Statistical Wavelet Analysis of functional Images of the Brain

Manuela P. Feilner

Thèse N° 2671 (2002)

*Thése presentée au département de microtechnique
pour l'obention du grade de docteur ès sciences
et acceptée sur propositon du jury:*

Prof. Hannes Bleuler,
Président

Prof. Michael Unser,
Directeur de thèse

Prof. Wulfram Gerstner,
Rapporteur

Dr. Fabiana Zama,
Rapporteur

Dr. Elena L. Piccolomini,
Rapporteur

Dr. François Lazeyras,
Rapporteur

Ecole de Polytechnique Fédérale de Lausanne - 2002

Contents

Abstract

Functional magnetic resonance imaging (fMRI) is a recent, non-invasive technique that allows the measurement of brain metabolism while a subject is performing specific motor or cognitive tasks. The practical difficulty is that the signal changes are very small and the noise is relatively high. To enhance the activation signal, conventional methods, such as SPM, apply a Gaussian filter to the data and perform a statistical analysis at the pixel-level. Gaussian filtering obviously suppresses high frequency information. To avoid this loss of resolution, we propose instead to apply a wavelet transform and to perform the statistical analysis in the wavelet domain. Our statistical analysis relies on a conservative decision strategy which ensures that the false detection rate is kept well under control.

In the first chapter, we introduce the fMRI technique and compare it with other modalities. We then characterize the noise of fMRI data and introduce the statistical tools for our analysis. Next, we describe different types of wavelet transforms and show how those can be applied to the analysis of fMRI data.

Different wavelet bases offer different compromises. To determine the wavelet properties that are the most beneficial for the detection of activation patterns in fMRI data, we develop a test procedure for the objective evaluation of analysis methods. This procedure allows us to compare various brands of wavelet transforms, including Daubechies wavelets (real and complex) and the newly-defined fractional splines. We observe that one of the transforms—dual fractional spline of degree 1.2—usually outperforms the others. We establish an interesting theoretical connection between this particular wavelet transform and the Gaussian filter recommended by SPM.

Traditional, separable wavelets are constrained to dyadic scale progressions (powers of two). To allow for finer scale progressions, we define new 2D and 3D fractional wavelets which use quincunx sampling. We propose an FFT-based implementation that turns out to be surprisingly fast. We also present some experimental examples where these quincunx wavelets offer some benefits.

Version abrégée

L'imagerie par résonance magnétique (IRM) est une récente technique non-invasive qui permet de mesurer des activités cérébrales lorsqu'un patient est soumis à une stimulation externe. En effet, les images obtenues sont très bruitées et les fluctuations des signaux à détecter sont habituellement très faibles. Pour augmenter la limite de détection, des méthodes conventionnelles telles que SPM appliquent un filtre Gaussien et une analyse statistique sur le niveau des pixels détectés. Mