \sum_{j=J_0}^{\infty}\sum_{k=0}^{2^j-1} b_{j,k}\psi_{j,k}(x),\ J_0, k \in \mathbb{Z} \tag{2.1}$$

and the wavelet coefficients are derived from the following wavelet transformation:

$$a_k = \int_{\mathbb{R}} \phi_{J_0,k}(t)f(t)dt\,, \qquad b_{j,k} = \int_{\mathbb{R}} \psi_{j,k}(t)f(t)dt \tag{2.2}$$

describing the global structure and the local detail of $f(x)$, respectively.

FMRI data are correlated in time, and this should be taken into account or diminished in the model. An important and attractive property of the wavelet transformation is that it reduces temporal correlation so that the transformed signal has the property of almost white noise (Stoev et al., 2005; Coifman et al., 1992; Atto et al., 2007). Flandrin (1992) shows that the within-scale wavelet coefficients are perfectly decorrelated. Moreover, the correlation between wavelet coefficients of long memory data decreases exponentially fast across scales and hyperbolically fast along time at a rate which is much faster than that of the autocorrelation of the data. If we define the vanishing moments L of wavelets as the number of zero moments of a particular wavelet function, that is,

$$\int_{\mathbb{R}} t^l \psi(t) dt \; = \; 0 \quad l = 0, 1, \ldots, L-1, \tag{2.3}$$

then wavelets with a higher number of vanishing moments achieve a greater decorrelation. Typically Daubechies wavelets (Daubechies, 1992) with $L = 2, 3, 4$ are a popular choice in many applications; we use them with $L = 2$ in our data analysis in Chapter 3.

Among the approaches used to shrink wavelet coefficients, the simplest nonparametric and nonlinear methods are to apply the hard or soft thresholds rules to the wavelets. We illustrate simply the hard thresholding rule:

$$\eta(b_{j,k}, \lambda) = \begin{cases} b_{j,k} & \text{if } |b_{j,k}| \geq \lambda \\ 0 & \text{if } |b_{j,k}| < \lambda, \end{cases} \tag{2.4}$$

where $\lambda \geq 0$. The hard thresholding rule sets the wavelet coefficients to zero when they are less than the predetermined threshold value, λ; therefore, significant coefficients from the whole wavelet set are extracted. Among the methods proposed for determining a threshold value λ, we choose the global threshold called VisuShrink (Donoho and Johnstone, 1994), which is defined as

$$\lambda = \sigma \sqrt{\frac{2 \log n}{n}},$$

where n is a sample size and σ/n is estimated by

$$\frac{\hat{\sigma}}{\sqrt{n}} = \frac{\text{median}_k(\{|\hat{b}_{J-1,k} - \text{median}_k \, \hat{b}_{J-1,k}| \; : \; k = 0, 1, \ldots, 2^{J-1} - 1\})}{0.6745},$$

where $J = \log_2 n$. By applying the hard thresholds rules to the wavelets, we keep the wavelet coefficients larger than λ and drop the wavelet coefficients less than λ.

2.2.2 Adaptive pivotal thresholding test

Park et al. (2011) propose the adaptive pivotal thresholding test based on wavelets to detect a trend in a function. The key idea is to convert the problem of testing a trend of a function into high dimensional normal mean inference in the wavelet domain. Using this approach we test each voxel to identify those with no trend (i.e., voxels that are non-task related) while controlling the false discovery rate to adjust for multiplicity (Benjamini and Hochberg, 1995).

The FDR procedure proposed by Benjamini and Hochberg (1995) is as follows. Assume that we perform m tests for the class of null hypotheses $\{H_0^1, H_0^2, \ldots, H_0^m\}$, where H_0^l indicates the lth null hypothesis for the lth test, and $l = 1, \ldots, m$. Also assume that r null hypotheses in the class of null hypotheses are true and the other $m - r$ are not. Here, information for what the number r is and where the true null hypotheses are located are not given. Define the class of the test statistics as $\{T_1, T_2, \ldots, T_m\}$. Using the class of test statistics, one can calculate m p-values and then sort those p-values from smallest to largest $\{p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}\}$. When r^* is the largest l satisfying that

$$r^* = \max \left\{ l : p_{(l)} \leq \frac{l}{m} q \right\}, \tag{2.5}$$

their proposal is to reject $H_0^1, H_0^2 \ldots, H_0^{r^*}$, where $1 \leq r^* \leq m$ and q is the false discovery rate. For instance, if $r^* = 1$, then only one hypothesis is rejected. If $r^* = m$, then all hypotheses are rejected. In case there are no l's satisfying (2.5), no rejection is made. Compared to traditional methods, their method provides a reasonable criterion when a very large number of tests are performed. We apply their approach to adjust for the multiplicity of our tests.

Let Y_{it} denote the BOLD signal value at the ith voxel and the tth time point. We can set up a regression model as follows. Suppose that

$$Y_{it} = f_i \left(\frac{t}{n} \right) + \varepsilon_{it}, \quad t = 1, \ldots, n,$$

where f_i is a trend function in the ith voxel, n is the total number of time points, and ε_{it} is the error term in the ith voxel time series. Our interest is to test $H_0 : f_i = 0$ for each i, which implies that the corresponding voxel is not active. The proposed approach transforms this null hypothesis to the wavelet domain. In order to simplify notation, we drop i from here on.

Let ϕ and ψ denote scaling and wavelet functions, respectively; the definitions of dilated and translated versions ($\phi_{j,k}(x)$ and $\psi_{j,k}(x)$, respectively) are given in Section 2.2.1, where j

is a scale and k is a location parameter. For a trend function $f \in L_2[0,1]$, f can be expanded with the wavelet basis functions. Let $f_n(x)$ be the projection of $f(x)$ on the span of the first $n = 2^{J_1}$ basis:

$$f_n(x) = \sum_{k=0}^{2^{J_0}-1} a_k \phi_{J_0,k}(x) + \sum_{j=J_0}^{J_1-1} \sum_{k=0}^{2^j-1} b_{j,k} \psi_{j,k}(x),$$

where J_0 is a positive integer and $J_1 > J_0$. In our analysis, we choose $J_0 = 3$ and $J_1 = 6$ by binning the data to avoid boundary effects. Let $\tilde{a}_k$ and $\tilde{b}_{j,k}$ be the naïve estimates of wavelet coefficients in (2.2) in that they are

$$\tilde{a}_k = \sum_{t=1}^{n} \tilde{Y}_i \int_{(t-1)/n}^{t/n} \phi_{J_0,k}(u)du \approx a_k + \frac{\sigma}{\sqrt{n}} Z_k,$$

$$\tilde{b}_{j,k} = \sum_{t=1}^{n} \tilde{Y}_i \int_{(t-1)/n}^{t/n} \psi_{j,k}(u)du \approx b_{j,k} + \frac{\sigma}{\sqrt{n}} Z_{j,k},$$

where Z_k and $Z_{j,k} \sim N(0,1)$. By denoting

$$\boldsymbol{\mu}^n = (\mu_1, \ldots, \mu_n) = (a_0, \ldots, a_{2^{J_0}-1}, b_{J_0,0}, \ldots, b_{J_0,2^{J_0}-1}, \ldots, b_{J_1-1,2^{J_1-1}-1}),$$

the null hypothesis $H_0 : f = 0$ becomes $H_0 : \boldsymbol{\mu}^n = \mathbf{0}$ (i.e., H_0 : all wavelet coefficients are zero vs. H_1: at least one wavelet coefficient is not zero). Here, define the loss function as

$$L_n = \sum_{k=1}^{n} (\mu_k - \hat{\mu}_k)^2,$$

and the variance estimate of the error term as

$$\hat{\sigma}_n^2 = 2 \sum_{k=(n/2)+1}^{n} \tilde{\mu}_k^2,$$

where $\tilde{\boldsymbol{\mu}}_k = (\tilde{\mu}_1, \ldots, \tilde{\mu}_n) = (\tilde{a}_0, \ldots, \tilde{b}_{J_1-1,2^{J_1-1}-1})$ and $\hat{\boldsymbol{\mu}}_k = (\hat{\mu}_1, \ldots, \hat{\mu}_n) = (\hat{a}_0, \ldots, \hat{b}_{J_1-1,2^{J_1-1}-1})$.

The hard thresholding rule introduced in Section 2.2.1 is applied on voxel BOLD signals with length of n for the wavelet transform to select significant coefficients among a total of n time points. Here, we consider the soft thresholding estimators $\hat{\mu}_k$ to increase the power of the test before applying this adaptive pivotal thresholding test, that are defined as

$$\begin{aligned} \hat{a}_k &= \tilde{a}_k, \\ \hat{b}_{j,k} &= \mathrm{sgn}(\tilde{b}_{j,k})(|\tilde{b}_{j,k}| - \lambda)_+. \end{aligned}$$

Here, $(\cdot)_+$ indicates that the negative parts of the values inside parentheses are set to zero. For λ, one can use either the universal threshold, $\lambda_j = \frac{\hat{\sigma}\sqrt{2\log n}}{\sqrt{n}}$, or the levelwise SureShrink rule (Donoho and Johnstone, 1995), which minimizes

$$S_n(\lambda_j) = \frac{\hat{\sigma}^2}{n} 2^{J_0} + \sum_{j=J_0}^{J_1-1} S_j(\lambda_j),$$

where

$$S_j(\lambda_j) = \sum_{k=0}^{2^j-1} \left[\frac{\hat{\sigma}^2}{n} - 2\frac{\hat{\sigma}^2}{n} I_{\{|\tilde{b}_{j,k}| \le \lambda_j\}} + \min(\tilde{b}_{j,k}^2, \lambda_j^2) \right], \qquad J_0 \le j \le J_1 - 1.$$

Genovese and Wasserman (2005) show that

$$\frac{\sqrt{n}(L_n - S_n(\hat{\boldsymbol{\lambda}}))}{\sqrt{2}\hat{\sigma}_n^2} \sim N(0,1),$$

which leads to reject $H_0 : \boldsymbol{\mu}^n = \mathbf{0}$ if

$$\sum_{k=1}^{n} \hat{\mu}_k^2 > \hat{\sigma}^2 \frac{z_\alpha}{\sqrt{n/2}} + S_n(\hat{\boldsymbol{\lambda}}),$$

where z_α stands for the $100(1-\alpha)\%$ percentile of the standard normal distribution.

When a large number of hypotheses are tested simultaneously, as is the case here, we must adjust for multiple testing (i.e., inflation of Type I error probabilities). Since classical procedures, such as the Bonferonni method, are critical in multiple testing problems for the justification of the multiplicity, they tend to reject hypotheses more than it should. Benjamini and Hochberg (1995) propose to control the false discovery rate for independent test statistics. The false discovery rate (FDR) is defined as the expected proportion of errors among the rejected hypotheses that is the mean of the number of false rejections divided by the number of total rejections.

2.2.3 K-MEANS CLUSTERING

The K-means method starts with specification of the number of clusters, K, in advance and assigns each voxel to the cluster that has the nearest centroid. The procedure of K-means clustering requires the following steps (Johnson and Wichern, 2002):

1. Partition the data arbitrarily into K initial clusters and calculate the cluster centroids.

2. Reassign each subject into the nearest cluster based on the Euclidean distance of each subject from the K centroids.

3. Update the cluster centroid and reassign each subject into the nearest cluster by recalculating the distance of each subject from the updated centroids.

4. Repeat steps 2 and 3 until there is no more reassignment.

The stability of the clustering can be checked by returning to the algorithm with a new initial partition, and the group differences can be explained with a table of the cluster centroids and within-cluster variances. There are various choices for the distance metric including the Euclidean, Minkowski, Mahalanobis, and correlation.

2.2.4 Principal component analysis

Principal component analysis (PCA) is widely used as a data dimension reduction method (Johnson and Wichern, 2002). The basic idea is that there may be linear relationships between variables in a data set, and then the computation will be effective only if uncorrelated variables are used. The purpose of PCA is to transform a number of linearly correlated variables into a smaller number of uncorrelated variables called principal components (PC). Much of the variability in a data set can be accounted for by a small number M of the PCs. The choice of M is made with a visual heuristic called the scree plot that shows the fraction of total variance in the data as explained or represented by each principal component in decreasing order.

Chapter 3

FMRI clustering analysis[1]

In this Chapter, we describe the data we analyze, propose our clustering procedure, and present the analysis results. Dyckman et al. (2007) evaluate the effect of context on behavior and brain activity during saccade tasks using our data set with one fMRI session. They apply a general linear model using data-driven regressors obtained by ICA (independent component analysis). The results from their study indicate that the superior parietal lobe (SPL), supplemental eye fields (SEF), and frontal eye fields (FEF) may be more important for supporting the complex behavioral response in prosaccades, whereas other brain regions, such as PFC, may be more involved in response selection and/or context updating in antisaccades. In the present study we evaluate the practice effects of the participants using two (pre- and post- tests) fMRI sessions. We take temporal correlation into account (the wavelet transform and thresholding), significantly reduce the high dimension of the data (the adaptive pivotal thresholding test), and produce clustered maps based on a model-free approach (PCA K-means clustering).

3.1 Introduction to saccadic data

Among eye movements, saccades are the rapid eye movements that redirect one's gaze to a newly appearing target. In this dissertation, we use eye movement data to evaluate changes in brain activity before and after daily exposure to a set of eye movement tasks. Two types of saccades and one eye movement are used: prosaccades, antisaccades, and fixation. "Prosaccades" are more reflexive, as a person simply redirects one's gaze to a newly appearing visual stimulus. "Antisaccades" are more complex and volitional, as a person redirects one's gaze to the mirror image location (same amplitude, opposite side) of the newly appearing

[1]Part of the results presented in this chapter have been published: Lee, J., Park, C., Dyckman, K. A., Lazar, N. A., Austin, B. P., Li, Q., and McDowell, J. E. (2013), Practice-related Changes in Neural Circuitry Supporting Eye Movements Investigated via Wavelet-based Clustering Analysis, *Human Brain Mapping*, 34(9):2276-2291.

visual stimulus. As such, a correct antisaccade response requires the inhibition of a glance toward the stimulus and the generation of a voluntary saccade to an unmarked location in the opposite visual field. "Fixation" is maintaining gaze on a target. Saccadic performance is supported by a network of subcortical and cortical regions, as identified via neuroimaging and other techniques (Camchong et al., 2008; Dyckman et al., 2007; McDowell et al., 2008, and references therein). While the basic circuitry is the same for prosaccades and antisaccades, the increased complexity of antisaccades is supported by increased activation of existing circuitry and/or the recruitment of additional neural regions into the circuitry. As such, the saccade circuitry provides a specific, well-studied system for understanding changes in brain activation associated with cognitive control and practice.

The present study uses fMRI to evaluate the changes across time in neural system activation after participants practice either specific or non-specific eye movement tasks. Participants are engaged in antisaccade performance during fMR image acquisition at two time points: 1) at pre-test before any exposure to the task and 2) at post-test after one week of daily practice of eye movement tasks. Each participant practices a single type of eye movement only: either a) antisaccades, b) prosaccades, or c) fixation, where fixation indicates that one maintains ones gaze on a target. It is hypothesized that the subjects who have task-consistent practice (i.e., the antisaccade practice group) would show changes in neural circuitry as measured by the blood oxygenation level-dependent, BOLD, signal that are not observed in the two task-inconsistent practice groups, the prosaccade and fixation groups, who practice eye movement control tasks generally, but not antisaccades specifically. It is further hypothesized that (i) for saccade tasks, saccade-related circuitry would show decreased BOLD signal over time as a result of the circuitry becoming more efficient and (ii) in the antisaccade practice group only, prefrontal cortex, frontal eye fields, and striatum would show decreased BOLD signal due to decreased need for a higher level of cognitive control (Munoz and Everling, 2004).

3.2 Data description

We perform the statistical analysis on the data collected by Dyckman (2007). The information about our fMRI data presented in this subsection comes from her dissertation. For more details, refer to her dissertation.

3.2.1 Participants

Data were collected from thirty-seven undergraduate women (mean age=19.5 years, SD=1.8), all right-handed. They were recruited from the Undergraduate Research Pool at the University of Georgia (UGA) or via flyers posted on campus. All participants were free (by self-report) of psychiatric illness, history of head injury, and drug and alcohol abuse. Participants provided written informed consent prior to the study, which was approved by the UGA Institutional Review Board (UGA IRB # 10387).

3.2.2 Procedure

FMRI data during antisaccade task performance were acquired at two time points: 1) pre-test (before any task exposure) and 2) post-test. Between pre- and post-tests fMRI sessions, participants practiced one type of randomly assigned eye movement daily for one week: either 1) antisaccades (glances away from a cue, n=12), 2) prosaccades (glances towards a cue, n=14), or 3) fixation (maintains gaze at a central target, n=11). All participants completed both test sessions.

3.2.3 FMRI Test Sessions

FMRI data were collected using a 1.5T GE Signal Horizon LX Scanner (GE Medical Systems, Waukesha, Wisconsin). Task instructions were given prior to entering the scan room. Once in the room, a participant was positioned on a gurney with her head stabilized with foam padding and a restraint strap across her forehead. A mirror placed over the participant's head allowed her to view the stimuli on a rear-projection screen placed near her feet. Eye movements were recorded using an fMRI compatible iView X System (SensoMotoric Instrument, Inc.; 60 Hz sampling rate) eye tracking system.

Two localizer images were taken at the beginning of each session to ensure optimal brain coverage for each participant. A high-resolution image was then obtained (SPGR-protocol: sagittal, 2 NEX .9375 $\times$ .9375 $\times$ 1.5 mm, 124 slices, TE 2.8 ms, TR=10.8 ms, flip angle = 20 degrees, scan time = 5 min 41 sec). After this structural image was acquired, functional imaging was conducted (spiral scan with 2 interleaves, 24 continuous axial slices, 3.75 $\times$ 3.75 $\times$ 4 mm, TE = 40 ms, TR = 1912 ms, flip angle = 77 deg, scan time = 5 min 1 sec).

The functional run consisted of 13 blocks (22.4 sec each) alternating between fixation and antisaccade trials (see Figure 3.1). Stimuli consisted of 1 deg circles filled with different colors. A fixation block was indicated by a purple stimulus at the central location; participants were instructed to fixate on the target for the duration of the block. An antisaccade block consisted of 7 individual trials, each of which started with a blue stimulus at the central location (2000 ms). The stimulus was extinguished and 200 ms later a cue presented 5 deg to the left or the right (half in each visual field) on the horizontal axis (1000 ms). Participants were instructed to look at the blue stimulus only when it was in the middle of the screen, and when it was presented in the periphery, as shown in Figure 3.1, to look at the mirror image of that stimulus (opposite direction, same amplitude) as quickly and accurately as possible.

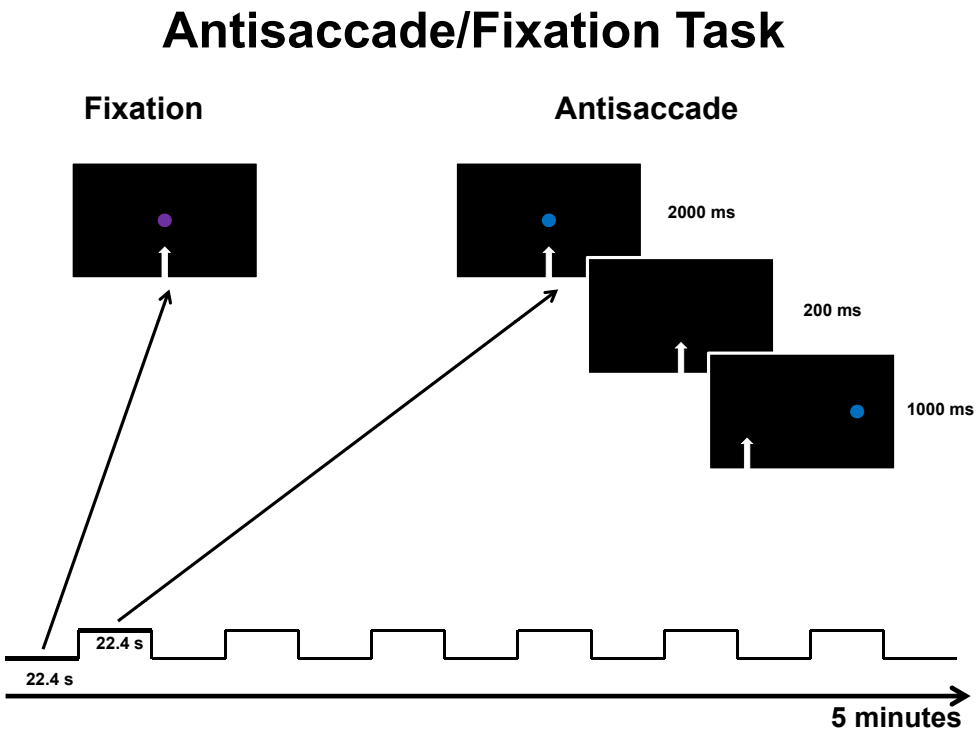


Figure 3.1: Saccade tasks in fMRI block design model. The functional run consists of 13 blocks (22.4 sec each) alternating between fixation and antisaccade trials. The white arrow indicates where the participant should be looking at each point in time. Stimuli consist of 1 deg circles filled with purple for fixation and blue for antisaccade tasks. The participants are instructed to fixate on the target for the duration of the fixation block. An antisaccade block consists of 7 individual trials, each of which starts with a blue stimulus at the central location (2000 ms). The stimulus is extinguished and is presented 5 deg to the left or the right (half in each visual field) 200 ms later on the horizontal axis (1000 ms). Participants are instructed to look at the blue stimulus only when it is in the middle of the screen, and when it is presented in the periphery, to look at the mirror image of that stimulus.

3.2.4 Behavioral Practice Sessions

Participants practiced their assigned tasks (antisaccade, prosaccade, or fixation) in the laboratory each of four days (the weekend excluded) beginning the day after the pre-test scanning session and ending the day before the post-test session. Refer to Figure 3.2. Practice sessions were conducted using a hand held device with an LCD screen known as "Fix-Train," which was designed for the purpose of providing practice with eye movement tasks (see Fischer et al., 2000).

The use of the Fix-Train device allowed for several participants to complete their training simultaneously. Each participant was assigned her own Fix-Train and her data were recorded on the device throughout the study. For each task, the goal of the participant was to determine the orientation of a "T" symbol and to press the corresponding arrow key. The bottom leg of the "T" symbol could be facing up, down, left, or right. This manual task served as a proxy for measuring eye movements because the location and duration of the "T" presentation were designed so that the participant must execute the correct eye movement to see the stimulus and discern its orientation.

For fixation training, the symbol remained in the center of the screen and changed direction two to five times before disappearing. Subjects pressed the arrow key corresponding to the final orientation of the symbol before it disappeared from the screen. For prosaccade training, the task was identical except that the symbol jumped to one side after changing orientation in the middle, and participants pressed the arrow key corresponding to the orientation of the symbol once it jumped to the periphery. For antisaccade training, a star appeared in the center of the screen to provide a fixation target. Then, the star jumped to one side of the screen and the "T" symbol appeared 70 ms later on the opposite side of the screen. Participants were instructed to look at the opposite side of the star and press the arrow key that corresponded to the orientation of the symbol. If participants made a saccade toward the star before looking away from it, they would miss the symbol as it was only visible for 100 ms.

Prosaccade and antisaccade training consisted of an equal number of trials to the left and right. Each participant completed 200 trials a day of her specific task (fixation, prosaccade, or antisaccade) and her data were recorded on the device. The number of correct trials and the mean reaction time for each practice session were recorded by the Fix-Train and downloaded to a computer for further analysis.

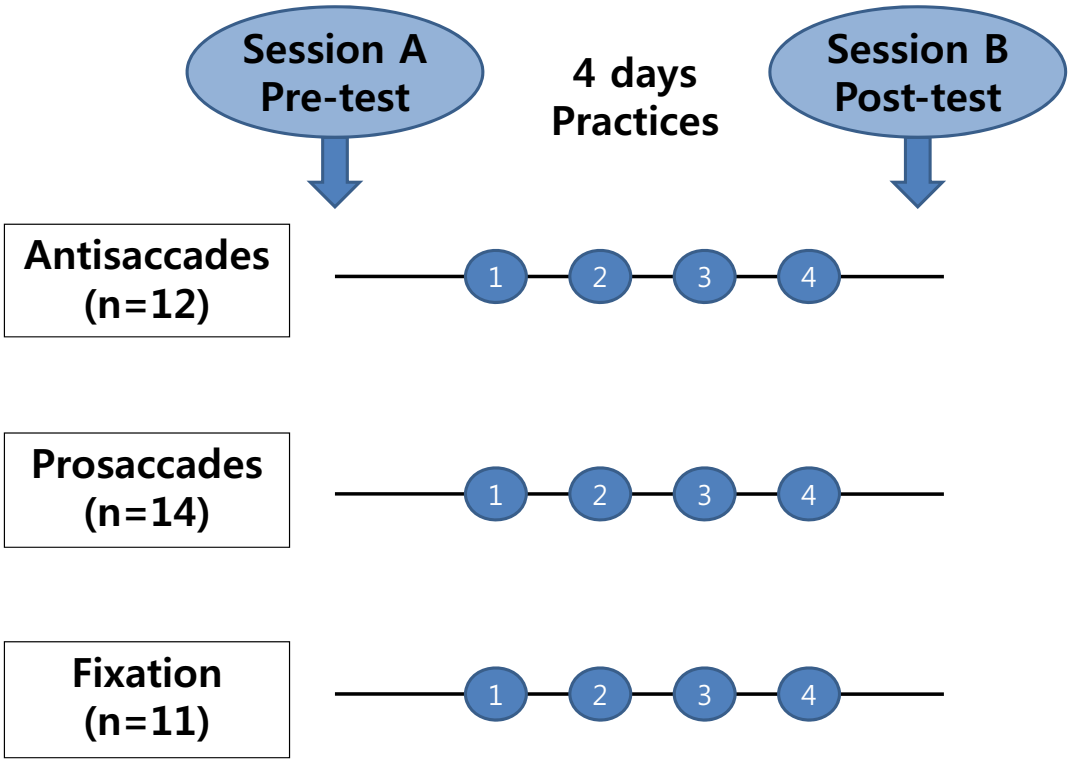


Figure 3.2: Study schedule. Data were collected from 37 undergraduate women. Participants were randomly divided into three task groups: antisaccade ($n = 12$), prosaccade ($n = 14$), and fixation ($n = 11$). On the first visit for the pre-test session, all participants had the same antisaccade task test and then practiced one type of assigned eye movement daily for one week (except for the test session day and weekend). On the second visit for the post-test session, all participants had the same antisaccade task test again. All participants completed both pre- and post-test sessions.

3.3 Preprocessing

FMRI test session data consist of eye movements recorded while the participants are engaged in antisaccade tasks. All 37 participants have MRI scans of 38 slices of size 64×64 voxels from the bottom to the top of the head during 75 scan time points. Each individual subject's data are preprocessed using standard methods in AFNI (Cox, 1996) including motion correction, slice timing correction, head motion correction, and spatial and temporal filtering. Regions of signal change can be described in three dimensions using standard normalized brain space (e.g., Talairach Atlas). The data are individually transformed into Talairach coordinates for standardization (Talairach and Tournoux, 1998), yielding slices of dimension 40×48.

Data preprocessed in AFNI are then exported to MAT-files (MATLAB formatted binary file) to be analyzed in MATLAB software (The Mathworks Inc., Natick, MA). Trials with blinks and trials with no saccades are eliminated. Saccades are scored for direction and reaction time.

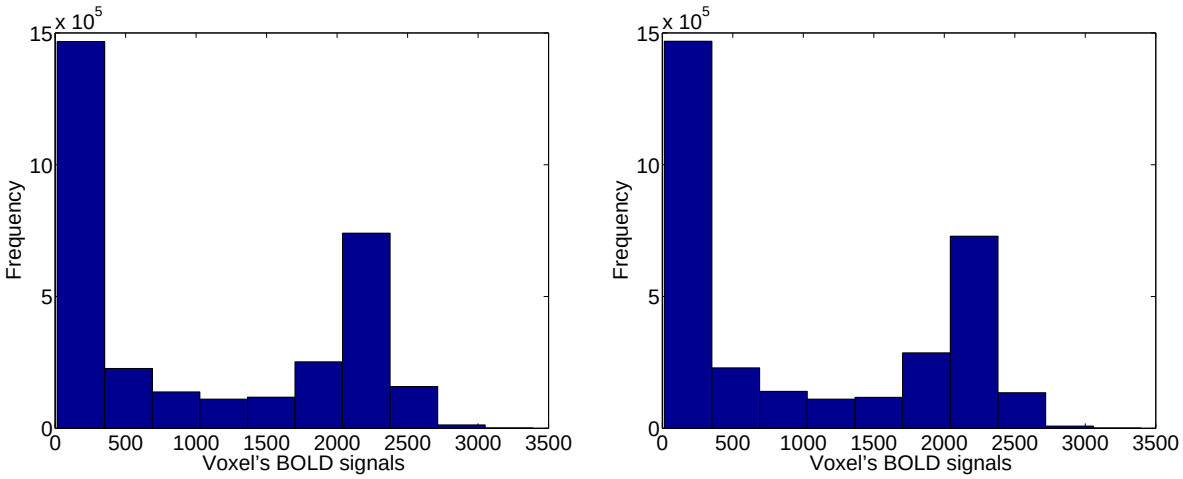
Activation occurs predominantly in the 11^{th} through 34^{th} slices when we inspect the data visually; therefore, we select this subset of the full volume for the remaining analyses. We exclude the first five scans from the analysis to allow for stabilization of the magnetization. Thus, the final data set contains 37 participants, each with data from 24 slices (40×48 voxels) and 70 scans.

3.3.1 Masking

The identification of brain tissue in a whole brain MR image is an essential first step in many neuroimaging studies (Rex et al., 2004). The first step after data are exported to MAT-files is to identify the location of the brain because it is not necessary to include the voxels falling outside the brain in clustering analysis. There are several techniques used for finding the location of the brain in sets of slices, for example, threshold masking (Friman et al., 2004), skull stripping (Mahapatra, 2012), MRI Defacer (Bischoff-Grethe et al., 2007), and Quickshear Defacing (Li, 2011).

In this analysis we adapt the threshold masking technique to remove the voxels falling outside the brain. The selection of the threshold value is critical in masking out the brain in the sense that if a threshold is set too low, it lowers efficiency by including too many voxels outside of the brain, whereas a high threshold might eliminate some voxels inside of the brain. We choose this threshold value using the histogram of voxel BOLD signals.

Figure 3.3 shows the histograms of voxel BOLD signals for the pre-test (Figure 3.3(a)) and the post-test (Figure 3.3(b)). The histograms are obtained by collapsing all voxel BOLD signals over all time points for all subjects and then divided by the number of subjects. The bottoms of the valleys in the histograms, which range from 1000 to 1500 for both the pre- and post-tests, are good candidates for the threshold values (Hawkes and Kazan, 1993). We apply two threshold values, 500 and 1000, to the clustering analysis, but there are no significant differences between the two results. Thus, we determine to mask the brain using the threshold value of 1000.



(a) The pre-test (b) The post-test

Figure 3.3: Histograms of voxel's BOLD signals. The histogram is obtained by collapsing all voxel BOLD signals over all time points for all subjects and then divided by the number of subjects. This histogram is used to choose the threshold value; the bottoms of the valleys in the histograms are good candidates. Here the bottoms of the valleys fall between 1000 and 1500 for both the pre- and post-tests, and thus, choosing a threshold between 1000 and 1500 would be reasonable.

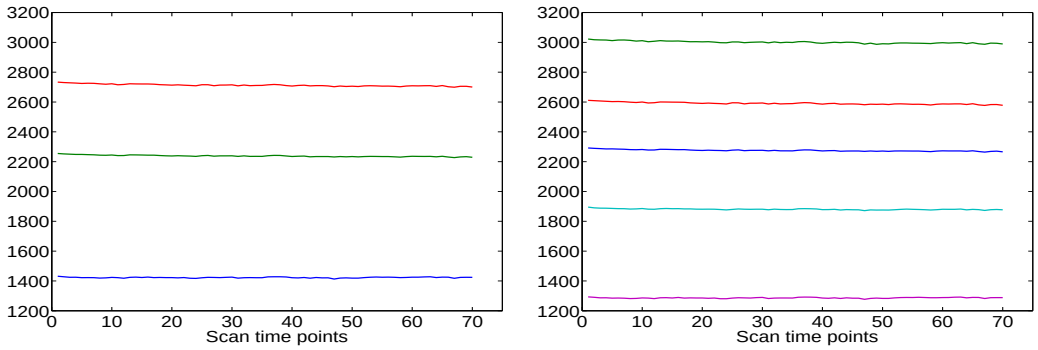
3.3.2 Detrending

The next step is to remove any deterministic trends that might affect the clustering analysis. For example, when grouping the raw data into three clusters and five clusters, we obtain the results in Figures 3.4(a) and 3.4(b), respectively. It is clear that some voxels have high BOLD signal responses and some voxels have low BOLD signal responses across all scan times. Since voxels should be grouped based on the similarities of their temporal patterns

rather than simply the levels of their BOLD signals, we remove the differences in level by subtracting the grand mean at each voxel, which is called the centered data. Then we group the centered data into three clusters and five clusters with results given in Figures 3.4(c) and 3.4(d), respectively. It is clear that linear trends are inherent in the voxel's time series.

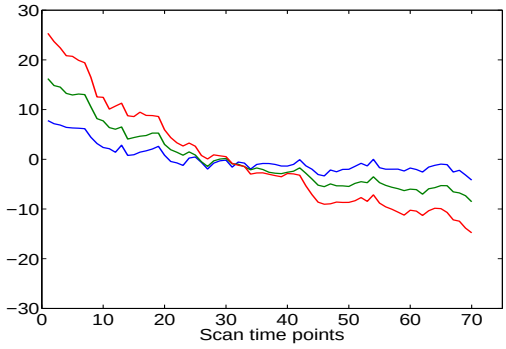
Lazar (2008) indicates that these linear trends are caused by system noise, which is inherent in MRI data because it comes from the MR system itself. It is also mentioned that an important form of system noise is trend in the voxel's BOLD signal, and the signal intensity gradually and systematically changes, as shown in Figures 3.4(c) and 3.4(d). Figures 3.4(e) and 3.4(f) show the clustering results of three and five clusters, respectively, after removing the linear trends from data. But clear quadratic trends still remain. Thus it motivates us to remove these quadratic trends by subtracting fitted second degree polynomials from the linearly detrended data and to use the linearly and quadratically detrended data instead of the raw data for further analysis.

Our findings are consistent with those of Skudlarski et al. (1999) who made an advanced improvement in efficiency of analysis by removing linear and quadratic trends on the voxel time series in a preprocessing step. They pointed out that the linear trend removal increases efficiency and that the removal of the quadratic polynomial makes a further improvement, whereas the cubic polynomial does not help beyond these.

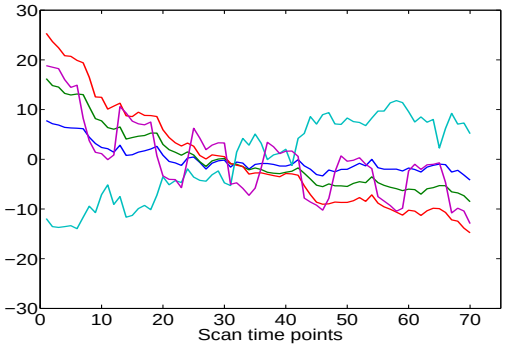


(a) Raw data into 3 clusters

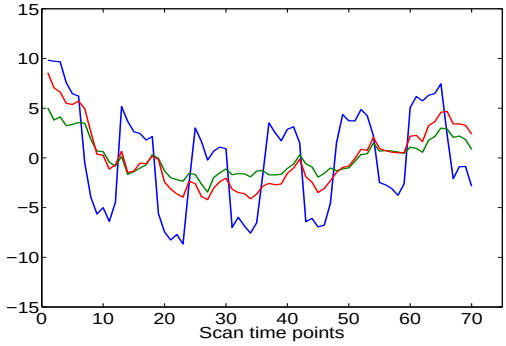
(b) Raw data into 5 clusters



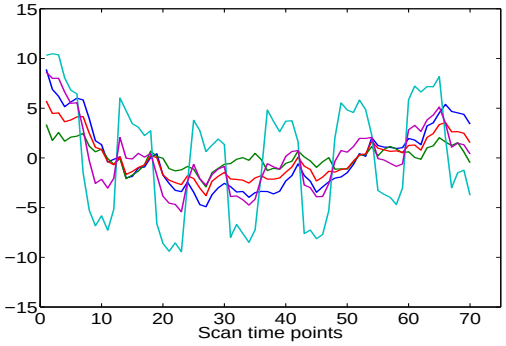
(c) Centered data into 3 clusters



(d) Centered data into 5 clusters



(e) Linearly detrended into 3 clusters



(f) Linearly detrended data into 5 clusters

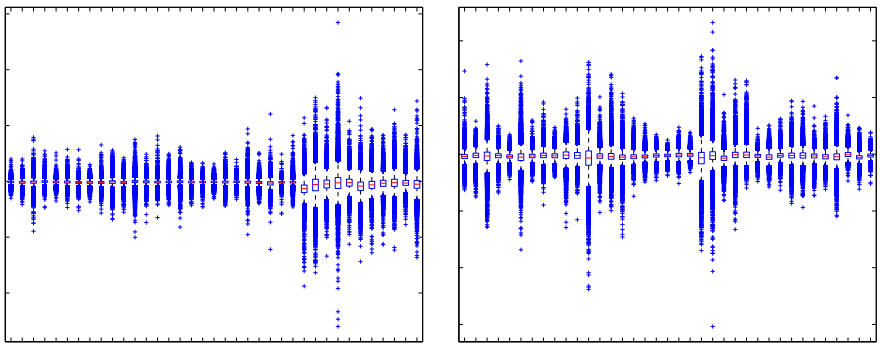
Figure 3.4: Clustering results showing temporal trends of the voxel's BOLD signals. Figures 3.4(a) and 3.4(b) are the clustering results when we group raw data into three clusters and five clusters, respectively. The x-axes indicate scan time points, and the y-axes are the voxel's BOLD signals. These figures show that the clusters are composed of the voxels which have similar BOLD signal amplitudes, but the similarities of their temporal patterns are not considered at all. We remove the overall differences in the BOLD signal levels by subtracting the grand mean, and the clustering results of three clusters and five clusters are shown in Figures 3.4(c) and 3.4(d), respectively. It is clear that linear trends are inherent, and these linear trends should be eliminated prior to grouping. Figures 3.4(e) and 3.4(f) are the clustering results after we remove linear trends from the data. There still remain clear quadratic trends, leading us to remove quadratic trends prior to further analyses.

3.3.3 Data aggregation

As we discuss at the beginning of Section 3.3, the final data have 37 participants $\times$ 24 slices $\times$ 70 scans $\times$ (40×48) voxels, which yields a total of 119, 347, and 200 voxels. This is beyond the capabilities of current clustering software (Mumford and Poldrack, 2007; Wanga et al., 2007). To make clustering more manageable, we must first reduce the size of the data. This has been a common concern in the literature. For example, Balslev et al. (2002) averaged the time series for each voxel over 18 subjects prior to clustering.

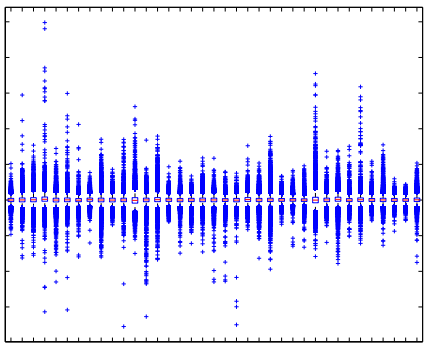
No extraordinary dissimilarity between subjects should be assumed to aggregatate properly the time series for each voxel over all subjects. In order to study the distributions of the 37 subjects on the purpose of validation of homogeneity of subjects, we evaluate two center measures, mean and median, and three variation measures, standard deviation, interquartile range and range, at each voxel for each subject. The distributions of the center measures are displayed in Figure 3.5. Here Figure 3.5(a) and Figure 3.5(b) present the distributions of voxel means and Figure 3.5(c) and Figure 3.5(d) present the distributions of voxel medians at pre- and post-tests, respectively. For example, we measure the mean of each voxel and draw a boxplot of the measured means for each subject. The Figure shows that the distributions do not look dramatically different across subjects, although some subject-to-subject variation does exist; in particular, though, there are no outlying subjects.

The distributions of variation measures, standard deviation, interquartile range and range, are displayed in Figure 3.6. The distributions of these measures across all subjects are similar, although there are some isolated cases suspected as outliers. Therefore, we conjecture that the aggregation of 37 subjects using a representative statistic would not cause a severe problem in our analysis. In conclusion, we combine the subjects for each group at each voxel and each time point separately using a representative value to circumvent the computational burden. Since we find some outliers within each subject, we choose the median instead of the mean as a representative value. We also try to group the voxels using the 10% and 20% trimmed means, and the clustering results are similar to the other summary measures. The complete results can be found in Section 3.5.3. An advantage of using the median is that it reduces the effect of heavy outliers found in some voxels. Another advantage of this aggregating approach is that it increases the signal-to-noise ratio; hence it produces sharper clustering brain maps. After aggregation we have a single image per group, for each of the 70 scan time points; the images are 24 slices $\times 40 \times 48$ voxels in size.

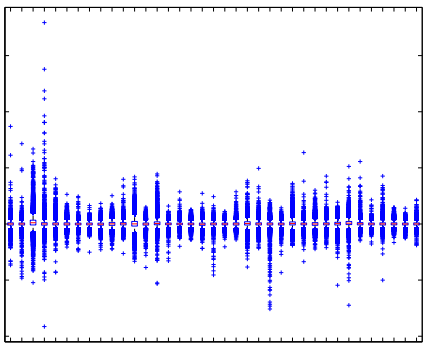


(a) mean (pre-test)

(b) mean (post-test)

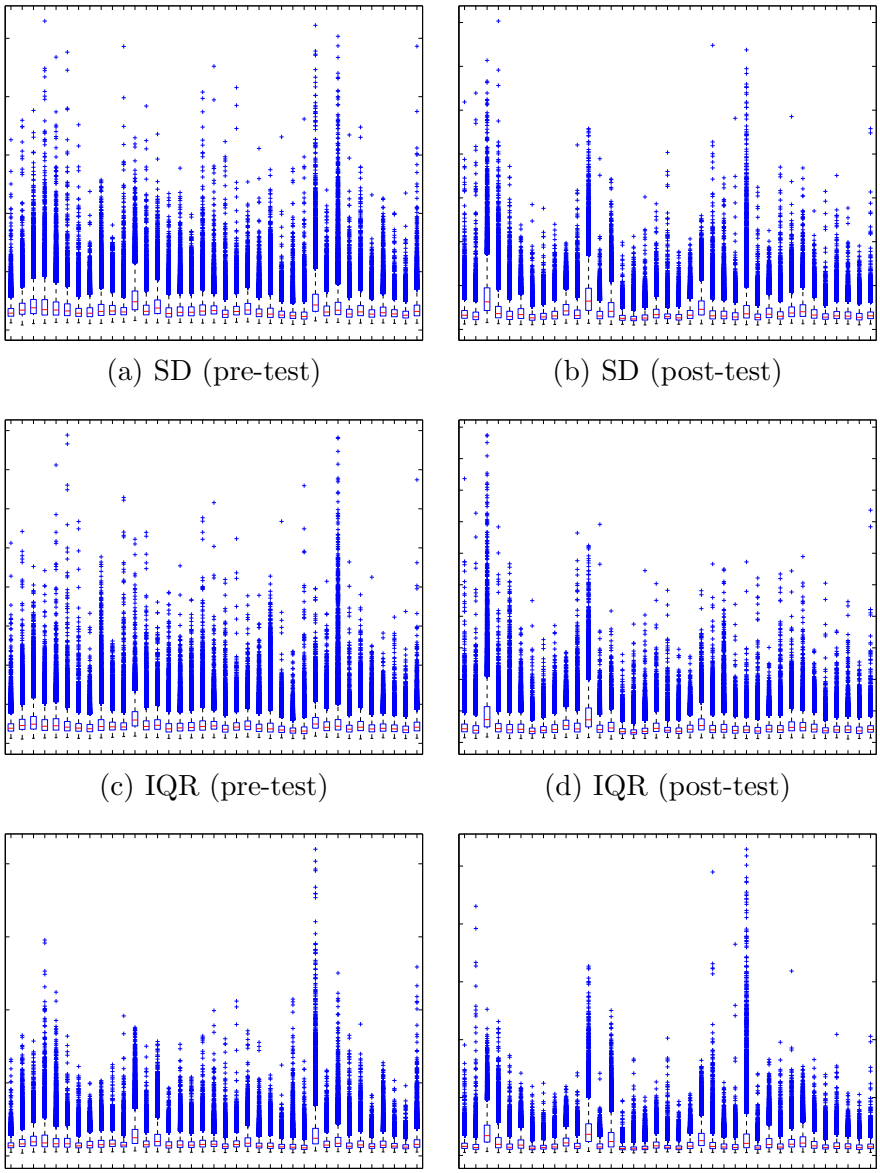


(c) median (pre-test)



(d) median (post-test)

Figure 3.5: Distributions of voxel center measures at the pre- (left) and post-test (right), respectively. Figures 3.5(a) and 3.5(b) present the distributions of voxel's mean, and Figures 3.5(c) and 3.5(d) present the distributions of median. The x-axes indicate subject's IDs.



(a) SD (pre-test) (b) SD (post-test)

(c) IQR (pre-test) (d) IQR (post-test)

(e) Range (pre-test) (f) Range (post-test)

Figure 3.6: Distributions of voxel variation measures at the pre- (left) and post-test (right), respectively. Figures 3.6(a) and 3.6(b) present the distributions of standard deviation, Figures 3.6(c) and 3.6(d) present the distributions of interquartile range, and Figures 3.6(e) and 3.6(f) present the distributions of voxel's range. The x-axes indicate subject's IDs.

3.3.4 Decorrelation

FMRI data are highly correlated temporally since stimuli are given continuously or periodically over time. Hence, the response to the stimulus at the current scan time point will be influenced by the stimulus at the previous scan time point and the response to it (Lazar, 2008). A typical approach to address this issue in clustering analysis is to assume a temporal model for characterizing the dependence structure in fMRI time series (e.g., Worsley, 2003). This temporal correlation has been evaluated using different types of time series models ranging from autoregressive of low order (see, for example, Woolrich et al., 2001; Luo and Nichols, 2003; Gautama and Hulle, 2005) to long-range dependence (see, for example, Bullmore et al., 2003; Park et al., 2010).

Friman et al. (2004) indicated that fMRI data have abundant temporal trends, and if not accounted for, these trends cause pronounced autocorrelation in the residual noise process. The dependence structure in fMRI time series makes clustering analysis more challenging, and it should be accounted for because it is nontrivial to distinguish deterministic patterns from dependence artifacts. Instead of modeling the temporal correlation explicitly, we use the wavelet transformation to take advantage of its decorrelation and denoising properties. One important advantage of using the wavelet transformation in fMRI analysis is that it removes the temporal correlation so that the transformed data have less serial correlation (i.e., have the property of nearly white noise) in the wavelet domain for large vanishing moments (L) (for example, see Stoev et al., 2005). Refer to Equation (2.3). If we then cluster the voxels in the wavelet domain rather than in the time domain, the temporal correlation inherent in fMRI data is reduced.

Veitch and Abry (1999) pointed out, however, that the theoretical improvement with large vanishing moments is balanced by an increase in the number of wavelet coefficients influenced by the boundary effect, due to the finite length of the time series. They used three vanishing moments for Internet traffic data to reduce strong serial correlation and also remove possible linear and quadratic trends in the data. In our analysis, since we remove those trends in advance, we use a Daubechies wavelet function (Daubechies, 1992) with two vanishing moments, which would increase the number of available wavelet coefficients, and apply the wavelet transformation to each median-valued voxel. We also compare clustering maps to check significant changes from L=2 to 4 and find them similar to each other.

In addition, we apply a wavelet hard thresholding technique to obtain a sparse representation of a signal by setting the wavelet coefficients to zero, less than the threshold

suggested by Donoho and Johnstone (1994). This denoising process amplifies important signals and removes noise in the observed data. We use the Wavelab package, available at `http://stat.stanford.edu/~wavelab/`, for the wavelet analysis.

3.3.5 Adaptive pivotal thresholding test

The size of the data is reduced considerably by applying the masking and data aggregation methods described in Sections 3.3.1 and 3.3.3. However, the areas related to the BOLD signal of interest involve relatively small portions of the brain. Clustering techniques can not play a perfect role in detecting the activated voxels embedded in a mass of inactive voxels. As a solution for this ill-balanced problem, and to improve the efficiency of computation, one can exclude voxels whose signals are not significantly different from zero. Goutte et al. (1999) applied a loose statistical test such as the F-test as a reduction tool. Fadili et al. (2000) removed white noise voxels and further reduced the number of voxels by considering only gray matter.

Park et al. (2011) proposed the adaptive pivotal thresholding test based on wavelets to detect any kind of trend in a time series. We apply this adaptive pivotal thresholding test based on wavelet coefficients (a_k and $b_{j,k}$ in Equation (2.2)) to filter out inactivated voxels. This test is more rigorous than a simple F-test as it accounts for temporal correlation. The key idea is to convert the problem of testing a trend of a function into high dimensional normal mean inference in the wavelet domain. We conduct this test at each voxel to identify voxels with no trend (i.e., voxels that are non-task related). Since each voxel is tested using this approach, adjustment for multiple testing is necessary and we apply the false discovery rate (FDR) with q=0.15. This is a relatively generous threshold to help ensure that truly interesting voxels are not removed for the rest of the analysis.

As indicated in Section 3.3 we start with 46,080 ($24 \times 40 \times 48$) voxels for each session and each group. After the masking procedure, 14,743 $\sim$ 15,319 voxels are categorized as brain. Application of the adaptive pivotal thresholding test with FDR for multiplicity retains 1,852 $\sim$ 3,329 voxels that are categorized as antisaccade task-related, categorizing the other voxels as possibly non-task related. This is a remarkable reduction because the number of voxels used in the clustering can be decreased up to 96% of the original number.

3.4 Clustering algorithms

We cluster the wavelet coefficients of the remaining voxels after the adaptive pivotal thresholding test in Section 3.3.5, using both K-means and PCA K-means clustering methods. Among the various choices for a distance metric for K-means clustering, we use the Euclidean metric for computing the distance between points and cluster centroids, a typical choice, because the data are suitably standardized by detrending and are approximately independent because of the wavelet transformation. Refer to Section 2.2.3 for details.

PCA (Jolliffe, 2002) is widely used as a data dimension reduction method. The basic idea is that if linear relationships exist between variables, analysis will be effective if only a smaller number of uncorrelated variables, called principal components (PC), are used. The procedure of PCA K-means clustering is as follows:

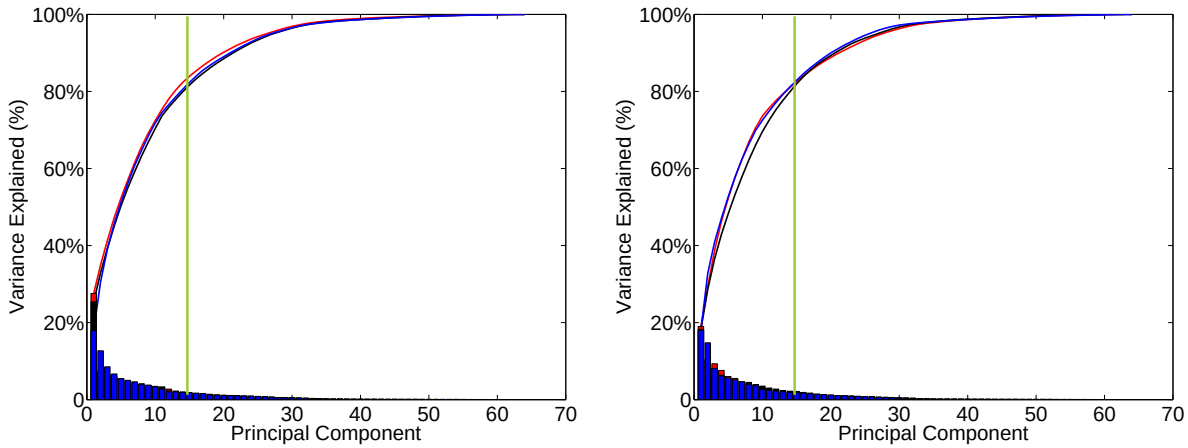
1. For wavelet transformed voxels $X = (x_1, \ldots, x_n)^T$, obtain the n PCs, labeled PC_1 to PC_n, via PCA. These PCs are ordered by decreasing order of contribution to total variance.

2. Draw a scree plot in which the horizontal axis contains the PCs sorted by decreasing fraction of total variance explained, and find the point of separation in fraction of total variance (sometimes called the elbow) where the most important components cease and the less important components begin. Identify the M most important PC scores using the elbow point. If the elbow happens at the $(M+1)$-st PC, then take the first M PC scores.

3. Apply K-means clustering to the M selected PC scores of the wavelet-transformed voxels.

The scree plots for the pre- and post-tests for the three different groups are shown in Figures 3.7(a) and 3.7(b), respectively. In our analysis 15 PCs are selected based on the scree plots, a noticeable decrease from the 70 original time points. These 15 components explain at least 80% of the total variance for each test session and each practice group. The results for the three groups are similar to one another so that they are indistinguishable at each time point (pre- and post-tests).

In both K-means and PCA K-means methods, the choice of K is an important issue. It can be determined using a criterion such as the silhouette value (Ye, 2003), the likelihood approach with cross-validation (Balslev et al., 2002), the hierarchical approach (Stanberry



(a) The pre-test

(b) The post-test

Figure 3.7: Scree plots of PCA for the pre- (left) and post-tests (right) and each practice group. Red, blue, and black curves represent the antisaccade, prosaccade, or fixation practice groups, respectively. The heights of the vertical bars on the bottom represent the contributions of each PC for that group. Note that there are little differences across the groups so that lines are indistinguishable. Based on these plots, we select 15 PCs (green vertical bar) because they explain at least 80% of the total variance for each test session and each practice group.

et al., 2003), and unsupervised fuzzy c-means analysis (Fadili et al., 2000). In our analysis, we try different numbers of clusters, K = 3, 4, 5, 7, and choose K =3 because the larger numbers of clusters only break the noise cluster into multiple clusters and do not provide additional information.

Another issue in K-means clustering is that initial partitions can affect the final clusters. We therefore run the clustering algorithm multiple times for the same number of clusters and obtain similar results to those presented in Section 3.5.

We summarize our clustering analysis strategy in Figure 3.8. First, we mask out the data standardized in Talairach space to delineate the contours of the brains using a threshold value of 1000 for each subject at each slice. Second, we remove the linear and quadratic trends in the masked data. Third, we aggregate the 37 subjects using the median values for each group at each voxel and each time point separately. Fourth, we apply the wavelet transformation to each median-valued voxel using Daubechies wavelet functions with two vanishing moments and apply a wavelet hard thresholding technique. Fifth, we apply the adaptive pivotal thresholding test based on wavelet coefficients to filter out inactivated voxels. Finally, we cluster the thresholded wavelet coefficients of the remaining voxels after the adaptive pivotal thresholding test using K-means and PCA K-means clustering methods.

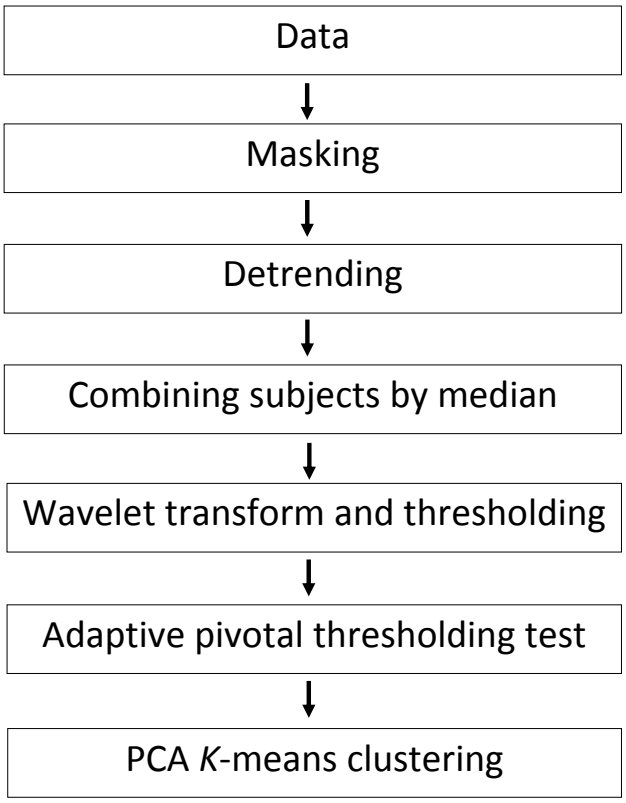


Figure 3.8: The proposed clustering analysis strategy.

3.5 Clustering results

For each test session (pre- and post-tests) the brain is partitioned into two, three, five or seven clusters, which are K=2, 3, 5, and 7, using K-means and PCA K-means clustering methods. We present the clustered brain maps in Section 3.5.1 and their average time series plots in Section 3.5.2. The results with other aggregation methods are provided in Section 3.5.3. Finally, we focus on the PCA K-means clustering results with $K = 3$ for two test sessions and for three practice groups in Section 3.5.4 and show their statistical differences using ANOVA in Section 3.5.5.

3.5.1 The clustered brain maps with K= 2, 3, 5, and 7 for test Sessions

We present the clustered brain maps in the following order:

- K-means method for the pre-test for slices 13 – 19 in Figure 3.9.
- K-means method for the pre-test for slices 20 – 26 in Figure 3.10.
- K-means method for the pre-test for slices 27 – 33 in Figure 3.11.
- PCA K-means method for the pre-test for slices 13 – 19 in Figure 3.12.
- PCA K-means method for the pre-test for slices 20 – 26 in Figure 3.13.
- PCA K-means method for the pre-test for slices 27 – 33 in Figure 3.14.
- K-means method for the post-test for slices 13 – 19 in Figure 3.15.
- K-means method for the post-test for slices 20 – 26 in Figure 3.16.
- K-means method for the post-test for slices 27 – 33 in Figure 3.17.
- PCA K-means method for the post-test for slices 13 – 19 in Figure 3.18.
- PCA K-means method for the post-test for slices 20 – 26 in Figure 3.19.
- PCA K-means method for the post-test for slices 27 – 33 in Figure 3.20.

The clusters having the largest to the smallest amplitude are displayed in colors ranging from green (largest) to blue, red, cyan, purple, olive, and black (smallest). The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

From the clustered brain maps, we notice several remarkable points. First, the clustering results by K-means and PCA K-means are similar. Second, partitioning brains into two

clusters (i.e., K=2) fails to detect some major areas known as antisaccade task-related regions (e.g., Dyckman et al. (2007)). For example, see slices 27 – 32 in Figure 3.11. Some of the antisaccade related regions shown in green and blue when $K = 3$ are not detected when $K = 2$. Note that we interpret only green areas as the activated region for $K = 2$ by combining the information from the time series plots in Figures 3.21(a) and 3.21(b). The cluster displayed in blue in the average time series plot for K=2 appears to be noise. Third, K greater than three merely breaks the noise cluster into multiple clusters and so does not provide additional information. For example, compare K=3 with K=5 and K=7 in Figure 3.11. The major areas known as antisaccade task-related regions are detected for K=3, 5, and 7, but the non-task related areas for K=3 (shown in red) are broken into several areas for K=5 and 7 (i.e., noise clusters are displayed in several dark colors, but not in green or blue).

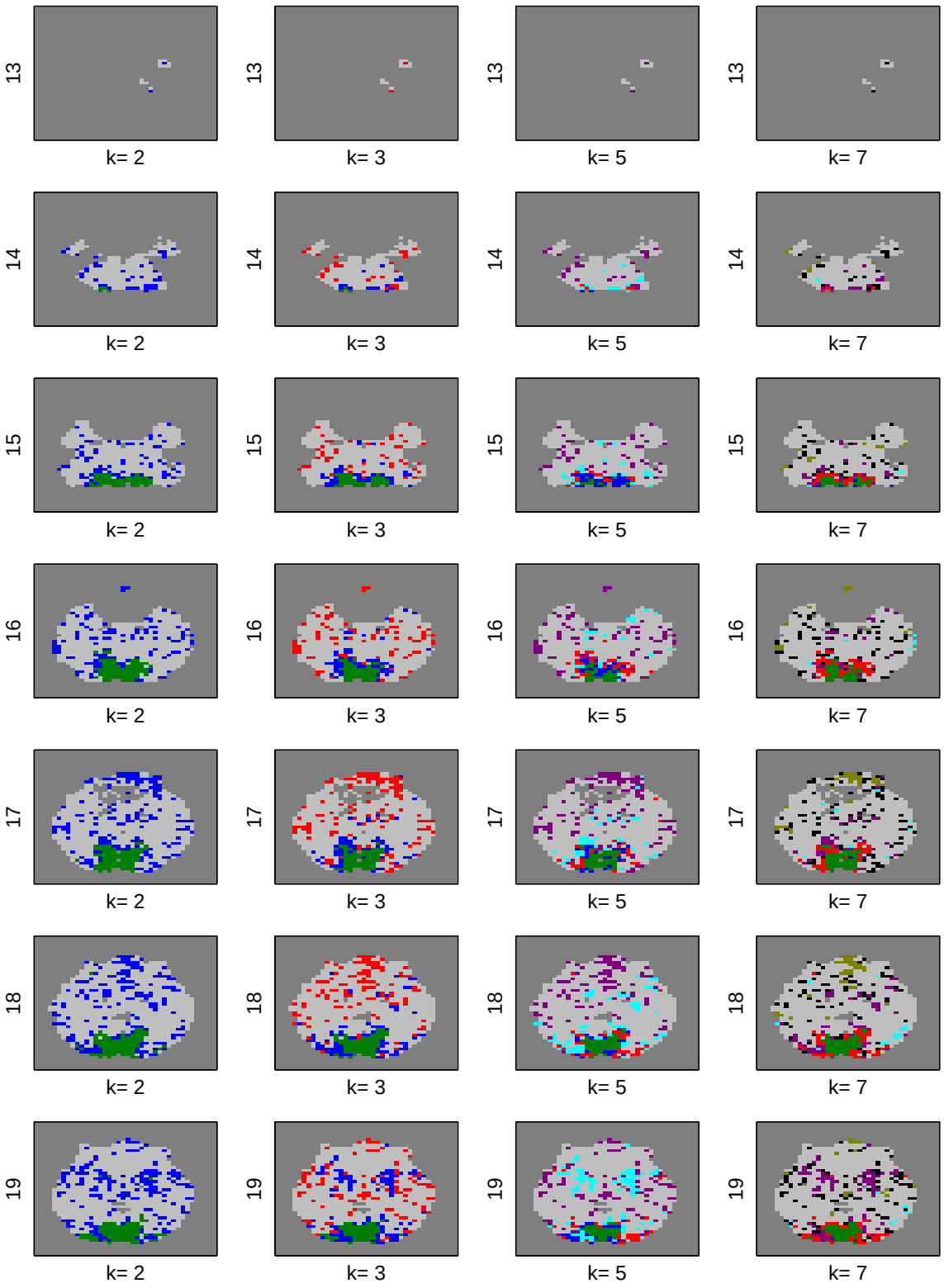


Figure 3.9: The clustered brain maps for the pre-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

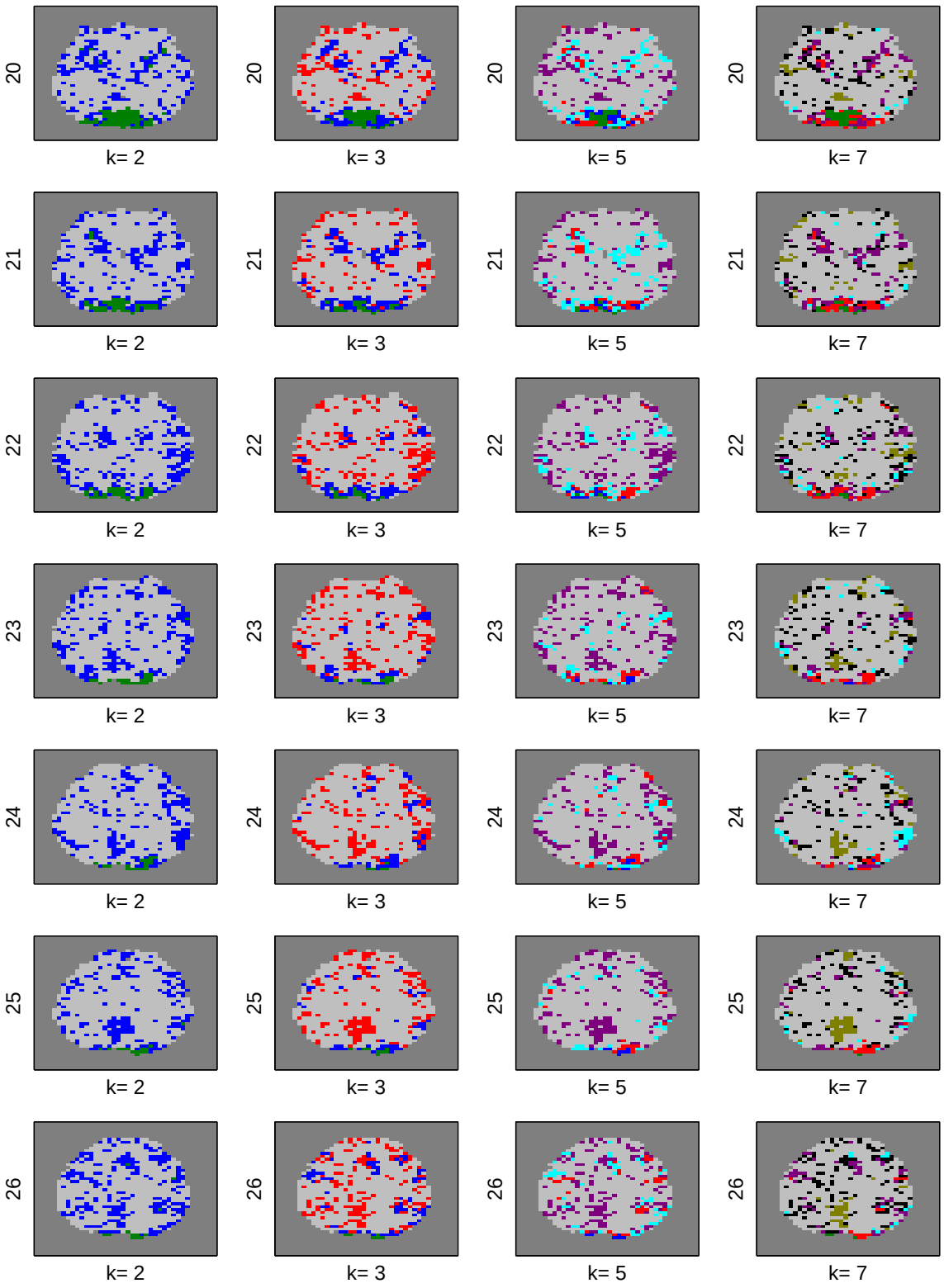


Figure 3.10: The clustered brain maps for the pre-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

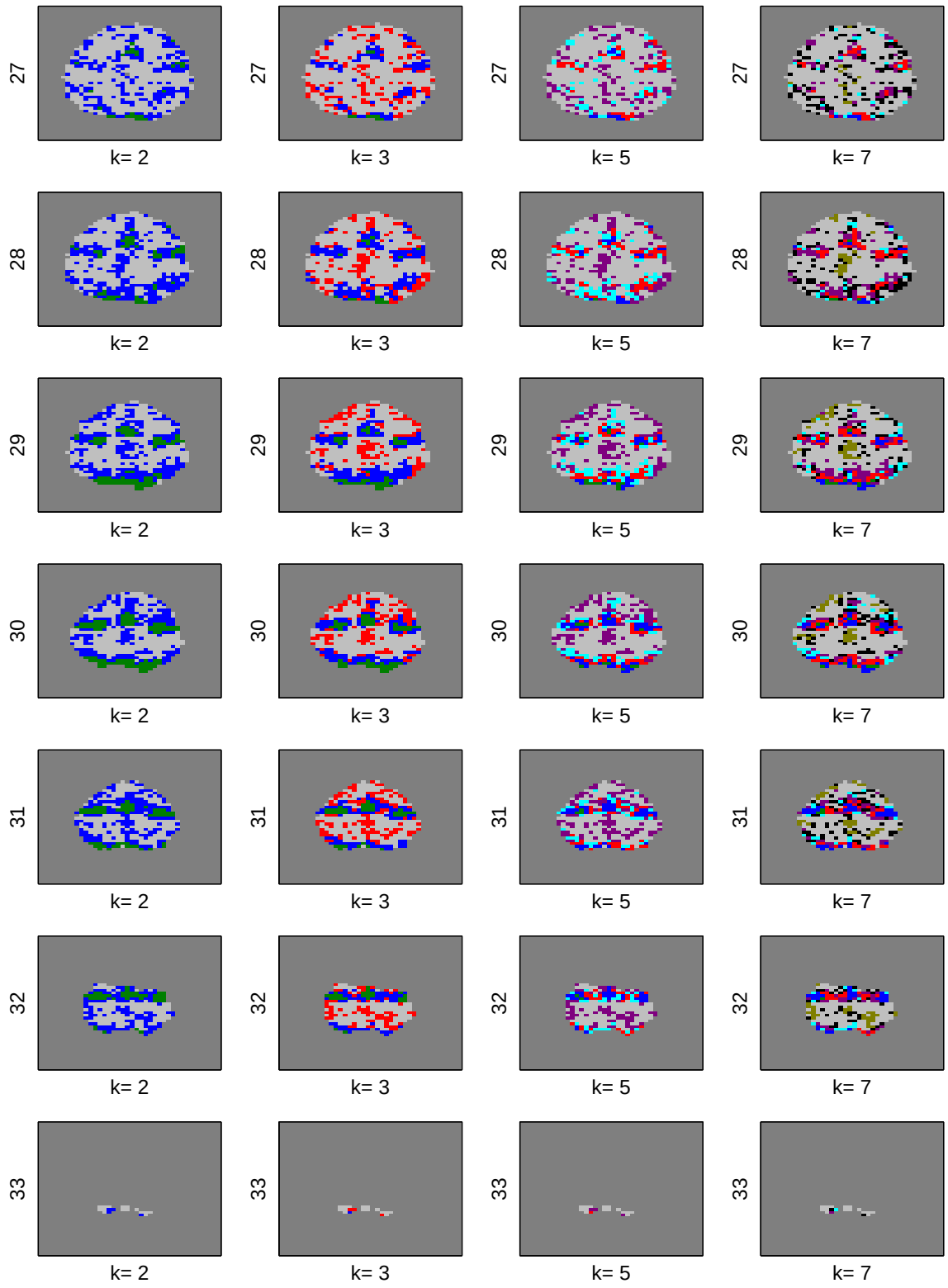


Figure 3.11: The clustered brain maps for the pre-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

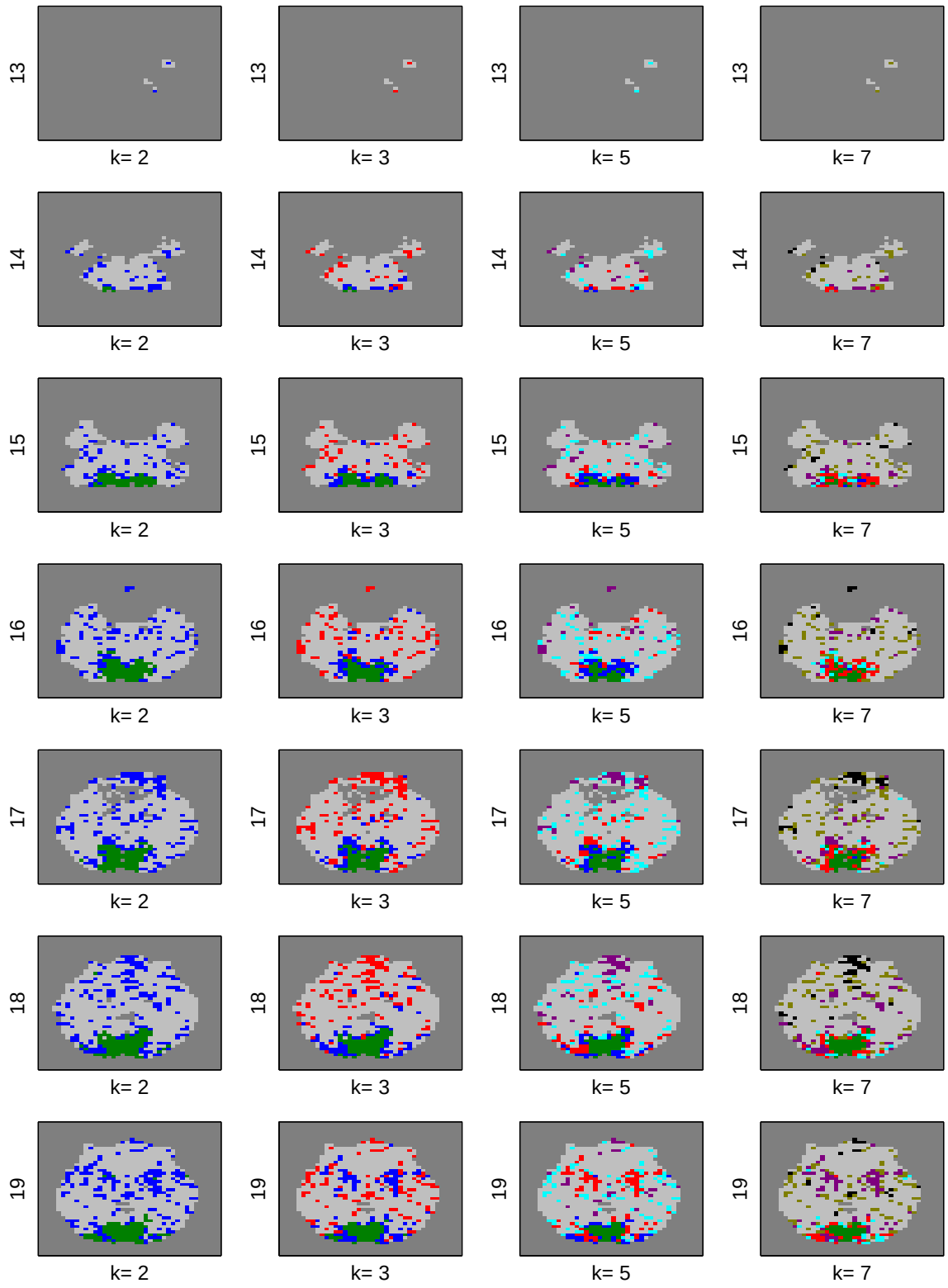


Figure 3.12: The clustered brain maps for the pre-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

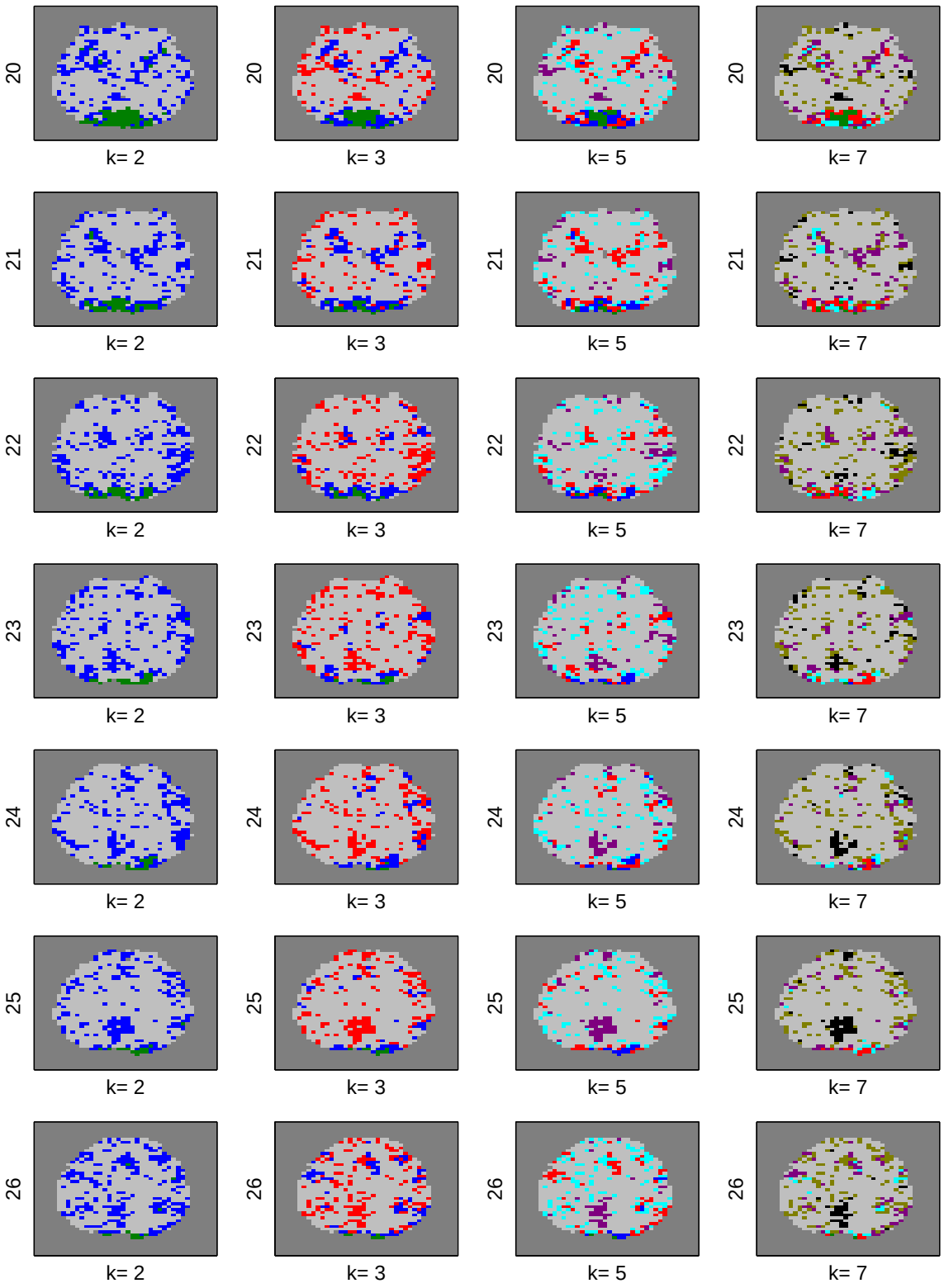


Figure 3.13: The clustered brain maps for the pre-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

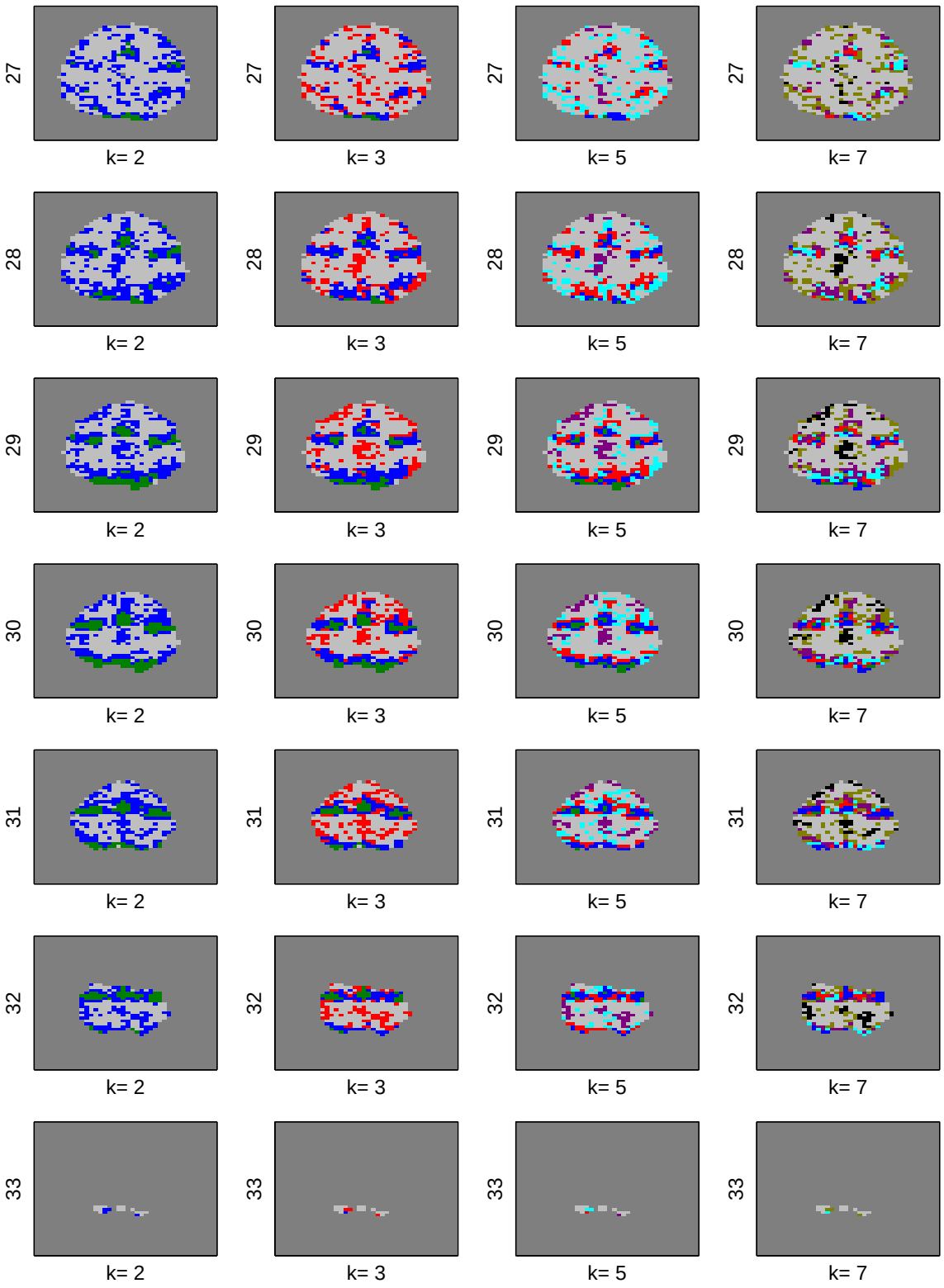


Figure 3.14: The clustered brain maps for the pre-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

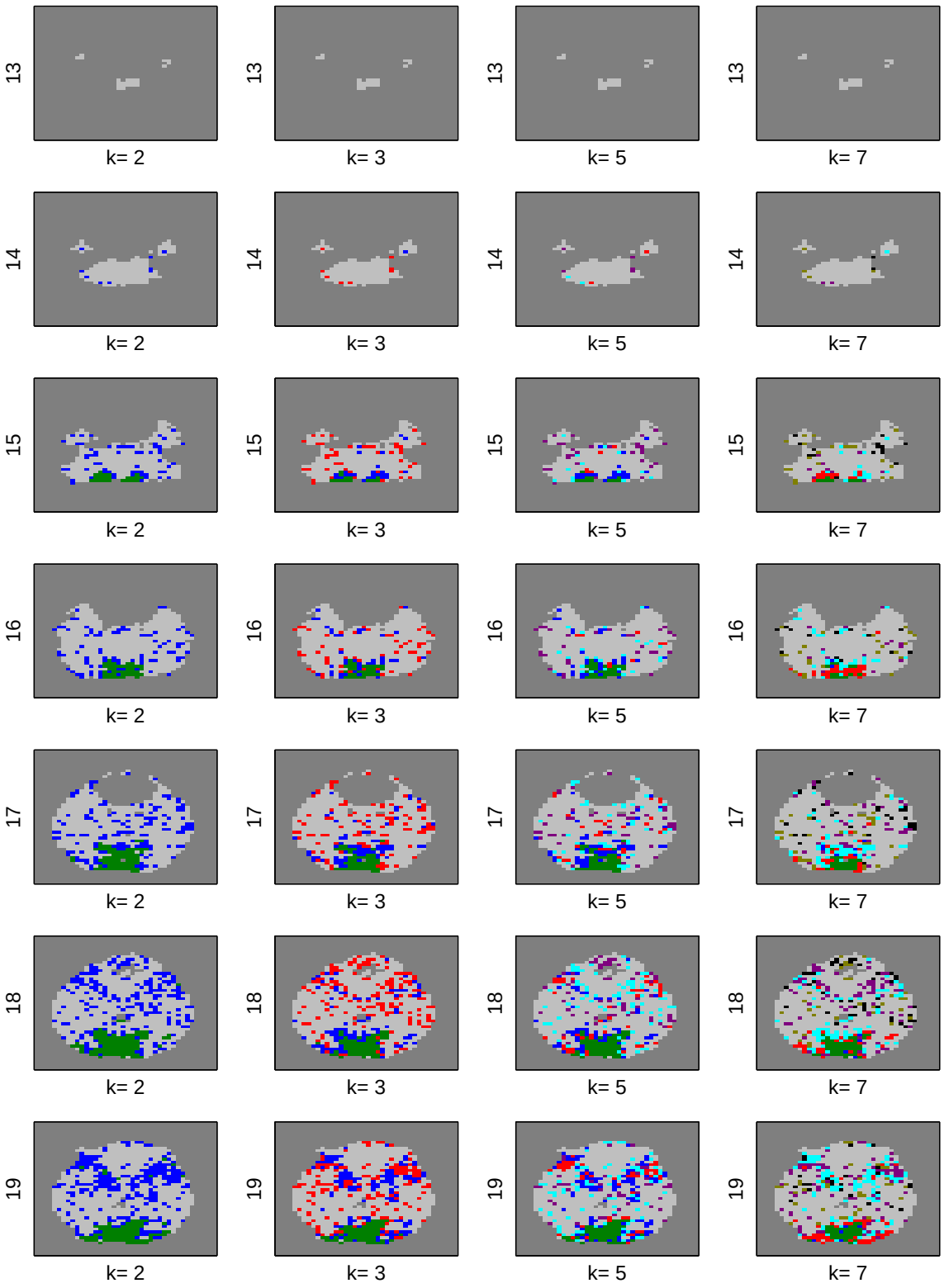


Figure 3.15: The clustered brain maps for the post-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

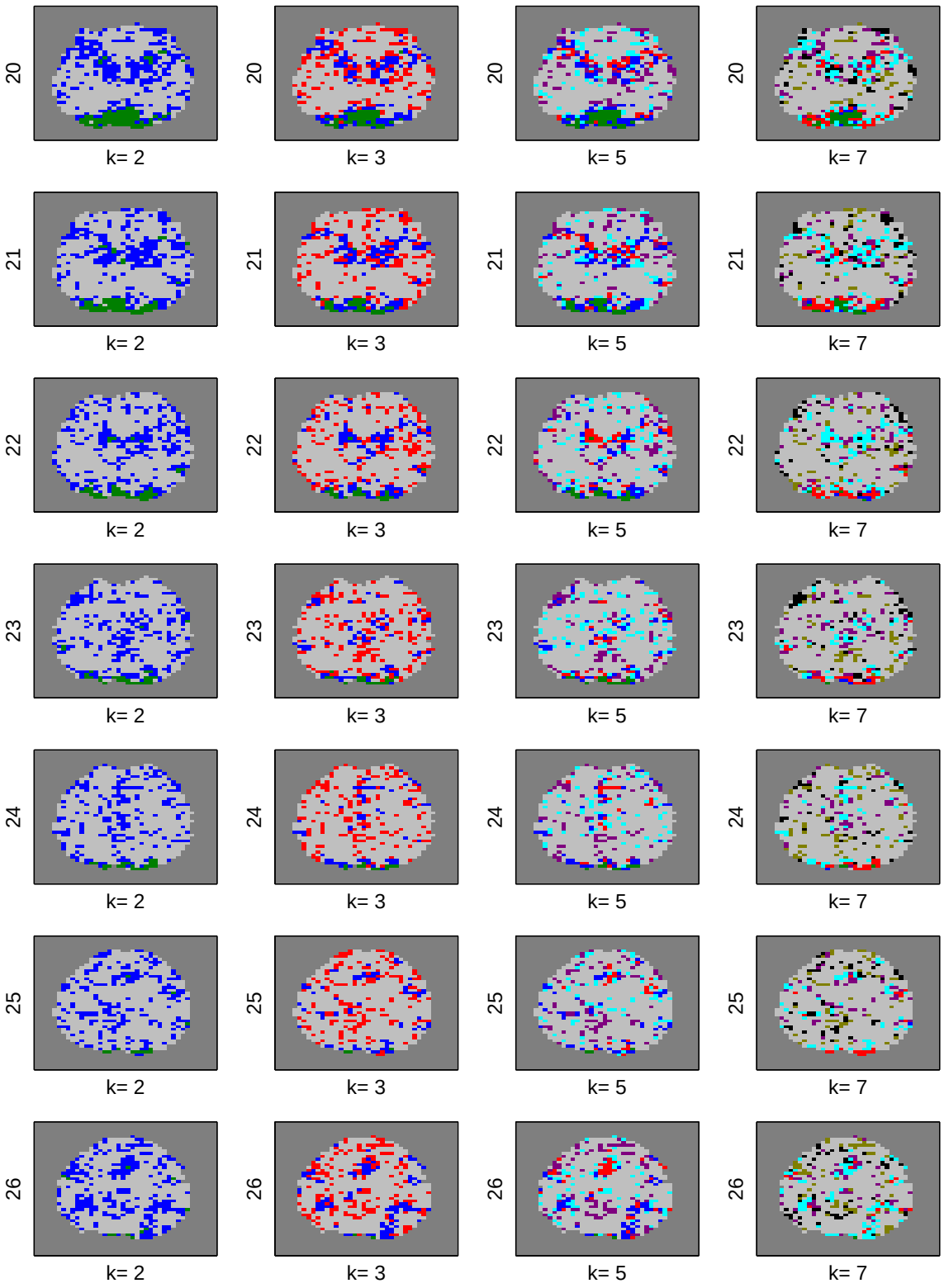


Figure 3.16: The clustered brain maps for the post-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

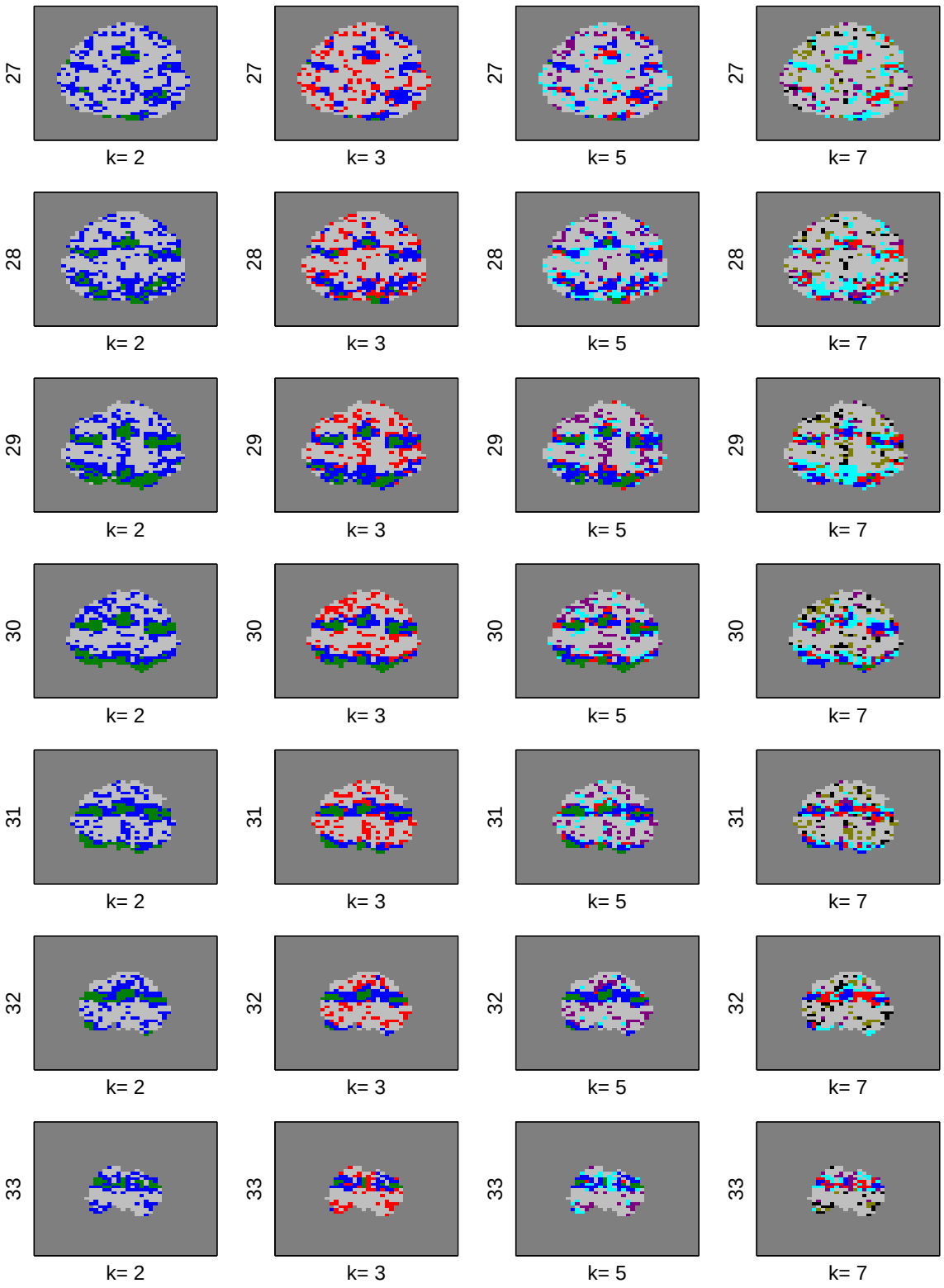


Figure 3.17: The clustered brain maps for the post-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

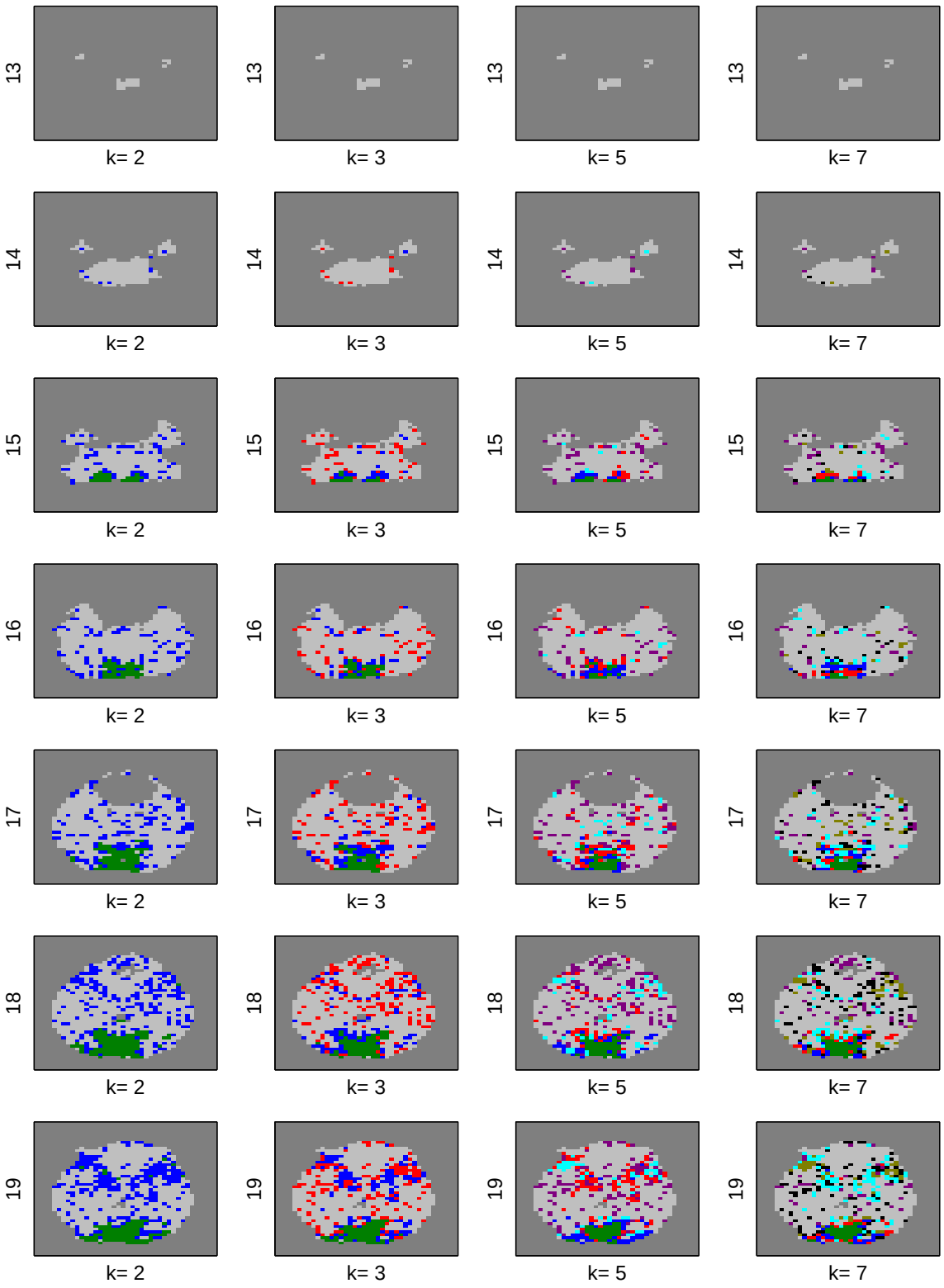


Figure 3.18: The clustered brain maps for the post-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

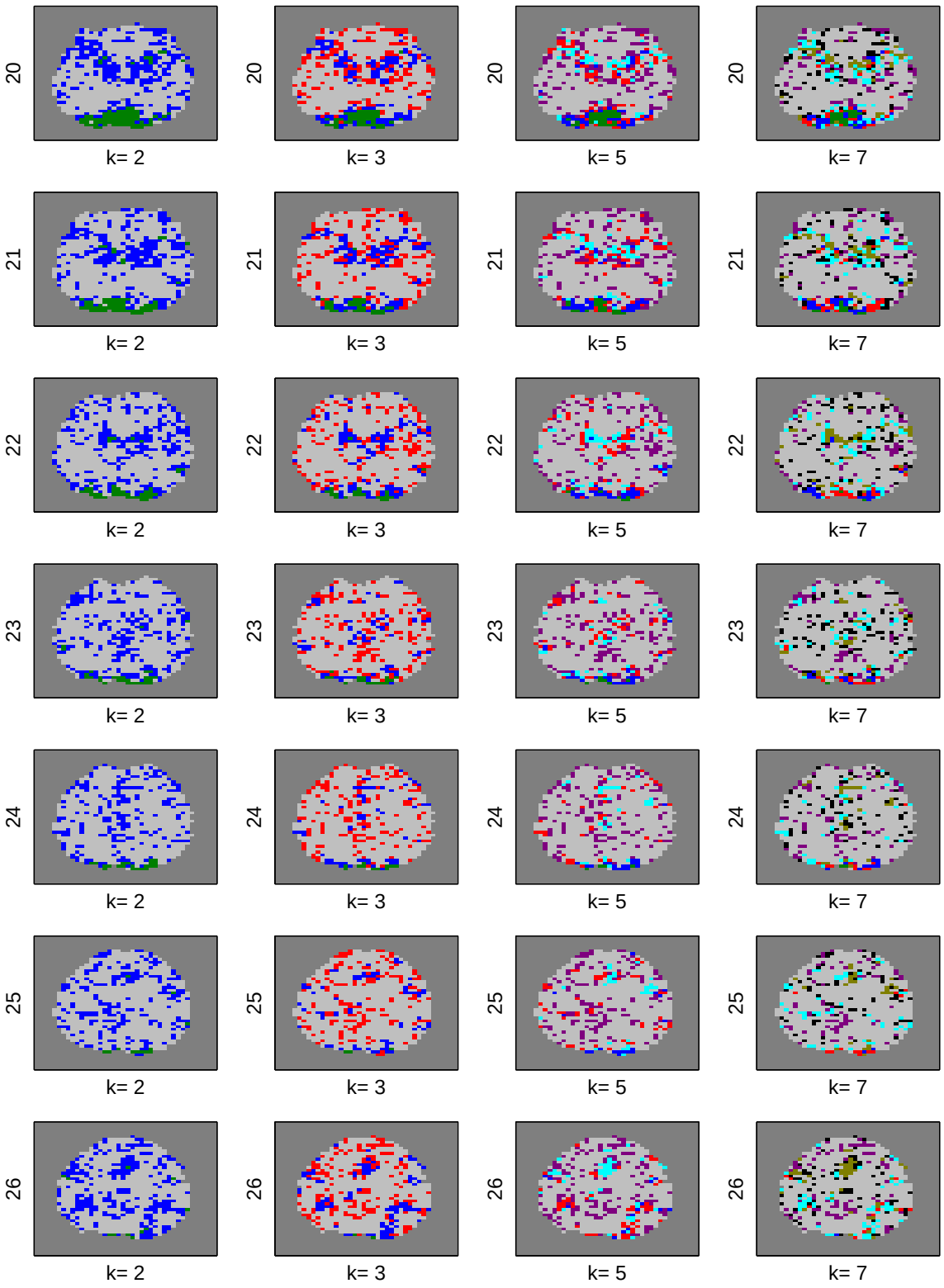


Figure 3.19: The clustered brain maps for the post-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

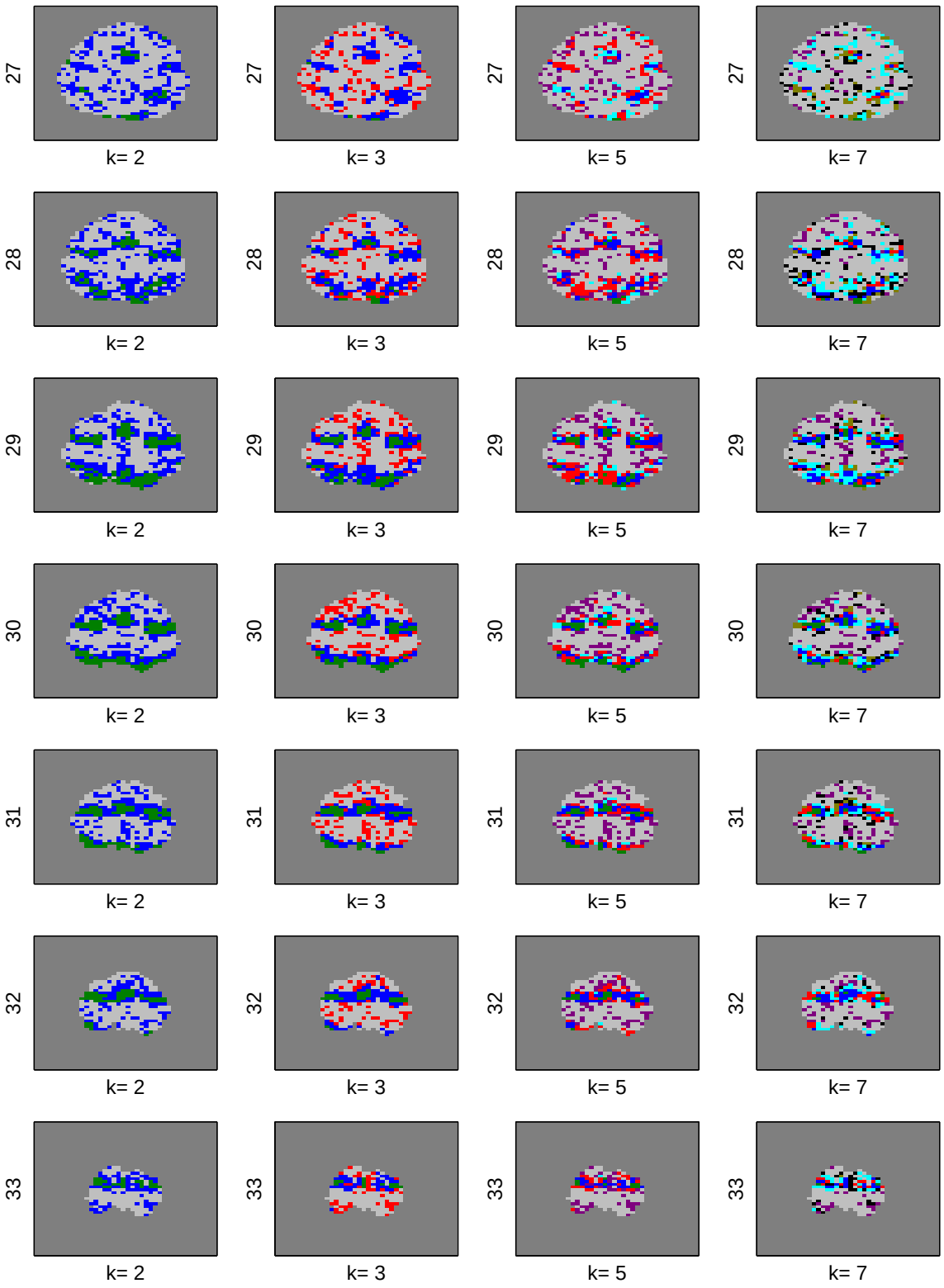
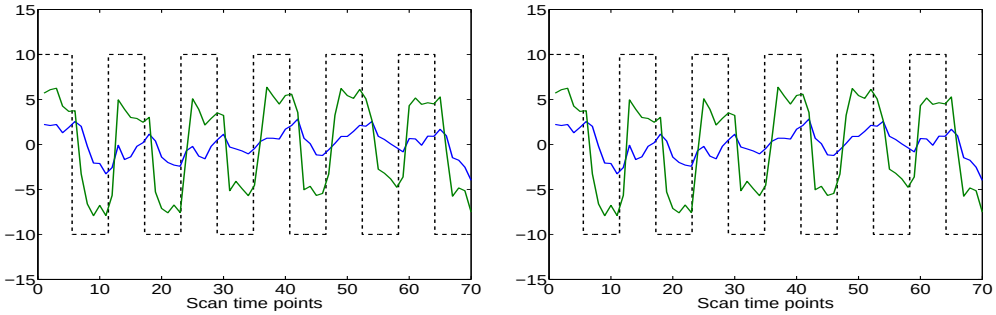


Figure 3.20: The clustered brain maps for the post-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

3.5.2 Average time series plot with K= 2, 3, 5, and 7 for test Sessions

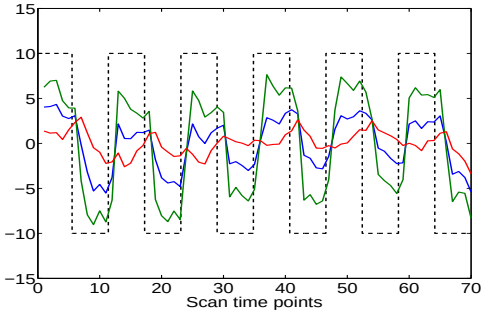
We present the average time series plots in each cluster for the pre-test in Figure 3.21 and for the post-test in Figure 3.22. The K-means and the PCA K-means results are displayed in the left and right panels, respectively. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. For reference, the stimulus presentation is in dashed black lines, depicting alternating blocks of fixation and the antisaccade task. From the largest to the smallest amplitude, clusters are displayed in green, blue, red, cyan, purple, olive, and black. The clusters displayed in these colors correspond to the same colored clusters in the brain maps in Section 3.5.1.

From these average time series plots, we find several noticeable points. First, we can clearly see that the clustering results by K-means and PCA K-means are similar as discovered in the clustered brain maps. Second, for each plot at least one cluster (shown in green) displays a box-car shape, suggesting the presence of antisaccade task-related voxels. Third, we again observe that the larger numbers of clusters K only break the noise cluster into multiple clusters and do not provide additional information. For example, see Figures 3.21(d) and 3.21(f). The clusters having the first two largest amplitudes in Figure 3.21(d) shown in green and blue look similar to the green and blue clusters in Figure 3.21(f). However, the cluster having the smallest amplitude (shown in red) in Figure 3.21(d) is divided into the clusters shown in red, cyan, and purple in Figure 3.21(f).

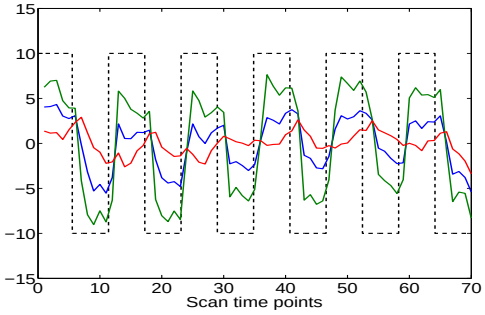


(a) K-means with K=2

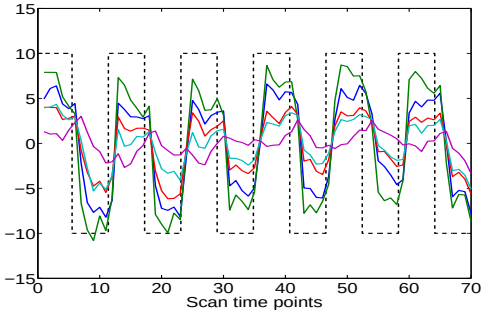
(b) PCA K-means with K=2



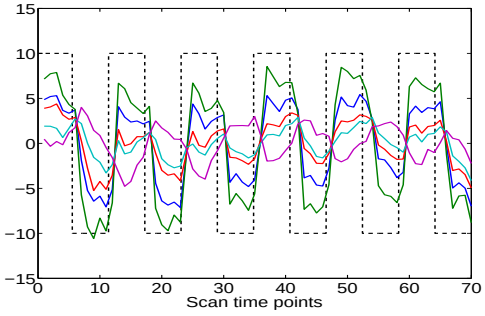
(c) K-means with K=3



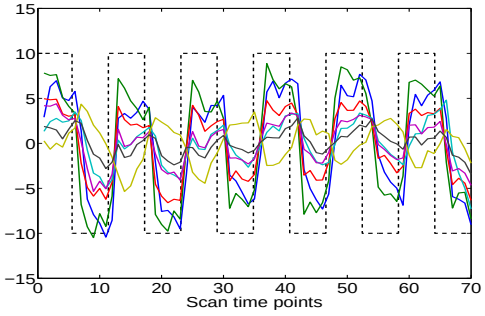
(d) PCA K-means with K=3



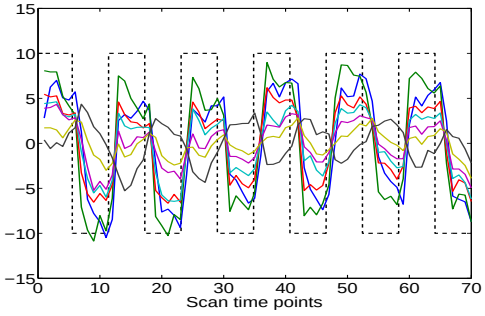
(e) K-means with K=5



(f) PCA K-means with K=5

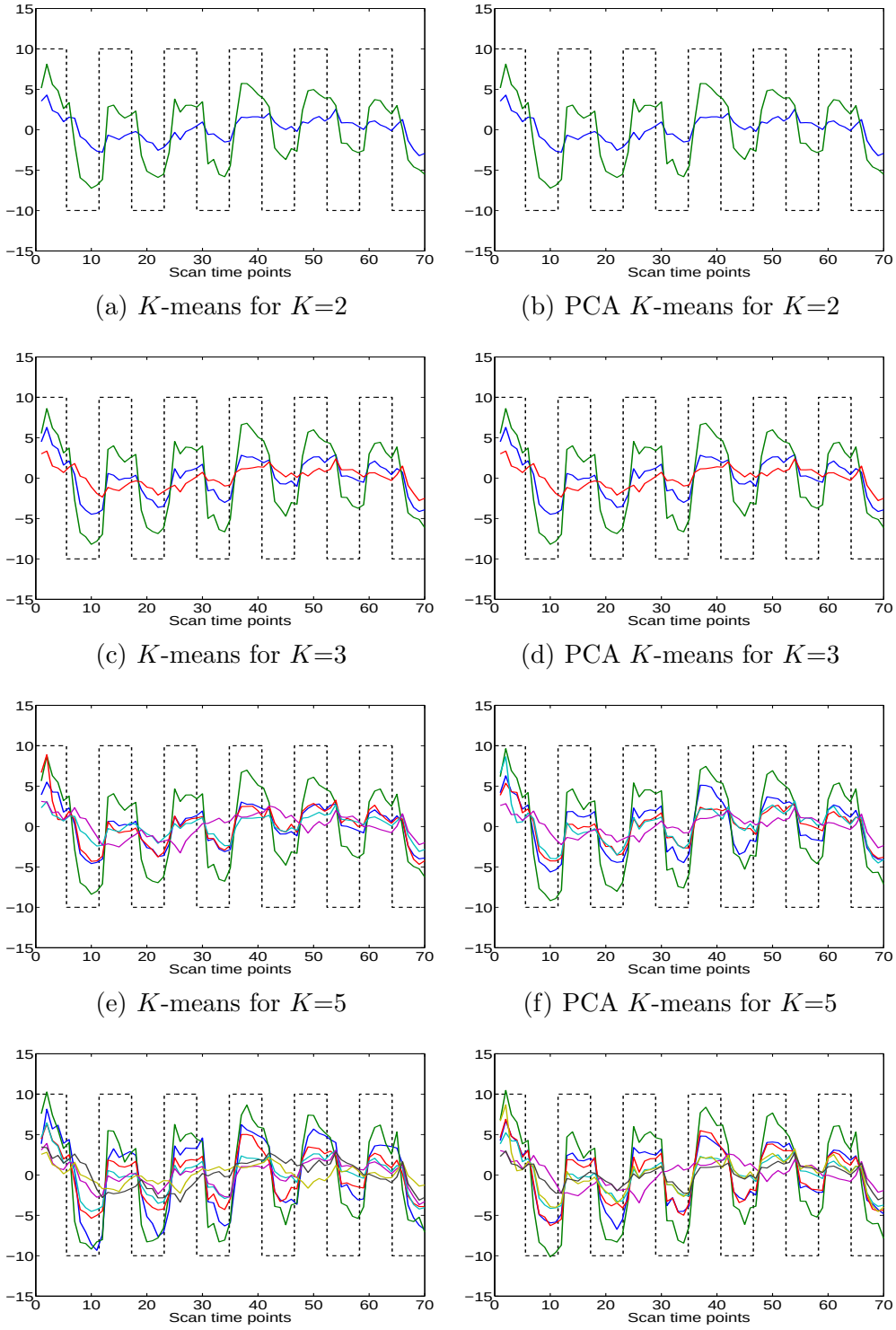


(g) K-means with K=7



(h) PCA K-means with K=7

Figure 3.21: Average time series plots for the pre-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing.



(a) K-means for K=2

(b) PCA K-means for K=2

(c) K-means for K=3

(d) PCA K-means for K=3

(e) K-means for K=5

(f) PCA K-means for K=5

(g) K-means for K=7

(h) PCA K-means for K=7

Figure 3.22: Average time series plots for the post-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing

3.5.3 Comparison of results by median, mean, and trimmed mean

As mentioned in Section 3.3.3, we also aggregate the data using other center measures such as mean, and 10% and 20% trimmed means. In this subsection we compare their clustering results with those from the median using $K = 3$. Figures 3.23 - 3.28 show the clustered brain maps when we group voxels into three clusters using PCA K-means. Each Figure has four columns, and each column represents the results by using mean, 10% and 20% trimmed means, and median, respectively. In the plots, the y labels indicate the slice numbers and the x labels indicate the center measures used for data aggregation. The upper and lower areas of each slice correspond to the anterior and posterior of the brain, respectively, and the larger slice number is toward the top of the head. As we mention in Section 3.3.3, the clustered brain maps do not show significant differences across different center measures for either the pre- (Figures 3.23-3.25) or post-tests (Figures 3.26-3.28).

We note that the aggregated data with median have fewer noise voxels after the adaptive pivotal thresholding test than other measures, while keeping a similar number of activated voxels. For example, see the four superior slices (slices 29 – 32) in Figure 3.28, which clearly contain the well known saccadic circuitry (Camchong et al., 2008; Dyckman et al., 2007, and references therein). The median brain maps show smaller red areas (noise voxels) but similar green and blue areas (activated areas) compared to the other measures.

Figure 3.29 displays the average time series plots by using these four center measures. Cluster 1 (green), cluster 2 (blue), and cluster 3 (red) represent the green, blue, and red areas in Figures 3.23-3.28, respectively. The first and the second columns in Figure 3.29 indicate the average time series plots for the pre- and post-tests, respectively. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. For reference, the stimulus presentation is in bashed black lines, depicting alternating blocks of fixation and the antisaccade task. It is evident that the clustering results by the four measures are similar to one another for both the pre- and post-tests.

We aggregate the data using the median because there exist some outliers within the voxels for each subject, and the adaptive pivotal thresholding test more effectively filters out the noise voxels when the median is used than when other measures are used.

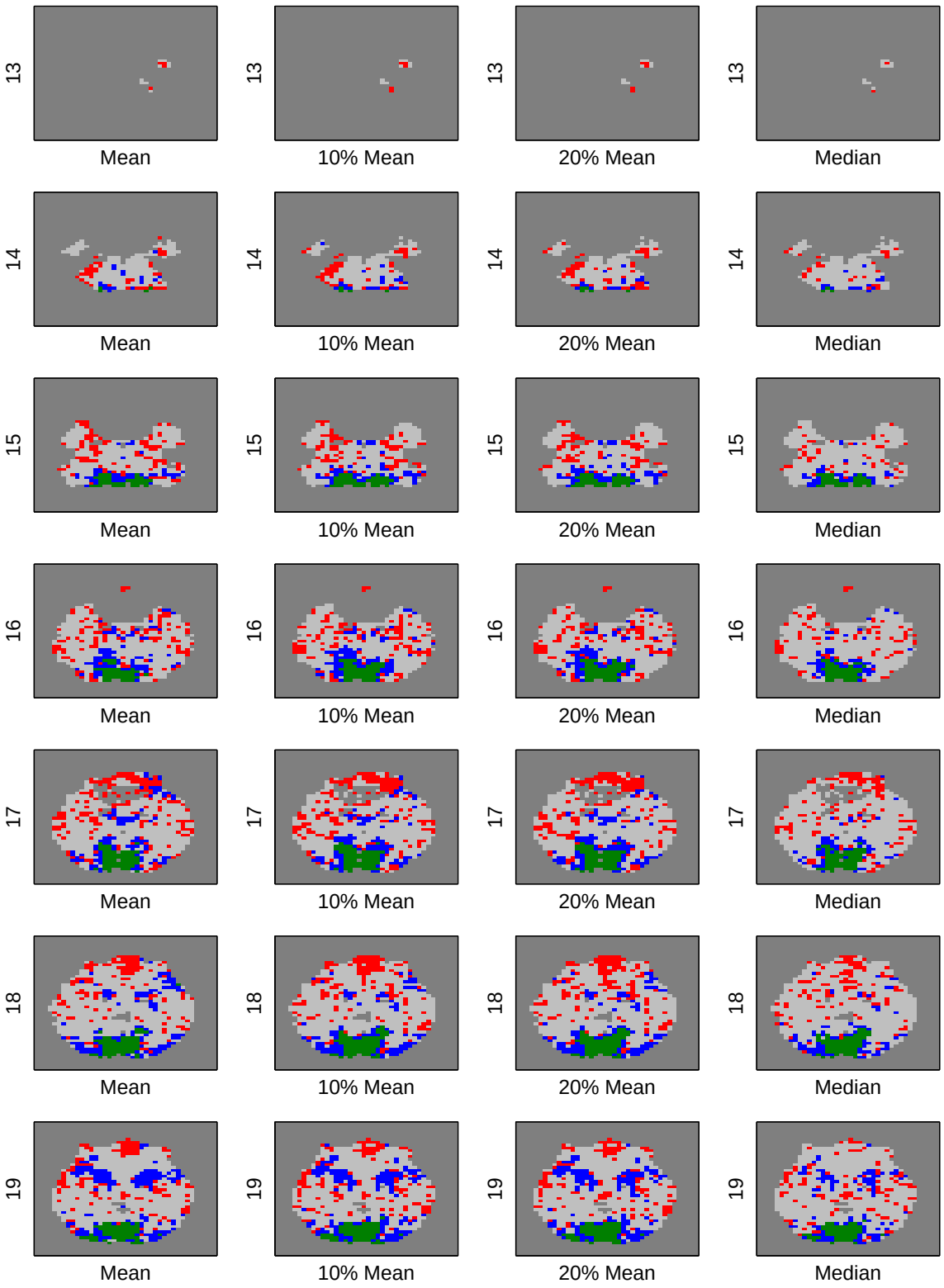


Figure 3.23: The clustered brain maps using other center measures for the pre-test for slices 13 – 19. These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the center measures used for data aggregation: mean, 10% trimmed mean, 20% trimmed mean, and median. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

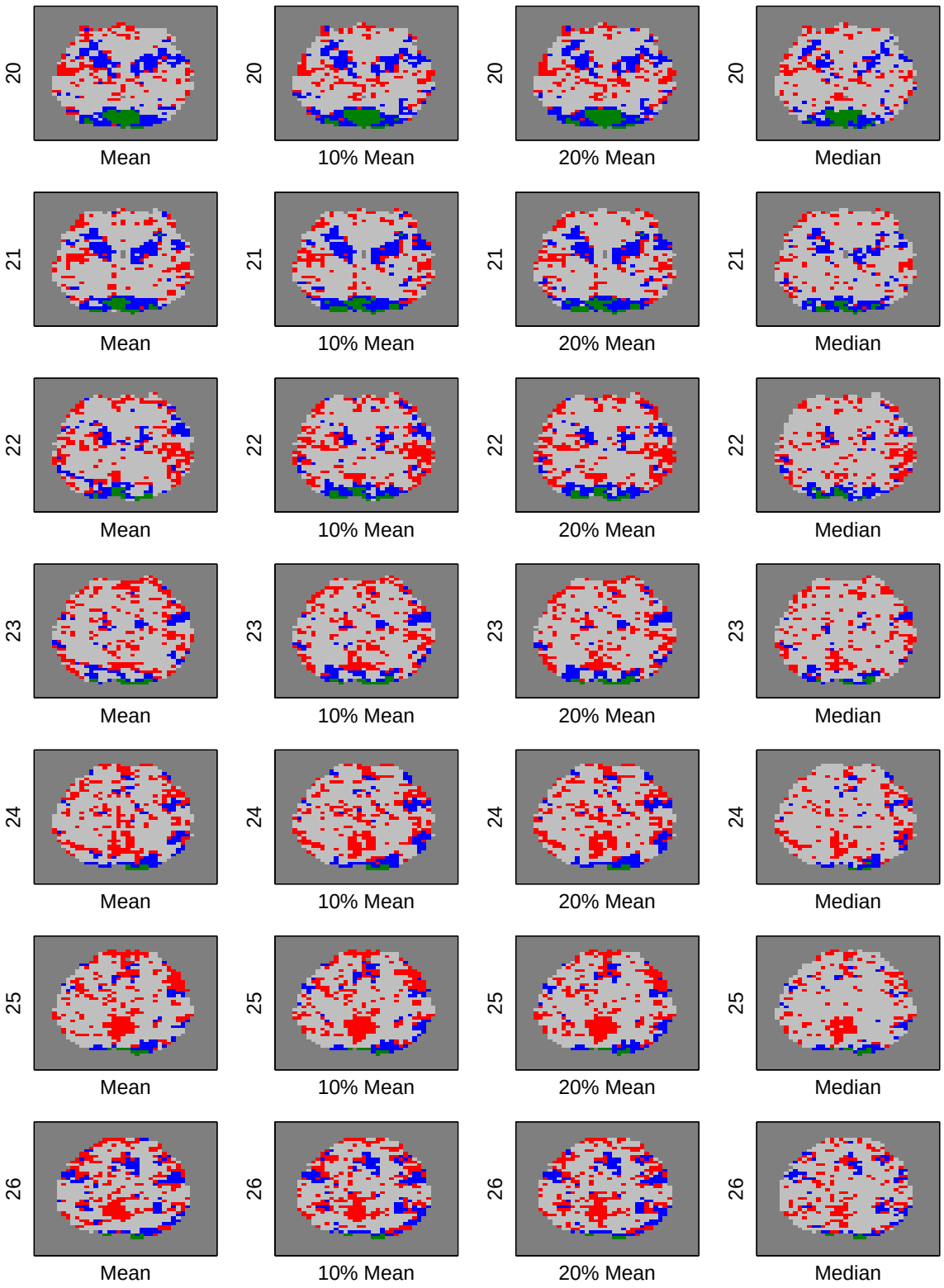


Figure 3.24: The clustered brain maps using other center measures for the pre-test for slices 20 – 26. These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the center measures used for data aggregation: mean, 10% trimmed mean, 20% trimmed mean, and median. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

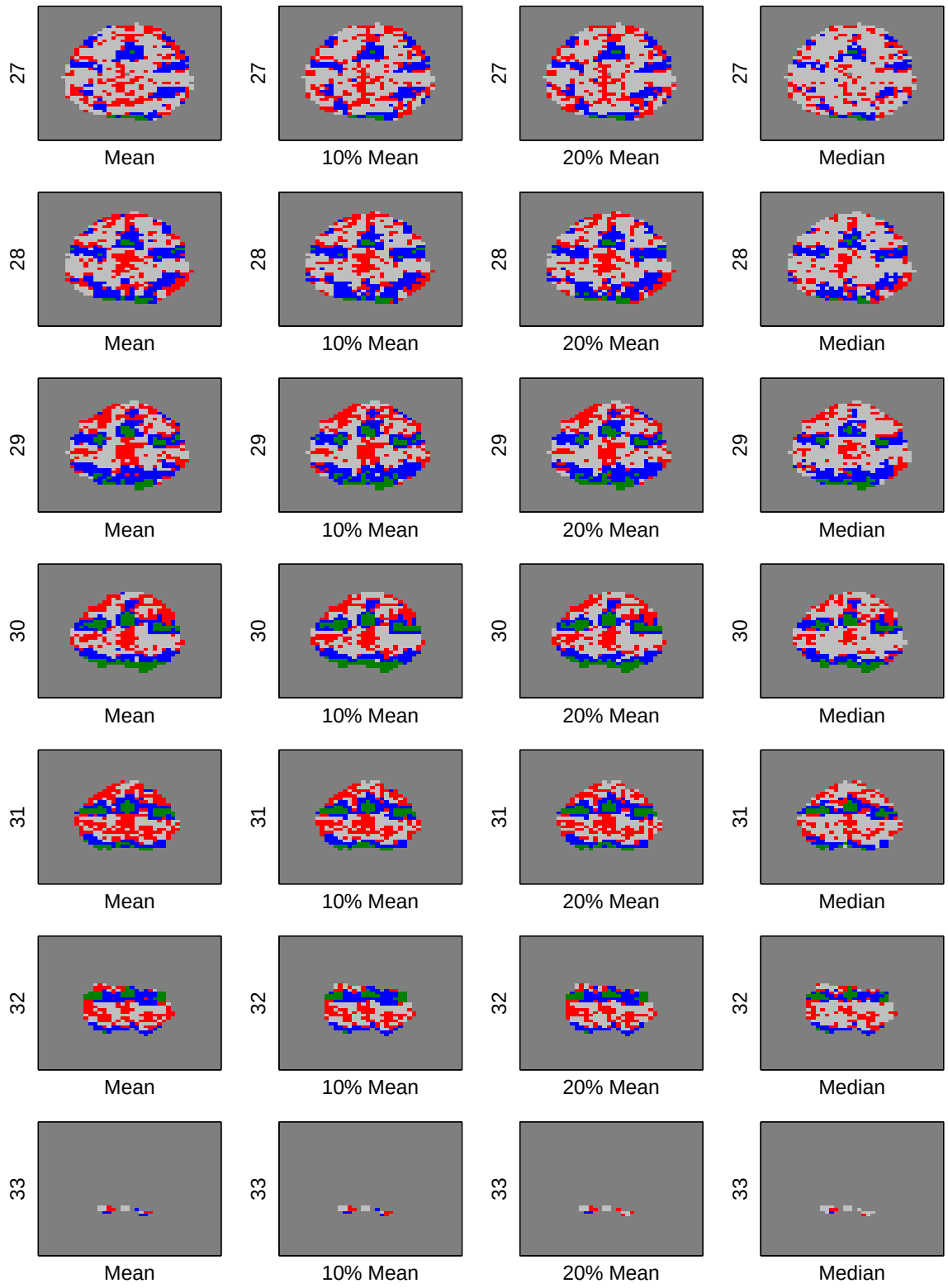


Figure 3.25: The clustered brain maps using other center measures for the pre-test for slices 27 – 33. These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the center measures used for data aggregation: mean, 10% trimmed mean, 20% trimmed mean, and median. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

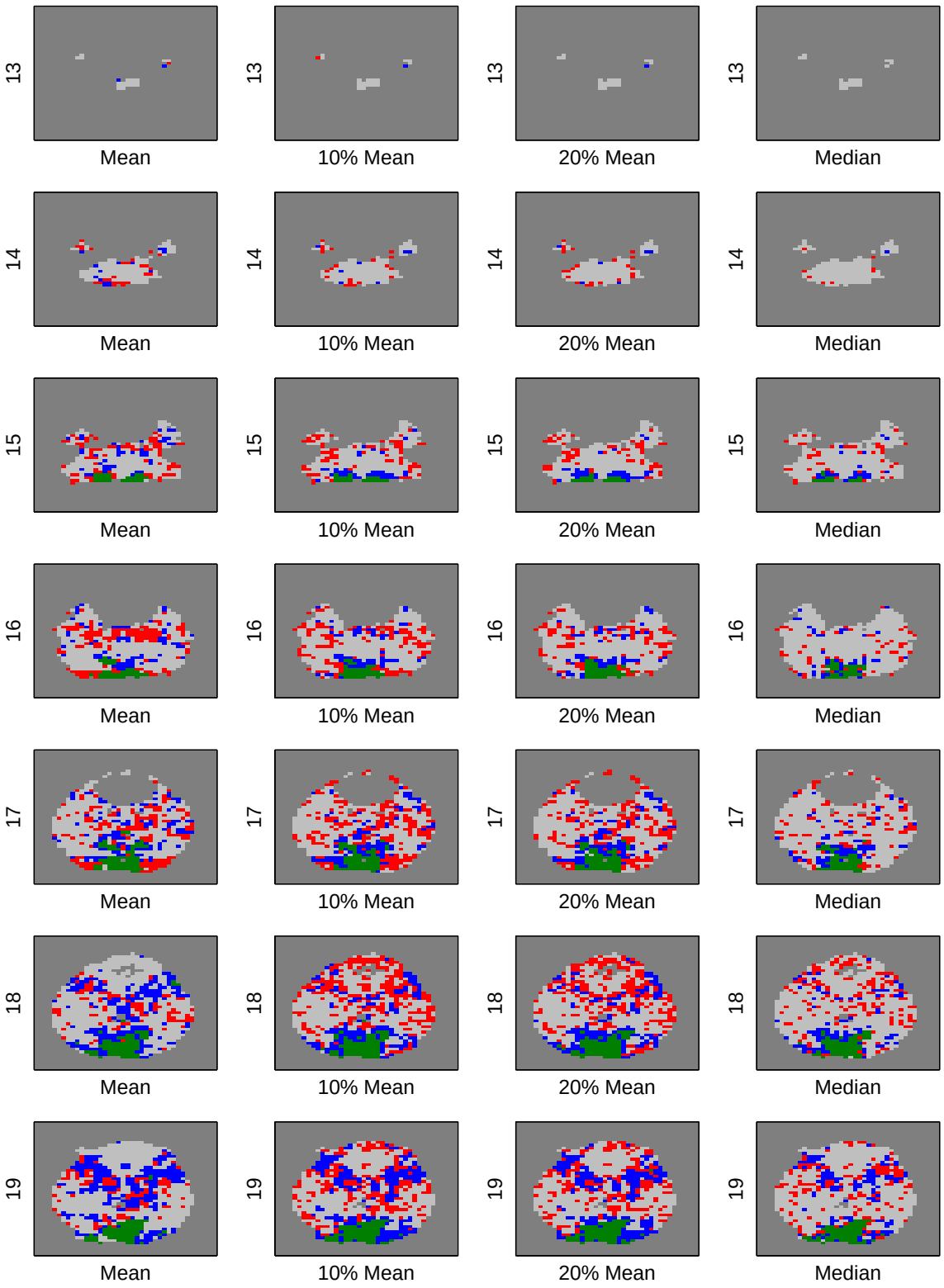


Figure 3.26: The clustered brain maps using other center measures for the post-test for slices 13 – 19. These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the center measures used for data aggregation: mean, 10% trimmed mean, 20% trimmed mean, and median. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

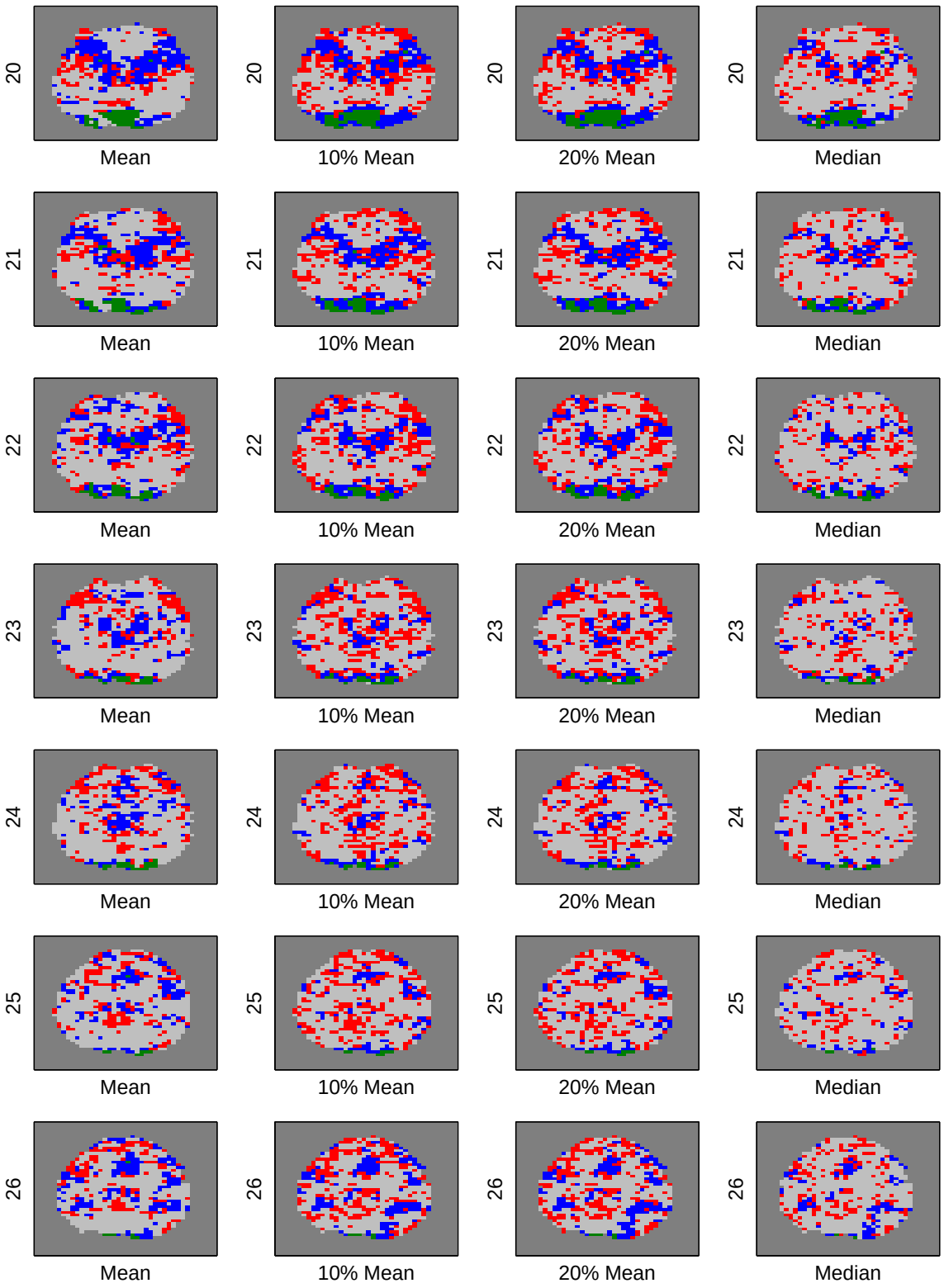


Figure 3.27: The clustered brain maps using other center measures for the post-test for slices 20 – 26. These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the center measures used for data aggregation: mean, 10% trimmed mean, 20% trimmed mean, and median. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

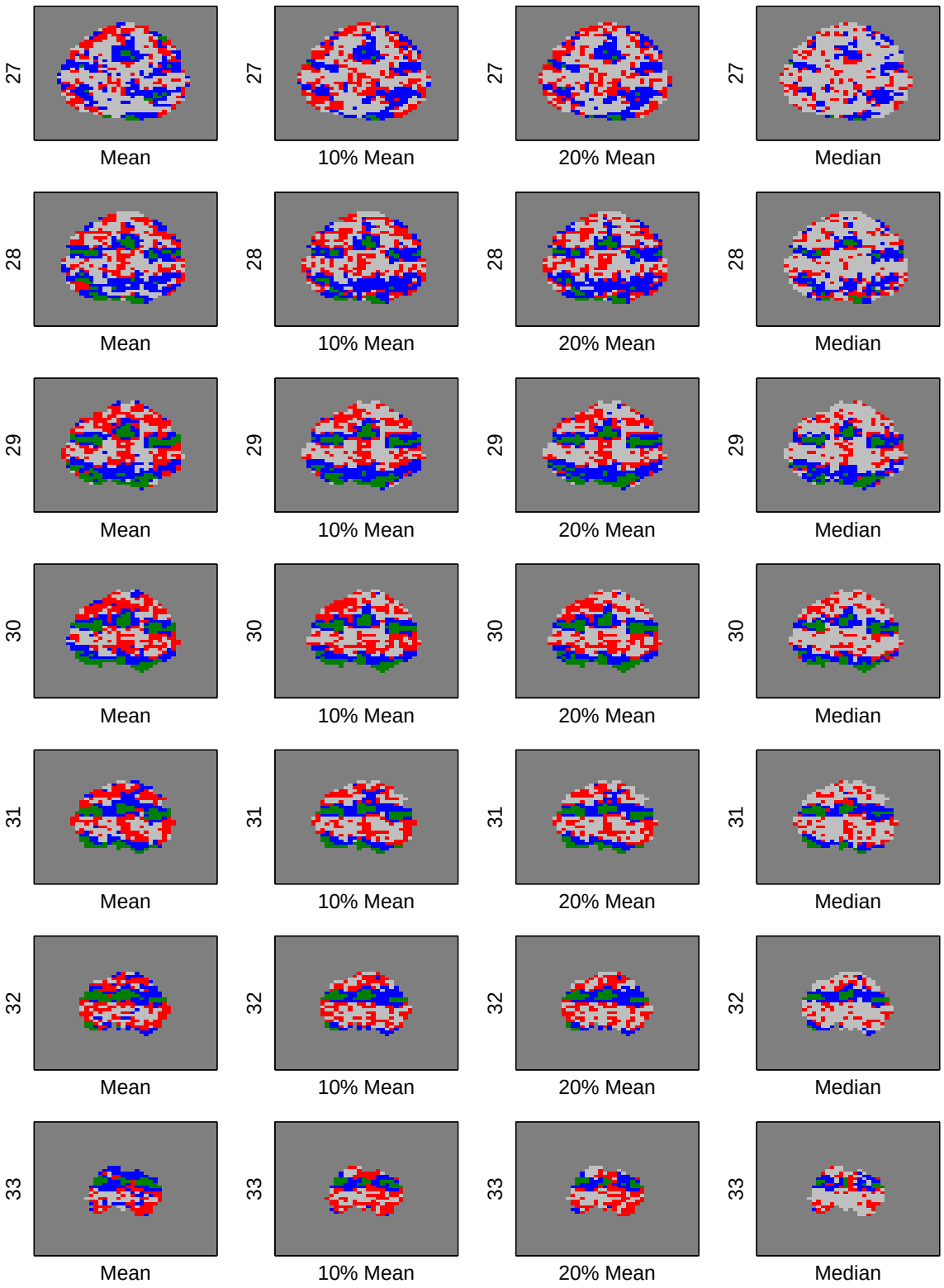
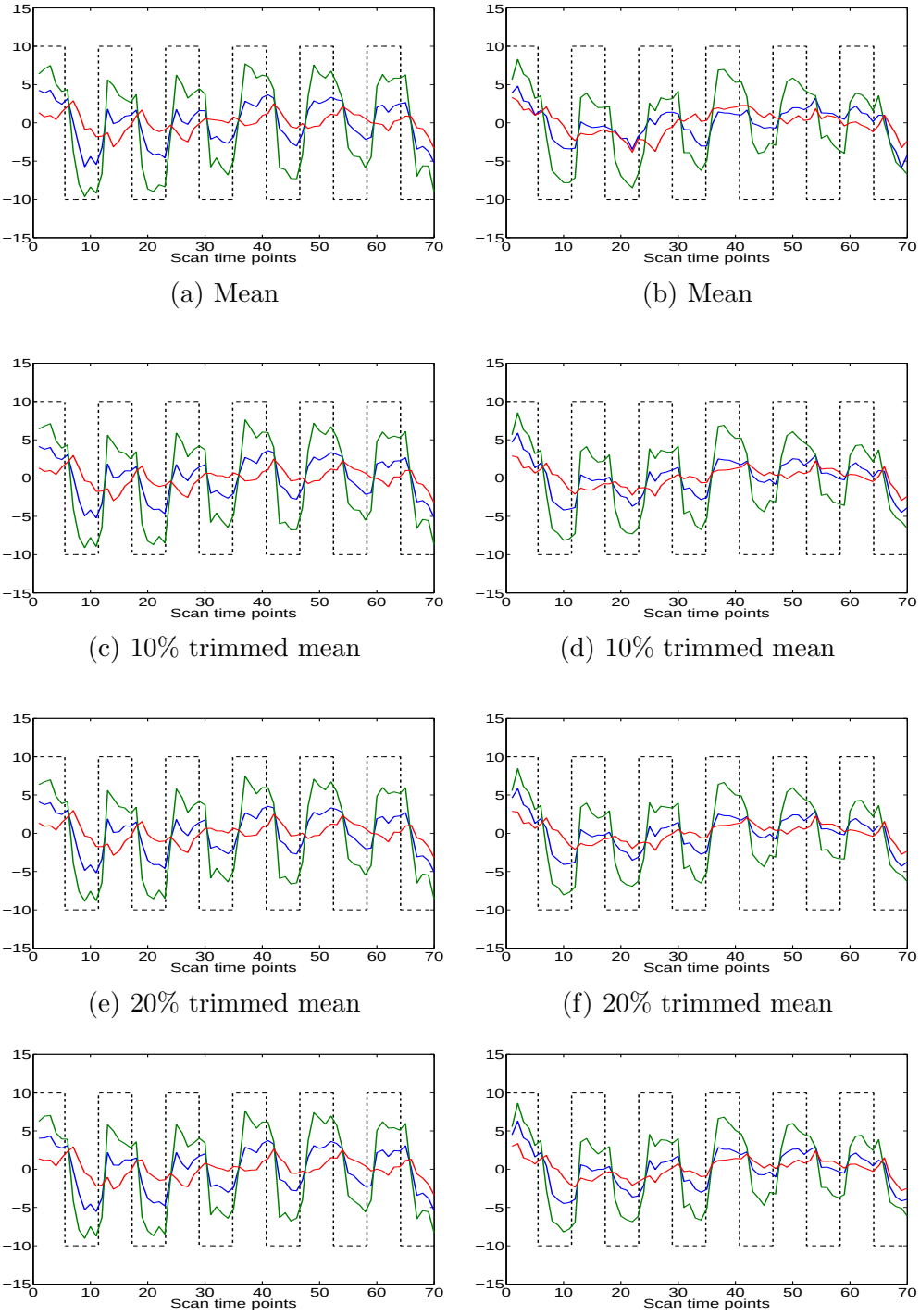


Figure 3.28: The clustered brain maps using other center measures for the post-test for slices 27 – 33. These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the center measures used for data aggregation: mean, 10% trimmed mean, 20% trimmed mean, and median. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.



(a) Mean (b) Mean

(c) 10% trimmed mean (d) 10% trimmed mean

(e) 20% trimmed mean (f) 20% trimmed mean

(g) Median (h) Median

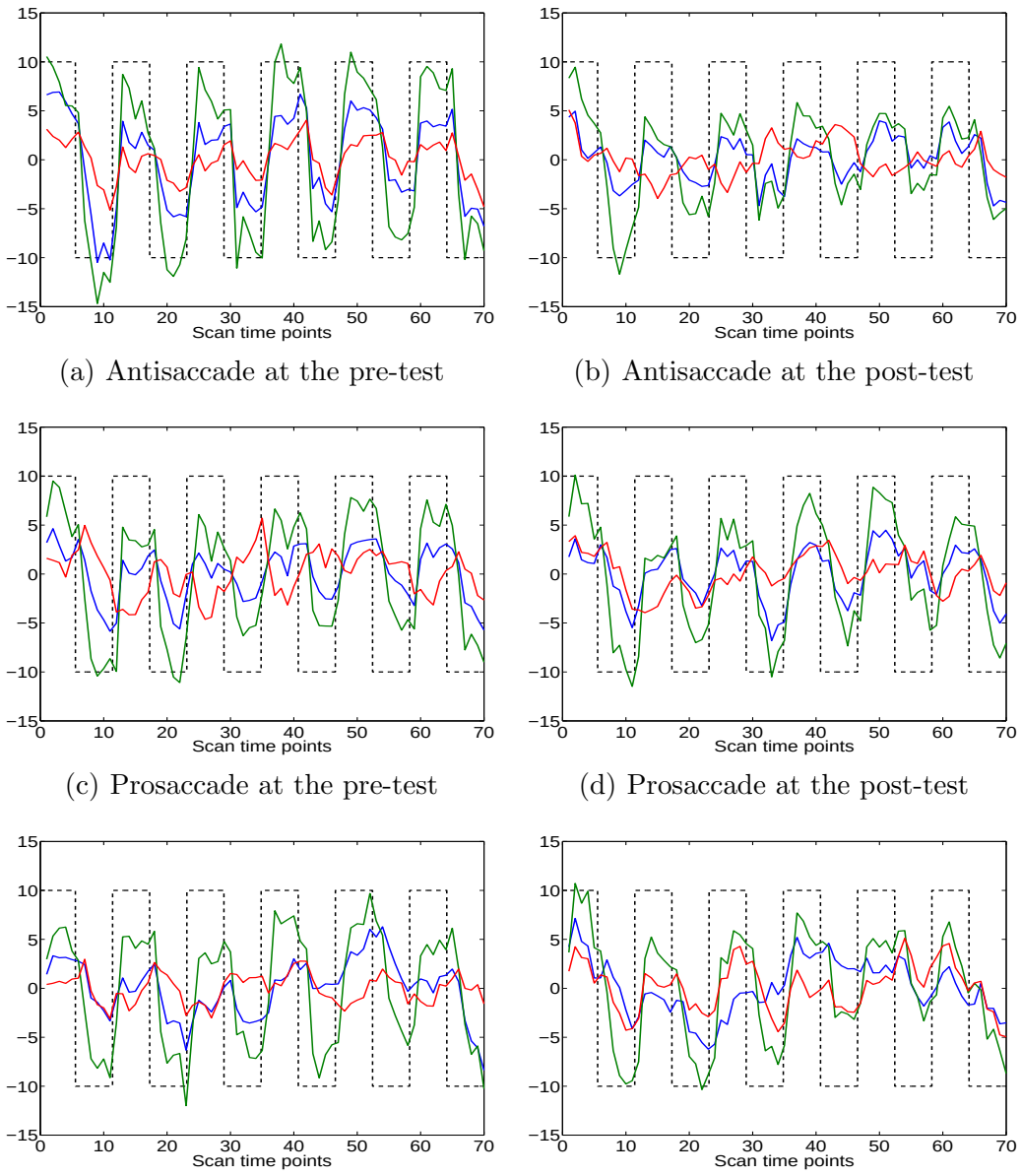
Figure 3.29: Average time series plots using other center measures for the pre- (left) and post-test (right). The clusters 1 (green), 2 (blue) and 3 (red) correspond the green, blue and red areas in the previous brain maps (Figures 3.23 – 3.28), respectively. These results are produced by PCA K-means clustering for K=3. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing.

3.5.4 Comparison of results for test Sessions by task groups

For the comparison of practice groups (antisaccade, prosaccade, or fixation) across time points (pre- and post-test), we partition the brain into two, three, five, or seven clusters using K-means and PCA K-means clustering methods for each practice group and for each test session. Since the clustering results from the two methods are similar, we present the results using PCA K-means clustering methods with $K = 3$ for each test session and for each practice group in this subsection. The complete results for antisaccade, prosaccade, and fixation task groups can be found in Appendix A, B and C, respectively.

Figure 3.30 shows the average time series plots in each cluster indicated by cluster 1 in green, cluster 2 in blue, and cluster 3 in red at the pre- and post-test for the antisaccade (Figures 3.30 (a) and (b)), prosaccade (Figures 3.30 (c) and (d)), and fixation groups (Figures 3.30 (e) and (f)). For each plot, at least one cluster, shown in green, closely follows the shape of the stimulus, suggesting the presence of antisaccade task-related voxels. Note that the amplitude of this primary cluster looks attenuated at the post-test compared to the pre-test, especially for the antisaccade group. This could be due to the fact that the participants in the antisaccade group practice the same task for one week and thus become less responsive to the stimulus.

The relevant areas to the antisaccade task are known as prefrontal cortex (PFC), frontal eye fields (FEFs), and striatum, which can be observed in superior slices (Camchong et al., 2008; Dyckman et al., 2007). To make the comparison more relevant to the task, we present the clustered brain maps in Figure 3.31 for four superior slices (slices 29 – 32). The green, blue, and red areas in the clustered brain maps correspond the cluster 1, 2, and 3, respectively, in Figures 3.30. Note that the blue area decreases at the post-test (denoted as Post) compared to the pre-test (denoted as Pre) for the antisaccade group if we compare the first row with the second row, where the blue area corresponds to cluster 2 in Figure 3.30. But the green area (cluster 1 in Figure 3.30) increases for all three groups. The red area (cluster 3 in Figure 3.30), which is a noise region, decreases at the post-test compared to the pre-test for the antisaccade group.



(a) Antisaccade at the pre-test

(b) Antisaccade at the post-test

(c) Prosaccade at the pre-test

(d) Prosaccade at the post-test

(e) Fixation at the pre-test

(f) Fixation at the post-test

Figure 3.30: Average time series plots for each practice group and for the pre- (left) and post-test(right). These results are produced by PCA K-means clustering for K=3. Three clusters (cluster 1 is shown in green, cluster 2 is shown in blue and cluster 3 is shown in red) for the pre- and post-test for the antisaccade (a, b), prosaccade (c, d), and fixation groups (e, f). Dashed black lines represent the stimulus timing. The x-axes indicate scan time points, and the y-axes indicate the voxel's detrended BOLD signal.

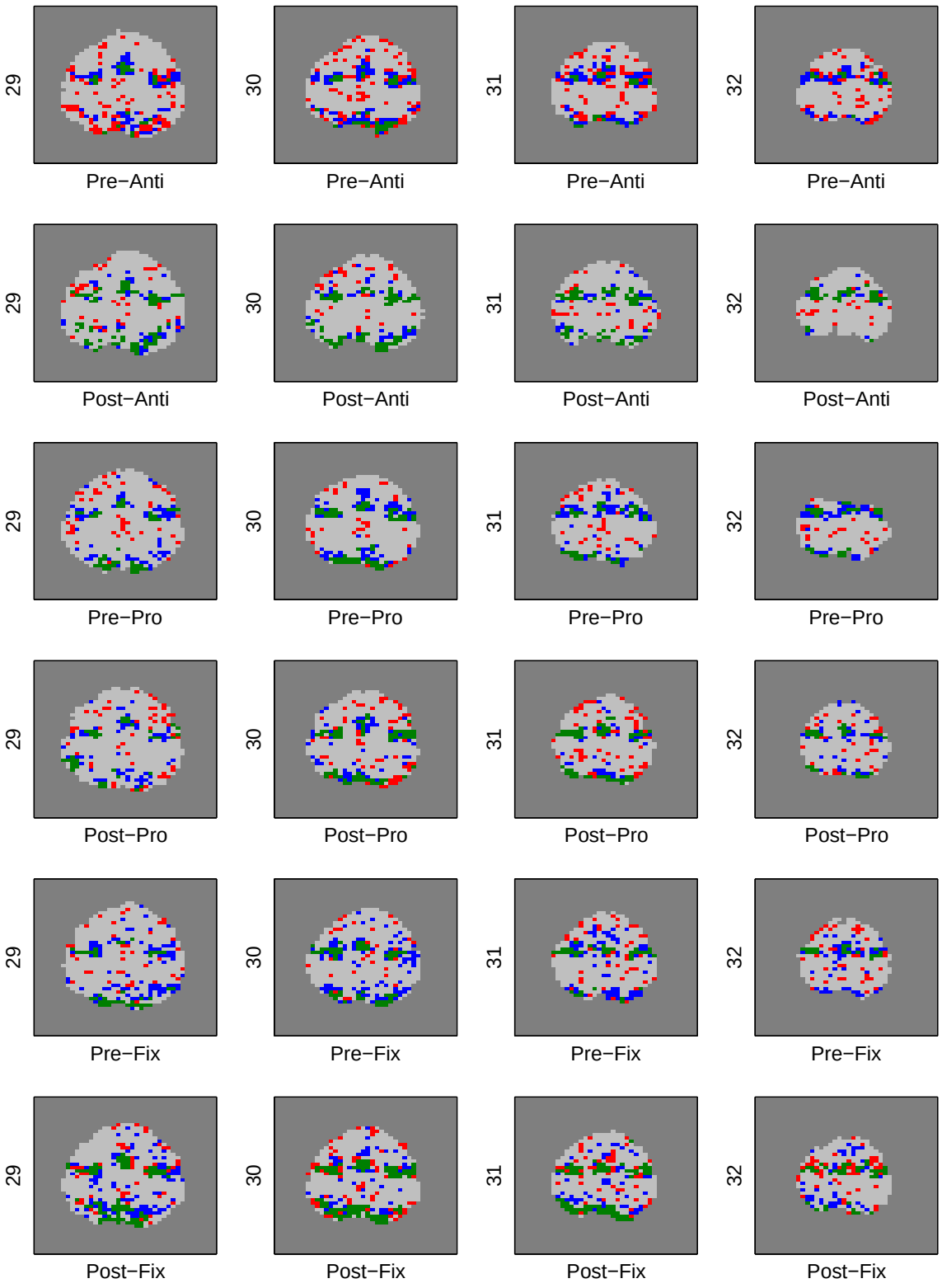
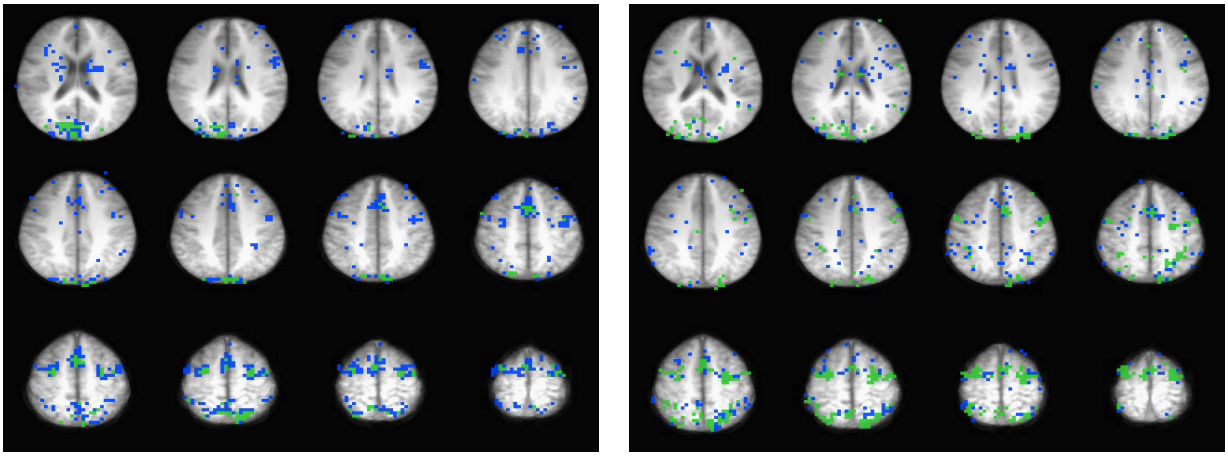


Figure 3.31: The clustered brain maps for the comparison of practice groups (antisaccade, prosaccade or fixation) across time points (pre- and post-test). These results are produced by PCA K-means clustering for K=3. The y labels indicate the slice number and the x labels indicate the task group and the test session, where the pre- and post-test are denoted by Pre and Post in x labels. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

To evaluate where in the brain the task-related clusters are located, and to determine whether the amplitude changes between the pre- and post-test occur in regions associated with saccadic performance, the voxels making up the two clusters showing the greatest task-related activation (green and blue in Figure 3.30) are plotted on the brain map. Figure 3.32 illustrates the voxels comprising these two clusters for the antisaccade group, which we refer to as activated voxels given their task-related BOLD activation pattern. The antisaccade group shows clear identification of the well known circuitry supporting saccadic performance that includes frontal eye fields (FEF), supplementary eye fields (SEF), and posterior parietal cortex (PPC).



(a) Pre-test (b) Post-test

Figure 3.32: Brain maps of the pre- (left) and post-test (right) for the antisaccade practice group with two primary clusters. Voxels shown in green represent cluster 1 and those shown in blue represent cluster 2. The slices are arranged starting from inferior (slice 21, upper left) to superior (slice 32, lower right).

Next we consider in more detail the four superior slices (slices 29 – 32) in order to investigate changes from the pre- to post-test across different practice groups. The locations of the activated voxels at each time point are colored to identify those showing antisaccade-related activity at the pre-test only (red), at both the pre- and post-test (orange) or at the post-test only (yellow) in Figure 3.33. The numbers of voxels in each category in the four slices are tabulated in Table 3.1. Note that the number of voxels activated at the post-test (in bold) is smaller for the antisaccade practice group (i.e., more red and less yellow) compared to the other groups. This is consistent with the observation in Figure 3.30(b) that

the box-car shaped time series is attenuated for the antisaccade at the post-test. Figures 3.33 (a) – (c) show the changes from the pre- to post-test by presenting the locations of the activated voxels. However, these results from Table 3.1 and Figure 3.33 do not display how the amplitude of voxels activated at both the pre- and post-test is changed.

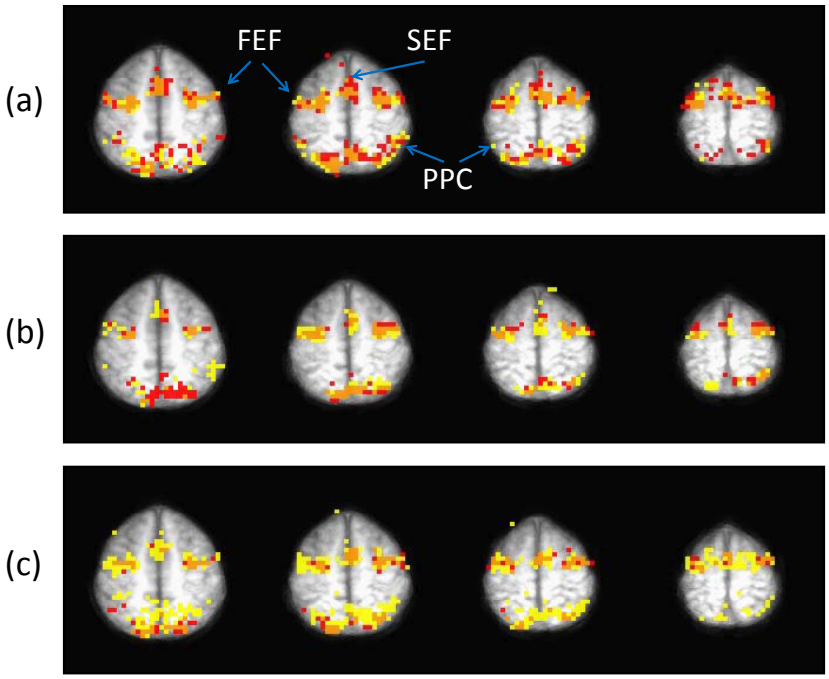


Figure 3.33: Overlay plots for (a) antisaccade, (b) prosaccade, and (c) fixation groups for four slices inferior to superior (slices 29 – 32). Voxels in red are activated at the pre-test but not at the post-test; those shown in orange are activated at both the pre- and post-test, and those shown in yellow are activated only at the post-test. Activation is observed in typical saccadic circuitry at this level of the brain which includes bilateral frontal eye field (FEF), supplementary eye field (SEF) at the midline and bilateral posterior parietal cortex (PPC).

Figure 3.34 shows the changes from the pre- to post-test for the voxels activated at both the pre- and post-test for the antisaccade practice group (orange voxels in Figure3.33 (a)). In the plot, solid blue and red lines indicate the pre- and post-test, respectively. Dashed black lines represent the stimulus timing. The x-axis indicates scan time points, and the y-axis indicates the voxel's detrended BOLD signal. The attenuation at the post-test within the regions activated at both the pre- and post-test is found for the antisaccade practice group.

	Pre-test only (Red)			Pre- and Post-test (Orange)			Post-test only (Yellow)		
Slice No.	Anti	Pro	Fix	Anti	Pro	Fix	Anti	Pro	Fix
29	43	41	16	51	16	31	**39**	33	92
30	57	15	16	61	46	41	**29**	34	83
31	38	16	9	43	22	33	**33**	40	63
32	34	16	6	24	18	14	**15**	22	54
Total	172	88	47	179	102	119	116	129	292

Table 3.1: Number of activated voxels by slice for the pre-test only (red), both the pre- and post-test (orange) and the post-test only (yellow) for the three practice groups: antisaccade ("Anti"), prosaccade ("Pro"), and fixation ("Fix") corresponding to Figure 3.33. The number of voxels activated at the post-test (in bold) is smaller for the antisaccade practice group compared to the other groups (i.e., more red and less yellow).

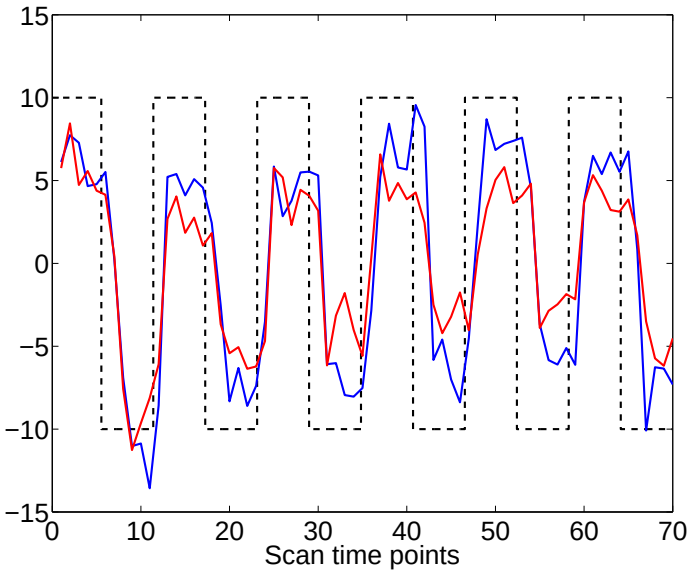
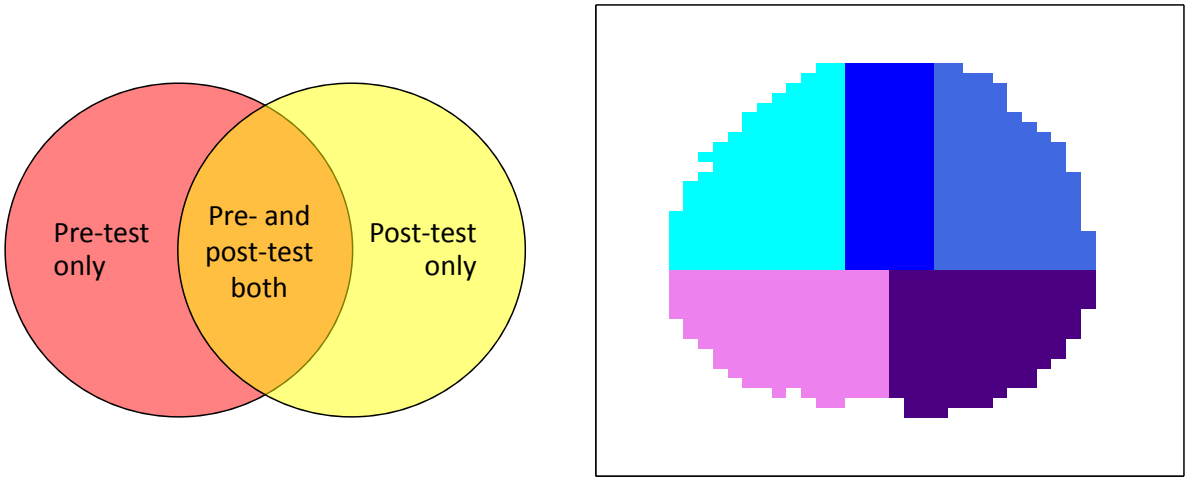


Figure 3.34: Average time series plots for the antisaccade group for four slices inferior to superior (slices 29 – 32). Solid blue and red lines indicate the pre- and post-test, respectively. Dashed black lines represent the stimulus timing. The x-axis indicates scan time points, and the y-axis indicates the voxel's detrended BOLD signal.

3.5.5 ANOVA RESULTS

In order to evaluate the visual impressions from the clustering analysis, we perform an ANOVA using the proportion of activated voxels based on Table 3.1. We define the proportion of activated voxels as the number of activated voxels in the pre- or post-test divided by the total number of activated voxels. For example, the proportion in the pre-test is defined as the number of activated voxels in the pre-test (red and orange regions in Figure 3.35(a)) divided by the number of total activated voxels (red, orange, and yellow regions in Figure 3.35(a)). We partition the brain into five regions: left frontal, middle frontal, right frontal, left posterior, and right posterior. See the partitioned brain map in Figure 3.35(b).

We use the proportion of activated voxels as the response variable, the day (pre- and post-test) and the practice condition (antisaccade, prosaccade, fixation) as factors, and the locations (left frontal, middle frontal, right frontal, left posterior, and right posterior brain) as a block variable.



(a) Voxel counts (b) Brain regions

Figure 3.35: (a) The Venn diagram shows the division of activated regions. The activated voxels are divided into three regions, which are pre-test only (red), pre- and post-test both (orange) and post-test only (yellow). (b) The brain is divided into five regions: the left frontal (cyan), middle frontal (blue), right frontal (royal blue), left posterior (violet) and right posterior (indigo). These five locations serve as a block variable in ANOVA.

Results of the ANOVA are shown in Table 3.2. There is a main effect of day (p-value = 0.0268) showing more activated voxels at the post-test, but practice condition (p-value = 0.2811) and location (p-value = 0.1618) are not statistically significant. In addition, there

is a significant interaction between day and practice condition (p-value < 0.0001). Post-hoc testing shows that the antisaccade group has fewer activated voxels at the post-test than at the pre-test, while the other groups have more. Figure 3.36 shows the interaction plot between day and practice condition. Only the antisaccade group has a decreased proportion of activation at the post-test.

Source	DF	Seq SS	Adj MS	F	p-value
day	1	0.13373964	0.13373964	5.06	**0.0268**
task	2	0.06799930	0.03399965	1.29	0.2811
location	4	0.17724253	0.04431063	1.68	0.1618
day*task	2	2.66801521	1.33400761	50.44	**<.0001**
task*location	8	0.12818030	0.01602254	0.61	0.7710
day*location	4	0.12058587	0.03014647	1.14	0.3423
Error	98	2.59163243	0.02644523		
Total	119	5.88739528			

Table 3.2: ANOVA using the proportion of activated voxels based on Table 3.1.

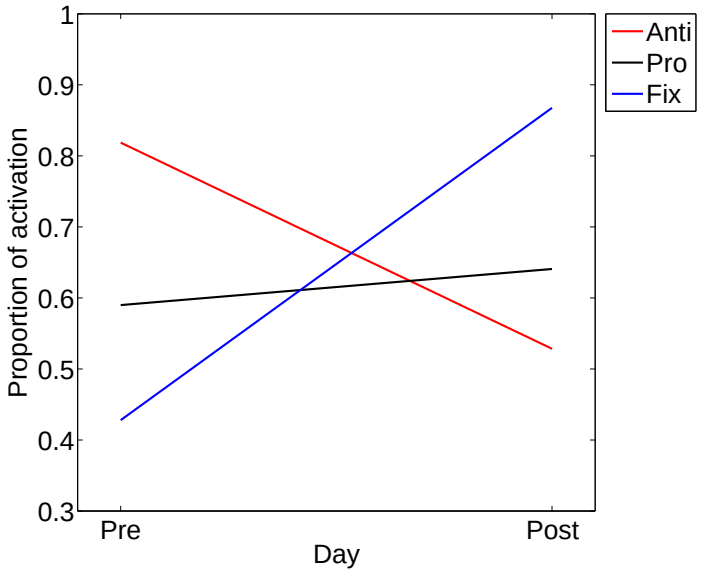


Figure 3.36: Interaction plot between day (pre- and post-test) and proportion of activated voxels by the practice group: antisaccade in red, prosaccade in black and fixation in blue.

From this analysis, we see that only the antisaccade practice group shows decreased activity from the pre-test to post-test. This pattern may be associated with increased effi-

ciency such that fewer neural resources are necessary to support the response due to task-specific practice. The two groups that practice test-irrelevant tasks (prosaccade and fixation) show increased activation at the post-test. This pattern may be indicative of increased difficulty with antisaccade performance at the post-test. Participants in the prosaccade group practice following a target as quickly as possible, which is the equivalent of priming the error response during an antisaccade test. Participants in the fixation practice group keeping their eyes stable on the target, which may have strengthened the neural circuitry that inhibits rapid responses away from the fixation target (Leigh and Zee, 2006).

Chapter 4

Simulation Study[1]

The temporal patterns of activation are assumed to be established in a spatially coherent fashion such that clustering, as a means of identifying brain systems, extracts the main temporal patterns and partitions the brain by grouping similarly behaved functions together. In order to detect and characterize the coherent patterns of activity, we propose a wavelet-based clustering method to aggregate voxels into a small number of clusters based on the patterns of voxel time courses in Chapter 3. Our clustering method successfully finds brain activation changes across two fMRI sessions associated with practice-related cognitive control during eye movement tasks. In this chapter, we attempt to empirically evaluate the performance of our wavelet-based clustering approach using simulated data. For the exact evaluation, we generate simulated data that mimic task-related BOLD signals by adding artificial fMRI signals to a resting-state brain, and then apply the proposed clustering approach. In particular, we compare the clustering results in the original time domain and in the wavelet domain using K-means and PCA K-means clustering methods. We describe our simulation design in Section 4.1 and provide the results in Section 4.2.

4.1 Simulation design

Multiple regions are connected functionally, and identification of this functional connectivity is often achieved by exploring resting-state connectivity. The term resting-state data refers to data from subjects who do not perform any experimental task while they lie in the MRI scanner (for more details, refer to Huettel et al., 2009; Biswal et al., 1995, and references therein). Park et al. (2010) perform a multiscale analysis using resting-state fMRI data. They collect the data on a single subject by asking the subject not to execute any task while he lies in the MR scanner for approximately one hour. We conduct our simulation study using

[1]The results presented in this chapter have been published: Lee, J. and Park, C. (2012), Assessment of a Wavelet-Based Clustering Analysis on FMRI Time Series, Quantitative Bio-Science, 31, 113–120.

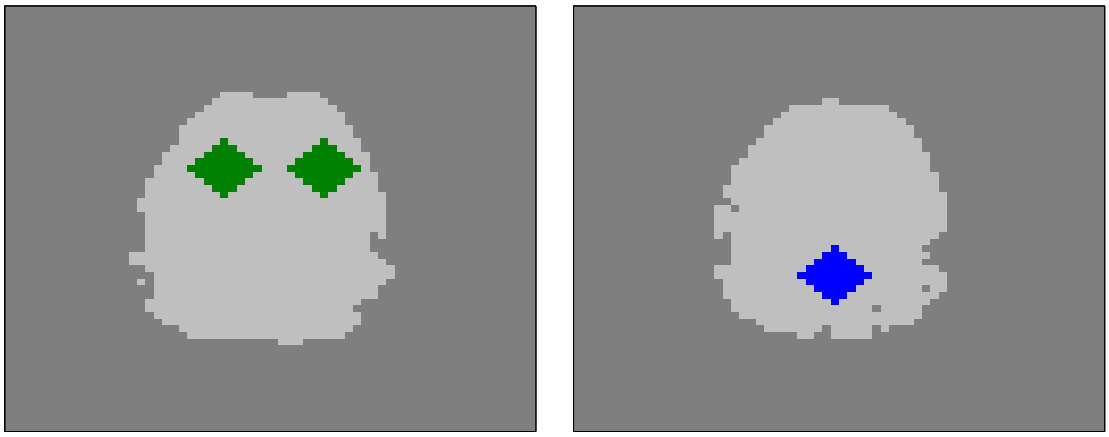
their resting-state fMRI data. The original data comprise 1500 scan time points with three slices of size 64 × 64 each. In our simulation study, we use only two slices and randomly select five sequences of length 64 from the original 1500 scan time points. We treat these five random samples as if they were from five subjects.

On these two slices, we create two types of activated clusters in the following way. We select two areas (green in Figure 4.1(a)) from slice 1 for cluster 1 and one area (blue in Figure 4.1(b)) from slice 2 for cluster 2, which yields 82 voxels for cluster 1 and 41 voxels for cluster 2. Note that the number of voxels in cluster 1 is twice as many as the number of voxels in cluster 2. We mimic task-related BOLD signals in a real example by adding artificial box-car shaped signals (shown in Figure 4.2) to these two areas. These two signals have the same pattern, but the amplitude of Signal 1 (shown in a green dashed line and indicated by 'a') is twice that of Signal 2 (shown in a blue solid line and indicated by '0.5a'). We plant Signal 1 on the selected area for cluster 1 (green in Figure 4.1(a)), and Signal 2 on the selected area for cluster 2 (blue in Figure 4.1(b)).

We determine 'a', the amplitude of Signal 1, by controlling the signal to noise ratio (SNR), which can be defined as the ratio of the task-related and noise variability. In calculating SNR, we treat Signal 1, which has the highest amplitude, as *signal* and the noise voxels, which are filtered out from the adaptive pivotal thresholding test, as *noise*. We select $a = 47$ and 62, which corresponds to SNR=1 and 1.5, respectively.

In summary, the simulated data consist of five samples (mimicking subjects); each sample has two slices, all the voxels in each slice are from the resting-state data, and each voxel contains 64 time points. In order to create the activated regions, the box-car shaped artificial signal with the amplitude of 'a' ('0.5a') is added to the resting-state voxels for cluster 1 (cluster 2). Therefore, it would be more challenging to correctly identify cluster 2 than cluster 1.

In this simulation study, we follow the procedures of the real data analysis proposed in Chapter 3 as closely as possible. First, we select five random samples, each of which has 64 time scan points. Second, we remove linear and quadratic trends at each voxel. Third, we plant the artificial task-related signals on the pre-selected areas. Fourth, we aggregate the five samples using a median value at each voxel and each time point separately, which yields the aggregated data. Fifth, we apply the adaptive pivotal thresholding test based on wavelet coefficients to filter out inactivated voxels. Sixth, we partition the remaining non-inactive voxels into three clusters using the K-means or PCA K-means clustering algorithm with



(a) Slice 1 (b) Slice 2

Figure 4.1: True cluster locations in the brain. (a) Two green areas are selected from slice 1 for cluster 1 and (b) one blue area from slice 2 for cluster 2. The volume of the selected area for cluster 1 is twice that of the area for cluster 2.

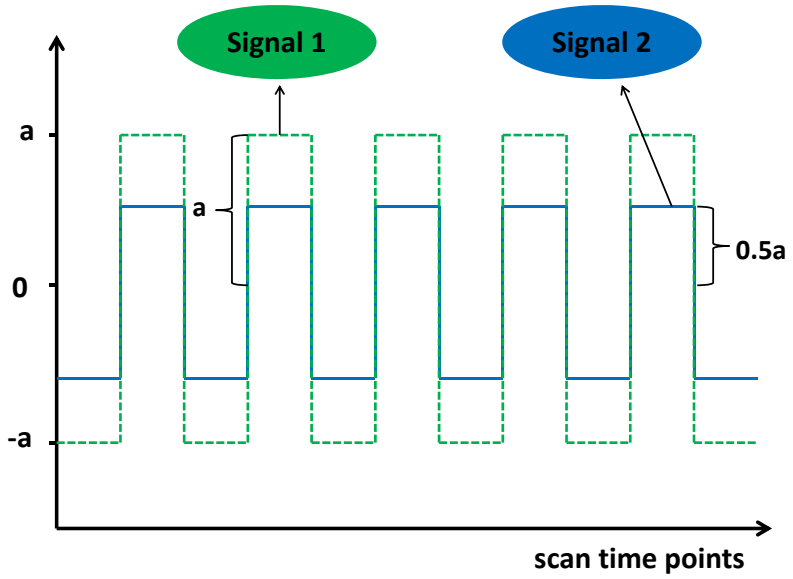


Figure 4.2: Two task-related signals sharing the same pattern but with different amplitudes are created. Signal 1 (a green dashed line) is implanted in the green area in slice 1 for cluster 1 in Figure 4.1(a). Signal 2 (a blue solid line) is implanted in the blue area in slice 2 for cluster 2 in Figure 4.1(b).

$K = 3$. In the sixth step, we have two clustering results using the detrended raw voxel values and the thresholded wavelet coefficients for each clustering algorithm. Finally, we repeat this simulation procedure 100 times. We choose $K = 3$, because some noise voxels might survive even after the adaptive pivotal test. We note that in the fifth step, all the voxels in both clusters 1 and 2 remain after the adaptive pivotal test, and hence they are included in the actual clustering in the sixth step. All the simulation steps are summarized in Figure 4.3.

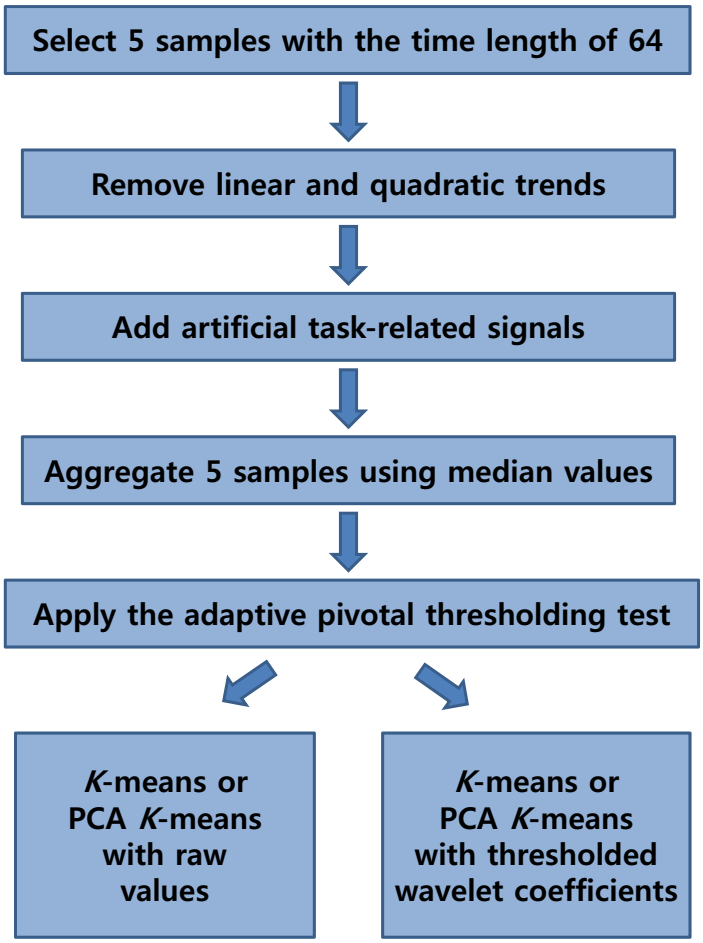


Figure 4.3: Simulation scheme.

4.2 Simulation results

In this section, we present the 100 repetitions of the simulation results using K-means and PCA K-means clustering algorithms for the two different SNRs. We compare the clustering results using the raw data and thresholded wavelet coefficients data. Table 4.1 shows an example of a clustering result when a clustering algorithm works perfectly. Note that cluster 1 is composed of 82 voxels and cluster 2 is made up of 41 voxels out of 8192 total voxels. Thus, there are four types of misclassification: voxels that are falsely assigned to cluster 1 or 2, and voxels that are falsely not assigned to cluster 1 or 2.

Cluster 1			
		Origin	
		C1	Not C1
Allocation	C1	82	0
	Not C1	0	8110

Cluster 2			
		Origin	
		C2	Not C2
Allocation	C2	41	0
	Not C2	0	8151

Table 4.1: Correct allocation. *Origin* indicates the true cluster labels of voxels before clustering and *Allocation* indicates the assignments of voxels after clustering. *C1* and *C2* stand for cluster 1 and 2, respectively. Cluster 1 is composed of 82 voxels and cluster 2, of 41 voxels out of 8192 total voxels.

Table 4.2 summarizes the clustering results using the raw data for SNR=1.0 and SNR=1.5. For each cell, the average counts (over the 100 repetitions) of the classified voxels are presented along with standard errors in parentheses. *Origin* indicates the true cluster of voxels before partitioning, and *Allocation* indicates the assignment of voxels after partitioning. *C1* and *C2* stand for cluster 1 and 2, respectively. Comparing the average counts of misclassified voxels, one can see that K-means clustering performs slightly better for SNR=1.0, while PCA K-means clustering produces more accurate results for SNR=1.5. However, the overall misclassification average counts are rather high, which is evident later in the comparison with the clustering results using the thresholded wavelet coefficients.

For illustration purposes, we present one of the 100 clustering results using the raw data with SNR=1.0 (SNR=1.5) in Figure 4.4 (Figure 4.5, respectively). In the plots, the clustering

results of K-means and PCA K-means methods with K=3 are displayed in the left and right panels, respectively.

For SNR=1.0 (Figure 4.4), the top panel draws the average time course of three clusters where green corresponds to cluster 1, blue to cluster 2, and red to the noise cluster. As can be seen, K-means correctly captures the shape and amplitude of each cluster, but PCA K-means fails to do so for cluster 2. The brain maps in the middle (slice 1) and bottom panels (slice 2) show the three assigned clusters using the same color scheme as in Figures 4.4(a) and 4.4(b). According to the clustered maps, K-means manages to identify true locations of clusters 1 and 2 with only a few errors, but PCA K-means completely misclassifies cluster 2 (*Origin*) as cluster 1 (*Allocation*) as can be seen in Figure 4.4(f).

For SNR=1.5 (Figure 4.5), the time series plots from both clustering methods (Figures 4.5(a) and 4.5(b)) depict an incorrect shape for cluster 2 (blue), and in the clustered maps (Figures 4.5(e) and 4.5(f)), they both classify cluster 2 (*Origin*) as cluster 1 (*Allocation*). This clearly shows the limitation of the clustering approach using the raw data.

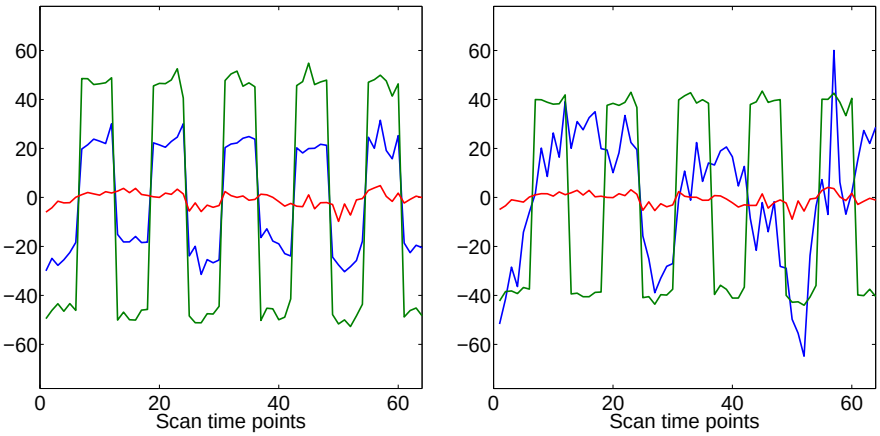
SNR=1.0 ($a = 47$)

K-means				PCA K-means			
		Origin				Origin	
		C1	Not C1			C1	Not C1
Allocation	C1	81.11 (0.56)	14.29 (1.96)	Allocation	C1	81.89 (0.11)	16.71 (2.01)
	Not C1	0.89 (0.56)	8095.71 (1.96)		Not C1	0.11 (0.11)	8093.29 (2.01)
		Origin				Origin	
		C2	Not C2			C2	Not C2
Allocation	C2	25.45 (1.99)	36.58 (7.43)	Allocation	C2	23.83 (2.02)	45.85 (7.74)
	Not C2	15.55 (1.99)	8114.42 (7.43)		Not C2	17.17 (2.02)	8105.15 (7.74)

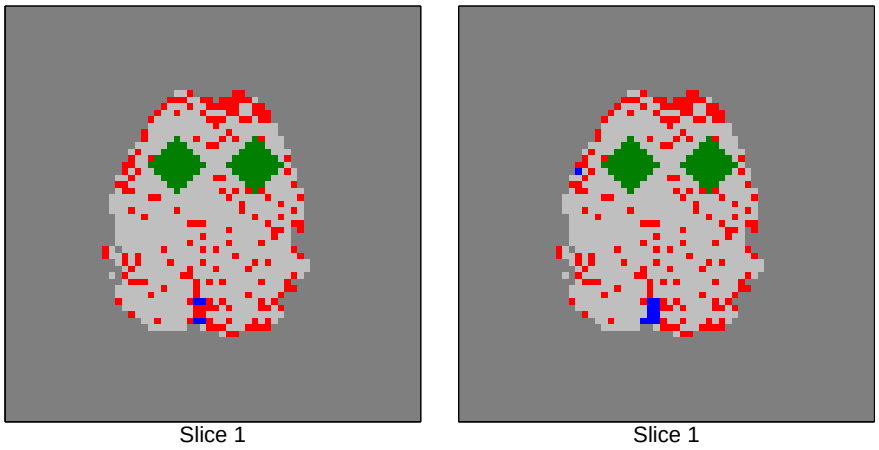
SNR=1.5 ($a = 62$)

K-means				PCA K-means			
		Origin				Origin	
		C1	Not C1			C1	Not C1
Allocation	C1	81.50 (0.22)	18.48 (2.05)	Allocation	C1	82.00 (0.00)	10.65 (1.81)
	Not C1	0.50 (0.22)	8091.52 (2.05)		Not C1	0.00 (0.00)	8099.35 (1.81)
		Origin				Origin	
		C2	Not C2			C2	Not C2
Allocation	C2	20.48 (2.06)	50.75 (7.90)	Allocation	C2	30.35 (1.81)	30.91 (7.16)
	Not C2	20.52 (2.06)	8100.25 (7.90)		Not C2	10.65 (1.81)	8120.09 (7.16)

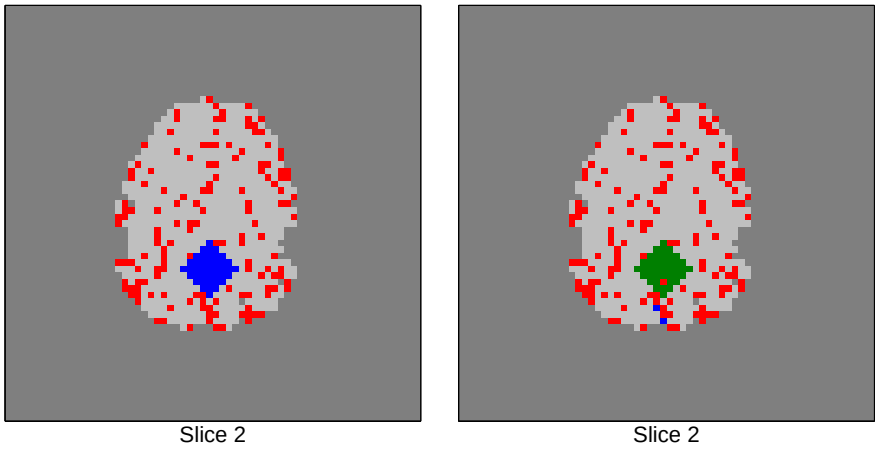
Table 4.2: Simulation results using the raw simulated data with two SNR values. Both K-means and PCA K-means clustering methods are applied. *Origin* indicates the true cluster labels of voxels before clustering, and *Allocation* indicates the assignments of voxels after clustering. *C1* and *C2* stand for cluster 1 and 2, respectively. For each cell, the average counts of the classified voxels are presented along with standard errors in parentheses. The values $a = 47$ and 62, the amplitude of the artificial task-related signal, correspond to SNR=1 and 1.5.



(a) K-means (b) PCA K-means

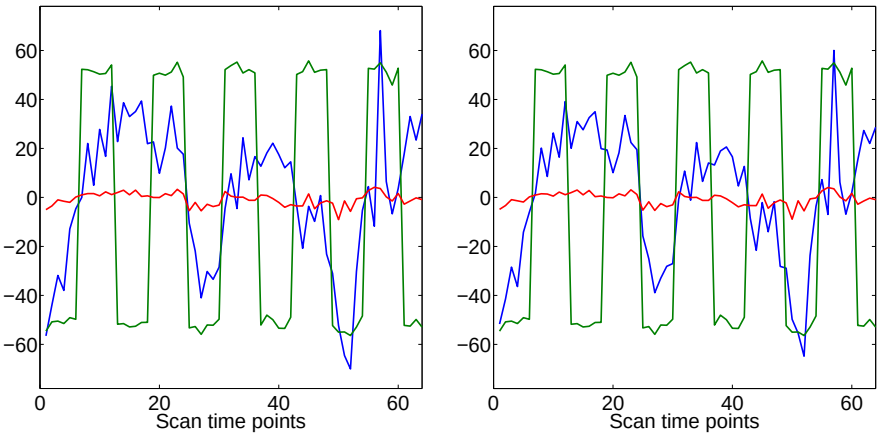


(c) K-means (d) PCA K-means

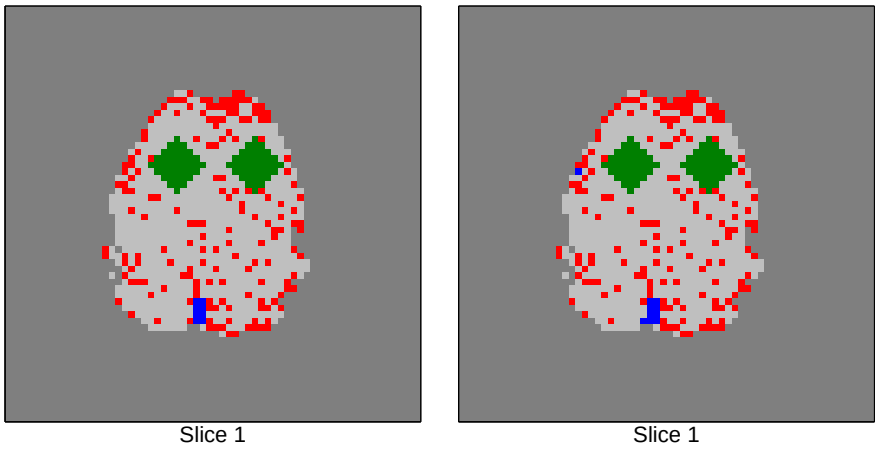


(e) K-means (f) PCA K-means

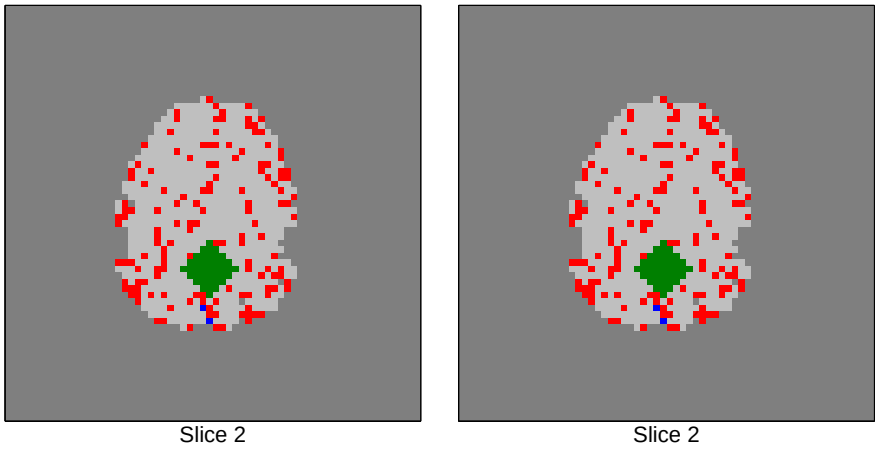
Figure 4.4: Example of average time series plots of three clusters and the clustered brain maps for SNR=1.0 (a=47) using the raw simulated data. Clusters 1 and 2 are indicated in green and blue, respectively. The red voxels are considered as noise as their average time series is rather flat.



(a) K-means (b) PCA K-means



(c) K-means (d) PCA K-means



(e) K-means (f) PCA K-means

Figure 4.5: Example of average time series plots for three clusters and the clustered brain maps for SNR=1.5 (a=62) using the raw simulated data. Clusters 1 and 2 are indicated in green and blue, respectively. The red voxels are considered as noise as their average time series is rather flat.

The simulation results using the thresholded wavelet coefficients are summarized in Table 4.3. For both SNR values, PCA K-means yields far superior results to K-means. All of the average counts of misclassified voxels that are falsely assigned to clusters 1 and 2 for K-means are at least seven times higher than those for PCA K-means. More importantly, in the comparison with the results from the raw data in Table 4.2, both clustering methods produce significantly smaller misclassification average counts for both SNR values when they use the thresholded wavelet coefficients. This demonstrates the effectiveness of the wavelet-based clustering procedure.

For illustration purposes, we again present one of the 100 clustering results using the thresholded wavelet coefficients for both SNR=1.0 (Figure 4.6) and SNR=1.5 (Figure 4.7). It can be seen from the average time series plots (Figures 4.6(a) and 4.6(b) for SNR=1.0, and Figures 4.7(a) and 4.7(b) for SNR=1.5) that both K-means and PCA K-means correctly capture the shape and amplitude of each cluster. It is evident that the clustering algorithms with the thresholded wavelet coefficients yield accurate results in the clustered brain maps. For example, all of the voxels in cluster 1 on slice 1 are successfully classified as cluster 1 in both figures. Likewise, all of the voxels in cluster 2 on slice 2 are also successfully classified as cluster 2 in both figures. This again shows that clustering analysis of fMRI data is more effective in the wavelet domain than in the time domain.

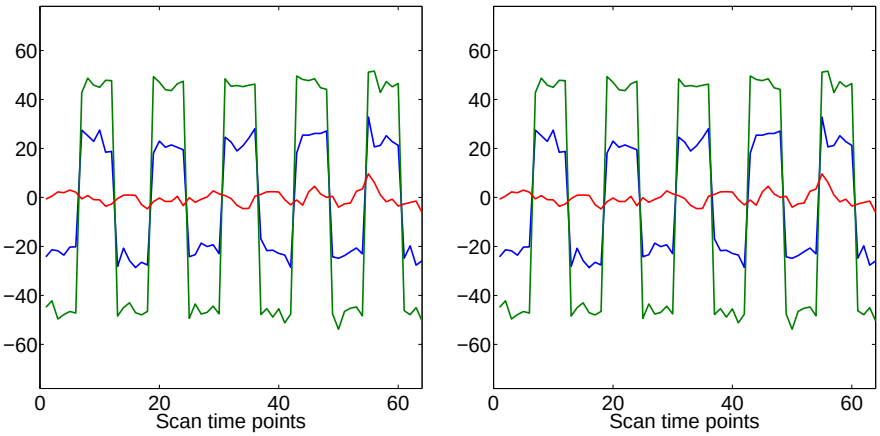
SNR=1.0 ($a = 47$)

K-means				PCA K-means			
		Origin				Origin	
		C1	Not C1			C1	Not C1
Allocation	C1	80.33 (0.58)	4.74 (1.13)	Allocation	C1	81.77 (0.06)	0.00 (0.00)
	Not C1	1.67 (0.58)	8105.26 (1.13)		Not C1	0.23 (0.06)	8110.00 (0.00)
		Origin				Origin	
		C2	Not C2			C2	Not C2
Allocation	C2	33.72 (1.42)	21.97 (4.77)	Allocation	C2	40.84 (0.05)	0.39 (0.07)
	Not C2	7.28 (1.42)	8129.03 (4.77)		Not C2	0.16 (0.05)	8150.61 (0.07)

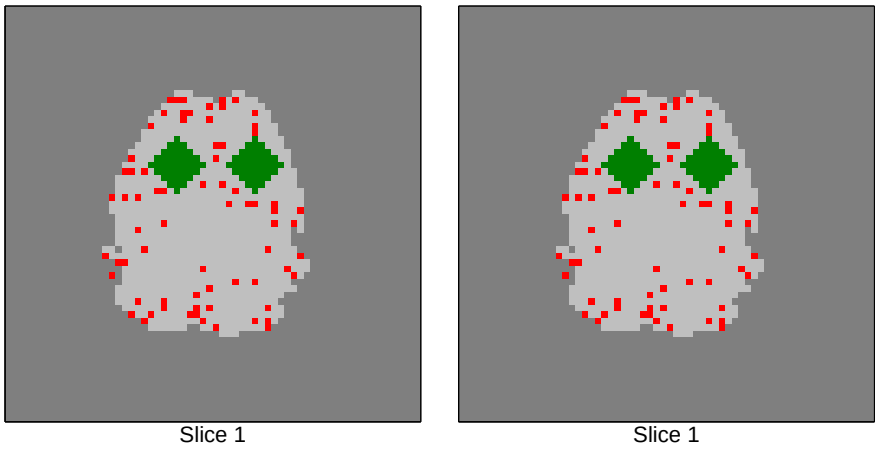
SNR=1.5 ($a = 62$)

K-means				PCA K-means			
		Origin				Origin	
		C1	Not C1			C1	Not C1
Allocation	C1	80.96 (0.60)	10.76 (1.78)	Allocation	C1	81.99 (0.01)	3.16 (1.08)
	Not C1	1.04 (0.60)	8099.24 (1.78)		Not C1	0.01 (0.01)	8106.84 (1.08)
		Origin				Origin	
		C2	Not C2			C2	Not C2
Allocation	C2	29.01 (1.84)	32.47 (6.13)	Allocation	C2	37.81 (1.09)	11.16 (4.28)
	Not C2	11.99 (1.84)	8118.53 (6.13)		Not C2	3.19 (1.09)	8139.84 (4.28)

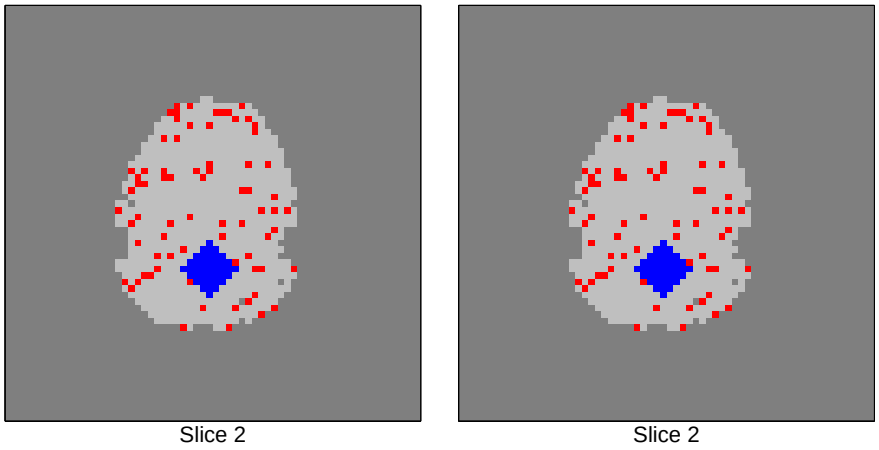
Table 4.3: Simulation results using the thresholded wavelet coefficients with two SNR values. Both K-means and PCA K-means clustering methods are applied. *Origin* indicates the true cluster labels of voxels before clustering, and *Allocation* indicates the assignments of voxels after clustering. *C1* and *C2* stand for cluster 1 and 2, respectively. For each cell, the average counts of the classified voxels are presented along with standard errors in parentheses. The values $a = 47$ and 62, the amplitude of the artificial task-related signal, correspond to SNR=1 and 1.5. The clustering results are improved compared to those using the raw data.



(a) K-means (b) PCA K-means

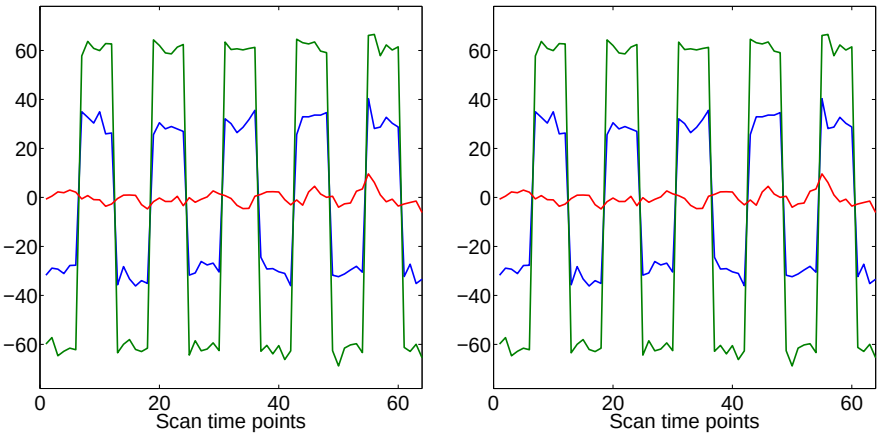


(c) K-means (d) PCA K-means

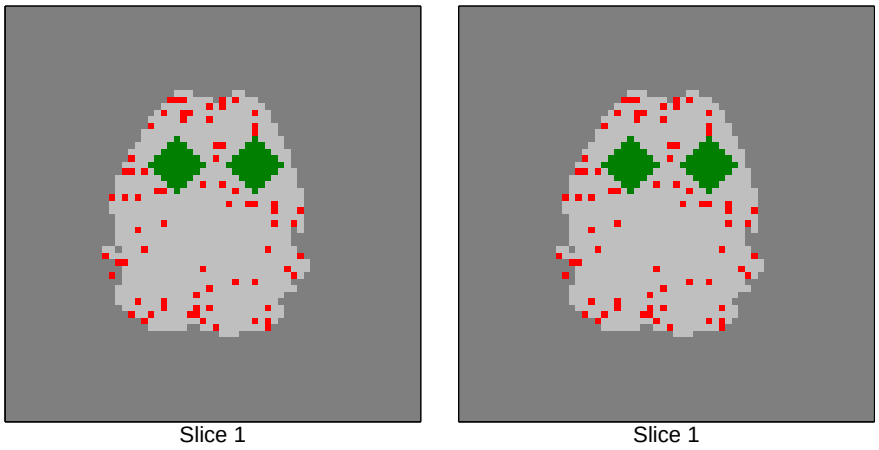


(e) K-means (f) PCA K-means

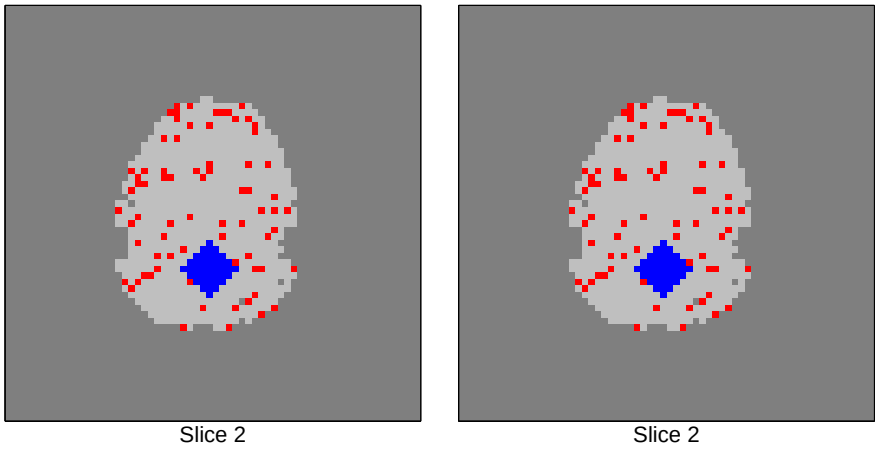
Figure 4.6: Example of average time series plots for three clusters and the clustered brain maps for SNR=1.0 (a=47) using the thresholded wavelet coefficients. Clusters 1 and 2 are indicated in green and blue, respectively. The red voxels are considered as noise among the voxels that remained after filtering out inactivated voxels using the adaptive pivotal thresholding test.



(a) K-means (b) PCA K-means



(c) K-means (d) PCA K-means



(e) K-means (f) PCA K-means

Figure 4.7: Example of average time series plots for three clusters and the clustered brain maps for SNR=1.5 (a=62) using the thresholded wavelet coefficients. Clusters 1 and 2 are indicated in green and blue, respectively. The red voxels are considered as noise among the voxels that remained after filtering out inactivated voxels using the adaptive pivotal thresholding test.

4.3 Conclusion

We perform this simulation study to evaluate the effectiveness of the wavelet-based clustering method proposed in Chapter 3. We generate simulated data by adding artificial fMRI signals to the real resting-state data and apply both K-means and PCA K-means clustering methods to the raw simulated data and the wavelet-transformed data separately to compare their clustering results. Due to the temporal correlation in fMRI time series, the clustering analysis using the raw data might yield improper clustering results because it is challenging to distinguish between deterministic patterns and dependence artifacts. In that sense, a wavelet-clustering approach is promising owing to the decorrelation property of the wavelet transform. In the comparison between K-means and PCA K-means clustering, PCA K-means is preferred when coupled with the thresholded wavelet coefficients. The simulation results show that the proposed clustering method is powerful and effective for fMRI data. Note that the results when using SNR=1.5 show fewer misclassification errors than the results when using SNR=1.0. Thus, the simulation study shows that the wavelet-based clustering approach provides more accurate clustering results, while PCA K-means clustering performs slightly better than K-means for the wavelet transformed data.

Chapter 5

Regions of Interest Analysis[1]

In Section 3.5.5 the ANOVA results reveal an attenuation at the post-test for the antisaccade practice group in the saccade circuitry constrained to the superior cortex (slices 29-32). This analysis, however, does not provide information about how specific neural regions within whole brain saccadic circuitry may have been differentially affected. Therefore, we conduct an analysis using 11 bilateral neural regions of interest (ROIs) previously identified using a similar task (Dyckman et al., 2007, and references therein). Note that the activated regions found by our clustering method are a subset of these ROIs. We open this chapter by introducing an ROI analysis (Section 5.1). To identify the specific ROIs that show attenuations at the post-test for the antisaccade group, we perform two statistical tests: a bootstrap resampling-based test (Section 5.2) and a mixed model (Section 5.3). And then this chapter ends with a conclusion (Section 5.4).

5.1 Introduction to ROI analysis

Our fMRI clustering method shown in Chapter 3 is based on a whole-brain voxelwise analysis. This is a common approach for exploring the brain areas related to specific cognitive processes using the whole brain. A voxelwise analysis, however, does not permit us to examine specific statistical hypotheses about the subregions of the brain. If brain regions involved in specific cognitive processes are detected in the primary research, then the research can be developed to investigate the detected brain regions or the specific task-related brain area, which is known as regions of interest (ROIs) analysis. Thus, an ROI analysis begins at the earlier discovery of the task-related areas within a brain, and extends a voxelwise analysis. Huettel et al. (2009, p. 363) clarify a difference between the inquiries invoked in a voxelwise analysis

[1]Part of the results presented in this chapter have been published: Lee, J., Park, C., Dyckman, K. A., Lazar, N. A., Austin, B. P., Li, Q., and McDowell, J. E. (2013), Practice-related Changes in Neural Circuitry Supporting Eye Movements Investigated via Wavelet-based Clustering Analysis, *Human Brain Mapping*, 34(9):2276-2291.

and in an ROI analysis. The inquiries invoked in a voxelwise analysis are condensed into "What brain regions evince a particular pattern of fMRI activation?" while an ROI analysis asks, "What pattern of activation occurs in a particular brain region?"

ROIs within a brain are generally established by two methods: an anatomical and a functional criterion (Poldrack, 2007). With an anatomical criterion, ROIs are selected visually based on the structure of the brain by constructing the borders of particular brain regions and then by including all voxels therein. With a functional criterion, only voxels that are activated by a particular stimulus are selected for analysis. One common method to find the voxels responding to a particular stimulus is to use a localizer task. A localizer task is a simple fMRI experiment to identify a set of voxels that are known to be functionally connected to a particular stimulus, in preparation for subsidiary analyses of that region in different patterns.

An ROI approach has both advantages and disadvantages in comparison to a voxelwise approach. It has two significant advantages: it uses relatively fewer voxels than a voxelwise analysis and increases the signal to noise ratio (SNR). One of the difficulties in understanding fMRI data through a voxelwise analysis is its huge data size. In our data, one slice has 1920 voxels (after being transferred to Talairach atlas), and there are 34 slices, which yields 65,280 voxels in total. Even if we choose the slices that show the entire shape of the brain and apply masking, there are still 14,743 $\sim$ 15,319 voxels before starting analysis. This requires us to address the need to correct the multiple testing problem and adjust the false discovery rate (Benjamini and Hochberg, 1995) due to a large number of tests in removal of the white noise area. Since the number of voxels used in an ROI analysis is much smaller than the number used in a voxelwise analysis, it reduces the burden of multiple comparisons remarkably. In our case, the number of total voxels pre-established in ROIs is 616 which is about 4 % of 15,319 voxels, the number of voxels remaining after we mask, and 0.94% of 65,280 voxels, the number of voxels before we start this analysis. Second, in an ROI analysis SNR is found to increase over a voxelwise analysis because each ROI combines data from many voxels, to the extent that the ROI is functionally connected (Huettel et al., 2009).

An ROI approach also has disadvantages compared to a voxelwise approach. First, establishing ROIs based on an anatomical criterion is difficult whether one uses automated programs or draw manually. There is a possibility of a discrepancy between anatomical and functional regions of the brain. In particular, complex tasks which intersect many anatomical brain regions, even with perfect mapping between anatomical and functional regions,

have a potential problem of connecting anatomical areas to functional regions. Moreover, a single anatomical area may have multiple functional roles. Second, Huettel et al. (2009) indicate that establishing ROIs based on a functional criterion is difficult when the boundaries of a (functionally significant) brain region cannot be readily identified by anatomical landmarks. A big improvement has been shown in automated programs used to divide a brain anatomically, but there are still problems, especially variations in shape and size between subjects. When ROIs are drawn manually, there is variation between evaluators, and good training is required. Caution is needed in the universal use of an ROI analysis, and due to the variability in function within any anatomical region, ROI approaches to an fMRI analysis should be combined with a voxelwise analysis (Huettel et al., 2009).

Dyckman et al. (2007) report that the saccade-related neural circuitry and performance on saccade tasks are affected by context. They carry out an ROI analysis to examine the features of the relationship between runs in the significantly activated brain regions. Using the same data we use, Dyckman et al. (2007) identify 20 ROIs based on BOLD activations and previous fMRI studies of saccadic performance. Most of their ROIs are divided into left and right: for example, left lateral frontal eye field and right lateral frontal eye field. We collapse these into single ROIs; hence, we use eleven ROIs for the additional analysis. They are listed in Table 5.1.

Regions of interest	Abbreviation	No. of voxels
Supplementary eye field	SEF	36
Frontal eye field	FEF	113
Superior parietal lobe	SPL	63
Cuneus	Cuneus	59
Thalamus	Thalamus	56
Inferior parietal lobule	IPL	59
Left PFC	PFC-L	28
Right PFC	PFC-R	36
Striatum (basal ganglia)	Striatum	67
Middle occipital gyrus	MOG	64
Right inferior frontal cortex	IFC	35
Total		616

Table 5.1: Eleven bilateral neural regions of interest (ROIs). 20 ROIs are established by Dyckman et al. (2007) based on BOLD activations observed in our data and previous fMRI studies of saccadic performance. We collapse these left and right ROIs into single ones, resulting in eleven in all.

5.2 Bootstrap resampling-based test

In this section, we use a bootstrap approach to identify the specific ROIs that show the difference between the pre- and post-tests. Bootstrap is a resampling approach used to infer some aspect of a group (i.e., a population), such as the mean and confidence interval, on the basis of a random sample drawn from that group.

5.2.1 Bootstrap resampling approach

The bootstrap resampling approach was invented by Efron (1979) to make inferences about the aspect of population distribution on the basis of the corresponding aspect of its empirical distribution. The key idea of bootstrapping starts from the perspective that a population and a sample drawn from that population are analogous (Mooney and Duval, 1993), and this analogy can be maximized by repetitive random resampling with replacement from the sample. In other words, samples, as subsets of a population, represent the population, and the properties of sampling distribution can substitute for the properties of population distribution.

Since it is almost impossible to know the whole population, it is natural at times to make an inference about the population from the sample without making unreliable assumptions about the population, which is a nonparametric method. A bootstrapping approach is used to draw conclusions that depend solely on the observed data, without distributional assumptions about the population, on the basis of the belief that a set of observed data comes from an independent and identically distributed population. The difference between bootstrapping and traditional parametric methods is shown by the fact that traditional parametric methods rely on assumptions about population distributions to make an inference about the population parameters. However, those assumptions are not always met, or the analytic solutions based on parametric assumptions could be unavailable. The estimates based on the violated assumptions bring out results that are different from what they should be, and the yielded misconclusions may be inefficient, seriously biased, or even misleading. To avoid these risks, applying nonparametric methods can be a solution, and nonparametric bootstrapping is a good alternative. Mooney and Duval (1993, p. 7) illustrate that nonparametric approximations may well be as good as or better than parametric approximation in certain situations.

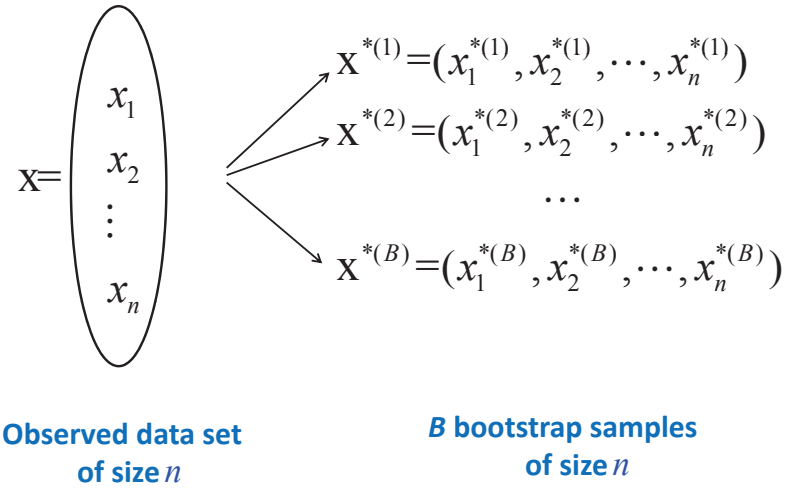


Figure 5.1: Bootstrap resampling method. The procedure starts with the observed data set $\mathbf{x}$ of size n. The key point is to draw resamples of size n with replacement from $\mathbf{x}$. The first resample drawn is denoted as $\mathbf{x}^{*(1)}$, the second resample drawn as $\mathbf{x}^{*(2)}$, and so on. By repeating this resampling procedure B times, B bootstrap samples are obtained, which are $\mathbf{x}^{*(1)}, \mathbf{x}^{*(2)}, \ldots, \mathbf{x}^{*(B)}$. The ith bootstrap sample is denoted as $\mathbf{x}^{*(i)} = (x_1^{*(i)}, x_2^{*(i)}, \ldots, x_n^{*(i)})$ where $i = 1, 2, \ldots, B$.

Bootstrapping involves drawing random samples repeatedly from the observed data with replacement, building the empirical distribution with the results of the resampling. For example, suppose that we are interested in estimating the median of a population using the observed sample. By employing a bootstrap approach in this case, we treat the observed data as a population and empirically build the sampling distribution of the sample median by a large number of random samples of size n (i.e., of equal size to the observed data points) with replacement, where n is the number of observed data points. For each resample, we calculate its median, creating a collection of resampled medians. Based on this, a sampling distribution for the median can be constructed.

Suppose we have a random sample $\mathbf{x}$ of size n from an unknown population distribution G and denote the observed data points of the sample $\mathbf{x}$ as $(x_1, x_2, \ldots, x_n)$. Then we can define a bootstrap sample $\mathbf{x}^*$ as a random sample of size n drawn from the empirical distribution $\hat{G}$, in which the randomly selected data points of the bootstrap sample $\mathbf{x}^*$ are denoted as $(x_1^*, x_2^*, \ldots, x_n^*)$. Since bootstrapping employs resampling with replacement, a bootstrap sample may have some data points more than once and others not at all. For example, suppose we have an observed sample of size 5 from unknown population G. Denote the observed data points of this original sample $\mathbf{x}$ as $(x_1, x_2, x_3, x_4, x_5)$; then the bootstrap sample $\mathbf{x}^*$ might be $\mathbf{x}^* = (x_2, x_1, x_1, x_5, x_5)$, which means $x_1^* = x_2, x_2^* = x_1, x_3^* = x_1, x_4^* = x_5, x_5^* = x_5$.

When we repeat this resampling procedure B times, we have B bootstrap samples, each of size n, denoted by $\mathbf{x}^{*(1)}, \mathbf{x}^{*(2)}, \ldots, \mathbf{x}^{*(B)}$, where $\mathbf{x}^{*(i)} = (x_1^{*(i)}, x_2^{*(i)}, \ldots, x_n^{*(i)})$ for the ith bootstrap sample and $i = 1, 2, \ldots, B$. See Figure 5.1. By drawing a large number of bootstrap resamples, we can build the empirical distribution $\hat{G}$ of a statistic of interest. Thus, the observed sample of size n drawn from unknown population distribution G is

$$G \to (x_1, x_2, \ldots, x_n),$$

and the bootstrap sample of size n drawn from the empirical distribution $\hat{G}$ is

$$\hat{G} \to (x_1^*, x_2^*, \ldots, x_n^*).$$

In case we are interested in estimating a parameter θ on the basis of a sample $\mathbf{x}$, which is a function of the unknown probability distribution G (i.e., $\theta = f(G)$), we evaluate an estimate $\hat{\theta} = f(\hat{G})$ from the sample $\mathbf{x}$. Sometimes it might not be a plug-in estimate, then generally denoted by $\hat{\theta} = s(\mathbf{x})$, where $s(\cdot)$ is a statistic of interest. For example, $s(\mathbf{x})$ might be a sample mean or a sample median. In a bootstrap approach, θ is estimated based on the empirical distribution $\hat{G}$, which is built with bootstrap resamples $\mathbf{x}^*$. Hence $\hat{\theta}^*$, an estimate for θ using a bootstrap sample $\mathbf{x}^*$, is calculated by

$$\hat{\theta}^* = s(\mathbf{x}^*),$$

where $s(\mathbf{x}^*)$ is obtained by applying the function $s(\cdot)$ to $\mathbf{x}^*$.

The procedures of the bootstrap resampling approach for estimating a parameter θ are listed below along with an example:

1. Draw a resample $\mathbf{x}^{*(1)} = (x_1^{*(1)}, x_2^{*(1)}, \ldots, x_n^{*(1)})$ of size n from the observed data set $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of size n.

2. Calculate a statistic of interest $s(\mathbf{x}^{*(1)})$ from $\mathbf{x}^{*(1)} = (x_1^{*(1)}, x_2^{*(1)}, \ldots, x_n^{*(1)})$. For example, in the case of estimating the population mean μ, $s(\mathbf{x}^{*(1)}) = \bar{\mathbf{x}}^{*(1)}$ which is $\frac{1}{n}\sum_{i=1}^{n} x_i^{*(1)}$.

3. Repeat Step 1 and 2 B times; $s(\mathbf{x}^{*(1)}), s(\mathbf{x}^{*(2)}), \ldots, s(\mathbf{x}^{*(B)})$ are obtained. In the estimating μ example, these are $\bar{\mathbf{x}}^{*(1)}, \bar{\mathbf{x}}^{*(2)}, \ldots, \bar{\mathbf{x}}^{*(B)}$.

4. Finally, estimate $\hat{\theta}^*$ by using $s(\mathbf{x}^{*(1)}, \mathbf{x}^{*(2)}, \ldots, \mathbf{x}^{*(B)})$. In the estimating μ example, $\hat{\mu}^* = \frac{1}{B}\sum_{j=1}^{B} \bar{\mathbf{x}}^{*(j)}$.

5.2.2 Confidence band based on Bootstrap resampling-based test

We perform a bootstrap resampling-based test (Efron, 1979) by building a 95% confidence band for the difference between the pre- and post-tests. As mentioned in Section 3.3, we exclude the first five scans from the analysis to allow for the magnetization's stabilization; consequently, 70 scans are used. As seen in Table 5.1, each ROI has a different size. On the basis of the idea that the voxels in the same ROI are similar to one another, we combine the voxels by taking an average of the voxels within the ROI. Thus, each subject in the given practice group has two average time course curves having 70 scan time points for each ROI, one for the pre-test and the other for the post-test. The procedures are as follows:

1. For a given ROI and a given practice group, obtain the difference between the pre-test and post-test time courses for each subject.

2. Obtain a bootstrap sample of the difference curves with the size corresponding to the number of subjects of the given practice group (e.g., $n = 12$ for the antisaccade practice group), and calculate the averaged curve (e.g, resulting in $\bar{\mathbf{x}}_{Anti}^{*(1)}$ for the antisaccade practice group).

3. Repeat Step 2 1,000 times. For instance, the obtained 1,000 averaged curves for the antisaccade practice group are $\bar{\mathbf{x}}_{Anti}^{*(1)}, \bar{\mathbf{x}}_{Anti}^{*(2)}, \ldots, \bar{\mathbf{x}}_{Anti}^{*(1000)}$.

4. Use the 1,000 averaged curves from Step 3 to construct a 95% confidence band.

5. Repeat Step 1– Step 4 for each practice group and each ROI.

The averaged curve $\bar{\mathbf{x}}^{*(1)}$ obtained at the second step in the procedures is equal to $s(\mathbf{x}^{*(1)})$ in the procedures of the bootstrap resampling approach described in Section 5.2.1.

For a given ROI, if the time courses of the pre- and post-tests show similar activation patterns, the confidence band for the difference curve is expected to include zero for most time points. If an attenuation occurs at the post-test, the confidence bands would miss zero at peaks (with a positive sign) and valleys (with a negative sign). The results of this bootstrap approach are displayed in Figures 5.2– 5.12. The time course for each ROI and each practice group is shown as a red line for the pre-test and a black line for the post-test. The time points where statistically significant attenuations occur at the post-test are identified by yellow bands. The converse of an attenuation at the post-test is also possible, and such an amplification effect is indicated by confidence bands that miss zero at peaks (with a negative sign) and valleys (with a positive sign). The time points that show statistically significant amplification at the post-test are identified by gray bands in the plots.

Application of this bootstrap approach suggests that the SEF, FEF, SPL, and cuneus show a strong attenuation at the post-test for the antisaccade practice group only. See Figures 5.2– 5.5. Many yellow bands appear in the average time series plots for these four ROIs at the antisaccade practice group, but the prosaccade and fixation practice groups have fewer yellow bands. Thus, visual inspection of these four ROIs shows that attenuations at the post-test are evident in these regions only for the antisaccade practice group.

Although the prosaccade and fixation practice groups show some periods of amplification on these four ROIs, there is little evidence of amplification in the antisaccade practice group at the post-test because it is rare to see gray bands for the antisaccade practice group on these four ROIs. Note that gray bands appear consistently for the fixation practice group for all ROIs. In particular, SPL shows only gray bands for the fixation practice group, which indicates a strong amplification at the post-test (see Figure 5.4).

It is interesting that the IPL shows strong amplifications at the post-test for all practice groups, and the average time series plot for the fixation practice group in particular contains many gray bands (see Figure 5.7). The fixation practice group also shows a strong amplification in the thalamus at the post-test (Figure 5.6). It is rather inconclusive in other ROIs.

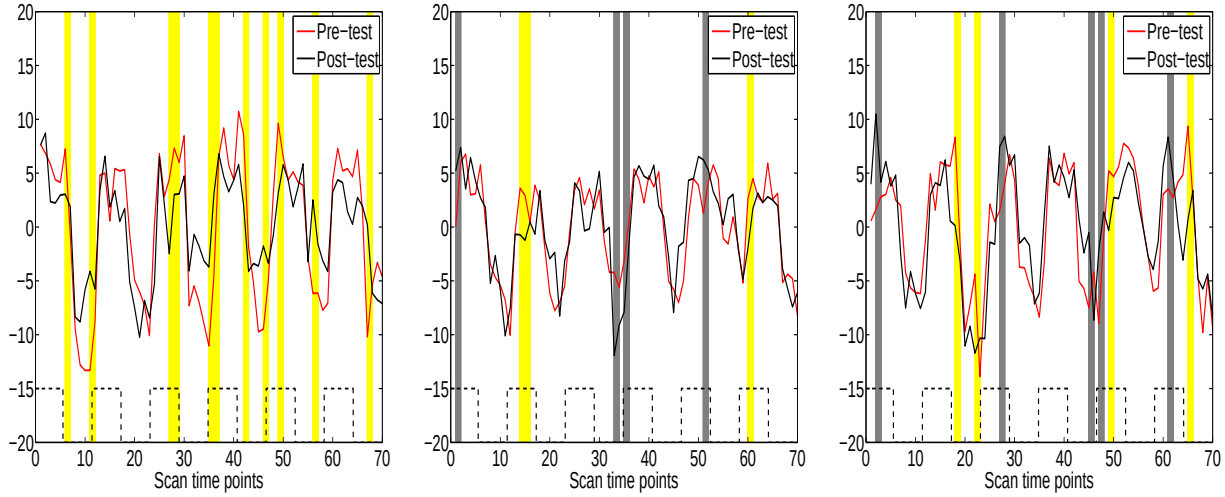
On the other hand, the prosaccade practice group shows amplifications and attenuations at the post-test with similar frequencies based on visual inspection. No notable points are found on the ROIs except on the SEF, FEF, SPL, cuneus, and IPL.

The findings from this ROI analysis are summarized in Table 5.2, which shows which trend in each ROI dominates at the post-test, either attenuation or amplification. The values

indicate differences between the proportion of time points (out of 70) that have significant attenuations (yellow bands) and the proportion of time points (out of 70) that have significant amplifications (gray bands) for each practice group at each ROI. The positive sign indicates that more attenuations occur at the post-test, whereas the negative sign indicates that more amplifications occur at the post-test. We highlight attenuations at the post-test if the difference values between the proportions are larger than 0.07, and amplifications at the post-test if the difference values between the proportions are smaller than -0.07.

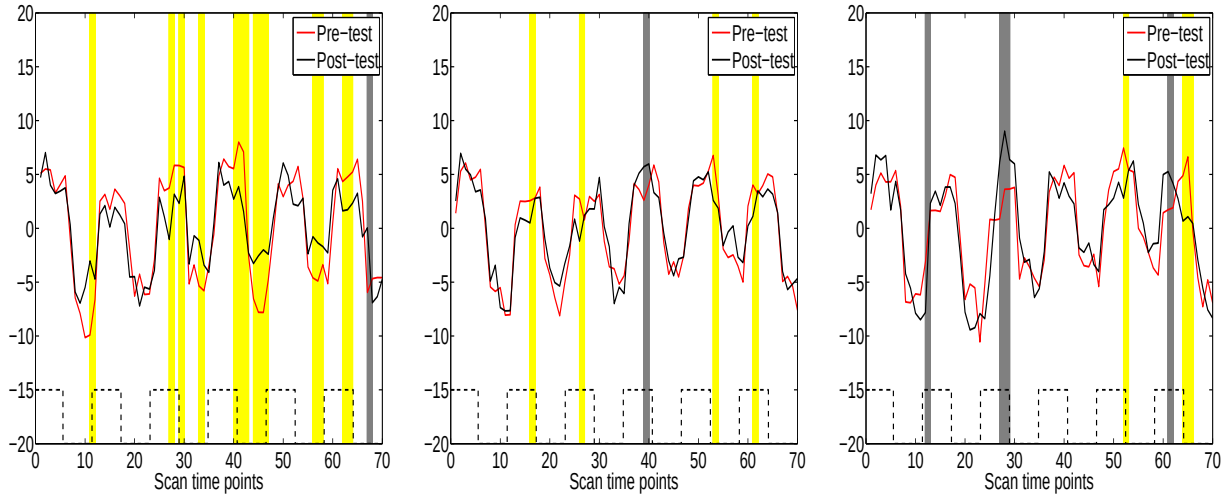
As seen in Table 5.2, attenuations at the post-test dominate on the SEF, FEF, SPL, and cuneus for the antisaccade practice group, and amplifications at the post-test dominate on the thalamus and IPL for the fixation practice group. Based on these criteria, in the prosaccade practice group, neither attenuations nor amplifications are dominant at the post-test.

A general pattern emerges showing that in the ROIs identified a priori as saccadic circuitry (Dyckman et al., 2007) there is a general attenuation (positive values showing less activation at the post-test) associated with task-specific practice (i.e., the antisaccade group). There appears to be little dramatic change in either direction in the prosaccade practice group. The only group that shows evidence of consistent amplification (negative values showing more activation at the post-test) across time is the fixation practice group. This pattern replicates that shown in the interaction plot (Figure 3.36) but with increased specificity and improved ability to evaluate which regions of the circuitry may be crucial for practice-related changes in performance.



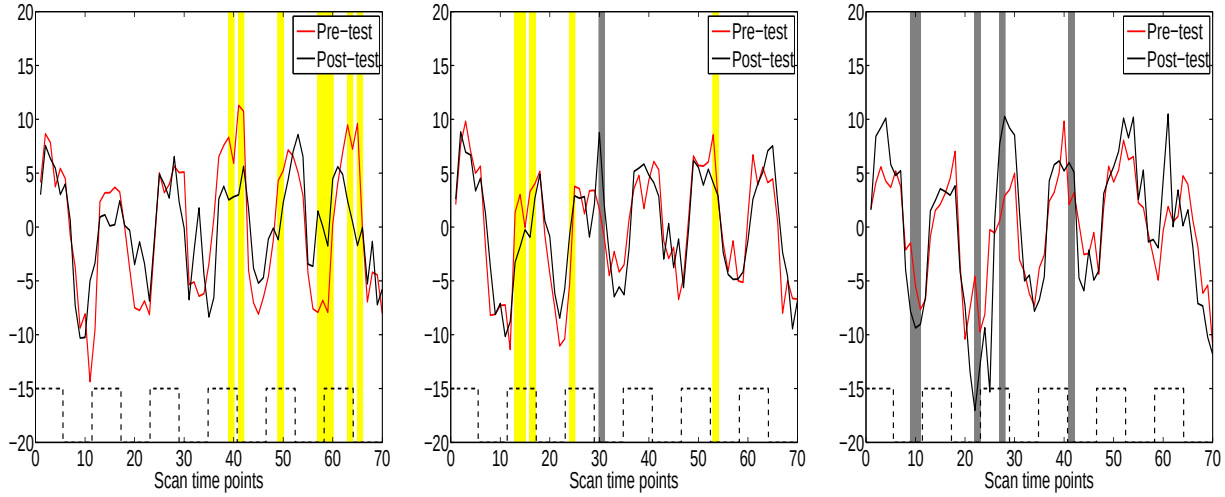
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.2: SEF: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



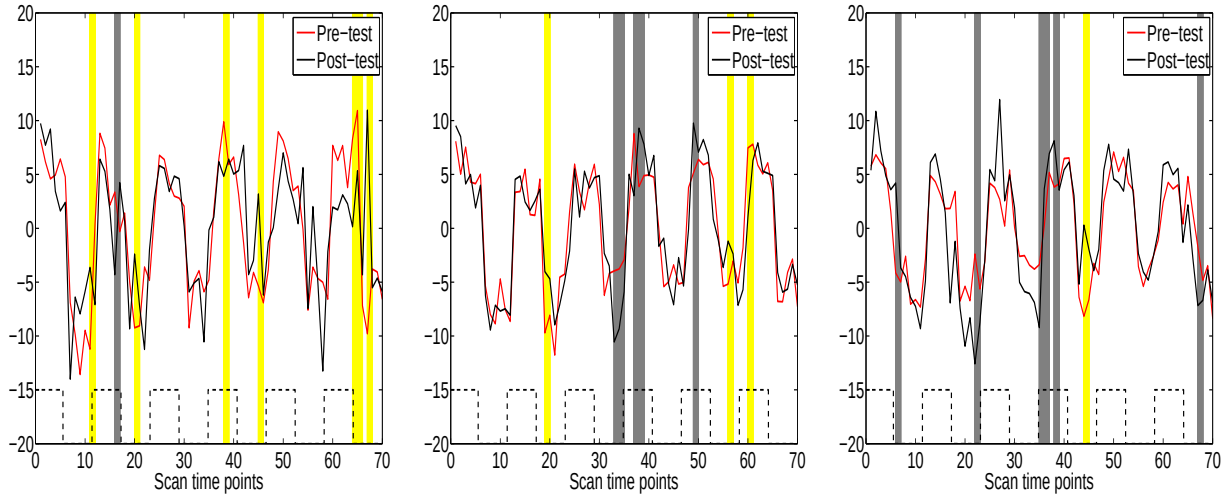
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.3: FEF: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



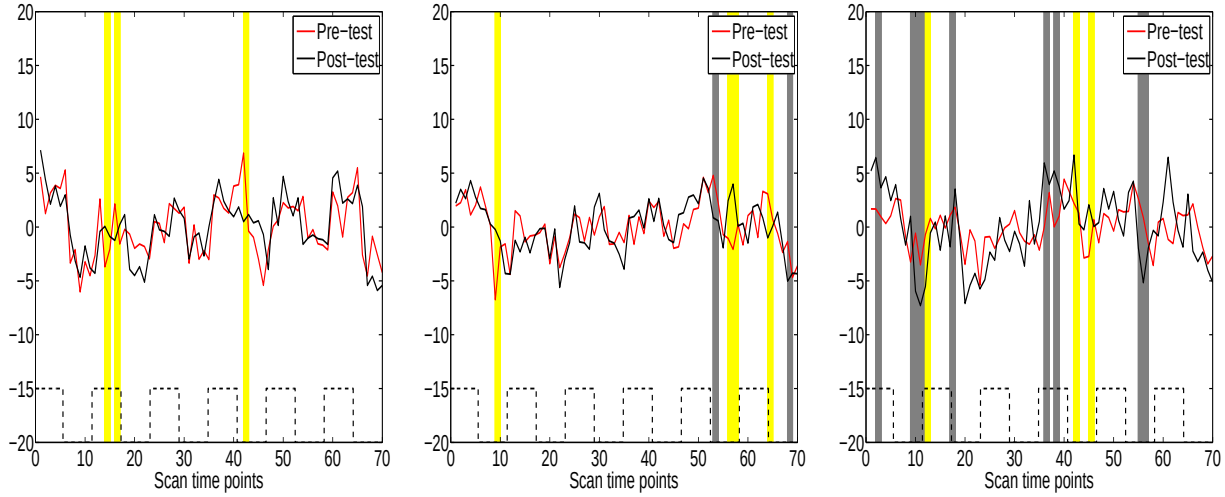
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.4: SPL: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



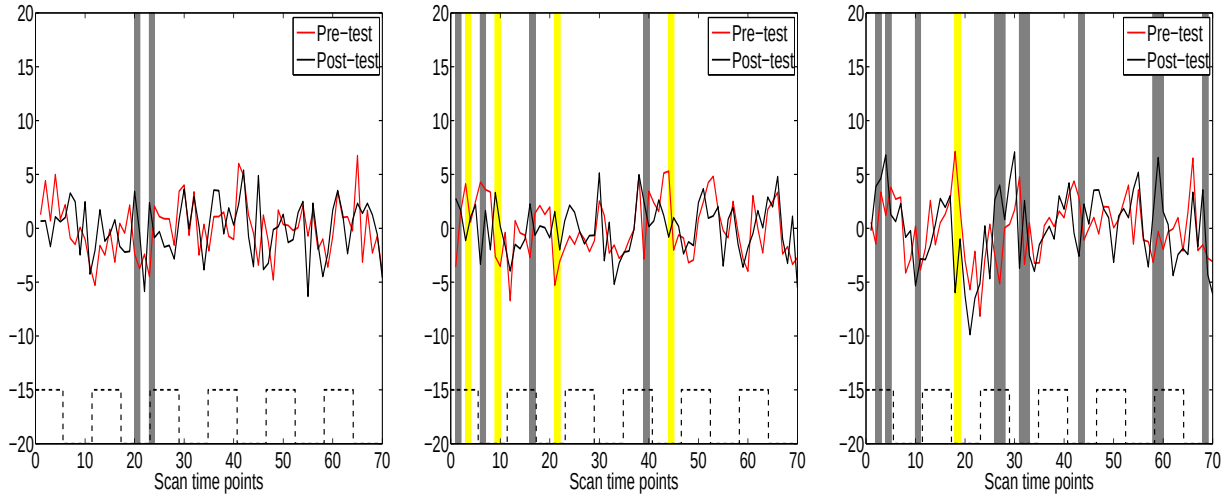
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.5: Cuneus: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



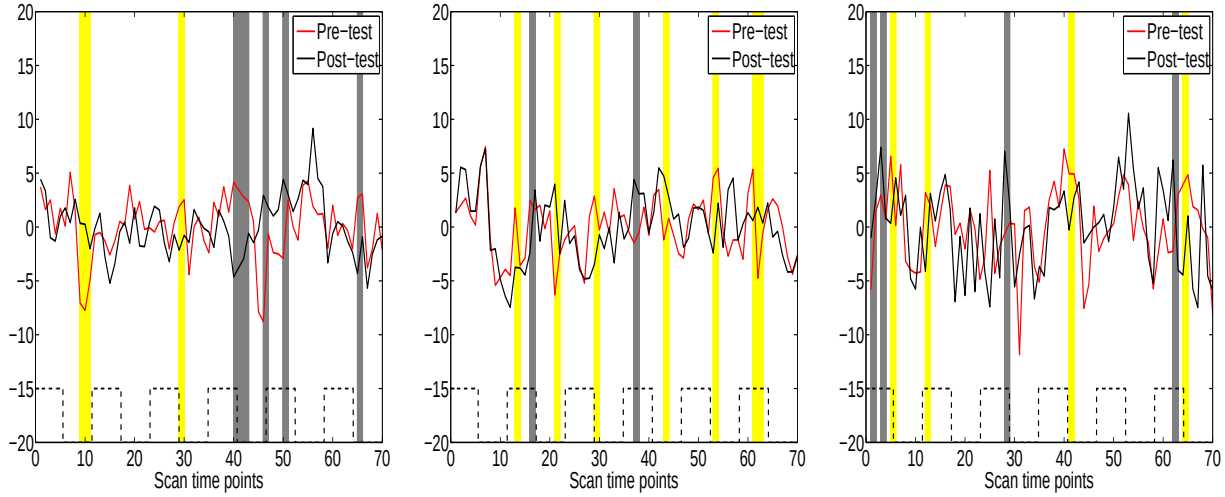
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.6: Thalamus: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



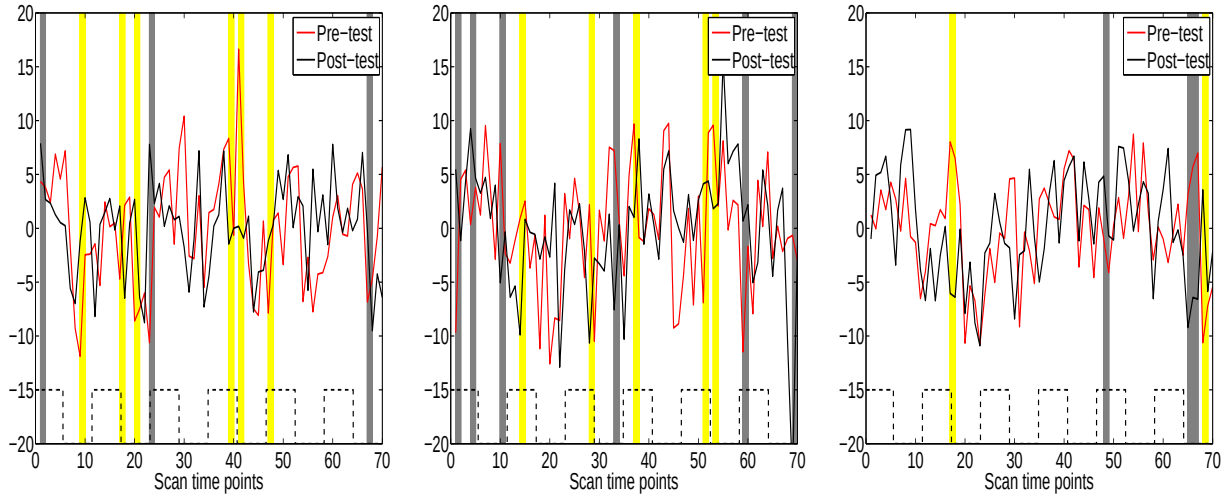
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.7: IPL: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



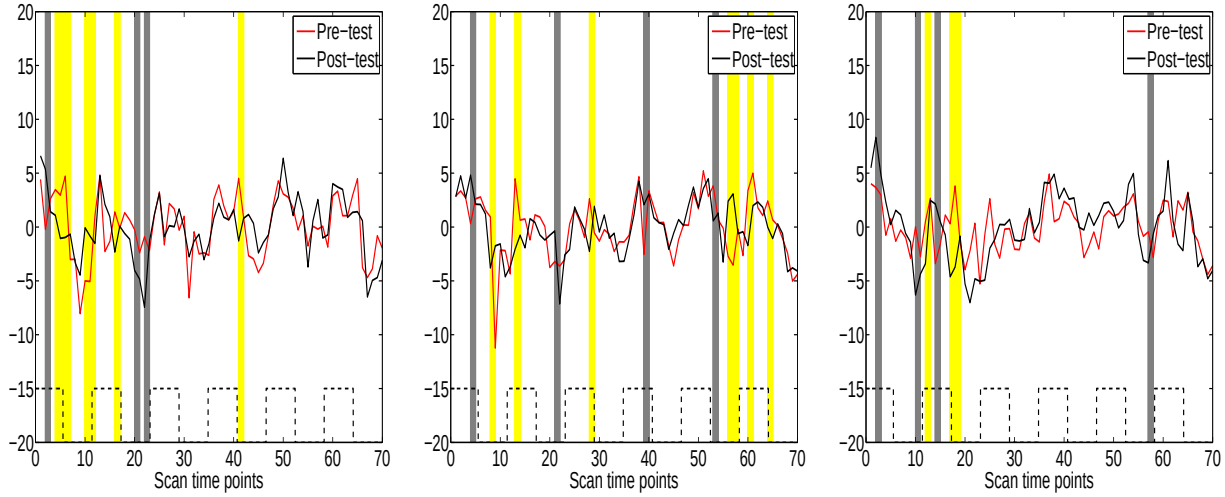
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.8: PFC-L: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



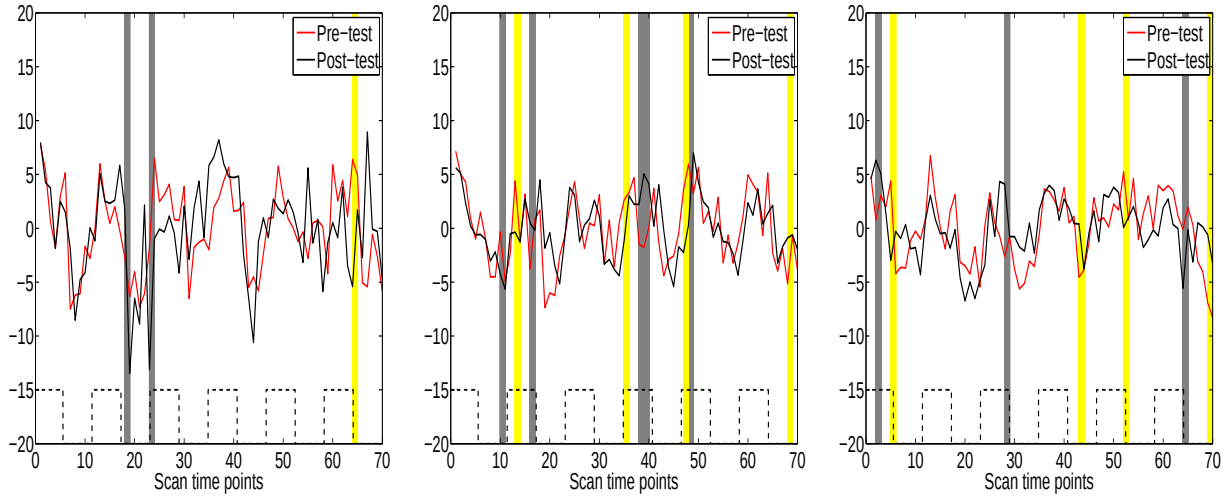
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.9: PFC-R: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



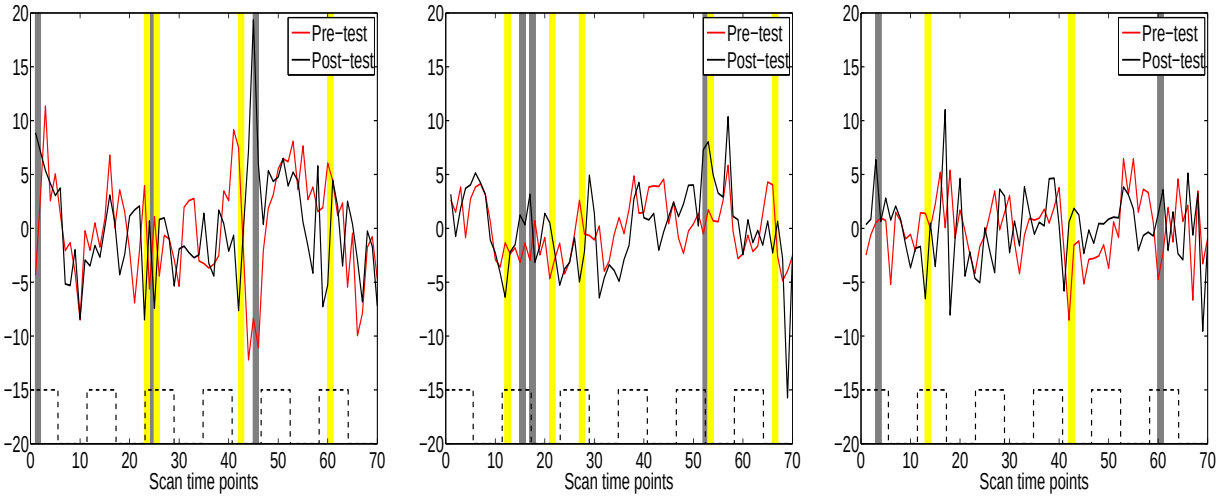
(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.10: Striatum: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.11: MOG: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.



(a) Antisaccade (b) Prosaccade (c) Fixation

Figure 5.12: IFC: The average time series plots for the pre-test (red) and post-test (black) are drawn for the three task groups. Dashed black lines at the bottom show the timing of antisaccade task blocks. Yellow bands represent time points where attenuations at the post-test are statistically significant, and gray bands represent time points where amplifications at the post-test are statistically significant by the bootstrap resampling method.

ROIs	Antisaccade	Prosaccade	Fixation
SEF	**0.16**	-0.01	-0.01
FEF	**0.19**	0.04	-0.01
SPL	**0.11**	0.06	-0.07
Cuneus	**0.09**	-0.03	-0.07
Thalamus	0.04	0.03	**-0.09**
IPL	-0.03	0.00	**-0.14**
PFC-L	-0.04	0.07	0.00
PFC-R	0.04	-0.01	-0.01
Striatum	0.06	0.04	-0.01
MOG	-0.01	-0.01	0.01
IFC	0.01	0.03	0.00

Table 5.2: Differences of the proportions of statistically significant attenuations (positive values) and amplifications (negative values) for the three groups at each ROI. These values are calculated as a proportion of time points attenuated - a proportion of time points amplified (i.e., (the number of time points significantly decreased at the post-test - the number of time points significantly increased at the post-test)/70). The values are denoted in bold if the absolute values are larger than 0.07.

5.3 Mixed model

In this section, we analyze ROIs using feature extraction and a mixed model to investigate attenuations at the post-test in the practice groups. The limitation of the bootstrap approach is that the test is done only within group, and thus the comparison between groups is done in an ad-hoc way by counting the numbers of significant attenuations and amplifications. Using a mixed model, we attempt to find statistical evidence in the difference among the three groups for some ROIs.

In our data, subjects are randomly divided into the three task group so that each subject has only one membership, which means that subjects are nested in the task groups. There is a nonnegligible likelihood that subjects belonging to the same task group behave similarly compared to subjects belonging to different task groups. This naturally leads us to assume that there is a common latent factor between subjects within the same task group. In addition, brain images from all subjects are scanned consecutively for more than five minutes so that each subject has 70 BOLD signals. From a different angle, this can be considered as repeated measures over time within each subject. Unlike a classical parametric approach, in which observations are frequently assumed to be randomly drawn, these repeated measures from a single subject are not independent of each other and affect the error term. In this section, we take the latent factor and the dependence from the repeated measures into account by introducing a mixed model.

5.3.1 Introduction to Mixed models

A mixed model (also known as a mixed effects model) is named for the statistical model containing both fixed effects and random effects. The term "fixed effects" indicates that the levels of the factor (i.e., the parameters of explanatory variables) are considered as fixed constants. Thus, the specific levels of the factor are of interest in and of themselves. Conversely, the term "random effects" indicates that the levels of the factor are considered random. In other words, the levels of a factor are drawn randomly from a population which we want to study so that they are representative of the population. Suppose, for instance, that we are interested in a clinical study investigating the effectiveness of a medication which is administrated at two different dose levels of interest, 20 mg and 40 mg, and the clinical study is conducted at 500 different clinics randomly selected in the United States. The two dose levels handled in this clinical study are the only dose levels being investigated, and

in using them there is no consideration for any other dose levels. On the other hand, the selected 500 clinics can be thought of as representative of all clinics existing in the United States because any clinic among all the clinics in the United States can be chosen. Hence, the factor of dose level is a fixed effect and the factor of clinic is a random effect. The most effective way to distinguish fixed and random factors depends on how the data are gathered and on which questions of interest are emphasized. Since the method of an analysis is affected by whether the factor is treated as fixed or random, inappropriate classification of factors may lead to an inaccurate result.

In a mixed model with nested groups, observations belonging to different groups (or clusters) are assumed to have different variance (i.e., heterogeneous variance), but observations belonging to the same group have equal variance (i.e., homogeneous variance). Hence, there is a need to consider two sources of variations; one is variation between groups (i.e., inter-cluster variation), and the other is variation within groups (i.e., intra-cluster variation).

Mixed models are used commonly in the analysis of longitudinal data, repeated measures, and hierarchical data. Longitudinal data are defined as data obtained repeatedly over time on each experimental unit. A mixed model is well suited for the analysis of longitudinal data, where each time series constitutes an individual curve (Demidenko, 2004). Repeated measures (or repeated measurements) are defined as data obtained repeatedly on each experimental unit. Longitudinal data and repeated measures are sometimes considered as the same, but they are slightly different in the sense that repeated measures do not have to be done over time. For instance, measurements at the different levels of a factor can be regarded as repeated measures. Hierarchical data are defined as data being structured as a one-to-many relationship. For example, one parent can have multiple children, but one child can have only one biological parent. Utilizing mixed models in those data formats is intended to increase sensitivity by making within-group comparisons (McCulloch and Searle, 2001). Demidenko (2004) states, "A mixed model is well suited for biological and medical data, which display notorious heterogeneity of responses to stimuli and treatment. An advantage of the mixed model is the ability to genuinely combine the data by introducing multilevel random effects" (p.1).

The general mixed model for the ith group (i.e., the ith cluster) can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, k, \tag{5.1}$$

where

- $\mathbf{y}_i$ is an $n_i \times 1$ vector of observed responses for the ith group.
- $\mathbf{X}_i$ is an $n_i \times p$ design matrix of fixed effects for the ith group.
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects coefficients associated with the design matrix $\mathbf{X}_i$ for all responses.
- $\mathbf{Z}_i$ is an $n_i \times q$ design matrix of random effects for the ith group.
- $\mathbf{a}_i$ is a $q \times 1$ vector of random effects associated with the design matrix $\mathbf{Z}_i$ for the ith group.
- $\boldsymbol{\epsilon}_i$ is an $n_i \times 1$ vector of error term (i.e., residual error) for the ith group.
- k is the number of groups and n_i is the number of observations in the ith group.

We assume that the random effects term $\mathbf{a}_i$ and the error term $\boldsymbol{\epsilon}_i$ are independent of each other for all i, $i = 1, \ldots, k$. The form of the mixed model for the ith group in Equation (5.1) can be condensed into one matrix format representing all groups (Demidenko, 2004):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\epsilon},$$

where

$$\mathbf{y}_{(n\times 1)} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{bmatrix}, \mathbf{X}_{(n\times p)} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_k \end{bmatrix}, \boldsymbol{\beta}_{(p\times 1)} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_p \end{bmatrix},$$

$$\mathbf{Z}_{(n\times kq)} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \ldots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{Z}_k \end{bmatrix}, \mathbf{a}_{(kq\times 1)} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_k \end{bmatrix}, \boldsymbol{\epsilon}_{(n\times 1)} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_k \end{bmatrix}.$$

Here, the dimension for each term is denoted inside parentheses located at the right bottom, and n indicates the total number of observations, which is $n = \sum_{i=1}^{k} n_i$. If $n_1 = n_2 = \ldots = n_k$; then it is said that the data are balanced. In classical statistics, the expected values for $\mathbf{y}$ in a linear model, for example, are denoted as $\mathrm{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$, which are considered as constants. The random effects term introduced by a mixed model is not a constant, whereas the fixed

effects term is still considered as a constant. Thus, the expected values for $\mathbf{y}$ in a mixed model are described in a different way than in a fixed model:

$$\mathrm{E}[\mathbf{y}|\mathbf{a}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}.$$

Since

$$\mathrm{var}[\mathbf{y}|\mathbf{a}] = \mathrm{var}[\boldsymbol{\epsilon}] = \mathbf{R},$$

and by assuming $\mathrm{E}[\mathbf{a}] = \mathbf{0}$ and $\mathrm{var}[\mathbf{a}] = \mathbf{G}$ (without any loss of generality), we have

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta},\ \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}). \tag{5.2}$$

Equation (5.2) shows that the marginal expectation of $\mathbf{y}$ is related to only a fixed effects term, but the marginal variance of $\mathbf{y}$ consists of two components from random effects and error terms; so these quantities are often called variance components. The best linear unbiased predictors (BLUP) and the best linear unbiased estimators (BLUE) from the mixed effects model are affected by variance components (Demidenko, 2004; McCulloch and Searle, 2001). Mixed models allow the researcher to make correct comparisons between groups using the correct source of variations, which are the variations between groups instead of the overall measurement errors.

5.3.2 ROI analysis using a mixed model

We employ a mixed model in our ROI analysis to compare practice groups by allowing heterogeneity in variance. As described in Section 5.2.2, we start with the difference between the pre- and post-test time courses for each subject as a response variable. Since subjects in our data are nested within the practice groups, it is reasonable to assume that the subjects belonging to the same practice group have equal variance (i.e., homogeneity in variance within the same group). On the other hand, it is legitimate to assume that the subjects belonging to different practice groups have different variance. We also assume that the repeated measures made on the same subject over a period of scan time points have a homogeneous correlation.

A feature extraction approach is associated with a detection of peak values in functions or curves. It adopts substitutes from detected peaks, and an analysis of these quantifications accompanies to determine which differ across defined populations or factors of interest (Morris, 2012; Morris et al., 2005). One merit of a feature extraction approach is that the

dimensions of the data can be reduced. Hence, we extract the features from the fMRI time series for ROI analysis and analyze them using a mixed model with repeated measurements. We use mean and median values to extract the features representing the stimulus-presented and -absent periods in our fMRI data.

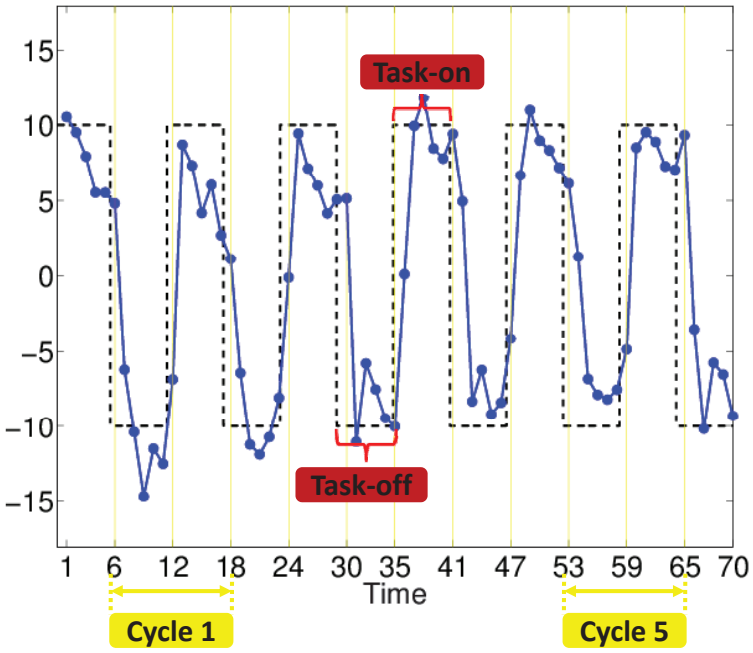


Figure 5.13: Cycle definition. The x-axis indicates scan time points ($1 \sim 70$) and the y-axis indicates the detrended BOLD signal. Dashed black lines represent the stimulus timing. The interval of stimulus timing is 5.86; thus the blocks of stimulus timing do not stand on the integer numbers in the x-axis. The first scan time point in each stimulus-presented and -absent period are denoted by solid yellow lines. The Dashed black lines do not overlap with the yellow lines standing on integer numbers in the x-axis. Task-on and -off indicate the condition when stimulus is presented and absent, respectively. By discarding the scan time points during the first stimulus-presented period and during the last stimulus-absent period, we have five blocks of experimental task alternation (indicated as "Cycle 1" and "Cycle 5") starting with a stimulus-absent period and ending with a stimulus-presented period.

In Figure 5.13, the x-axis indicates scan time points ($1 \sim 70$), and the y-axis indicates the detrended BOLD signal. Task-on and -off indicate the condition when stimulus is presented and absent, respectively. As can be seen in the figure, our data have six scan time points mostly in each stimulus-presented and -absent period, respectively, so that there are twelve scan time points in one block of experimental task alternation. Note that there are five scan

time points only in the first stimulus-presented period and the third stimulus-absent period. Dashed black lines represent the stimulus timing; the interval of stimulus timing is 5.86; thus the blocks of stimulus timing do not stand on the integer numbers in the x-axis. The first scan time point in each stimulus-presented and -absent period are denoted by solid yellow lines (e.g., the first scan time point in the first stimulus-absent period is 6, the first scan time point in the second stimulus-presented period is 12, and the first scan time point in the last stimulus-absent period is 65). Thus, the blocks shown in dashed black lines (non-integer numbers) do not overlap with the yellow lines (integer numbers).

By discarding the scan time points during the first stimulus-presented period and during the last stimulus-absent period, we have five blocks of experimental task alternation. A cycle is defined as one block of experimental task, starting from a stimulus-absent period and ending with a stimulus-presented period. The first cycle contains from the 6th to the 17th scan time points, and the last cycle contains from the 53rd to the 64th scan time points (indicated as "Cycle 1" and "Cycle 5" inside yellow text boxes in Figure 5.13).

In what follows, we explain our feature extraction procedure. Each cycle can be represented by one feature extracted value using the difference between the stimulus-presented period and the stimulus-absent period within the same cycle. For example, suppose that $f(t)$ indicates the detrended BOLD signal at the scan time point t and we use the mean value to extract the features representing the stimulus-presented and -absent periods. For better understanding, the magnified plot of the first two cycles is depicted in Figure 5.14. For the first cycle the stimulus-absent period has the scan time points $(t_6, \ldots, t_{11})$ and the stimulus-presented period has the scan time points $(t_{12}, \ldots, t_{17})$. The average value of $(f(t_6), \ldots, f(t_{11}))$ is $\frac{1}{6}\sum_{i=6}^{11} f(t_i)$; this value substitutes for the stimulus-absent period in the first cycle and is denoted by $\overline{f(t_{1,0})}$ in Figure 5.14. The average value of $(f(t_{12}), \ldots, f(t_{17}))$ is $\frac{1}{6}\sum_{i=12}^{17} f(t_i)$; this value substitutes for the stimulus-presented period in the first cycle and is denoted by $\overline{f(t_{1,1})}$ in Figure 5.14.

Hence, the representative of the first cycle can be condensed into one value by using the difference between those two averages (i.e., $\overline{f(t_{1,1})} - \overline{f(t_{1,0})}$). Likewise, the representative of the ith cycle can be condensed into $\overline{f(t_{i,1})} - \overline{f(t_{i,0})}$ for $i = 1, \ldots, 5$, where $\overline{f(t_{i,0})}$ indicates the average value in the stimulus-absent period for the ith cycle and $\overline{f(t_{i,1})}$ indicates the average value in the stimulus-presented period for the ith cycle.

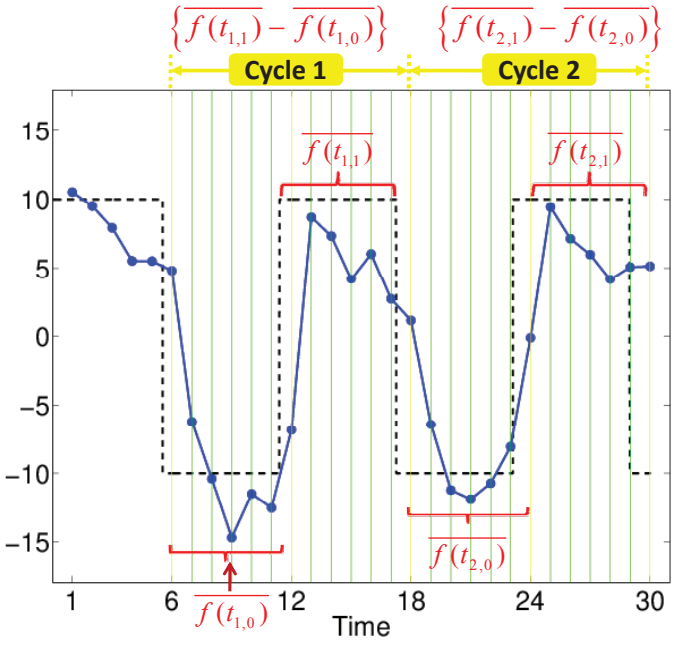


Figure 5.14: Feature extraction. This is the magnified plot of the first two cycles. The x-axis indicates scan time points, and the y-axis indicates the detrended BOLD signal. Dashed black lines represent the stimulus timing. The interval of stimulus timing is 5.86. The rooftop and the nadir in dashed black lines indicate when stimulus is presented and absent, respectively. The scan time points during the first stimulus-presented period are discarded. The scan time points used in our analysis are colored with solid green lines; solid yellow lines indicate the first scan time point in each stimulus-presented and -absent period. $\overline{f(t_{i,0})}$ and $\overline{f(t_{i,1})}$ denote the values of the extracted feature for the stimulus-absent and -present periods in the ith cycle, respectively. Finally, the ith cycle is represented by the difference between these two values of the extracted feature.

With visual inspection we notice that all subjects have roughly one time-lag until responding to a stimulus. For a more exact comparison between groups, we modify the cycle definition made above by shifting one scan time point in order to account for this response delay. As a method of extracting the features from the stimulus-presented and -absent periods in each cycle, we use mean and median values, and then we have representative values for cycles by using the difference between those two mean or median values. Thus, for a given subject in a specific practice group, each cycle, which usually has 12 responses, is represented by one summary value, that is the difference between means or medians

extracted from the original twelve values. Each subject has one average time course curve containing 70 detrended BOLD signals in the original data structure in a given practice group and in a given ROI. With the feature extraction approach, the data structure, finally, becomes such that each subject has five extracted values (i.e., one value for each cycle).

Denote the response variable for the j-th subject at the k-th cycle in the i-th practice group by y_{ijk}; then the mixed model form in (5.1) can be written as

$$y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + a_{j(i)} + e_{ijk}, \tag{5.3}$$

where μ is the grand mean, α_i is the fixed effect of the i-th practice group, β_k is the fixed effect of the k-th cycle, $(\alpha\beta)_{ik}$ is the fixed effect of the interaction between the i-th practice group and the k-th cycle, $a_{j(i)}$ is the random effect associated with the j-th subject in the i-th practice group, and e_{ijk} is the random error associated with the j-th subject at the k-th cycle in the i-th practice group for $i = 1, 2, 3$, $j = 1, \ldots, n_i$, $n_1 = 12$ (antisaccade), $n_2 = 14$ (prosaccade), $n_3 = 11$ (fixation), and $k = 1, \ldots, 5$. We assume that $a_{j(i)} \sim i.i.d.\ N(0, \sigma_i^2)$ and e_{ijk} is normally distributed with $E(e_{ijk}) = 0, V(e_{ijk}) = \sigma^2$ and $Cov(e_{ijk}, e_{ijk'}) = \sigma^2\rho$ for $k \neq k'$ for all i and j. Here, ρ, the intra-class correlation, is defined by $\sigma_i^2/[\sigma^2 + \sigma_i^2]$. Note that

$$\begin{aligned} E(y_{ijk}) &= \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik}, \\ Var(y_{ijk}) &= \sigma_i^2 + \sigma^2. \end{aligned}$$

As a result of the application of this mixed model based on the feature extraction approach, the interaction effect term between the task group and cycle is not significant at the significant level $\alpha = 0.1$ for all ROIs for both mean and median values. After dropping the interaction effect term $(\alpha\beta)_{ik}$ from the model (5.3), we find that the cycle effect β_k is also not significant for all ROIs for both mean and median values. Thus, our final model becomes

$$y_{ijk} = \mu + \alpha_i + a_{j(i)} + e_{ijk},$$

where the notations are the same as in the model (5.3). Hence,

$$\begin{aligned} E(y_{ijk}) &= \mu + \alpha_i, \\ Var(y_{ijk}) &= \sigma_i^2 + \sigma^2. \end{aligned}$$

The results of the application of this mixed model based on the feature extraction approach are summarized in Table 5.3. F-tests and p-values indicate whether the differences between the practice groups are significant for each ROI. Large F-values imply that

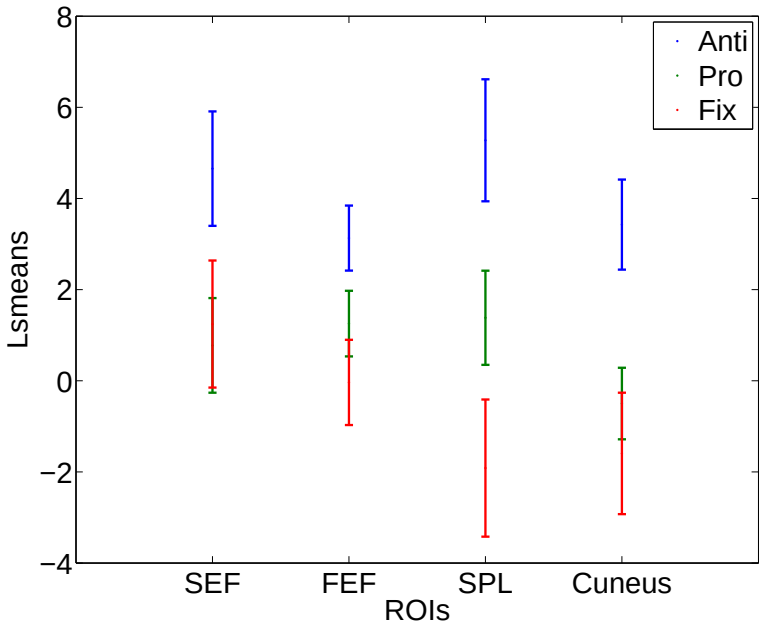
there are significant differences between practice groups. The test results show that SEF, FEF, SPL, and cuneus have significant differences at the post-test between the three practice groups at significance level $\alpha = 0.1$ for both mean and median values. These results are consistent with the results in Section 5.2.2.

For the four ROIs which show significant differences between the three practice groups, least squares means (LSmeans) (Hsu, 1999) are computed for each practice group to see where an attenuation or an amplification at the post-test occurs. These LSmeans results are shown in Figures 5.15 (a) and (b) in which the feature is extracted by using mean and median values, respectively. Blue, green, and red stand for the antisaccade, prosaccade, and fixation practice groups, respectively. The vertical lines indicate the 95% confidence intervals for each practice group. Positive values indicate attenuations at the post-test, and negative values indicate amplifications at the post-test. The significance of these attenuations and amplifications at the post-test are valid statistically only when their confidence intervals do not cover zero. The overlapped vertical lines evince that those overlapped groups do not have significant differences from one another.

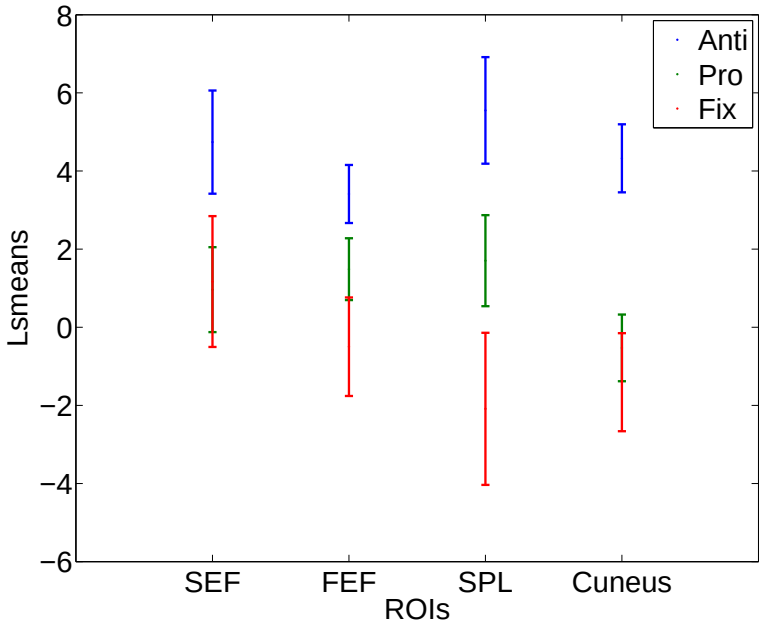
The antisaccade practice group has a strong attenuation at the post-test for SEF, FEF, SPL, and cuneus for both mean and median values. The vertical blue lines indicating the antisaccade practice groups lie far above zero. The prosaccade practice group (green) also has an attenuation at the post-test for FEF and SPL for both mean and median values. In contrast, the fixation practice group has a strong amplification at the post-test for the SPL and cuneus for both mean and median values. Blue, red, and green lines do not overlap in the SPL for both mean and median values, which means that a significant difference between the three practice groups is found in the SPL. Green and red lines overlap in the SEF, FEF, and cuneus for both mean and median values, which means no significant difference between the prosaccade and fixation practice groups is found in these regions.

5.4 Conclusion

We conduct ROI analyses with 11 bilateral neural ROIs using the bootstrap resampling approach and the mixed model with the feature extraction approach to study whether there is a general attenuation associated with task-specific practice. These 11 ROIs are identified a priori as saccadic circuitry by Dyckman et al. (2007) in the previous study using a similar task practice. We reach the same conclusion through these two analyses. The antisaccade practice group shows a general attenuation in brain activity at the post-test; in particular, significant



(a) Features extracted by using mean values



(b) Features extracted by using median values

Figure 5.15: Least squares means (LSmeans) plots for SEF, FEF, SPL, and cuneus reveal significant differences between the three practice groups. Blue, green, and red stand for the antisaccade, prosaccade, and fixation practice groups, respectively. The vertical lines indicate the 95% confidence intervals for each practice group. The positive and the negative values indicate attenuations and amplifications at the post-test, respectively. These indications of attenuations and amplifications at the post-test are valid statistically only when their confidence interval does not cover zero. The overlapping vertical lines mean those overlapping groups do not have significant differences from one another.

ROIs	Mean		Median	
	F value	p-value	F value	P-value
SEF	3.05	0.0511	2.67	0.0737
FEF	3.90	0.0230	3.95	0.0220
SPL	6.46	0.0022	5.47	0.0054
Cuneus	6.38	0.0024	10.57	<.0001
Thalamus	2.00	0.1403	2.07	0.1309
IPL	0.24	0.7872	0.17	0.8402
PFC-L	0.12	0.8883	0.04	0.9603
PFC-R	0.73	0.4824	0.32	0.7247
Striatum	1.08	0.3433	1.59	0.2090
MOG	0.26	0.7726	0.26	0.7682
IFC	0.94	0.3944	1.47	0.2339

Table 5.3: The results of the application of the mixed model with the feature extraction approach. SEF, FEF, SPL, and cuneus reveal significant differences between practice groups when we use mean and median values for feature extraction at significance level $\alpha = 0.1$.

attenuations are found in the SEF, FEF, SPL, and cuneus only for the antisaccade practice group. These four ROIs are confirmed as the regions related to task-specific practice in the previous study. On the other hand, the fixation group shows a general amplification in brain activity at the post-test. The prosaccade group shows that attenuation and amplification in brain activity at the post-test occur in a similar frequency.

Chapter 6

conclusion and future work

6.1 Conclusion

This dissertation evaluates brain activation associated with practice-related cognitive control during eye movement tasks. All participants are tested on antisaccade performance at the pre- and post-test while fMRI data are acquired. In between the pre- and post-test, participants engage in daily practice of one assigned eye movement task: antisaccade, prosaccade, or fixation. In order to detect changes of brain activation between the two test sessions, we propose a wavelet-based clustering approach. A notable component of our clustering method is the application of a novel combination of statistical techniques selected specifically to circumvent typically acknowledged methodological problems associated with the structure of the data generated via fMRI and/or the model-based GLM analyses as they are commonly applied to fMRI data. These problems include massive-sized data that are ill-balanced and influenced by temporal autocorrelations, which we successfully moderate via the integration and application of clustering routines and wavelet analyses. The difficulties inherent in accurate modeling of the BOLD response, which are standard in GLM analyses, are likewise avoided in this dissertation by using a data-driven technique. This model-free analysis successfully identifies the well-known brain activation patterns known to support antisaccade performance. Not only is the circuitry identified, but also the analysis of practice-related changes can be evaluated with greater confidence because these methods are more resistant to common problems. As such, the differences between practice groups across time are tested in a rigorous manner and reveal task-specific brain activation changes in cognitive control across time.

We demonstrate the effectiveness of our wavelet-based clustering method for fMRI clustering analysis via the simulated data by adding artificial signals. For precise verification, we follow the procedures identical to the ones used in the real data analysis after mimicking the conditions of real data. The clustering results are compared in the time and wavelet domains using K-means and PCA K-means clustering methods. The simulation results show that

the clustering analysis using the raw data yield improper clustering results for some cases, whereas PCA K-means clustering methods provide accurate and appropriate results using the wavelet transformed data. Therefore, PCA K-means is preferred when coupled with the thresholded wavelet coefficients.

Using the proposed clustering procedure, we find that over the series of analyses conducted, task-consistent practice is associated with decreased activation from the pre- to post-test. The antisaccade practice group is the only group to show consistently decreased antisaccade-related activation in saccadic circuitry, particularly evident in FEF, SEF, SPL, and cuneus. This pattern may be associated with increased efficiency such that fewer neural resources are necessary to support the response due to task-specific practice. A similar pattern has been reported in numerous studies showing decreased activity associated with learning (Garavan et al., 2000; Jansma et al., 2001; Koch et al., 2006; Poldrack et al., 2005).

A distinctly different pattern of neural activation response is demonstrated by the two groups who practice test-inconsistent tasks (prosaccade and fixation). The prosaccade practice group is most likely to show unchanged activation levels between the pre- and post-test. ROI analyses suggest some minor decreases in activation at the post-test, but they are neither as widespread nor as strong as those observed in the antisaccade practice group. Thus, the brain activation across time in the prosaccade group is relatively stable. The fixation practice group is the only one that demonstrates generally increased activation at the post-test. The persistent increases in the fixation group may suggest strengthening of the neural circuitry that inhibits rapid responses away from the fixation target (Leigh and Zee, 2006). Although both prosaccade and fixation practice groups constitute control conditions, their different task requirements would cause us to predict different brain activation patterns.

In summary, these data demonstrate that changes in brain activation patterns of the well-identified neural circuitry supporting antisaccade performance (McDowell et al., 2008; Munoz and Everling, 2004; Sweeney et al., 2007) differ between task-specific and task-irrelevant practice. Changes in brain activation are observed despite no quantifiable differences in antisaccade performance at the post-test, which may imply that brain activation patterns are more sensitive than behavioral measures as indications of plasticity. Of course, this intervention is merely one week in duration. It may take longer for changes in the behavioral component of the task to be manifest if they are a less sensitive measure of changes in brain function.

6.2 Future work

Our wavelet-based clustering method is powerful and effective for fMRI clustering analysis. Nevertheless, the approach considered in this dissertation, like many other clustering procedures in the literature of fMRI studies, still does not account for the spatial correlation in the data. We plan to apply the work by Zhang et al. (2008) to handle the spatial correlation in a wavelet-based clustering analysis of fMRI data. They use a hidden Markov random field to model the spatial dependence in the data and develop a maximum likelihood approach via the Expectation-Maximization algorithm with stochastic variation. The limitation of their approach, however, is that it can be applied to only one image, not a series of images. This shortcoming will be remedied in our future project.

The multiple subjects for each group are combined via median at each voxel and each time point separately in this work. However, taking the median does not fully utilize the information available in the raw data. We plan to develop a better aggregation method, for example, smoothing neighboring voxels instead of taking a simple statistic.

In the initial analysis, the simple hard thresholding rule is applied. To produce a more efficient clustering algorithm, we plan to apply a data-driven block thresholding rule using neighboring empirical coefficients (Park, 2010) and compare it with other thresholding rules.

Bibliography

Atto, A., Pastor, D., and Isar, A. (2007). On the statistical decorrelation of the wavelet packet coefficients of a band-limited wide-sense stationary random process. *Signal Processing*, 87(10):2320–2335.

Balslev, D., Nielsen, F., Frutiger, S., Sidtis, J., Christiansen, T., Svarer, C., Strother, S., Rottenberg, D., Hansen, L., Paulson, O., and Law, I. (2002). Cluster analysis of activity-time series in motor learning. *Human Brain Mapping*, 15(3):135–145.

Baumgartner, R. (2000). Comparison of two exploratory data analysis methods for fMRI: Fuzzy clustering vs. principal component analysis. *Magnetic Resonance Imaging*, 18(1):89–94.

Baumgartner, R., Windischberger, C., and Moser, E. (1998). Quantification in functional magnetic resonance imaging: Fuzzy clustering vs. correlation analysis. *Magnetic Resonance Imaging*, 16(2):115–125.

Baune, A., Sommer, F., Erb, M., Wildgruber, D., Kardatzki, B., Palm, G., and Grodd, W. (1999). Dynamical cluster analysis of cortical fMRI activation. *Neuroimage*, 9(5):477–489.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1):289–300.

Bischoff-Grethe, A., Ozyurt, I. B., Busa, E., Quinn, B. T., Fennema-Notestine, C., Clark, C. P., Morris, S., Bondi, M. W., Jernigan, T. L., Dale, A. M., Brown, G. G., and Fischl, B. (2007). A technique for the deidentification of structural brain MR images. *Human Brain Mapping*, 28(9):892–903.

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541.

Bullmore, E., Suckling, J., Zelaya, F., Long, C., Honey, G., Routledge, C., Ng, V., Fletcher, P., and Williams, S. (2003). Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical Methods in Medical Research*, 12(5):375–399.

Buračas, G. and Boynton, G. (2002). Efficient design of event-related fMRI experiments using m-sequences. *NeuroImage*, 16(3):801–813.

Camchong, J., Dyckman, K. A., Austin, B. P., Clementz, B. A., and McDowell, J. E. (2008). Common neural circuitry supporting volitional saccades and its disruption in schizophrenia patients and relatives. *Biological Psychiatry*, 64(12):1042–1050.

Chein, J. and Schneider, W. (2005). Neuroimaging studies of practice-related change: FMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, 25(3):607–623.

Coifman, R. R., Meyer, Y., and Wickerhauser, V. (1992). Wavelet analysis and signal processing. In *Wavelets and Their Applications*, pages 153–178. Johnson & Bartlett, Boston.

Cox, R. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3):162–173.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Donoho, D. and Johnstone, J. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Donoho, D. and Johnstone, J. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Dyckman, K. A. (2007). *FMRI/EEG study of changes in behavior and brain activity during and after saccade practice*. PhD thesis, The University of Georgia, Athens, GA, U.S.A.

Dyckman, K. A., Camchong, J., Clementz, B. A., and McDowell, J. E. (2007). An effect of context on saccade-related behavior and brain activity. *NeuroImage*, 36(3):774–784.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Fadili, M., Ruan, S., Bloyet, D., and Mazoyer, B. (2000). A multistep unsupervised fuzzy clustering analysis of fMRI time series. *Human Brain Mapping*, 10(4):160–178.

Filzmoser, P., Baumgartner, R., and Moser, E. (1999). A hierarchical clustering method for analyzing functional MR images. *Magnetic Resonance Imaging*, 17(6):817–826.

Fischer, B., Hartnegg, K., and Mokler, A. (2000). Dynamic visual perception of dyslexic children. *Perception*, 29(5):523–530.

Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38:910–917.

Ford, A., Goltz, C., Brown, G., and Everling, S. (2005). Neural processes associated with antisaccade task performance investigated with event-related fMRI. *Journal of Neurophysiology*, 94(1):429–440.

Friman, O., Borga, M., Lundberg, P., and Knutsson, H. (2004). Detection and detrending in fMRI data analysis. *NeuroImage*, 22(2):645–655.

Friston, K. J., Holmes, A., Worsley, K., Poline, J., Frith, C., and Frackowiak, R. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210.

Garavan, H., Kelley, D., Rosen, A., Rao, S., and Stein, E. (2000). Practice-related functional activation changes in a working memory task. *Microscopy Research Technique*, 51(1):54–63.

Gautama, T. and Hulle, M. V. (2005). Estimating the global order of the fMRI noise model. *NeuroImage*, 26(4):1211–1217.

Genovese, C. and Wasserman, L. (2005). Confidence sets for nonparametric wavelet regression. *The Annals of Statistics*, 33(2):698–729.

Gibbons, R. D., Lazar, N. A., Bhaumik, D. K., Sclove, S. L., Chen, H. Y., Thulborn, K. R., Sweeney, J. A., Hur, K., and Patternson, D. (2004). Estimation and classification of fMRI hemodynamic response patterns. *NeuroImage*, 22(2):804–814.

Golay, X., Kollias, S., Meier, D., Valavanis, A., and Boesiger, P. (1997). Fuzzy membership vs. probability in cross correlation based fuzzy clustering of fMRI data. *Proceedings of the 3rd International Conference on Functional Mapping of the Human Brain*, 3:S481.

Goutte, C., Hansen, L., Liptrot, M., and Rostrup, E. (2001). Feature space clustering for fMRI meta-analysis. *Human Brain Mapping*, 13(3):165–183.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F. A., and Hansen, L. K. (1999). On clustering fMRI time series. *NeuroImage*, 9(3):298–310.

Hartigan, J. (1975). *Clustering Algorithms*. John Wiley & Sons, Inc., New York.

Hawkes, P. and Kazan, B. (1993). *Advances in Electronics and Electron Physics*, volume 85. Academic Press.

Heller, R., Stanley, D., Yekutieli, D., Rubin, N., and Benjamini, Y. (2006). Cluster based analysis of fMRI data. *NeuroImage*, 33(2):599–608.

Hsu, J. C. (1999). *Multiple Comparisons: Theory and Methods*. Chapman & Hall/CRC: Boca Raton, Fla.

Huettel, S. A., Song, A. W., and McCarthy, G. (2009). *Functional Magnetic Resonance Imaging*. Sinauer Associates Inc., Sunderland, MA, second edition.

Jansma, J. M., Ramsey, N. F., Slagter, H. A., and Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *Journal of Cognitive Neuroscience*, 13(6):730–743.

Johnson, R. and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall, New York, fifth edition.

Jolliffe, T. (2002). *Principal Component Analysis*. Springer, New York, second edition.

Keedy, S., Ebens, C., Keshavan, M., and Sweeney, J. (2006). Functional magnetic resonance imaging studies of eye movements in first episode schizophrenia: Smooth pursuit,

visually guided saccades and the oculomotor delayed response task. *Psychiatry Research*, 146(3):199–211.

Kelly, A. and Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, 15(8):1089–1102.

Koch, K., Wagner, G., von Consbruch, K., Nenadic, I., Schultz, C., Ehle, C., Reichenbach, J., Sauer, H., and Schlosser, R. (2006). Temporal changes in neural activation during practice of information retrieval from short-term memory: An fMRI study. *Brain Research*, 1107(1):140–150.

Lazar, N. A. (2008). *The Statistical Analysis of Functional MRI Data*. Springer, New York.

Leigh, R. and Zee, D. (2006). *The Neurology of Eye Movements*. Oxford University Press, USA, fourth edition.

Li, Y. (2011). *Data and Applications Security and Privacy: 25th Annual IFIP WG 11.3 Conference, DBSec 2011, Richmond, Va, USA, July 11-13, 2011, Proceedings*. Lecture Notes in Computer Science. Springer.

Lindquist, M. (2008). The statistical analysis of fMRI data. *Statistical Science*, 23(4):439–464.

Luo, W. and Nichols, T. (2003). Diagnosis and exploration of massively univariate neuroimaging models. *NeuroImage*, 19(3):1014–1032.

Mahapatra, D. (2012). Skull stripping of neonatal brain MRI: Using prior shape information with graph cuts. *Journal of Digital Imaging*, 25(6):802–812.

Marco, G. (2009). Effective connectivity and brain modeling by fMRI. *Advanced Studies in Biology*, 1(3):139–144.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Inc.

McDowell, J. E., Dyckman, K. A., Austin, B. P., and Clementz, B. A. (2008). Neurophysiology and neuroanatomy of reflexive and volitional saccades: Evidence from studies of humans. *Brain and Cognition*, 68(3):255–270.

McKeown, M., Hansen, L., and Sejnowski, T. (2003). Independent component analysis of functional MRI: What is signal and what is noise. *Current Opinion in Neurobiology*, 13(5):620–629.

McKeown, M. and Sejnowski, T. (1998). Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping*, 6(5–6):368–372.

Mooney, C. Z. and Duval, R. D. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage Publications.

Morris, J. S. (2012). Statistical methods for proteomic biomarker discovery based on feature extraction or functional modeling approaches. *Statistics and Its Interface*, 5(1):117–136.

Morris, J. S., Coombes, K. R., Koomen, J., Baggerly, K. A., and Kobayashi, R. (2005). Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*, 21(9):1764–1775.

Mumford, J. and Poldrack, R. (2007). Modeling group fMRI data. *Social Cognitive and Affective Neuroscience*, 2(3):251–257.

Munoz, P. and Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, 5(3):218–228.

Muri, R., Heid, O., Nirkko, A., Ozdoba, C., Felblinger, J., Schroth, G., and Hess, C. (1998). Functional organisation of saccades and antisaccades in the frontal lobe in humans: A study with echo planar functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 65(3):374–377.

O'Driscoll, A., Alpert, M., Matthysse, W., Levy, L., Rauch, L., and Holzman, S. (1995). Functional neuroanatomy of antisaccade eye movements investigated with positron emission tomography. *Proceedings of the National Academy of Sciences of USA*, 92(3):925–929.

Park, C. (2010). Block thresholding wavelet regression using SCAD penalty. *Journal of Statistical Planning and Inference*, 140(9):2755–2770.

Park, C., Ahn, J., Hendry, M., and Jang, W. (2011). Analysis of long period variable stars with nonparametric tests for trend detection. *Journal of the American Statistical Association*, 106(495):832–845.

Park, C., Lazar, N. A., Ahn, J., and Sornborger, A. (2010). A multiscale analysis of the temporal characteristics of resting-state fMRI data. *Journal of Neuroscience Methods*, 193(2):1407–1433.

Paus, T. (1996). Location and function of the human frontal eye field: A selective review. *Neuropsychologia*, 34(6):475–483.

Poldrack, R., Sabb, F., Foerde, K., Tom, S., Asarnow, R., Bookheimer, S., and Knowlton, B. (2005). The neural correlates of motor skill automaticity. *Journal of Neuroscience*, 25(22):5356–5364.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, 2(1):67–70.

Raemaekers, M., Jansma, J., Cahn, W., der Geest, J. V., der Linden, J. V., Kahn, R., and Ramsey, N. (2002). Neuronal substrate of the saccadic inhibition deficit in schizophrenia investigated with 3-dimensional event-related functional magnetic resonance imaging. *Archives of General Psychiatry*, 59(4):313–320.

Rex, D. E., Shattuck, D. W., Wood, R. P., Narr, K. L., Luders, E., Rehm, K., Stolzner, S. E., Rottenberg, D. A., and Toga, A. W. (2004). A meta-algorithm for brain extraction in MRI. *Neuroimage*, 23(2):625–637.

Schneider, W. and Chein, J. (2003). Controlled and automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science*, 27(3):525–559.

Schneider, W. and Shiffrin, R. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1):1–66.

Skudlarski, P., Constable, R. T., and Gore, J. C. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. *Neuroimage*, 9(3):311–329.

Stanberry, L., Nandy, R., and Cordes, D. (2003). Cluster analysis of fMRI data using dendrogram sharpening. *Human Brain Mapping*, 20(4):201–219.

Stoev, S., Taqqu, M., Park, C., and Marron, J. (2005). On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of internet traffic. *Computer Networks*, 48(3):423–445.

Sweeney, J. A., Luna, B., Keedy, S. K., McDowell, J. E., and Clementz, B. A. (2007). fMRI studies of eye movement control: Investigating the interaction of cognitive and sensorimotor brain systems. *Neuroimage*, 36(S2):T52–T60.

Talairach, J. and Tournoux, P. (1998). *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - An Approach to Cerebral Imaging*. Thieme Medical Publishers, New York.

van Zijl, P. C. M., Hua, J., and Lu, H. (2012). The BOLD post-stimulus undershoot, one of the most debated issues in fMRI. *NeuroImage*, 62(2):1092–1102.

Veitch, D. and Abry, P. (1999). Wavelet analysis of long-range dependence traffic. *IEEE Transactions on Information Theory*, 44(1):2–15.

Venkataraman, A., Dijk, K., Buckner, R., and Golland, P. (2009). Exploring functional connectivity in fMRI via clustering. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 441–444.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc.

Wanga, Z., Childressc, A., Wanga, J., and Detrea, J. (2007). Support vector machine learning-based fMRI data group analysis. *NeuroImage*, 36(4):1139–1151.

Woolrich, W., Ripley, D., Brady, M., and Smith, M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage*, 14(6):1370–1386.

Worsley, K. (2003). Detecting activation in fMRI data. *Statistical Methods in Medical Research*, 12(5):401–418.

Ye, J., Lazar, N. A., and Li, Y. (2009). Geostatistical analysis in clustering fMRI time series. *Statistics in Medicine*, 28(19):2490–2508.

Ye, N. (2003). *The Handbook of Data Mining*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Zhang, X., Johnson, T., Little, R., and Cao, Y. (2008). Quantitative magnetic resonance image analysis via the EM algorithm with stochastic variation. *Annals of Applied Statistics*, 2(2):736–755.

Appendix A

Clustering results for the Antisaccade task group

In this appendix, we present the clustering results for the antisaccade task group. For each test session (pre- and post-test) the brain is partitioned into two, three, five or seven clusters, which are K=2, 3, 5, and 7, using K-means and PCA K-means clustering methods.
The clustered brain maps are listed in the following order:

- K-means method for the pre-test for slices 13 – 19 in Figure A.1.
- K-means method for the pre-test for slices 20 – 26 in Figure A.2.
- K-means method for the pre-test for slices 27 – 33 in Figure A.3.
- PCA K-means method for the pre-test for slices 13 – 19 in Figure A.4.
- PCA K-means method for the pre-test for slices 20 – 26 in Figure A.5.
- PCA K-means method for the pre-test for slices 27 – 33 in Figure A.6.
- K-means method for the post-test for slices 13 – 19 in Figure A.7.
- K-means method for the post-test for slices 20 – 26 in Figure A.8.
- K-means method for the post-test for slices 27 – 33 in Figure A.9.
- PCA K-means method for the post-test for slices 13 – 19 in Figure A.10.
- PCA K-means method for the post-test for slices 20 – 26 in Figure A.11.
- PCA K-means method for the post-test for slices 27 – 33 in Figure A.12.

The average time series plots corresponding to each cluster in the clustered brain maps are listed in the following order:

- The pre-test using K-means and PCA K-means method in Figure A.13.
- The post-test using K-means and PCA K-means method in Figure A.14.

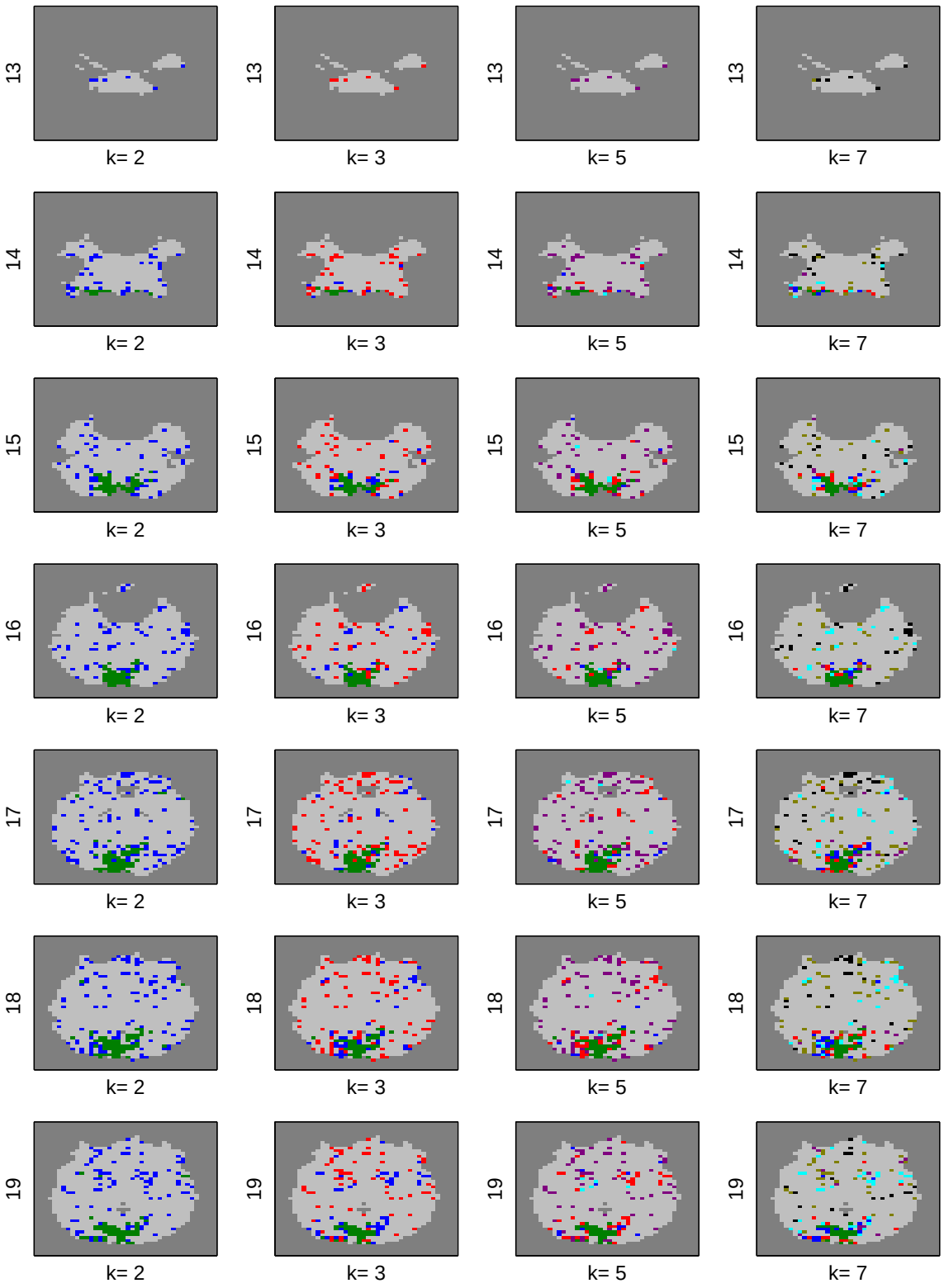


Figure A.1: The clustered brain maps for the antisaccade group for the pre-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

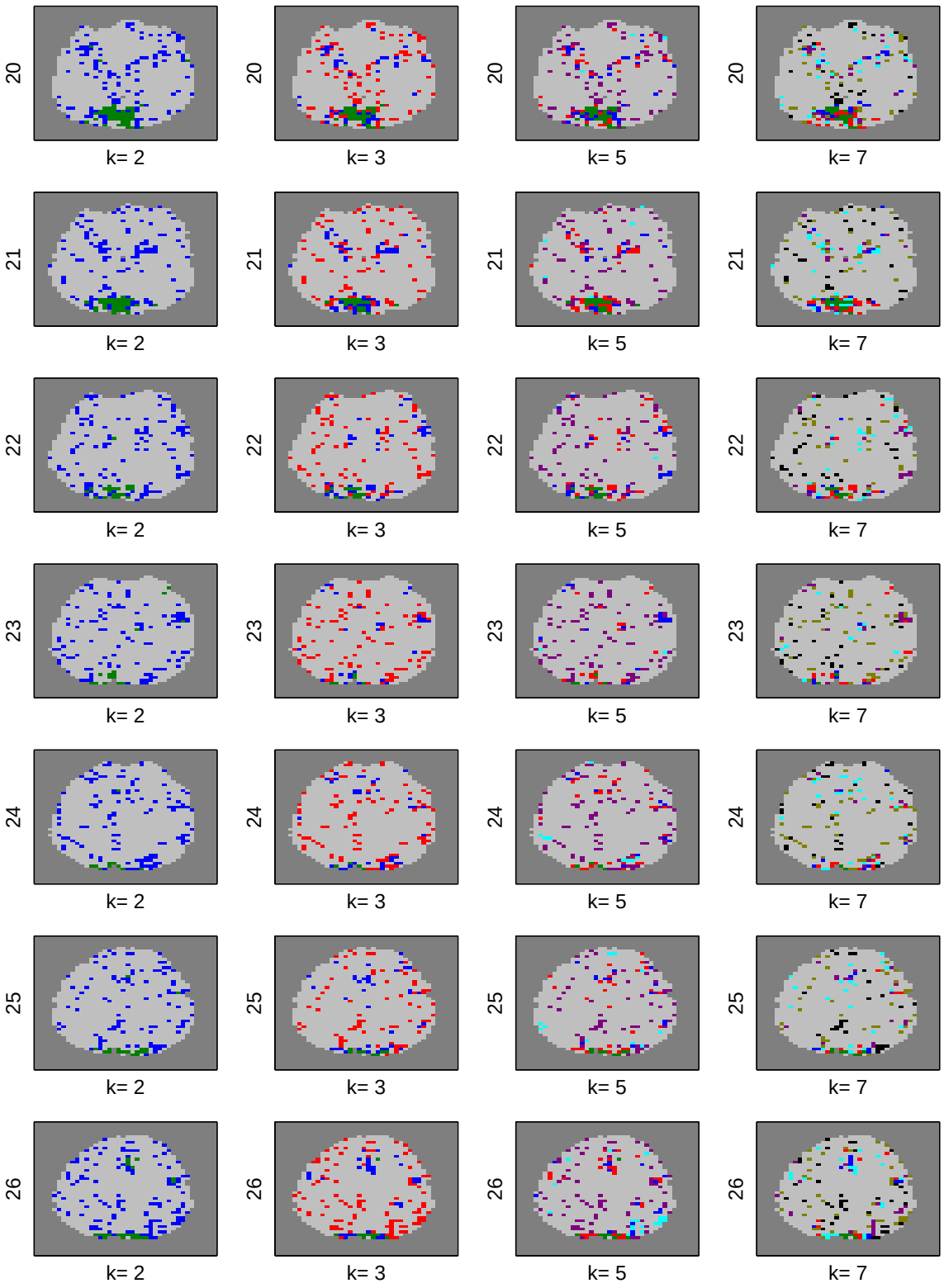


Figure A.2: The clustered brain maps for the antisaccade group for the pre-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

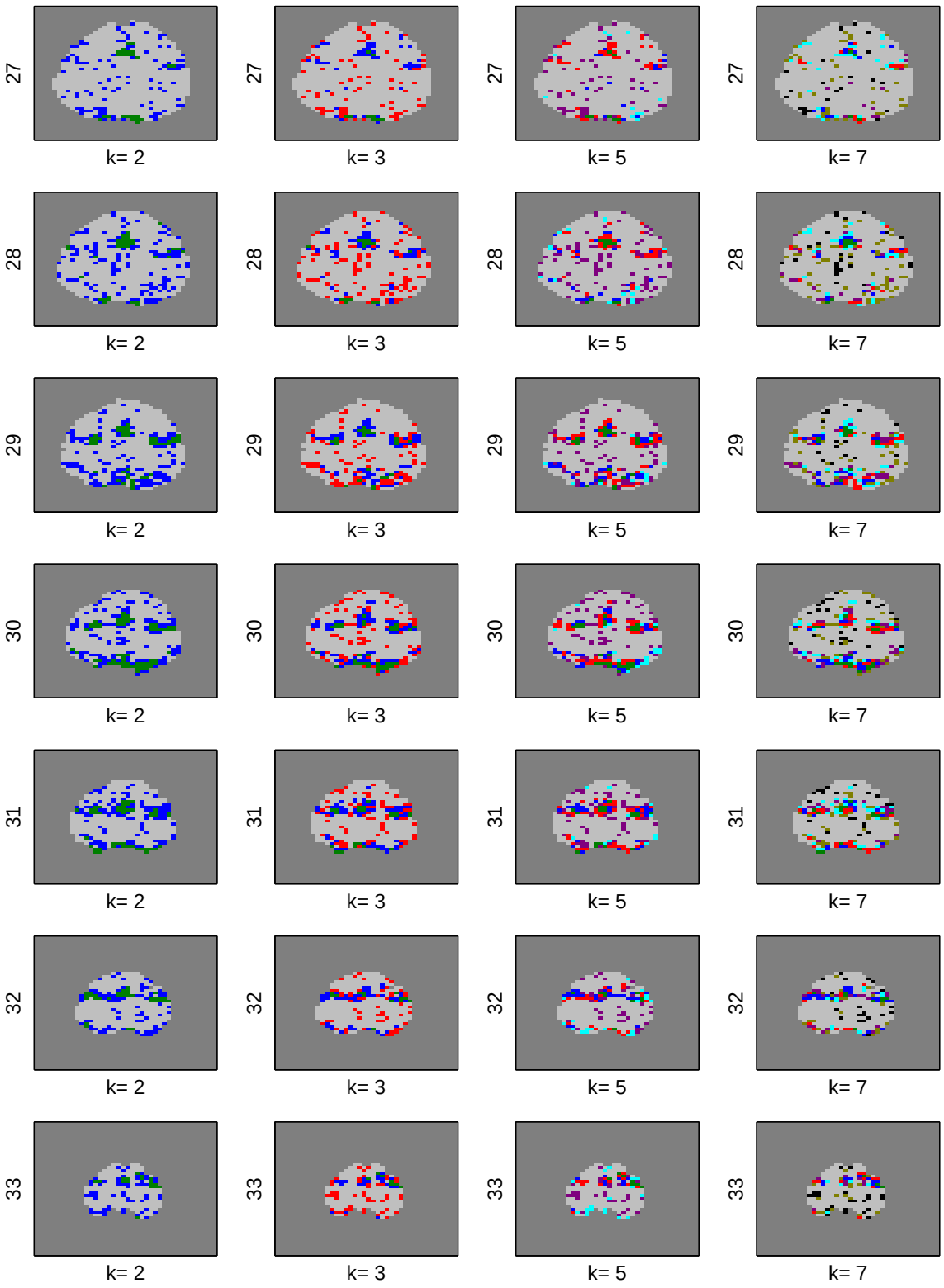


Figure A.3: The clustered brain maps for the antisaccade group for the pre-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

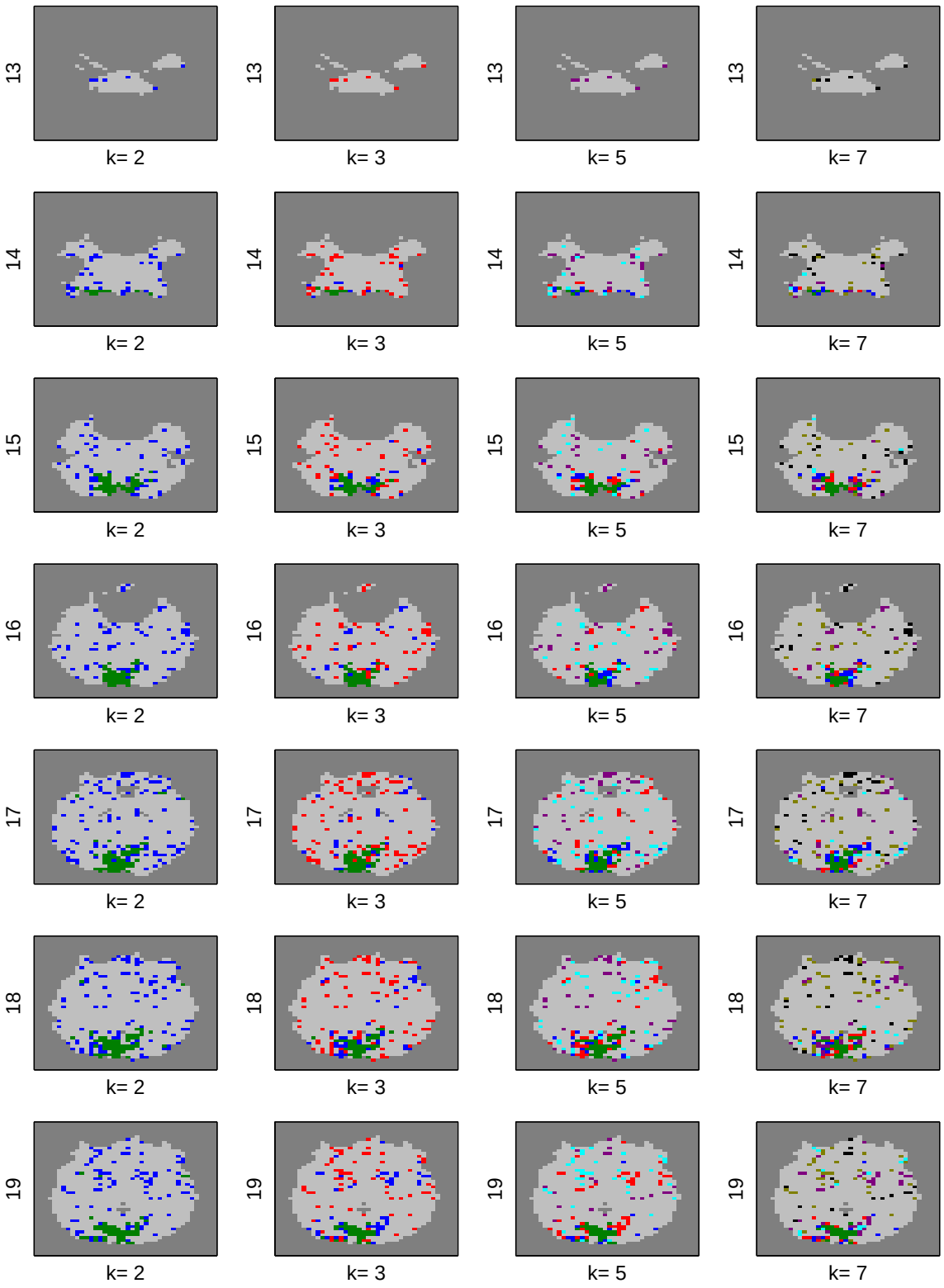


Figure A.4: The clustered brain maps for the antisaccade group for the pre-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

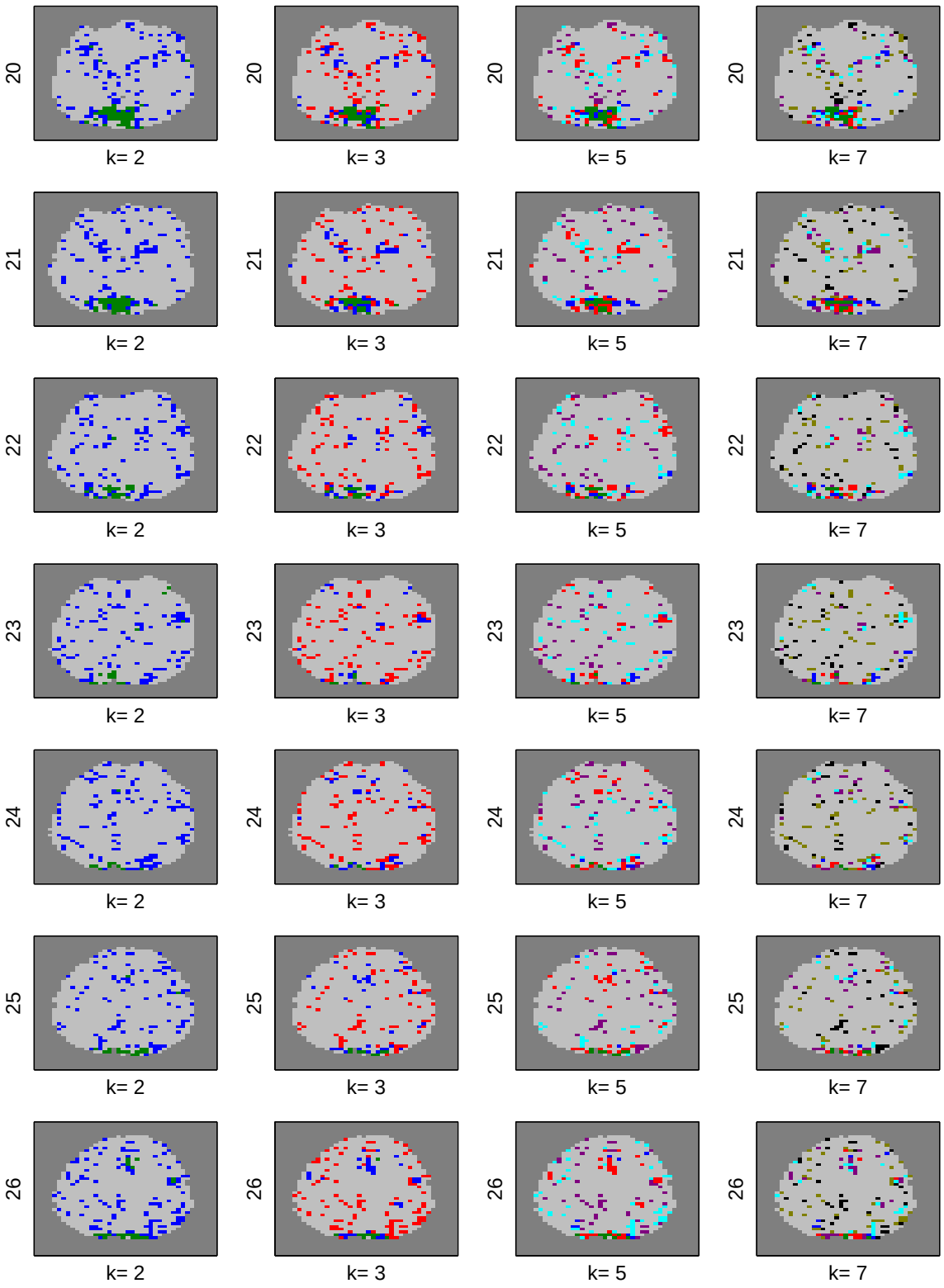


Figure A.5: The clustered brain maps for the antisaccade group for the pre-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

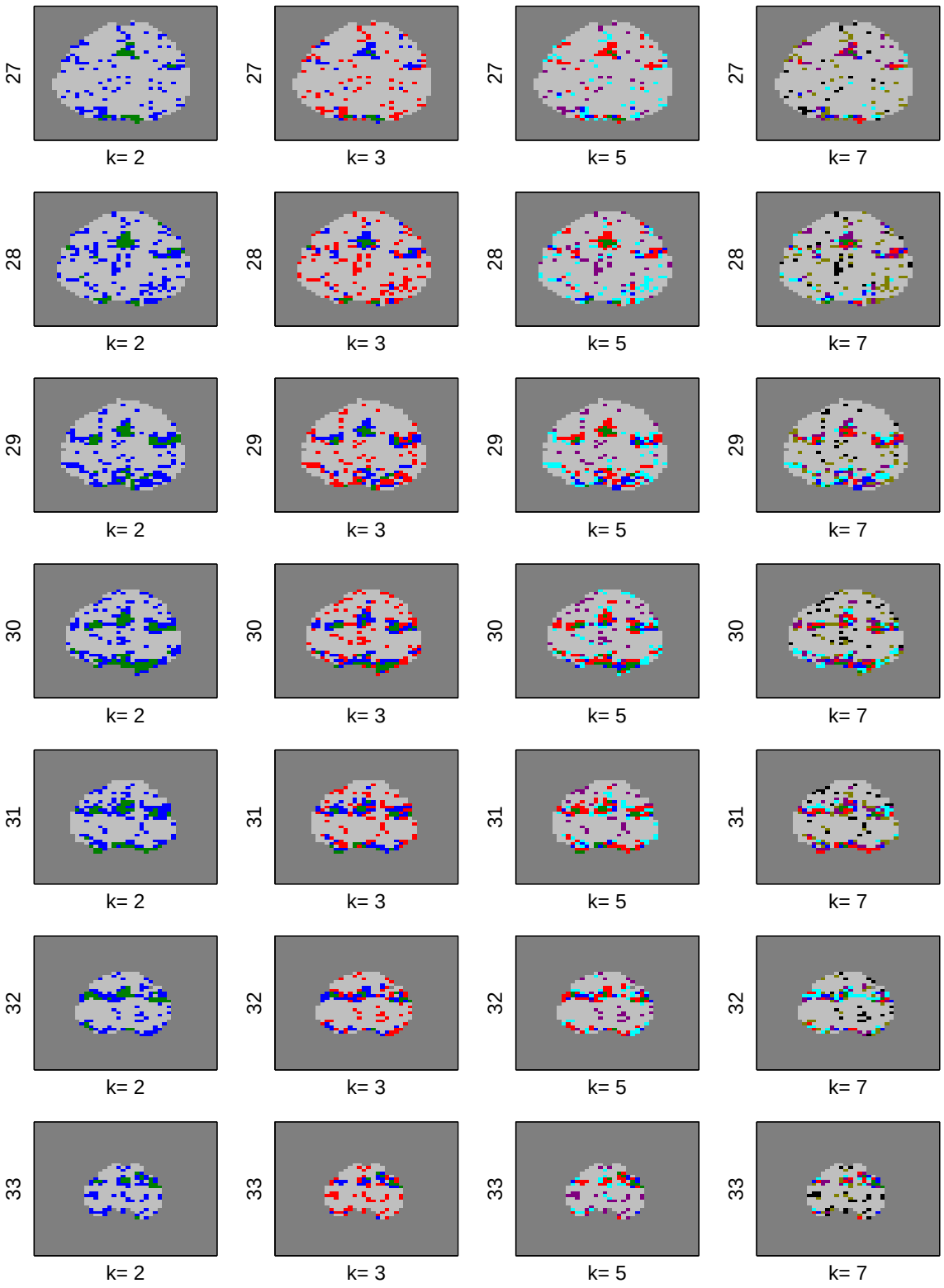


Figure A.6: The clustered brain maps for the antisaccade group for the pre-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

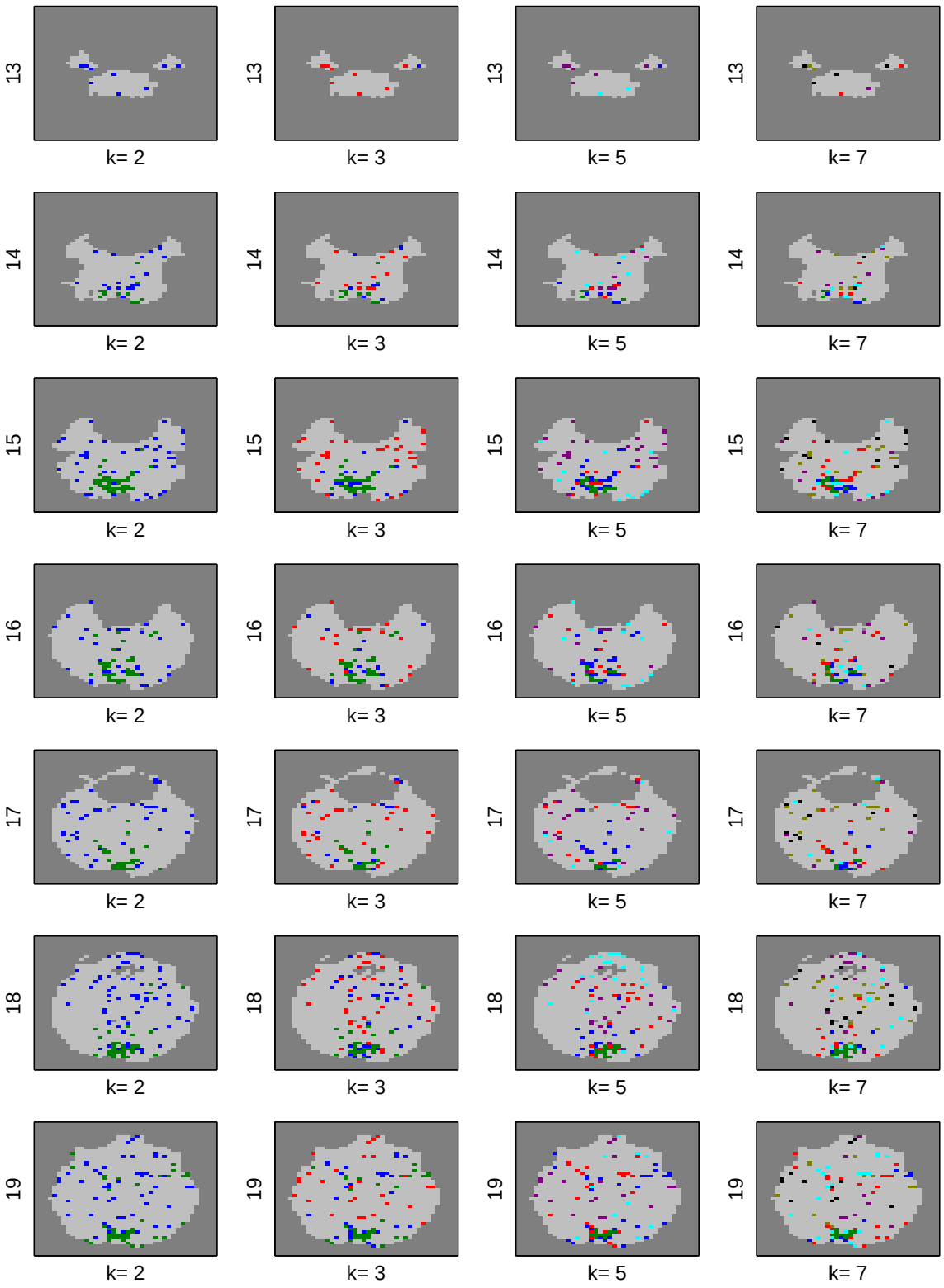


Figure A.7: The clustered brain maps for the antisaccade group for the post-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

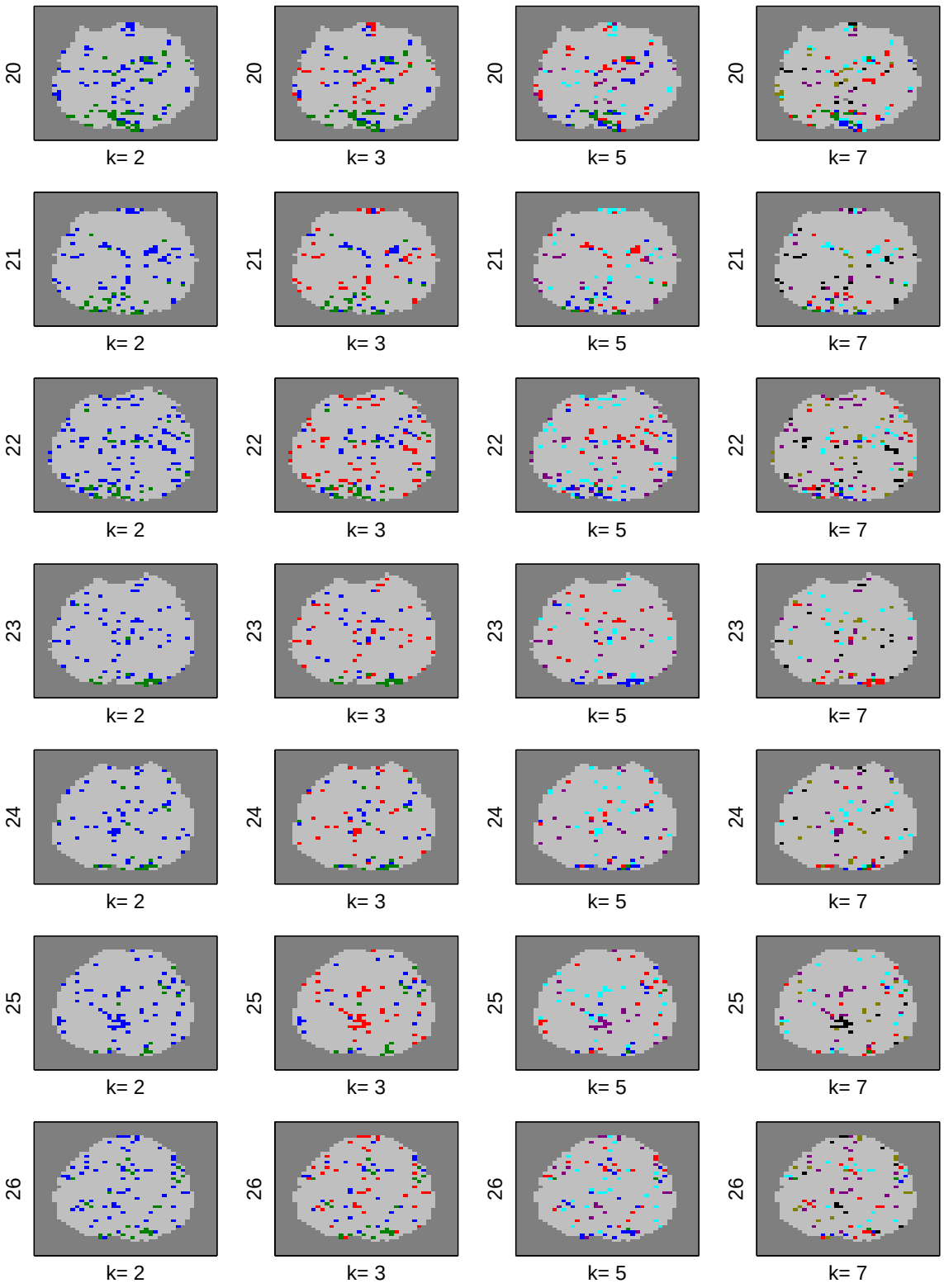


Figure A.8: The clustered brain maps for the antisaccade group for the post-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

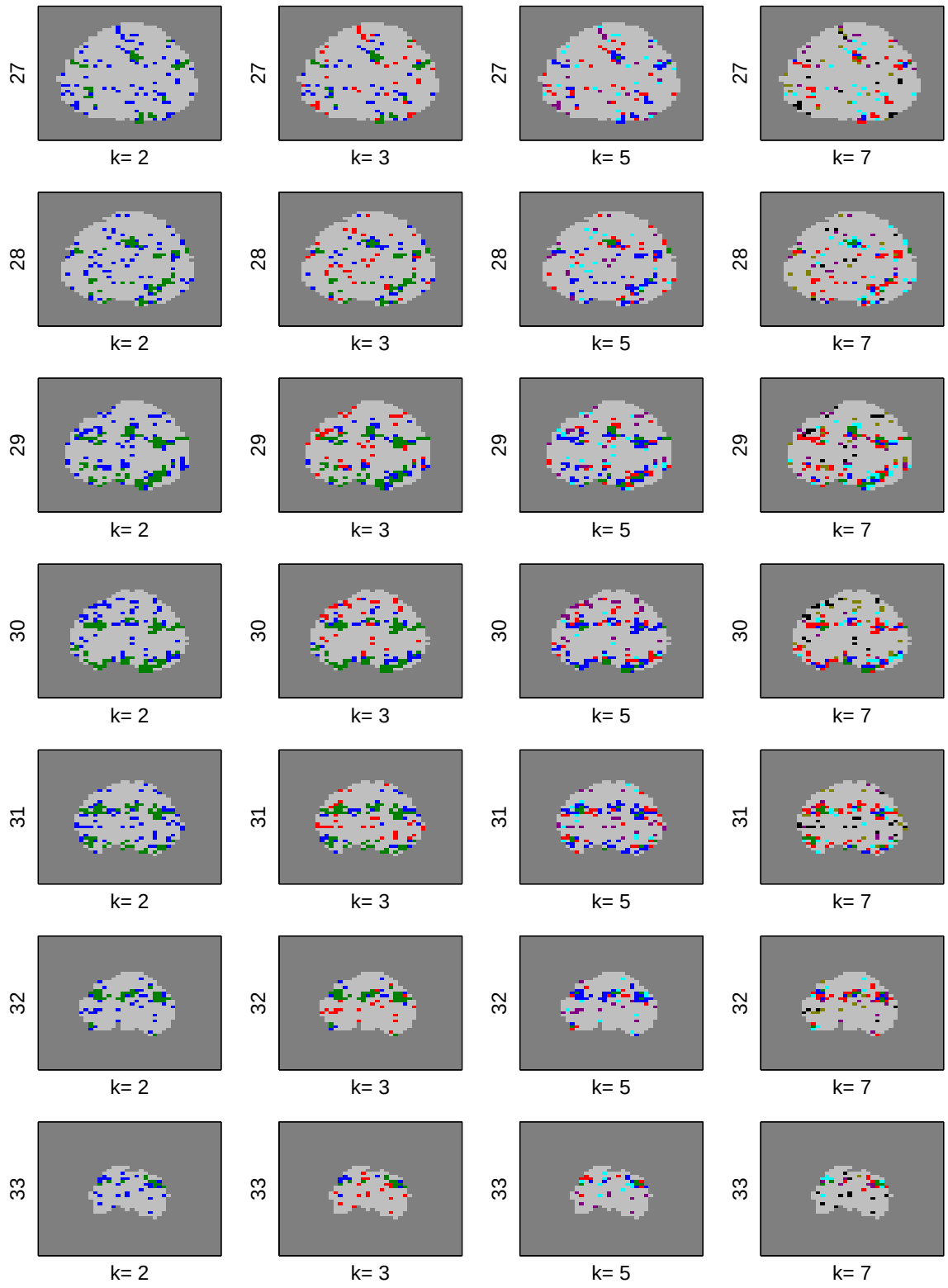


Figure A.9: The clustered brain maps for the antisaccade group for the post-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

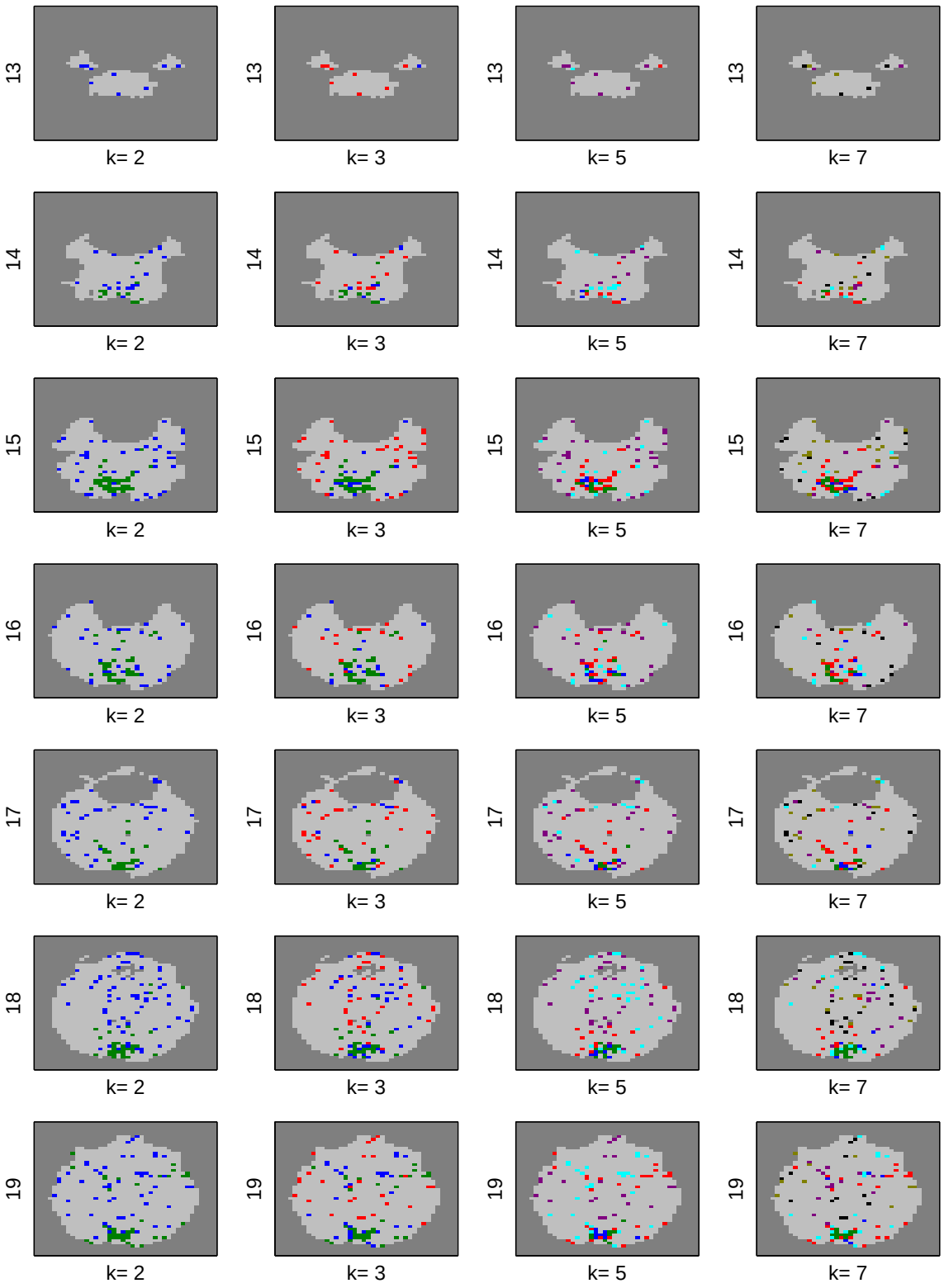


Figure A.10: The clustered brain maps for the antisaccade group for the post-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

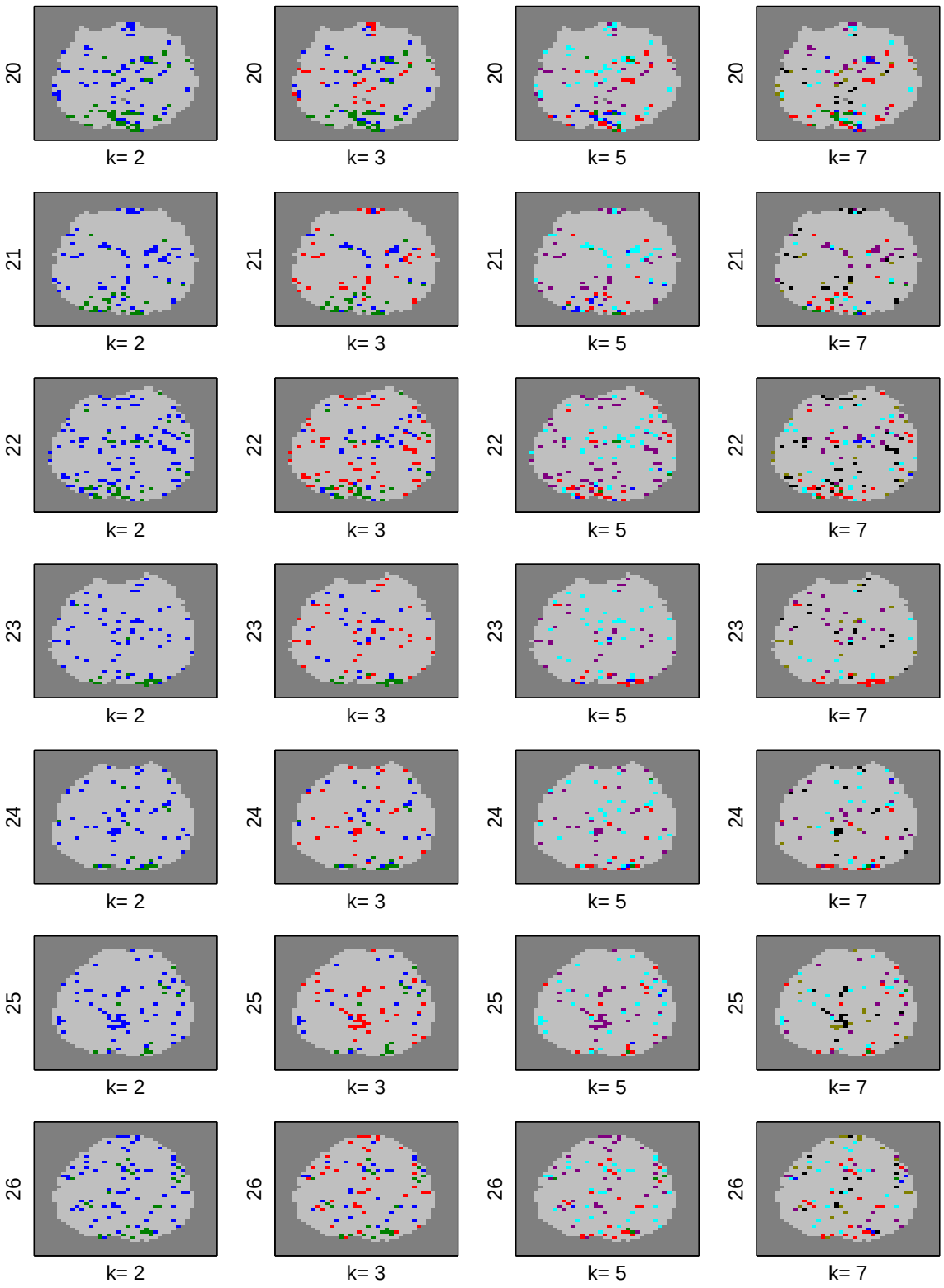


Figure A.11: The clustered brain maps for the antisaccade group for the post-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

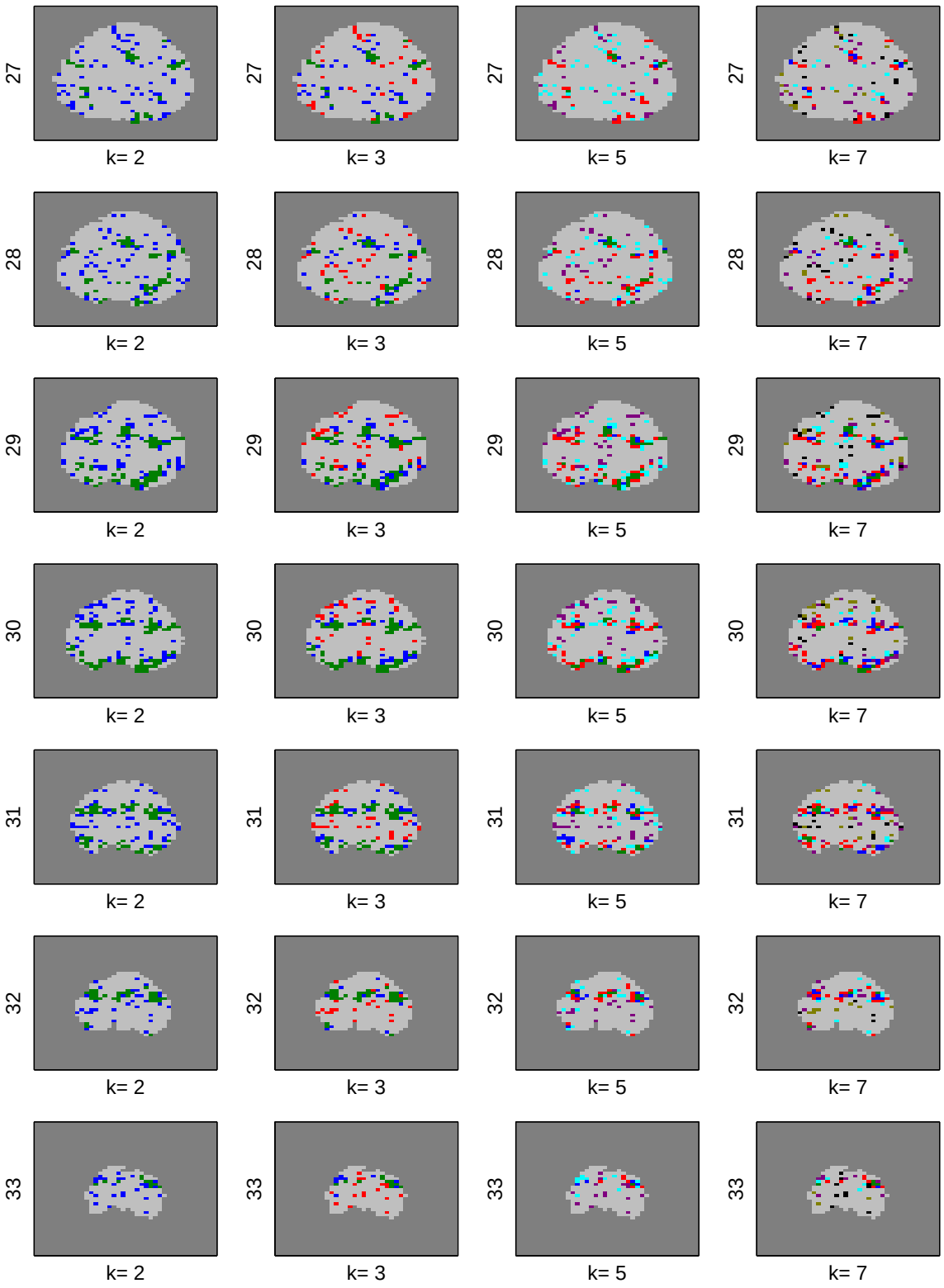
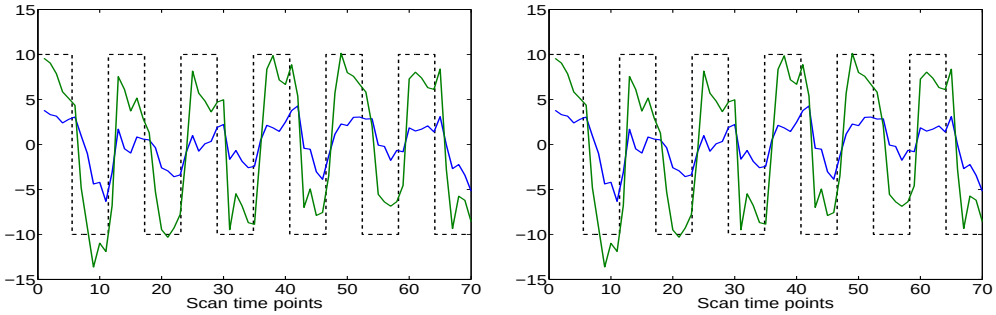
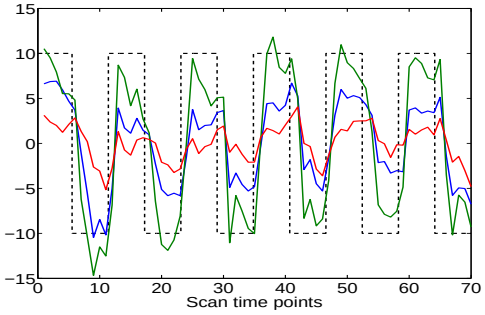


Figure A.12: The clustered brain maps for the antisaccade group for the post-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

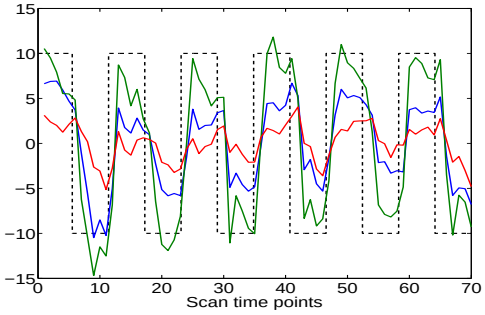


(a) K-means with K=2

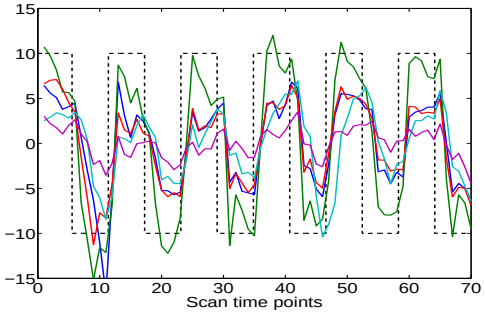
(b) PCA K-means with K=2



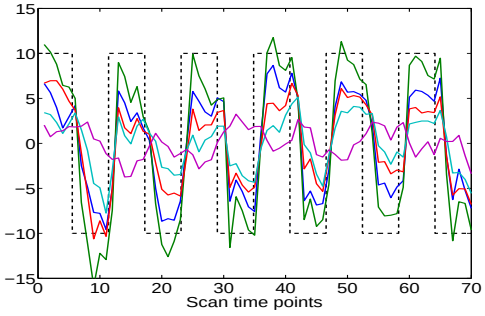
(c) K-means with K=3



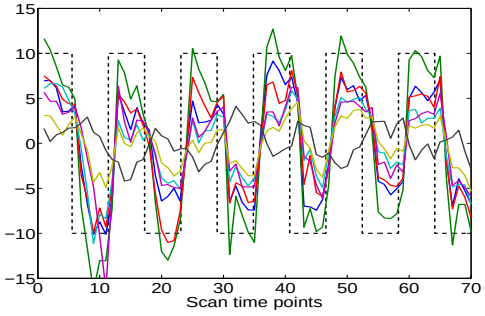
(d) PCA K-means with K=3



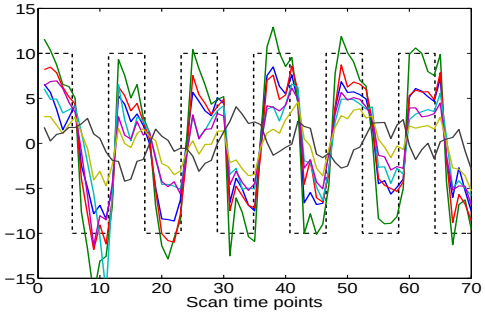
(e) K-means with K=5



(f) PCA K-means with K=5

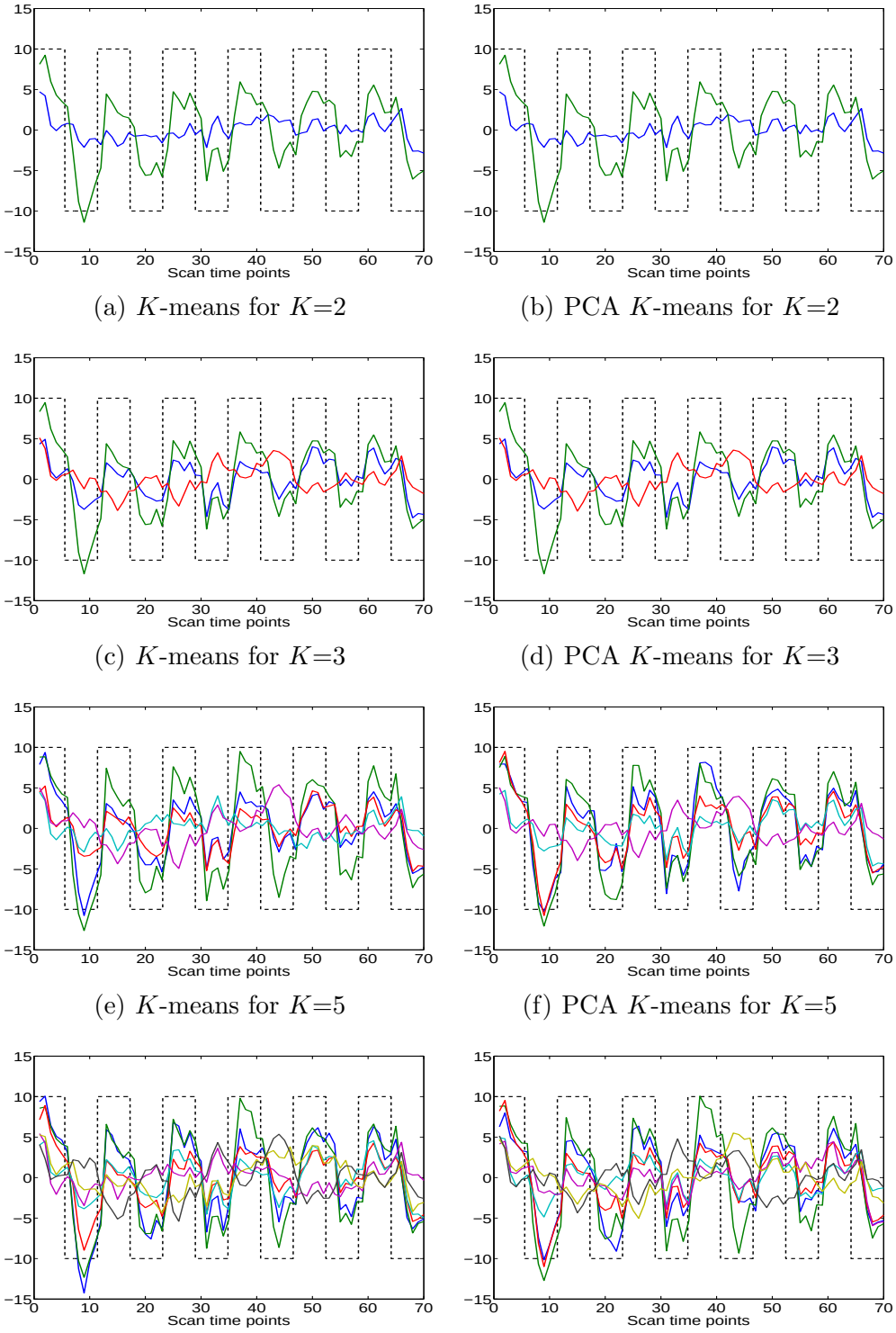


(g) K-means with K=7



(h) PCA K-means with K=7

Figure A.13: Average time series plots for the antisaccade group for the pre-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing.



(a) K-means for K=2

(b) PCA K-means for K=2

(c) K-means for K=3

(d) PCA K-means for K=3

(e) K-means for K=5

(f) PCA K-means for K=5

(g) K-means for K=7

(h) PCA K-means for K=7

Figure A.14: Average time series plots for the antisaccade group for the post-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing

Appendix B

Clustering results for the Prosaccade task group

In this appendix, we present the clustering results for the prosaccade task group. For each test session (pre- and post-test) the brain is partitioned into two, three, five or seven clusters, which are K=2, 3, 5, and 7, using K-means and PCA K-means clustering methods.
The clustered brain maps are listed in the following order:

- K-means method for the pre-test for slices 13 – 19 in Figure B.1.
- K-means method for the pre-test for slices 20 – 26 in Figure B.2.
- K-means method for the pre-test for slices 27 – 33 in Figure B.3.
- PCA K-means method for the pre-test for slices 13 – 19 in Figure B.4.
- PCA K-means method for the pre-test for slices 20 – 26 in Figure B.5.
- PCA K-means method for the pre-test for slices 27 – 33 in Figure B.6.
- K-means method for the post-test for slices 13 – 19 in Figure B.7.
- K-means method for the post-test for slices 20 – 26 in Figure B.8.
- K-means method for the post-test for slices 27 – 33 in Figure B.9.
- PCA K-means method for the post-test for slices 13 – 19 in Figure B.10.
- PCA K-means method for the post-test for slices 20 – 26 in Figure B.11.
- PCA K-means method for the post-test for slices 27 – 33 in Figure B.12.

The average time series plots corresponding to each cluster in the clustered brain maps are listed in the following order:

- The pre-test using K-means and PCA K-means method in Figure B.13.
- The post-test using K-means and PCA K-means method in Figure B.14.

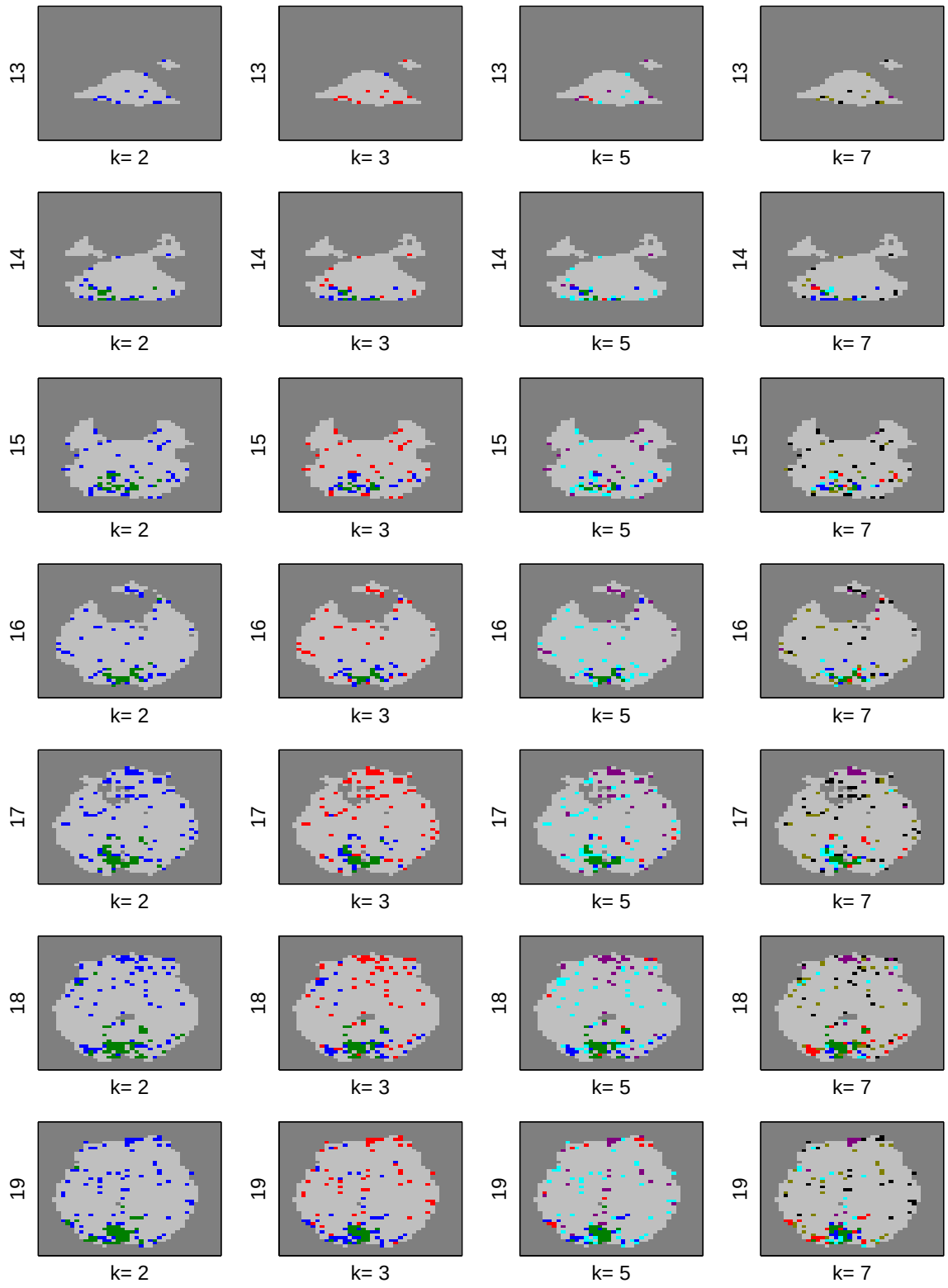


Figure B.1: The clustered brain maps for the prosaccade group for the pre-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

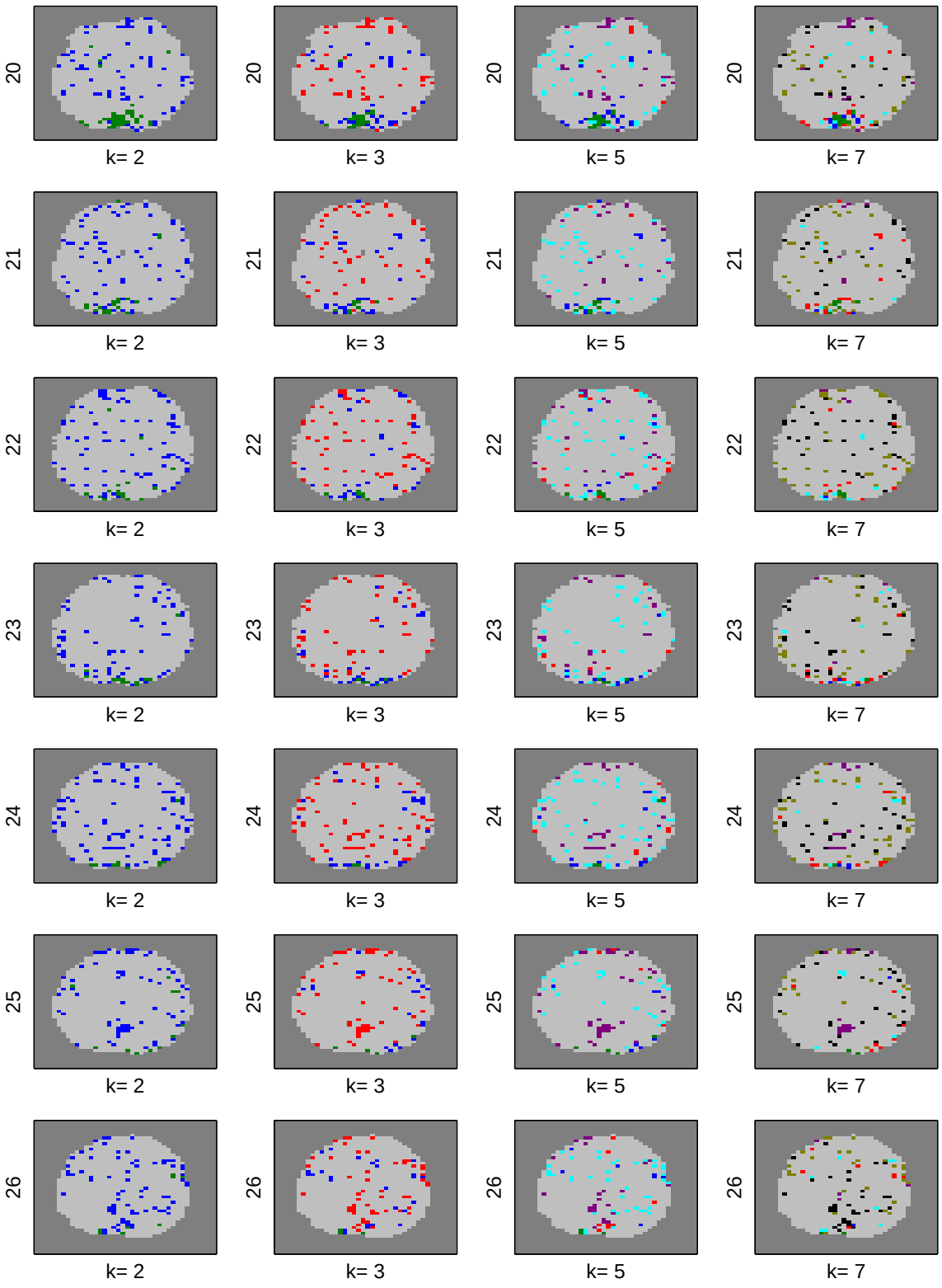


Figure B.2: The clustered brain maps for the prosaccade group for the pre-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

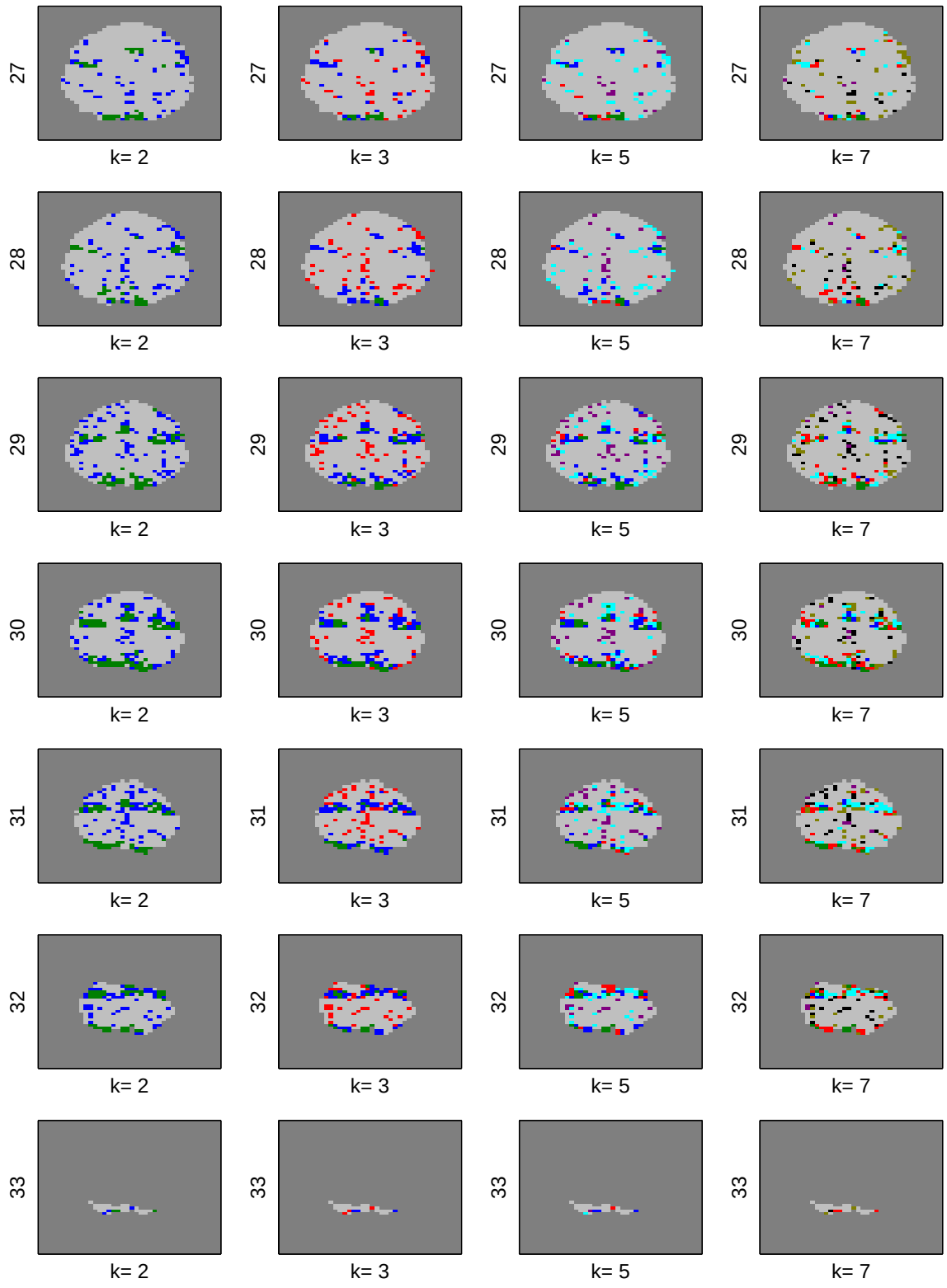


Figure B.3: The clustered brain maps for the prosaccade group for the pre-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

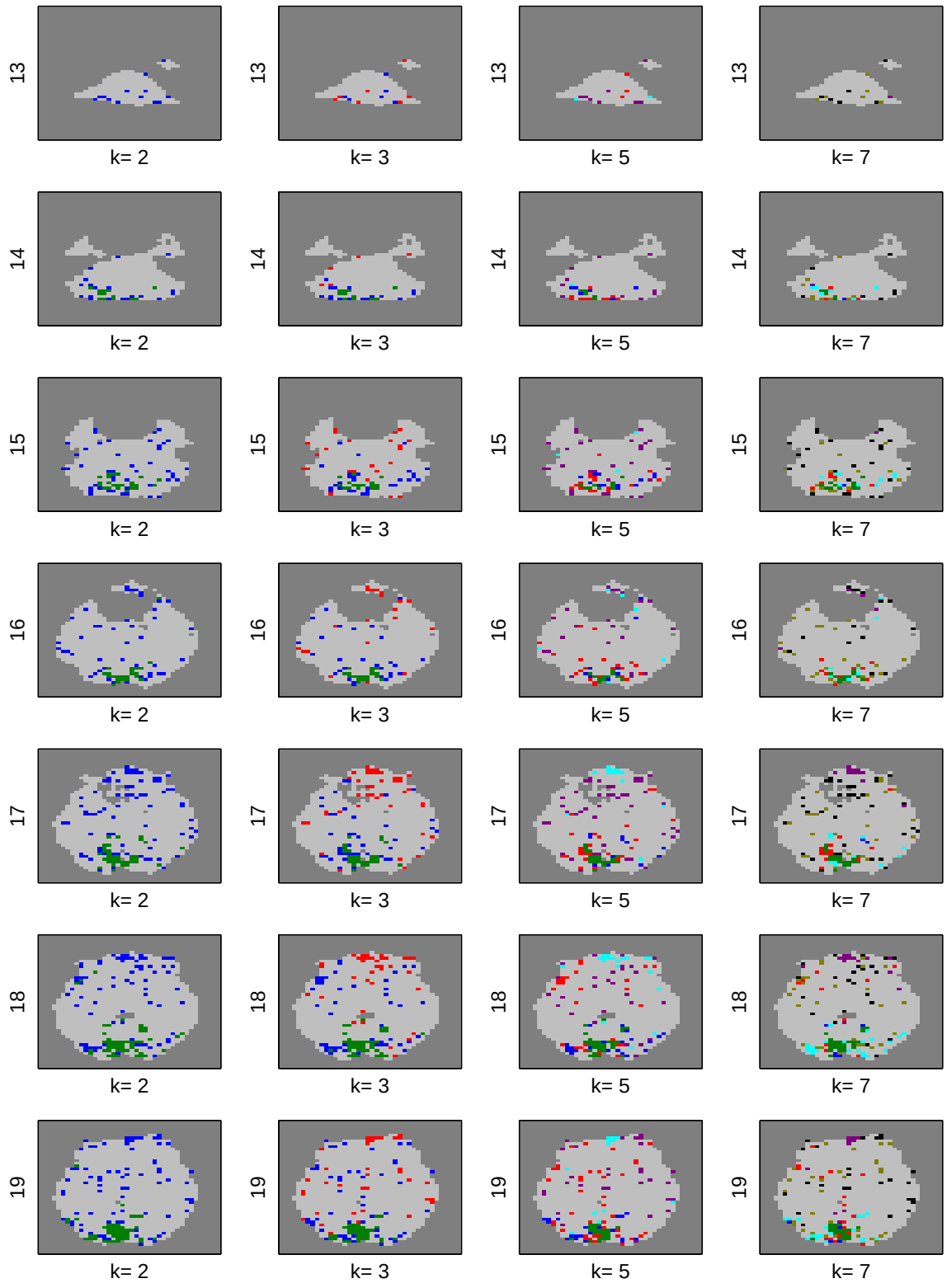


Figure B.4: The clustered brain maps for the prosaccade group for the pre-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

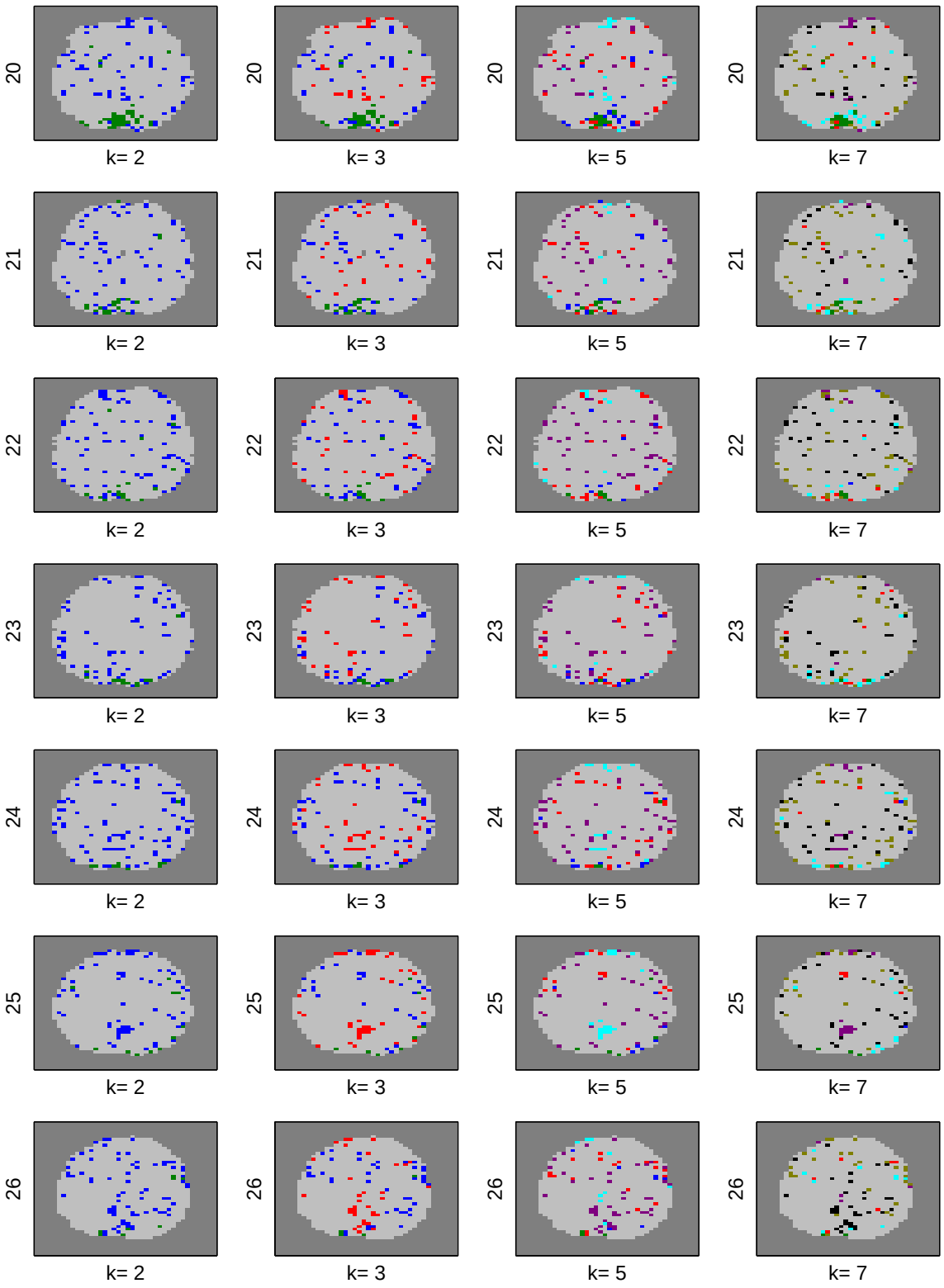


Figure B.5: The clustered brain maps for the prosaccade group for the pre-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

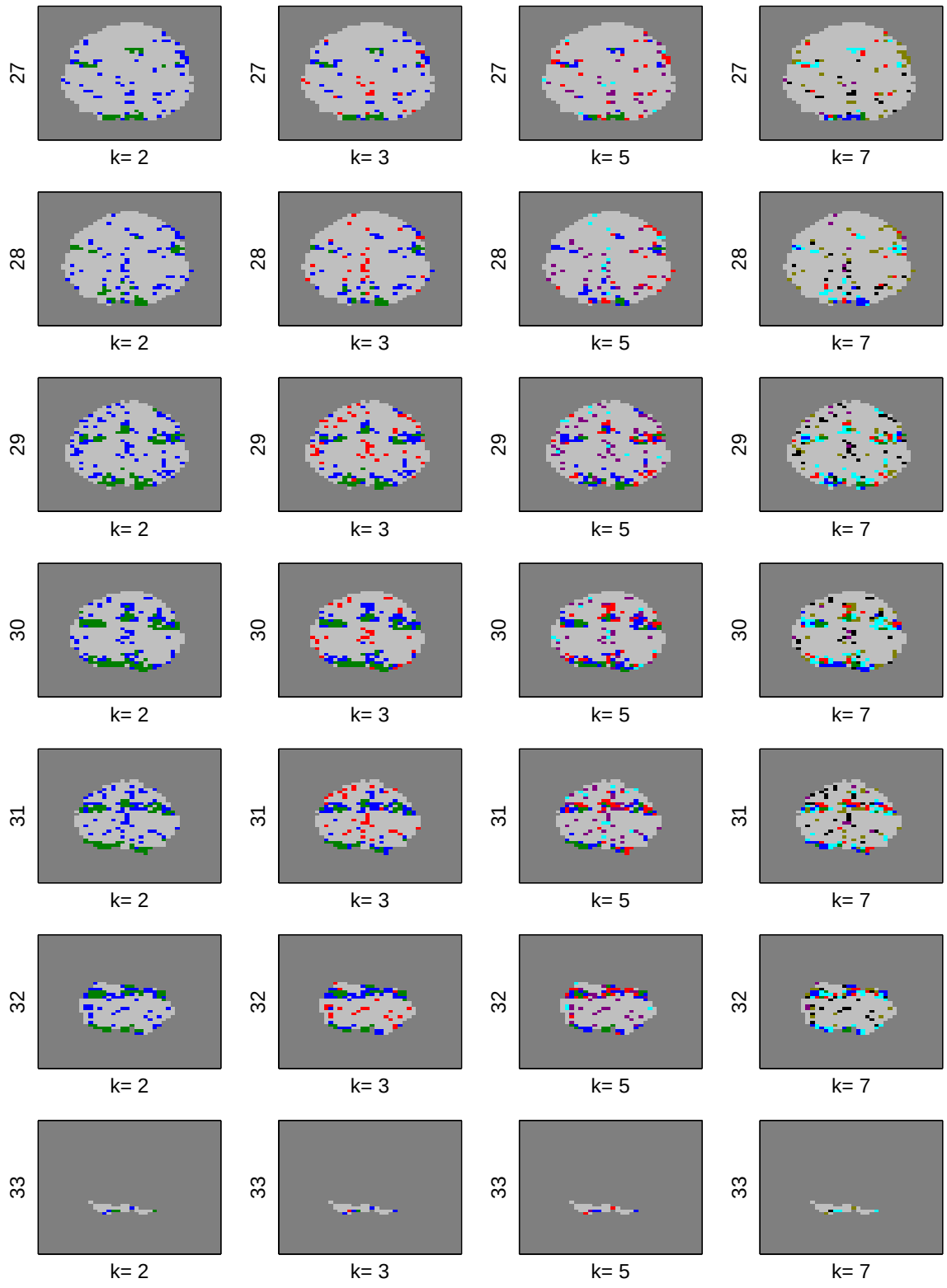


Figure B.6: The clustered brain maps for the prosaccade group for the pre-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

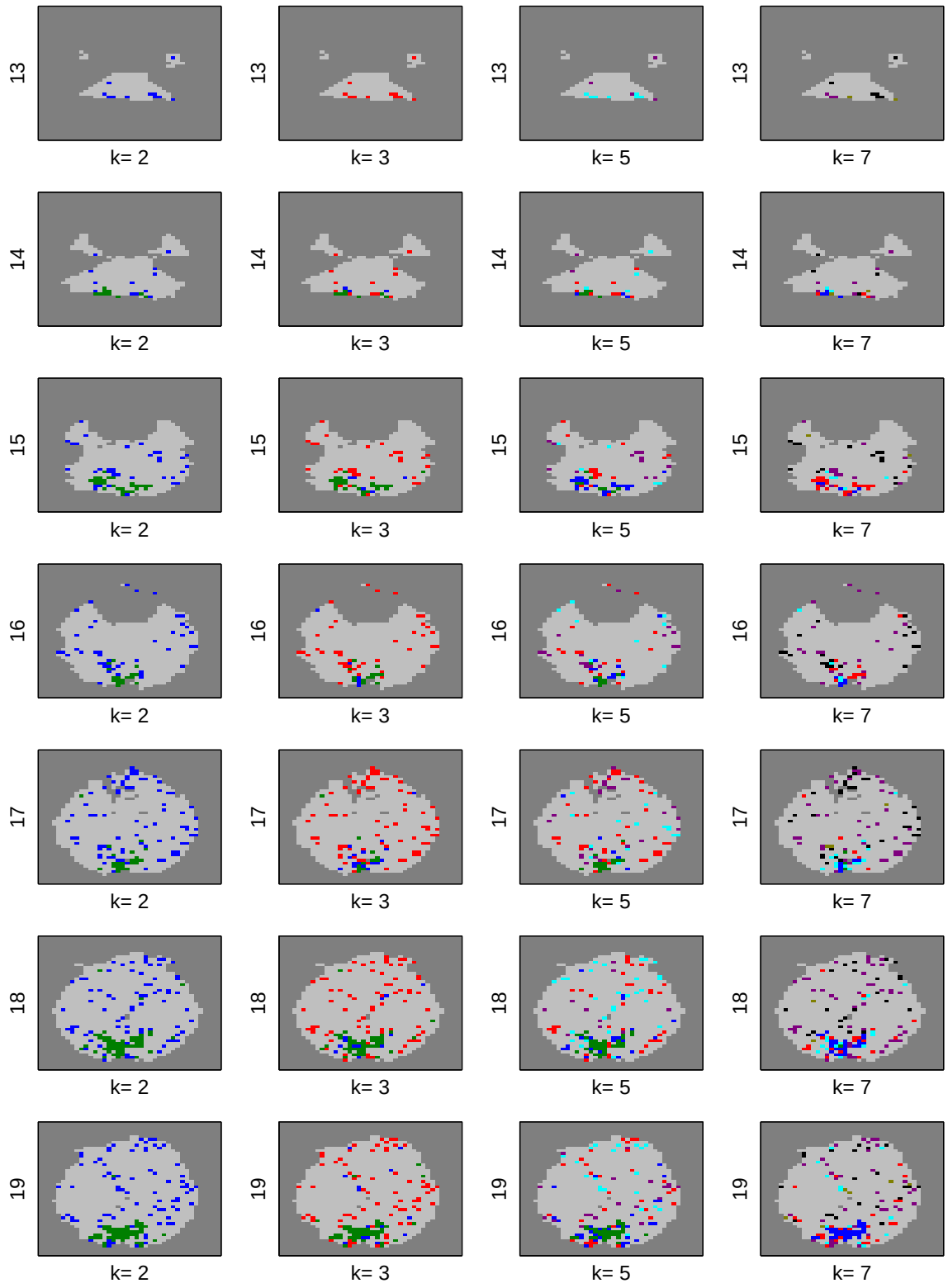


Figure B.7: The clustered brain maps for the prosaccade group for the post-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

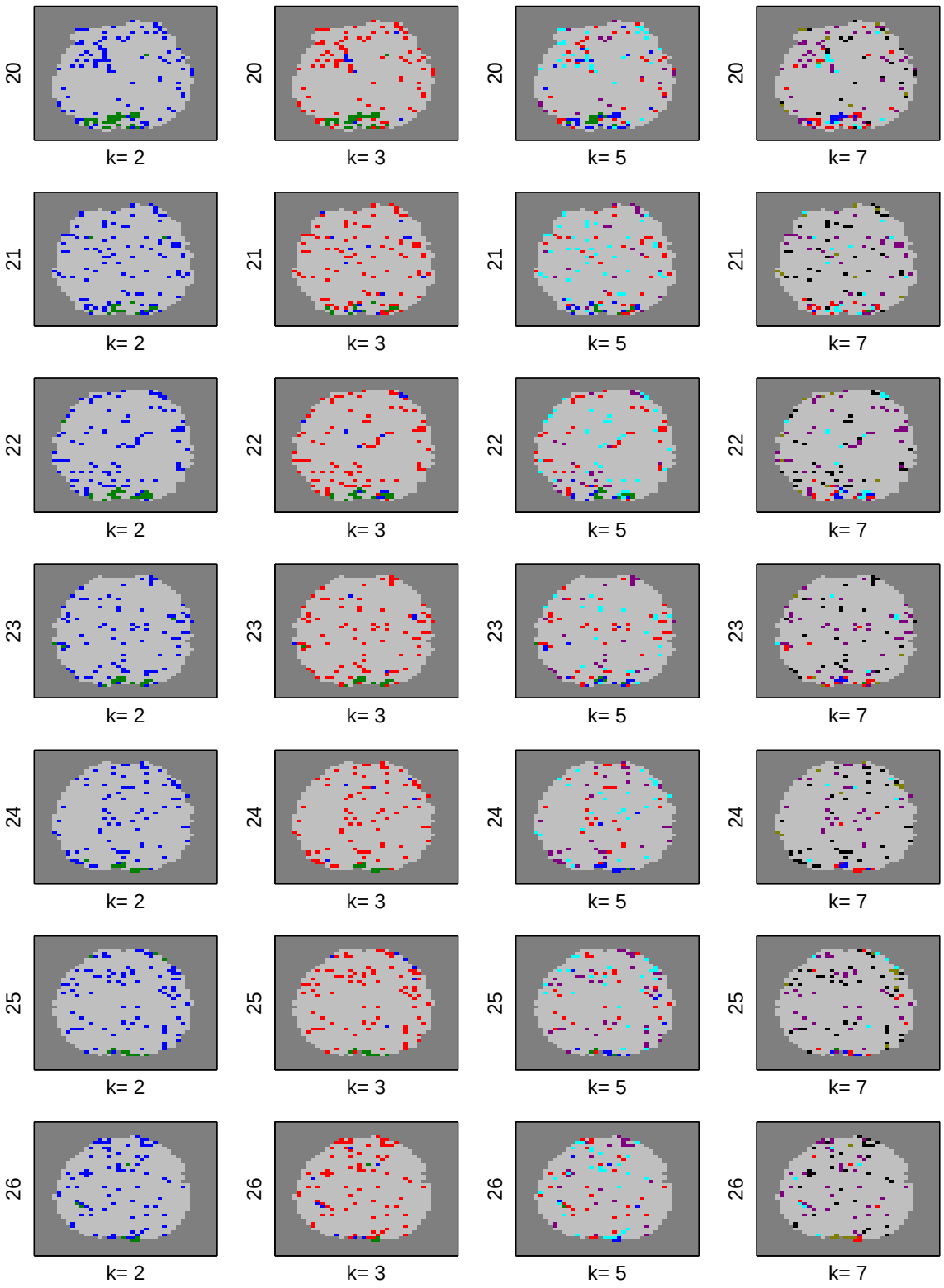


Figure B.8: The clustered brain maps for the prosaccade group for the post-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

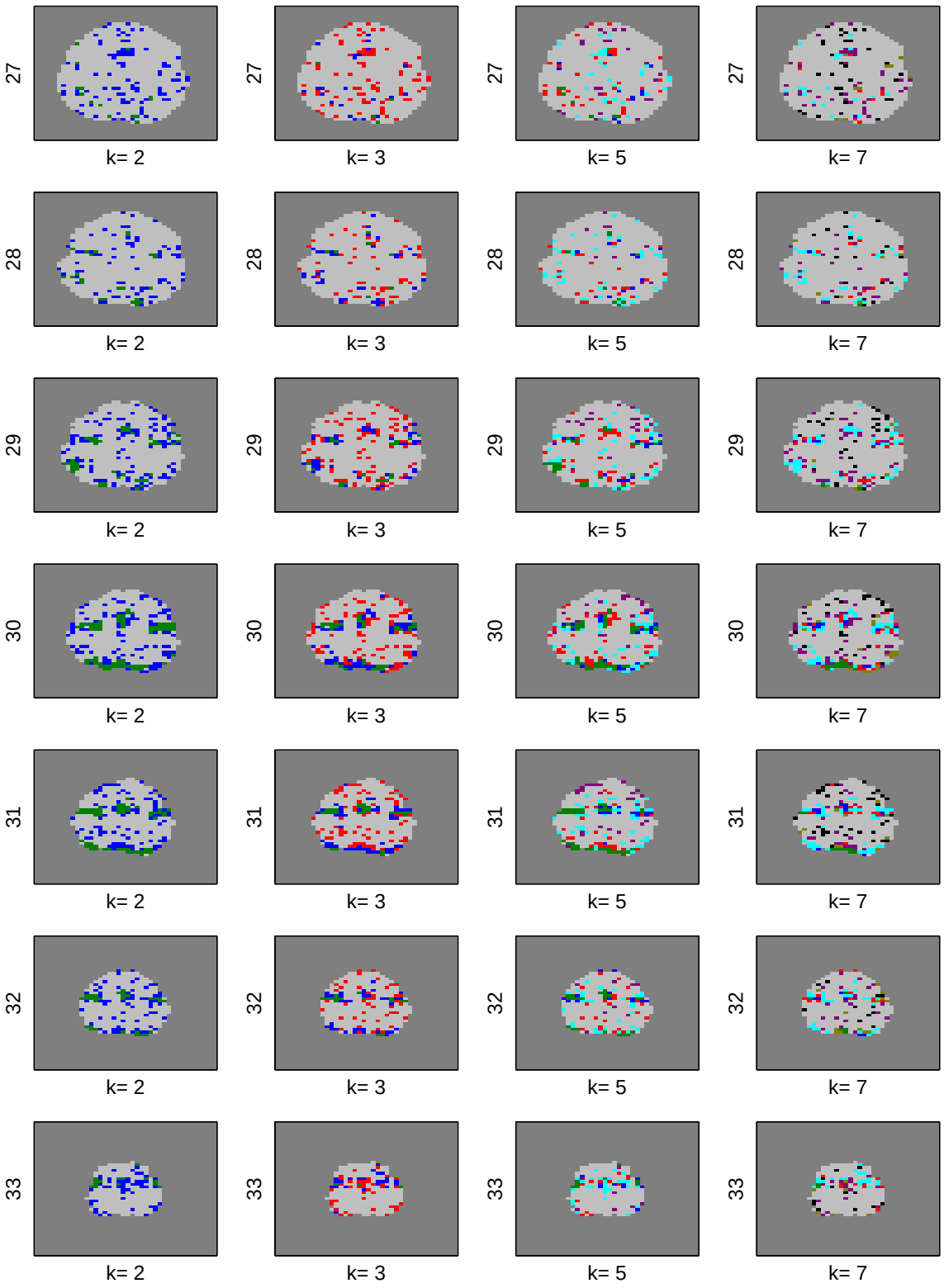


Figure B.9: The clustered brain maps for the prosaccade group for the post-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

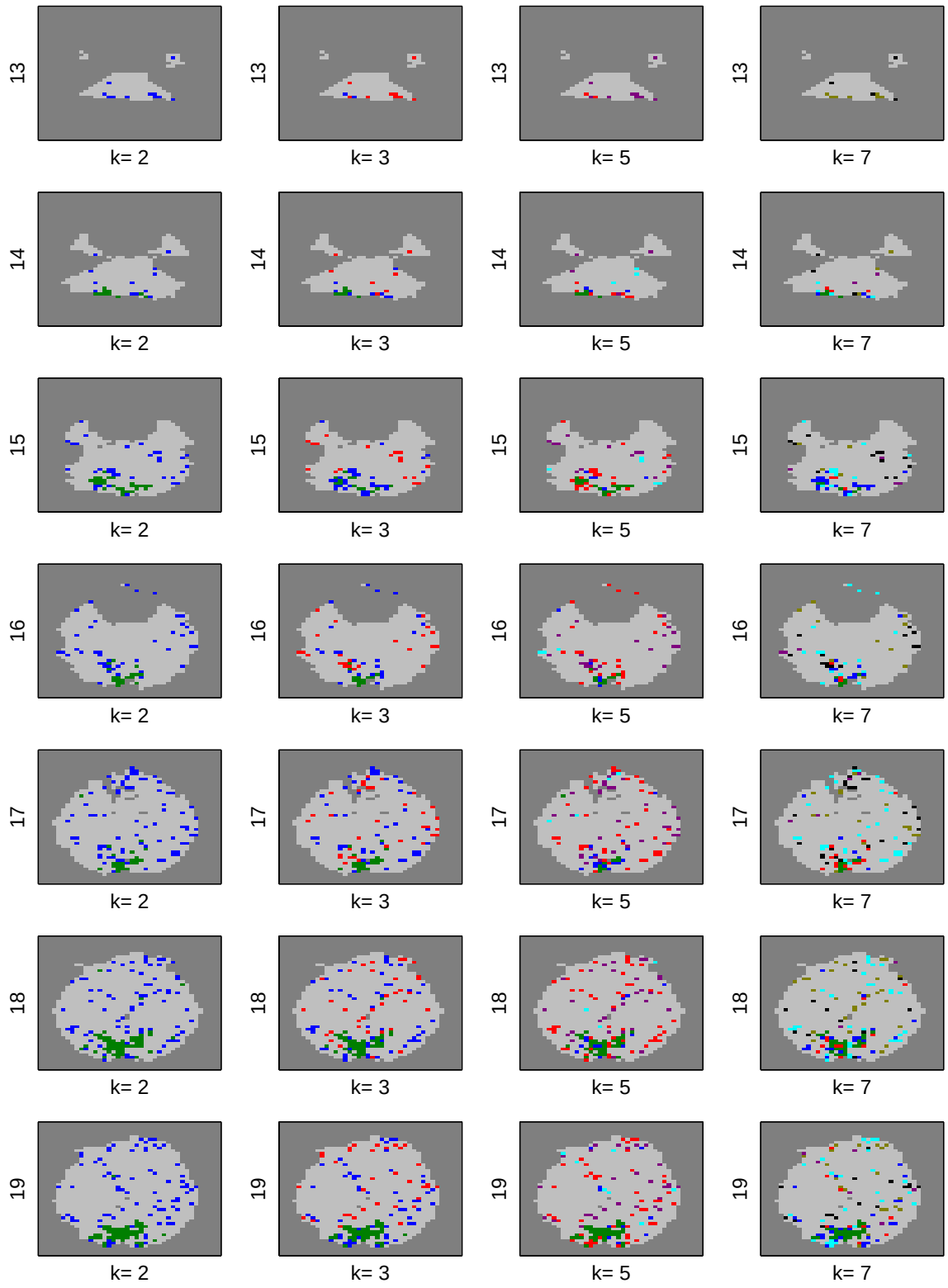


Figure B.10: The clustered brain maps for the prosaccade group for the post-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

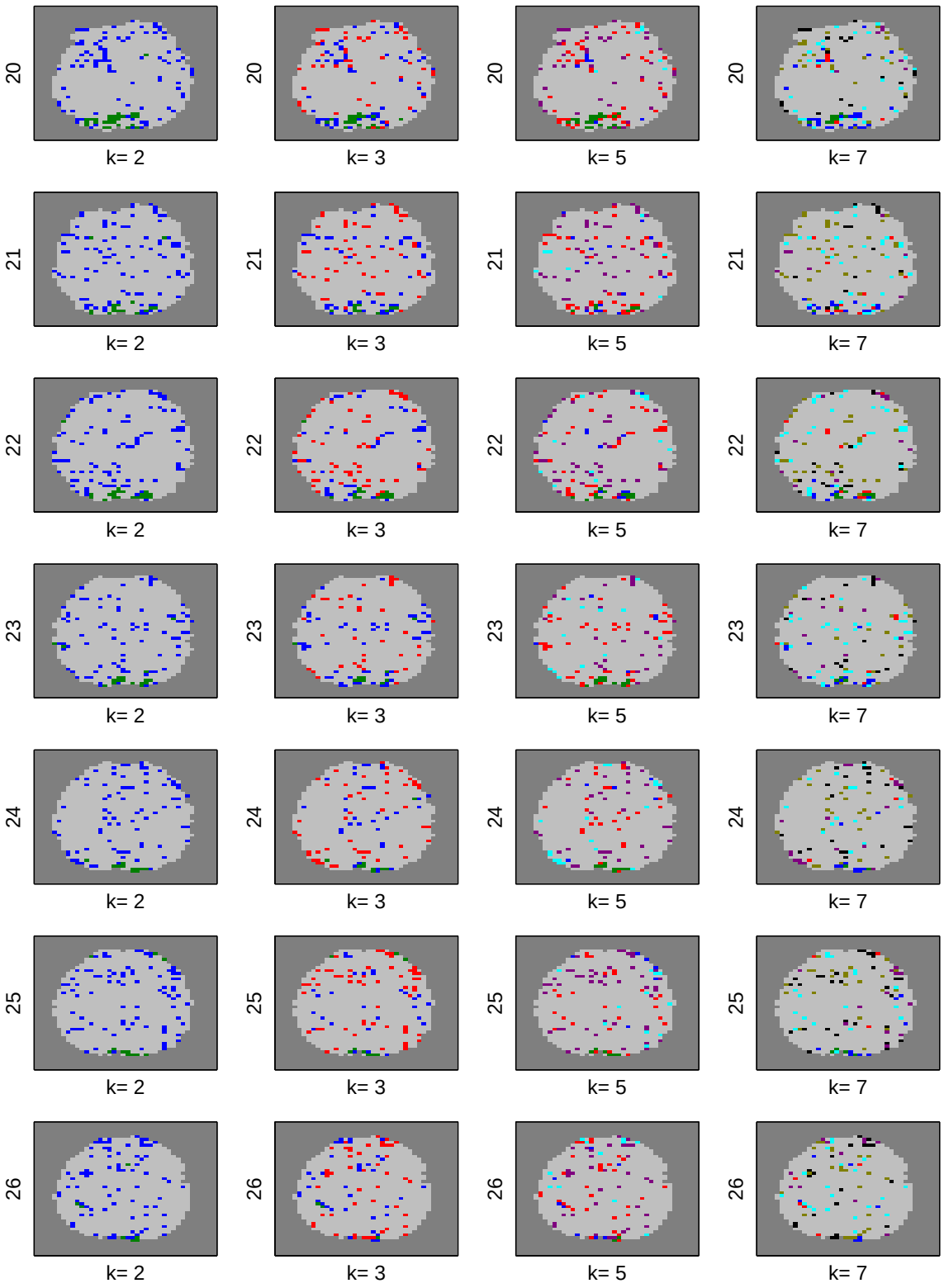


Figure B.11: The clustered brain maps for the prosaccade group for the post-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

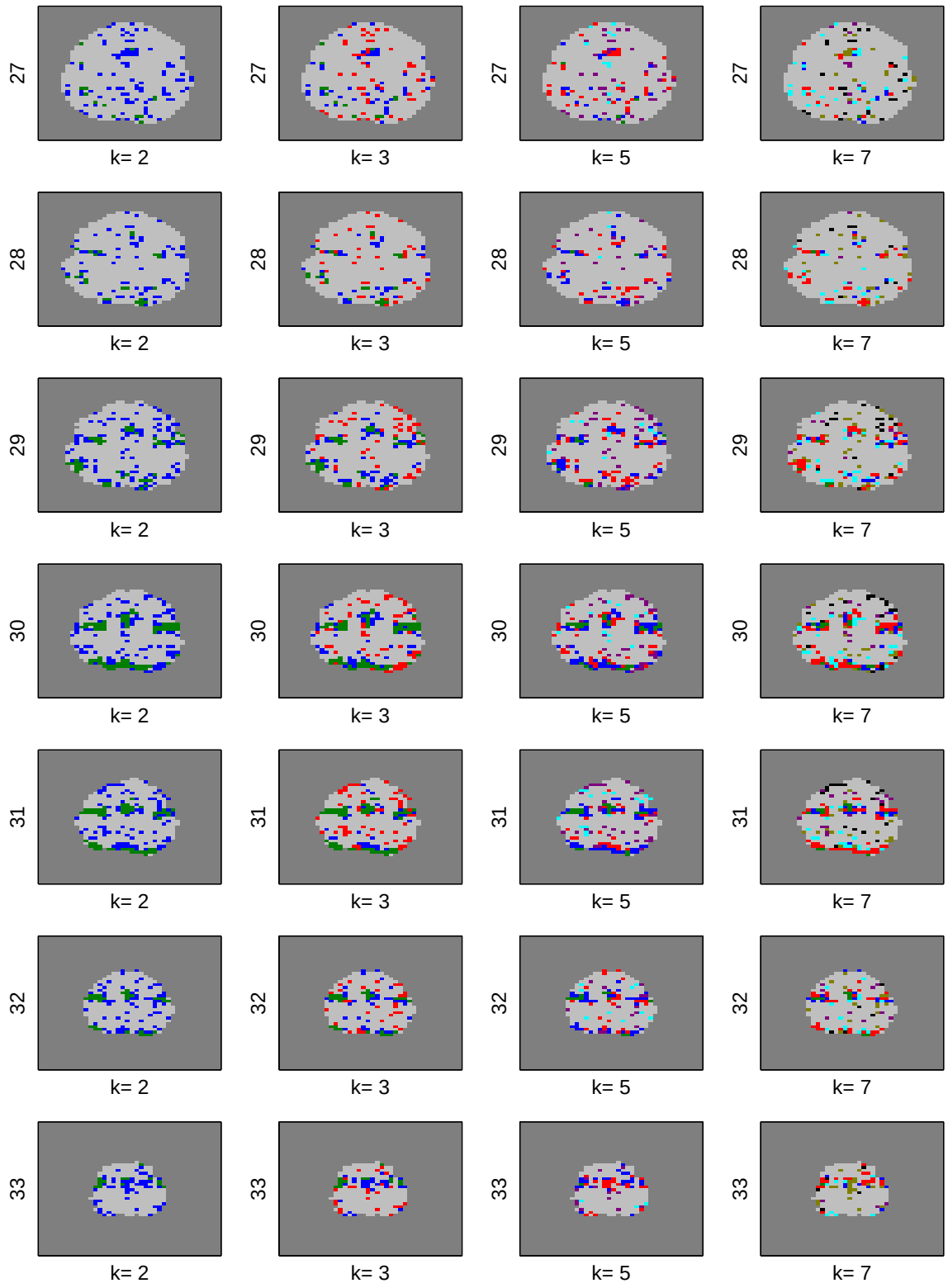
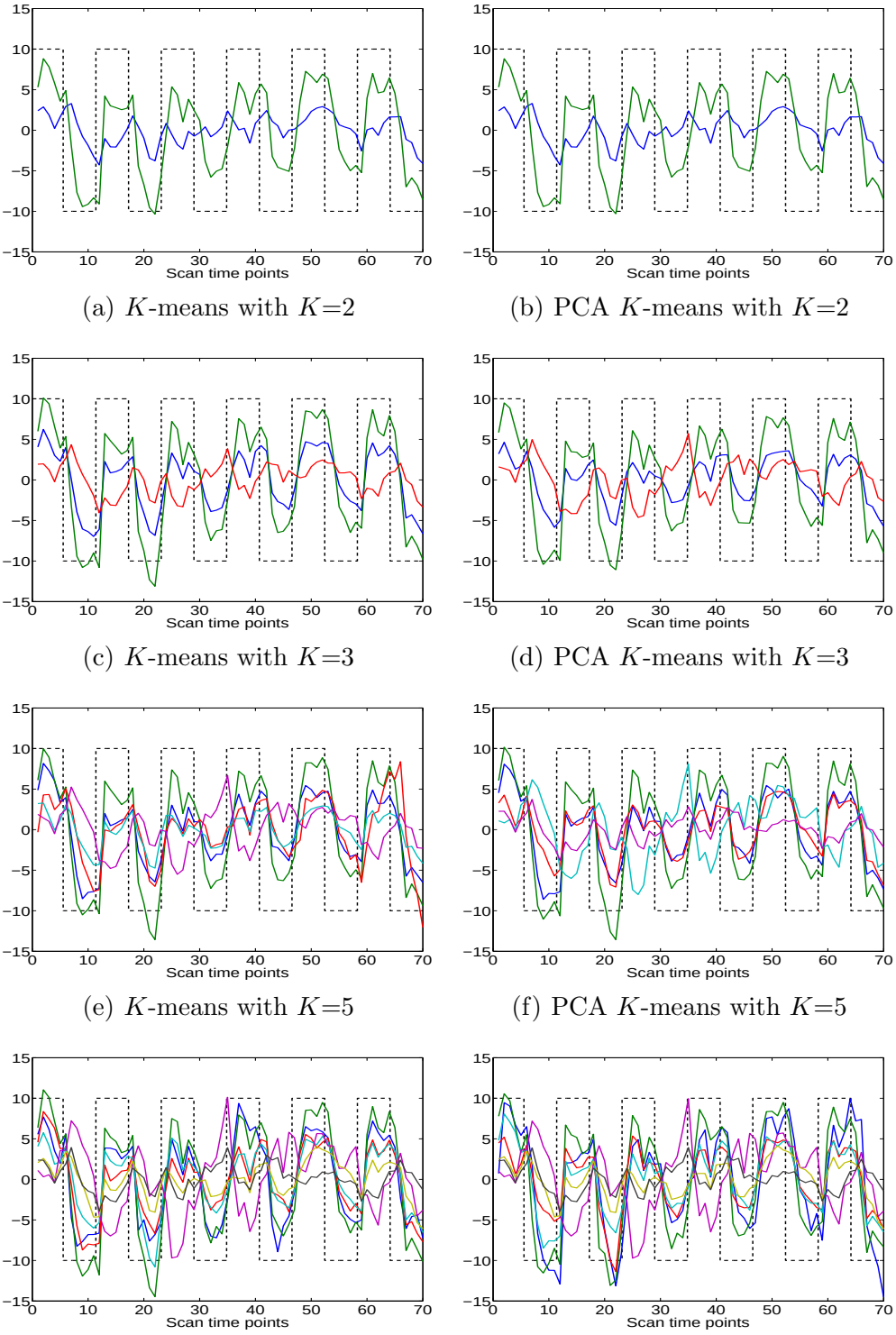


Figure B.12: The clustered brain maps for the prosaccade group for the post-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.



(a) K-means with K=2

(b) PCA K-means with K=2

(c) K-means with K=3

(d) PCA K-means with K=3

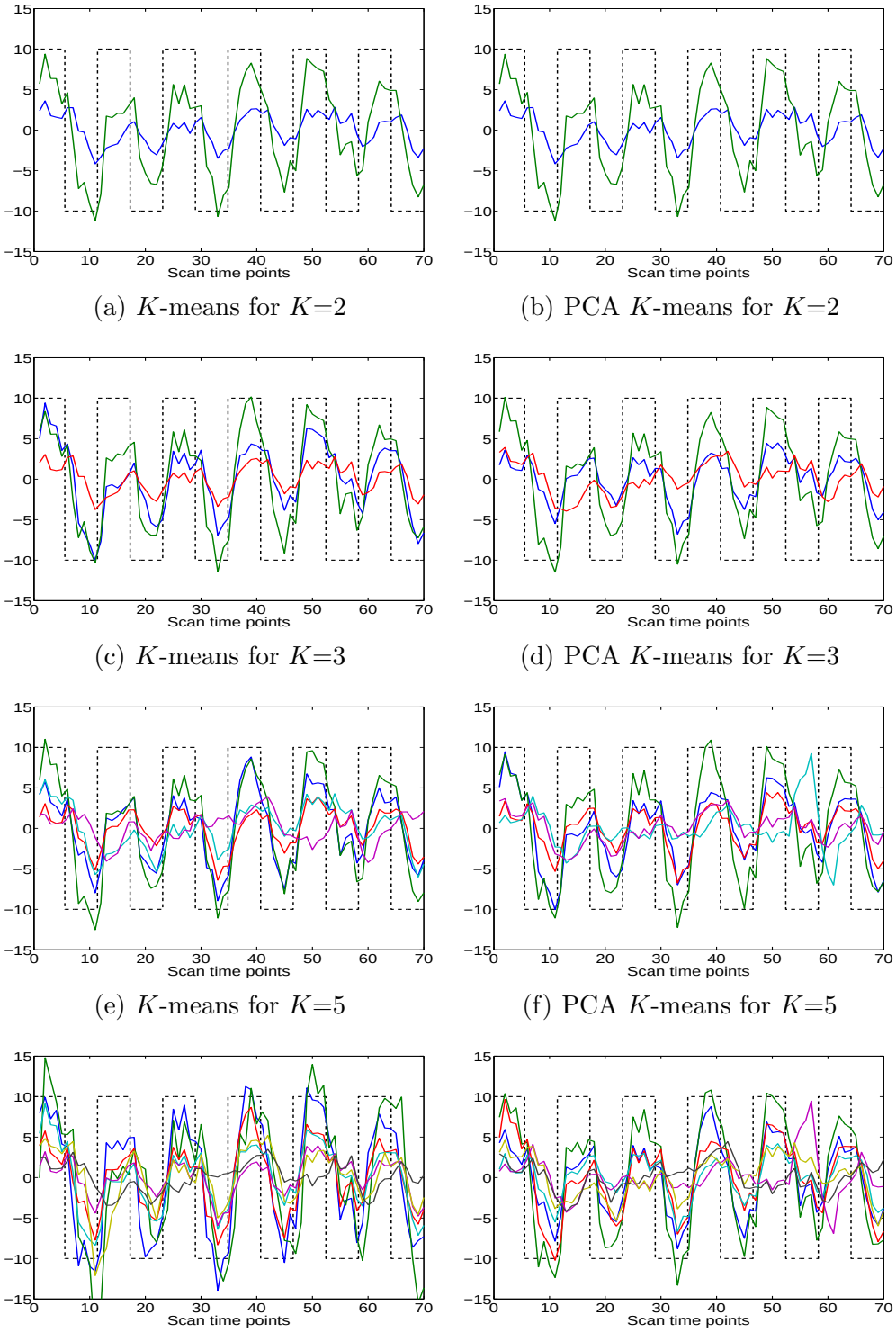
(e) K-means with K=5

(f) PCA K-means with K=5

(g) K-means with K=7

(h) PCA K-means with K=7

Figure B.13: Average time series plots for the prosaccade group for the pre-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing.



(a) K-means for K=2

(b) PCA K-means for K=2

(c) K-means for K=3

(d) PCA K-means for K=3

(e) K-means for K=5

(f) PCA K-means for K=5

(g) K-means for K=7

(h) PCA K-means for K=7

Figure B.14: Average time series plots for the prosaccade group for the post-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing

Appendix C

Clustering results for the Fixation task group

In this appendix, we present the clustering results for the fixation task group. For each test session (pre- and post-test) the brain is partitioned into two, three, five or seven clusters, which are K=2, 3, 5, and 7, using K-means and PCA K-means clustering methods.
The clustered brain maps are listed in the following order:

- K-means method for the pre-test for slices 13 – 19 in Figure C.1.
- K-means method for the pre-test for slices 20 – 26 in Figure C.2.
- K-means method for the pre-test for slices 27 – 33 in Figure C.3.
- PCA K-means method for the pre-test for slices 13 – 19 in Figure C.4.
- PCA K-means method for the pre-test for slices 20 – 26 in Figure C.5.
- PCA K-means method for the pre-test for slices 27 – 33 in Figure C.6.
- K-means method for the post-test for slices 13 – 19 in Figure C.7.
- K-means method for the post-test for slices 20 – 26 in Figure C.8.
- K-means method for the post-test for slices 27 – 33 in Figure C.9.
- PCA K-means method for the post-test for slices 13 – 19 in Figure C.10.
- PCA K-means method for the post-test for slices 20 – 26 in Figure C.11.
- PCA K-means method for the post-test for slices 27 – 33 in Figure C.12.

The average time series plots corresponding to each cluster in the clustered brain maps are listed in the following order:

- The pre-test using K-means and PCA K-means method in Figure C.13.
- The post-test using K-means and PCA K-means method in Figure C.14.

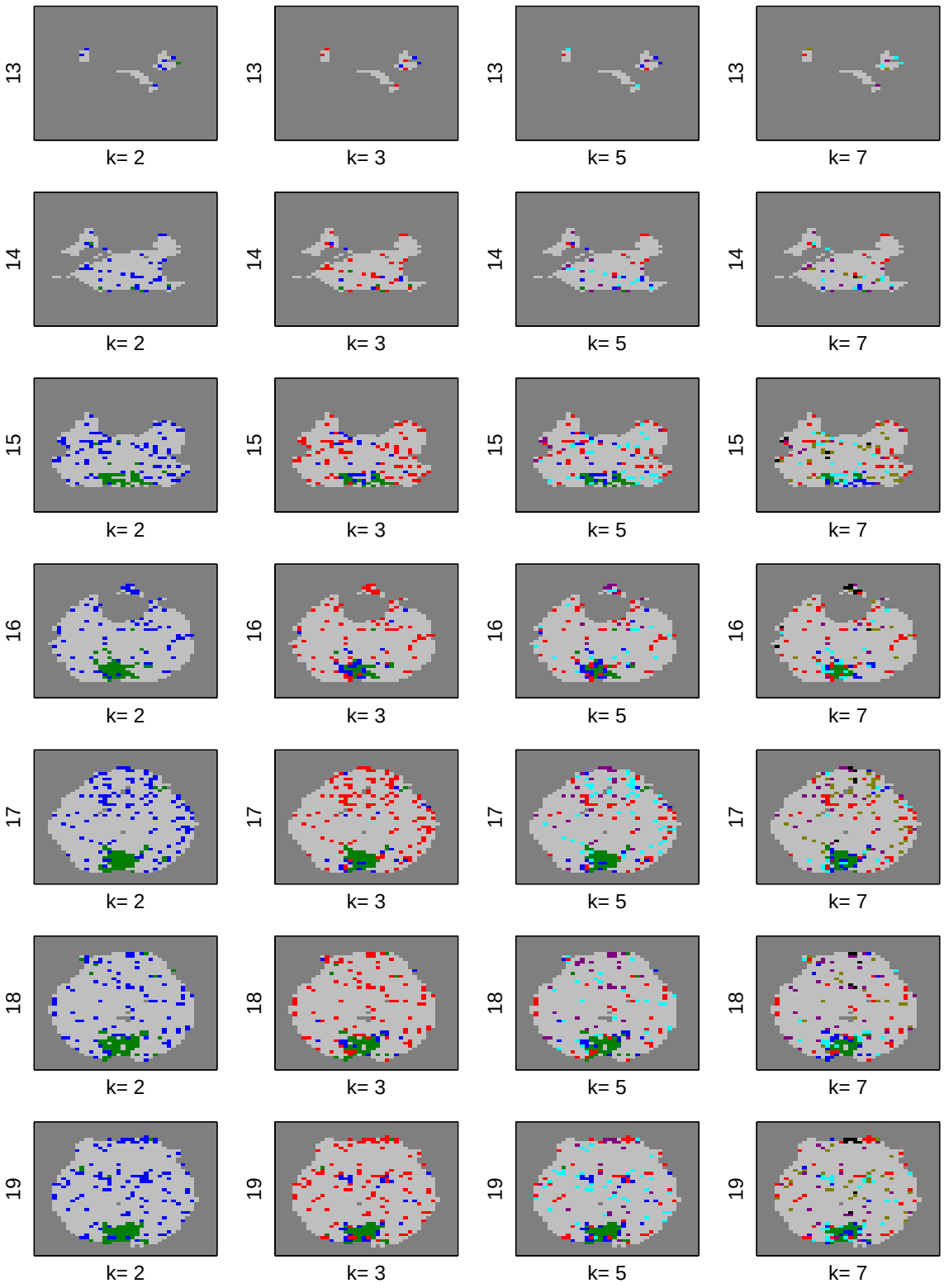


Figure C.1: The clustered brain maps for the fixation group for the pre-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

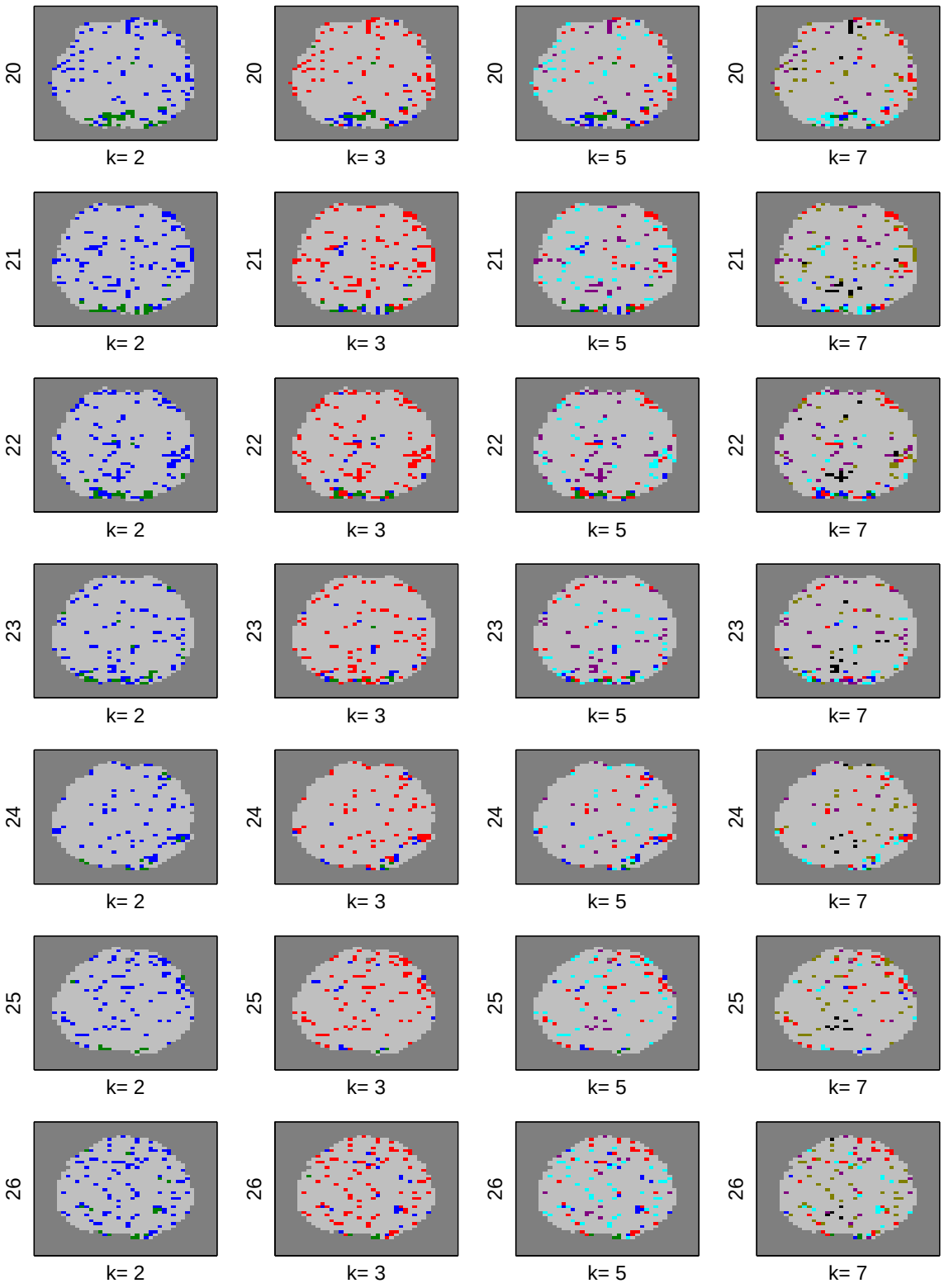


Figure C.2: The clustered brain maps for the fixation group for the pre-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

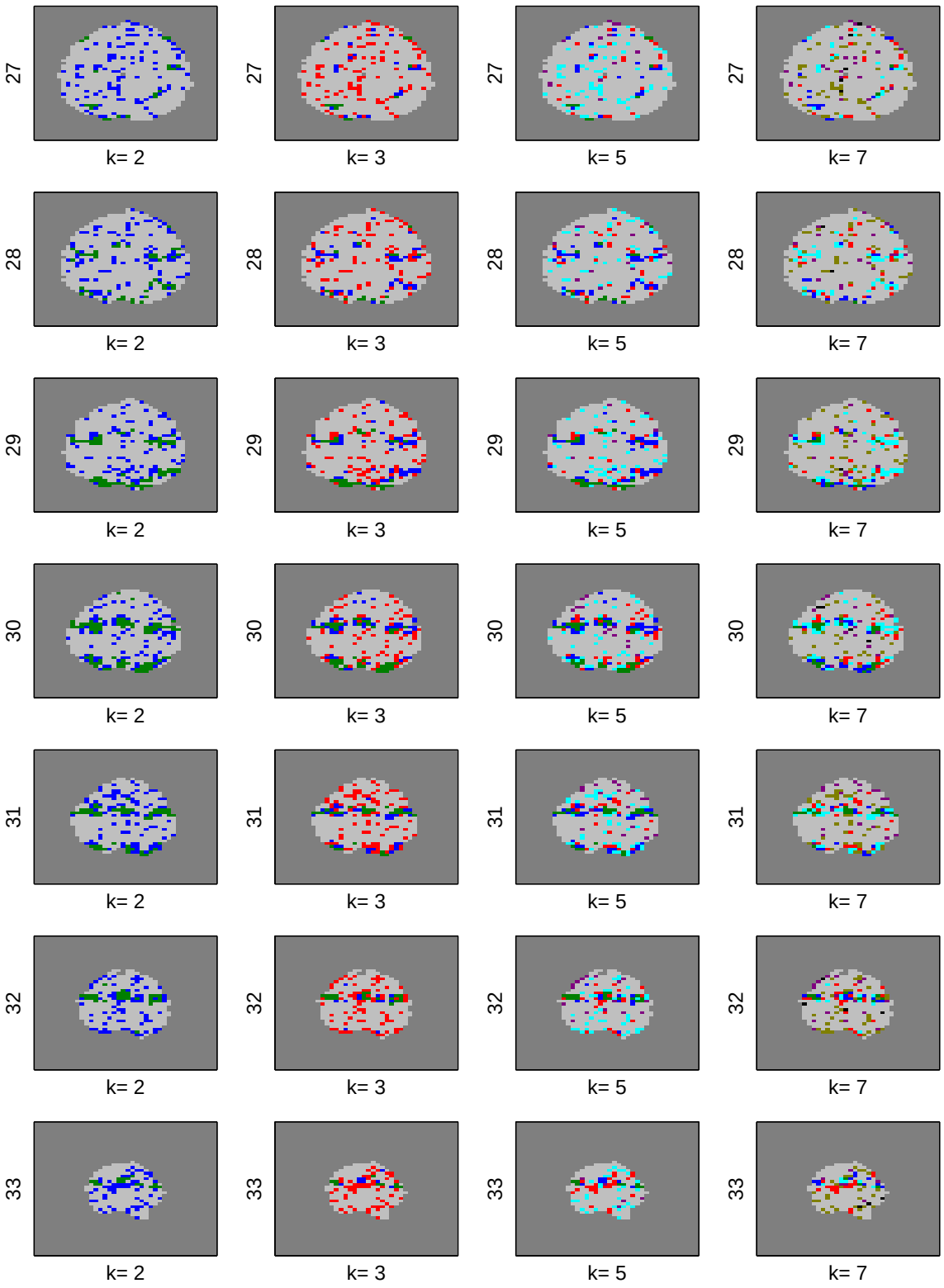


Figure C.3: The clustered brain maps for the fixation group for the pre-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

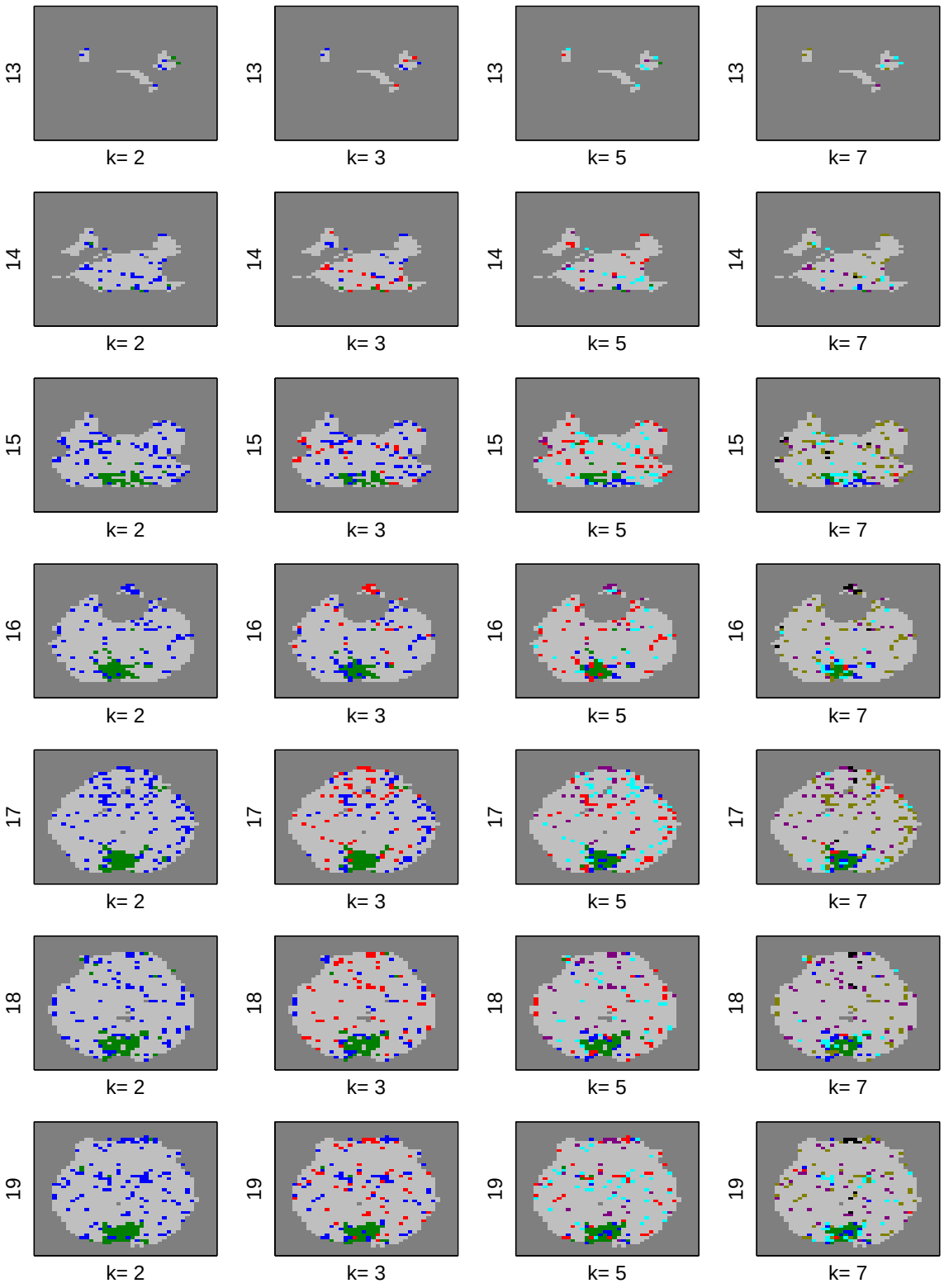


Figure C.4: The clustered brain maps for the fixation group for the pre-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

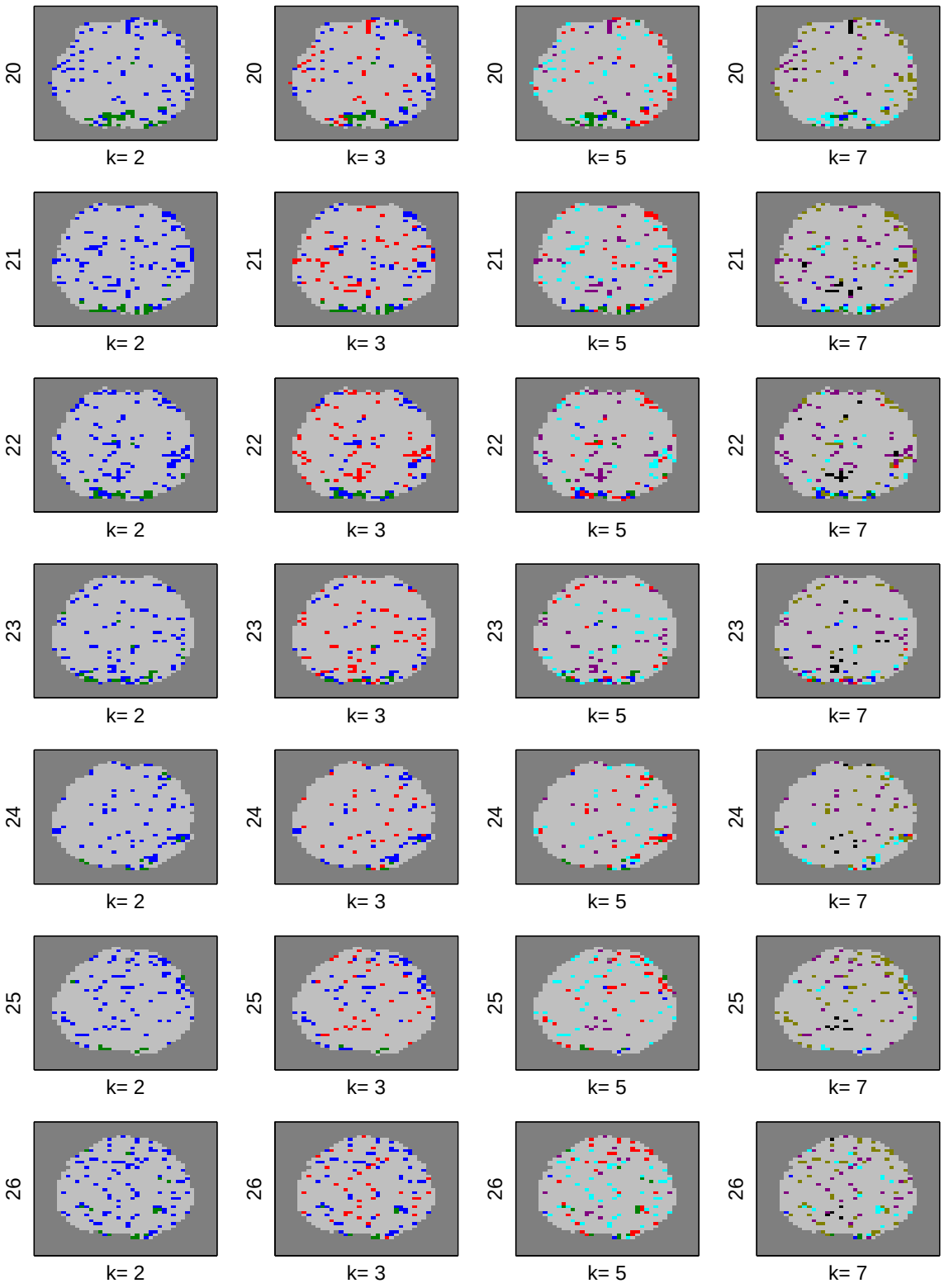


Figure C.5: The clustered brain maps for the fixation group for the pre-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

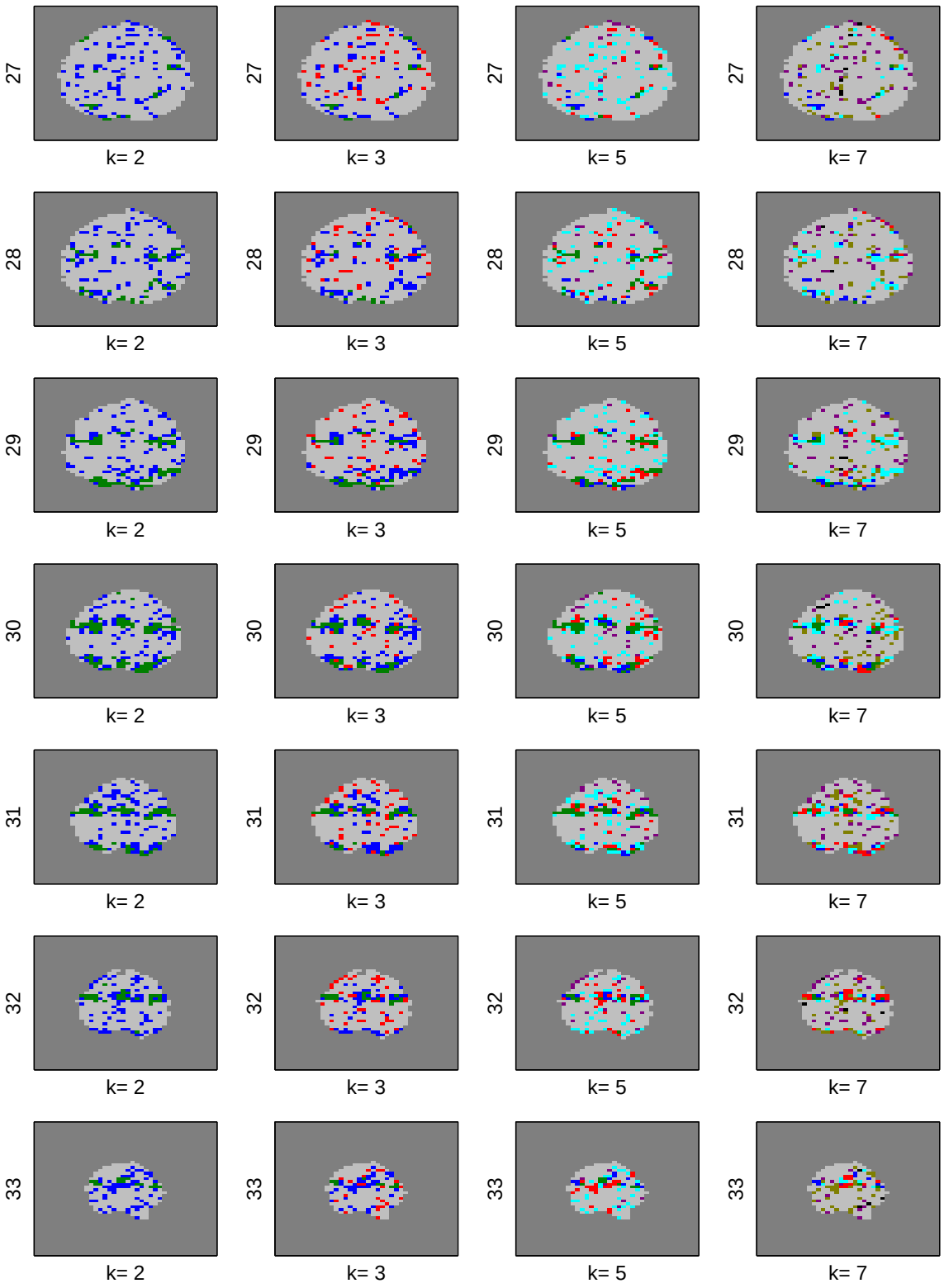


Figure C.6: The clustered brain maps for the fixation group for the pre-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

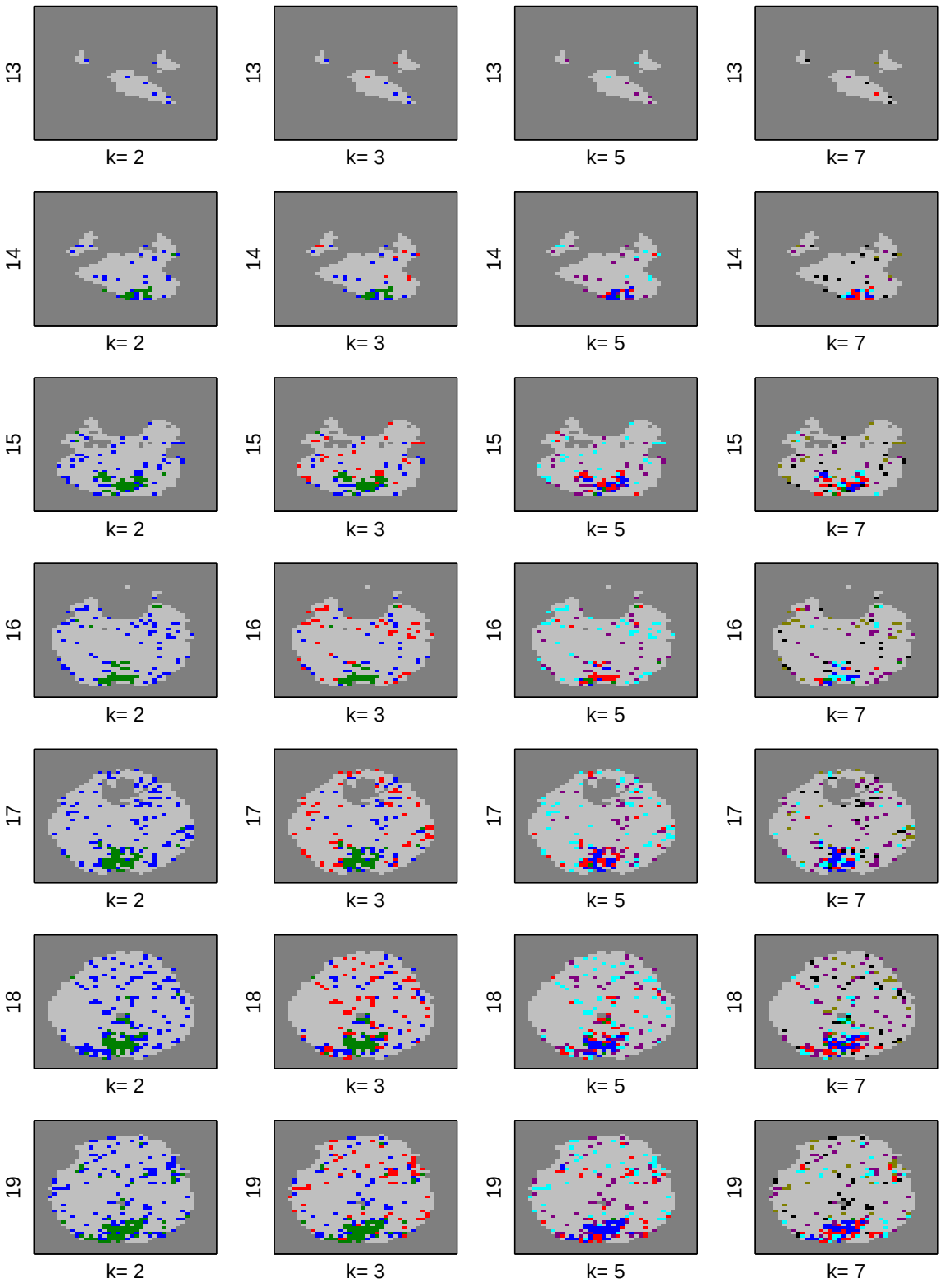


Figure C.7: The clustered brain maps for the fixation group for the post-test using K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

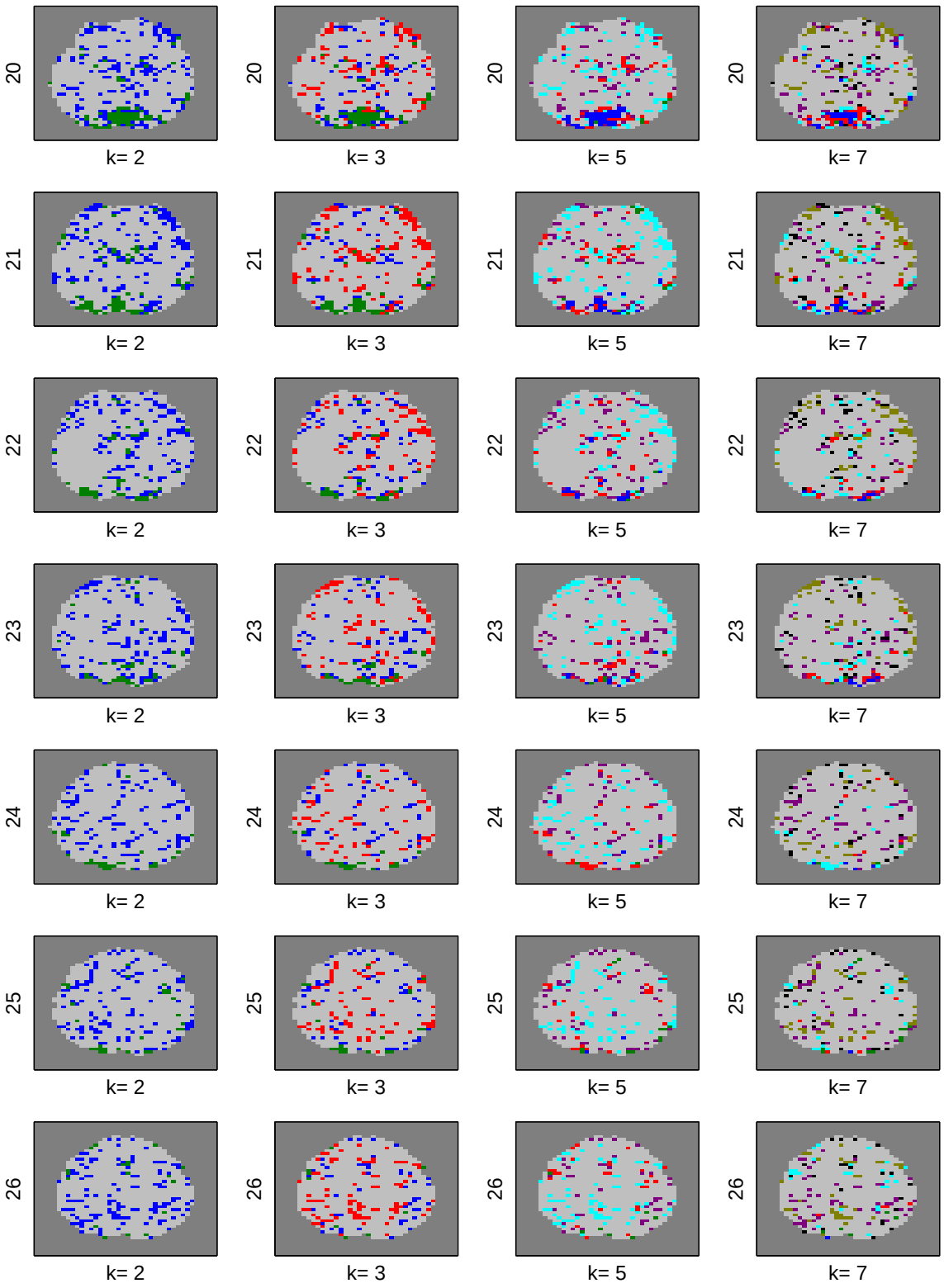


Figure C.8: The clustered brain maps for the fixation group for the post-test using K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

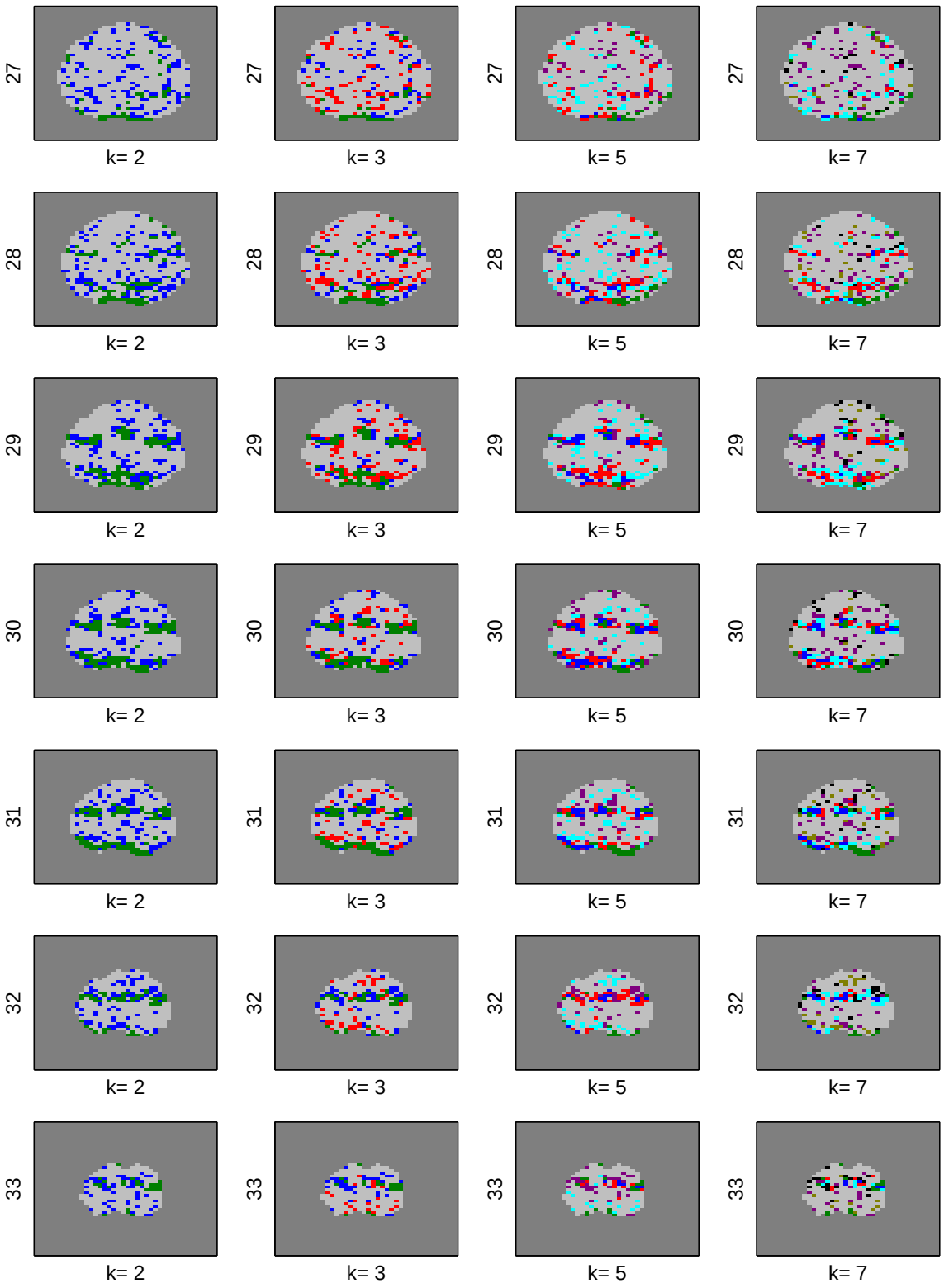


Figure C.9: The clustered brain maps for the fixation group for the post-test using K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

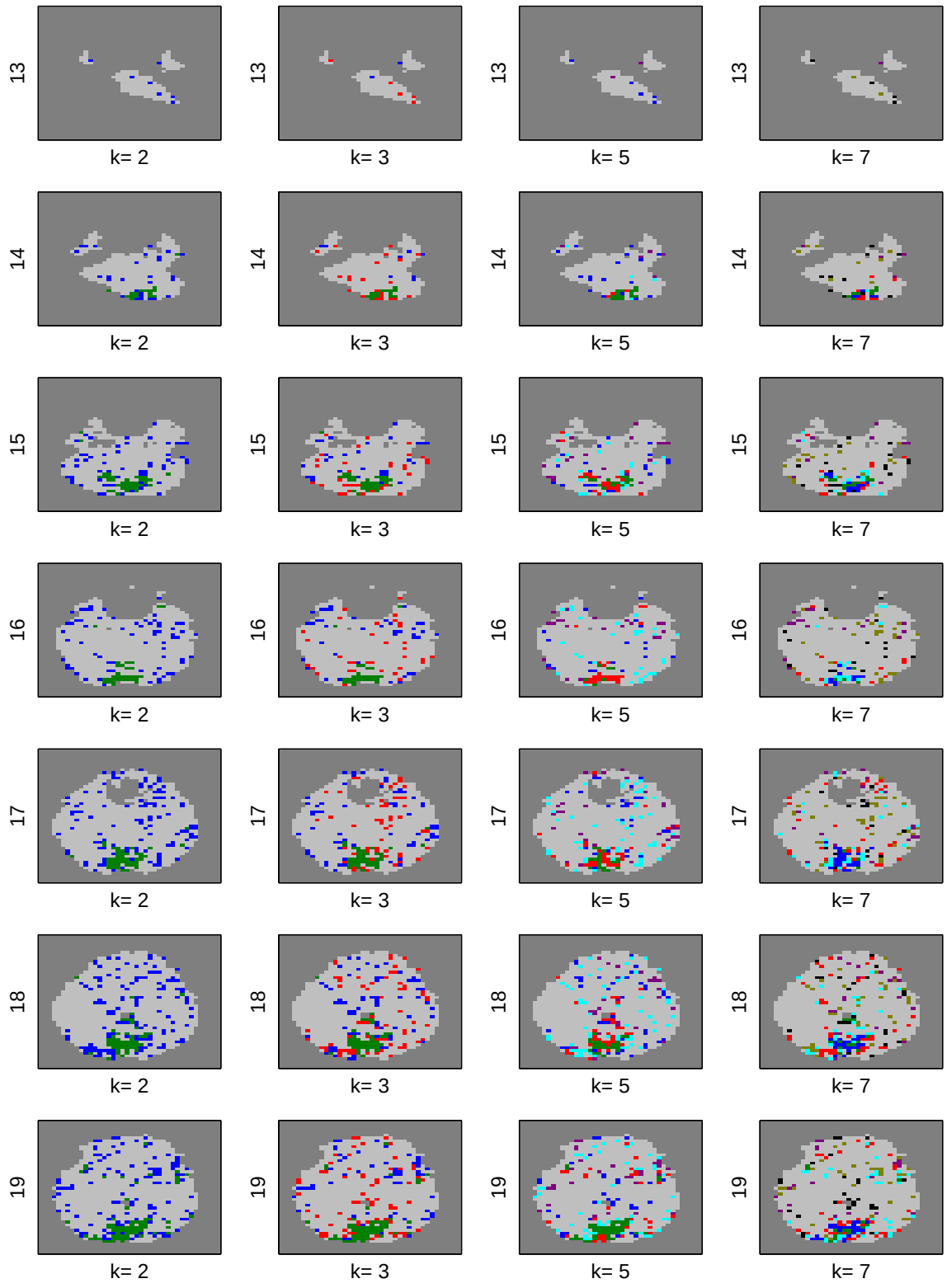


Figure C.10: The clustered brain maps for the fixation group for the post-test using PCA K-means for slices 13 – 19. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

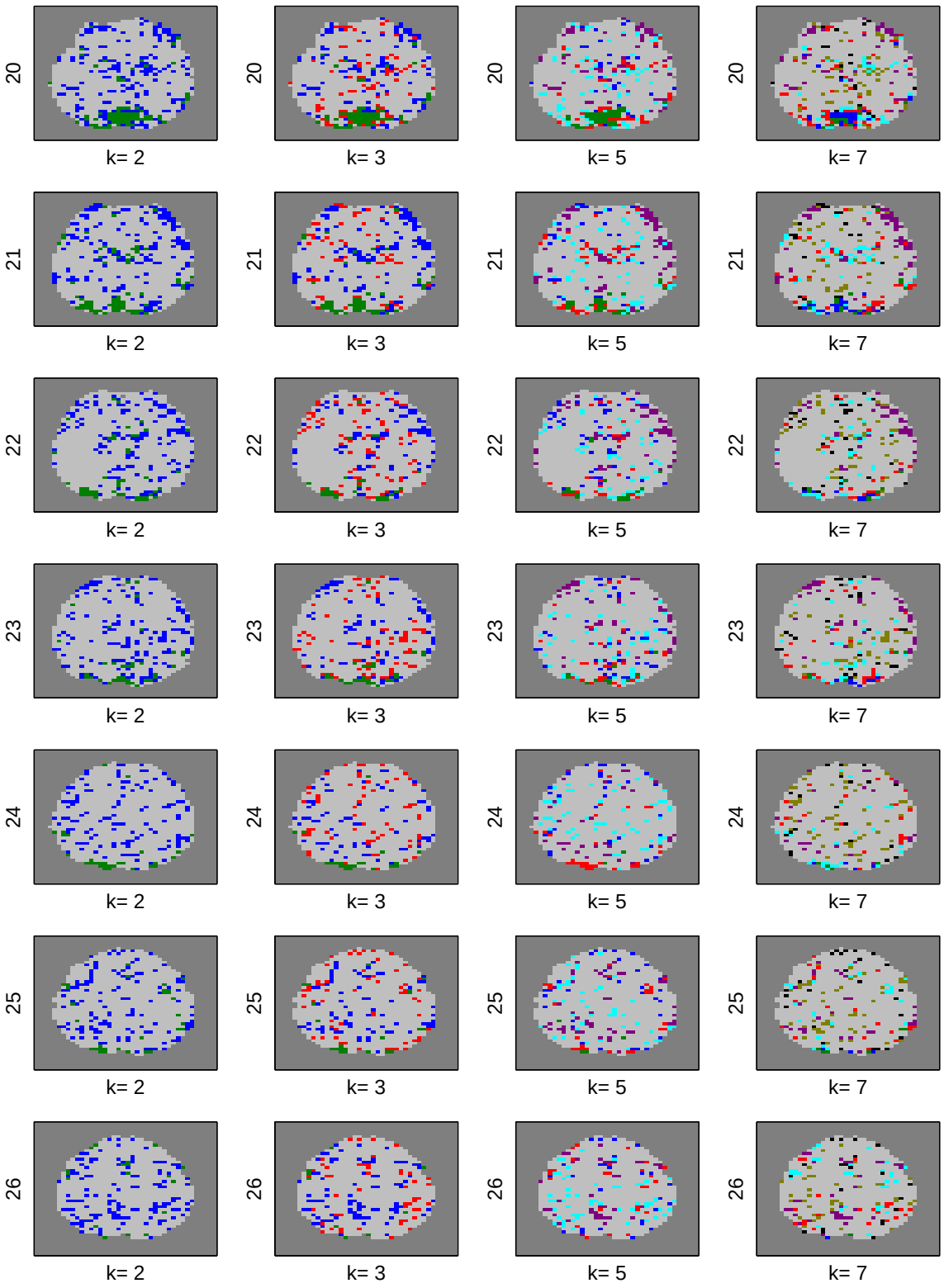


Figure C.11: The clustered brain maps for the fixation group for the post-test using PCA K-means for slices 20 – 26. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.

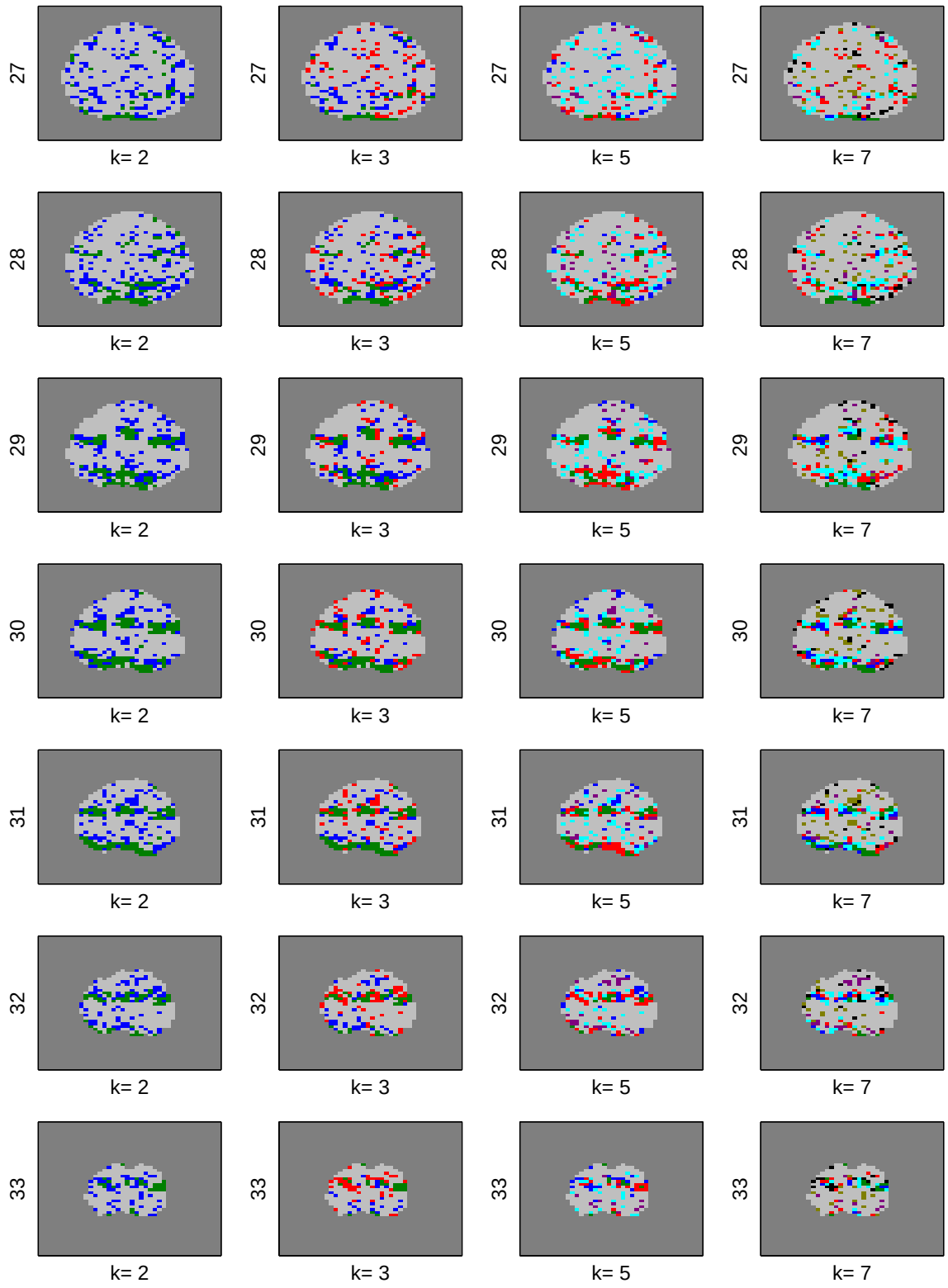
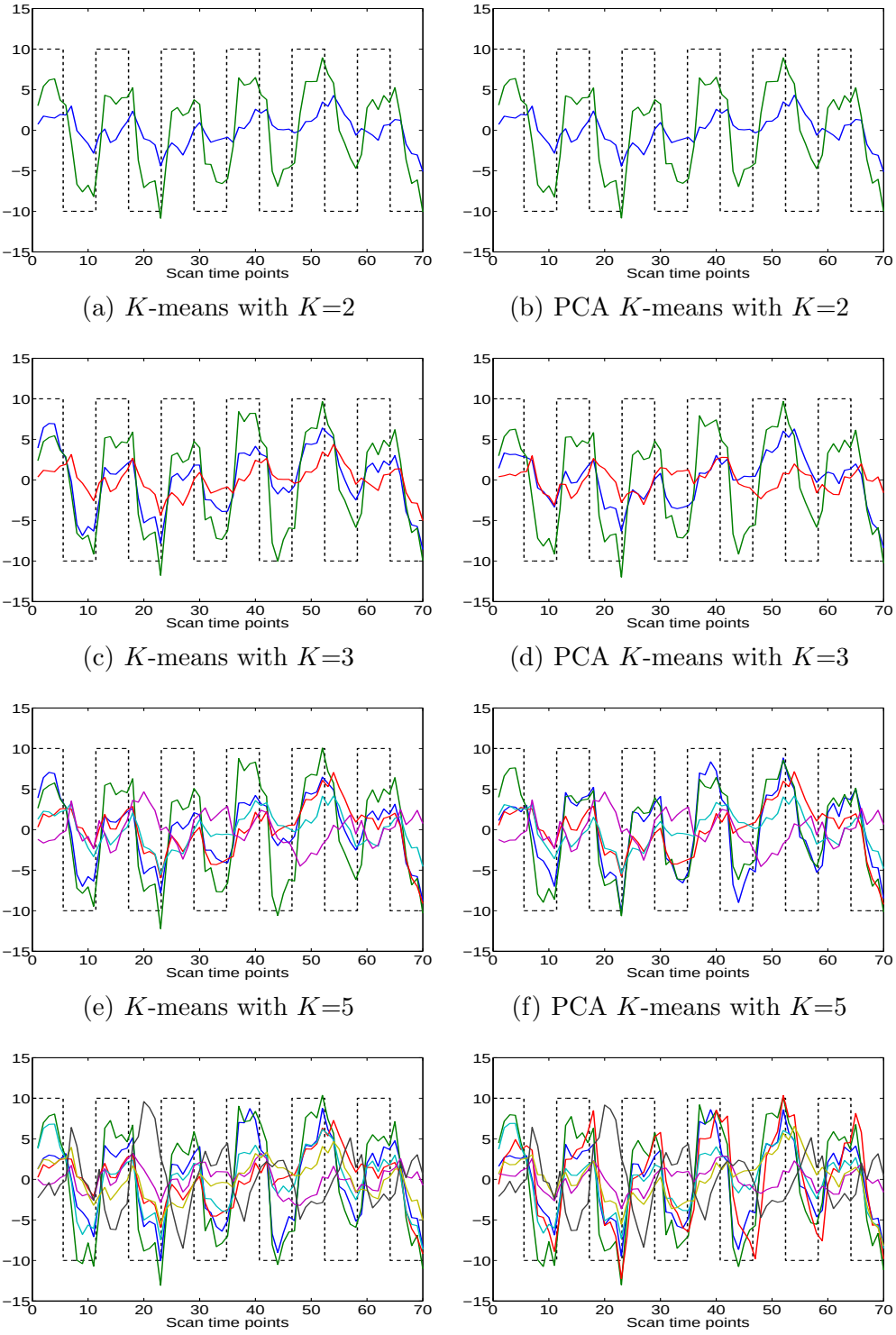


Figure C.12: The clustered brain maps for the fixation group for the post-test using PCA K-means for slices 27 – 33. The y labels indicate the slice number and the x labels are the number of clusters. The upper and lower areas of each slice correspond to anterior and posterior, respectively, and the larger slice number is toward the top of the head.



(a) K-means with K=2

(b) PCA K-means with K=2

(c) K-means with K=3

(d) PCA K-means with K=3

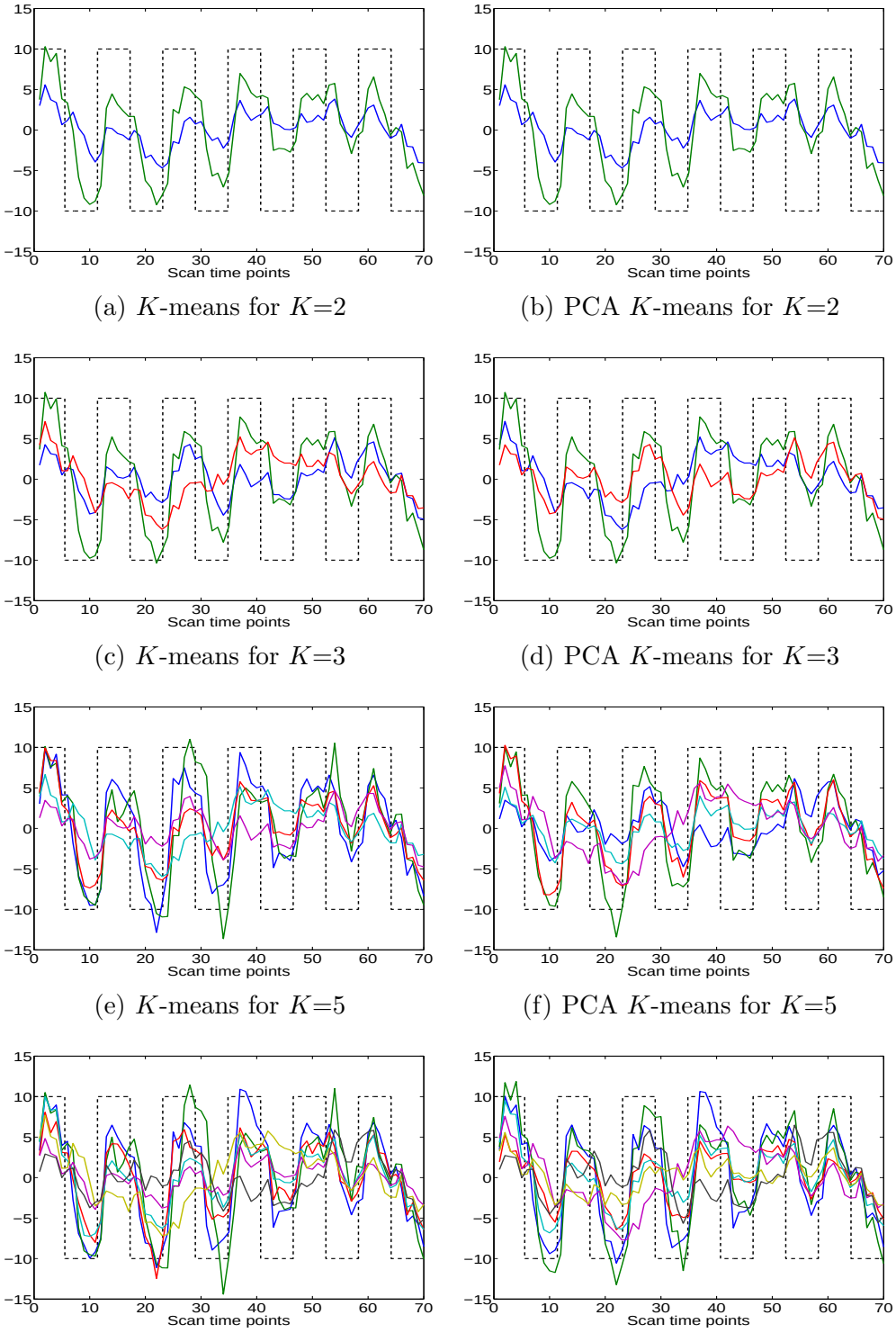
(e) K-means with K=5

(f) PCA K-means with K=5

(g) K-means with K=7

(h) PCA K-means with K=7

Figure C.13: Average time series plots for the fixation group for the pre-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing.



(a) K-means for K=2

(b) PCA K-means for K=2

(c) K-means for K=3

(d) PCA K-means for K=3

(e) K-means for K=5

(f) PCA K-means for K=5

(g) K-means for K=7

(h) PCA K-means for K=7

Figure C.14: Average time series plots for the fixation group for the post-test. The x-axes indicate scan time points and the y-axes indicate the voxel's detrended BOLD signal. Dashed black lines represent the stimulus timing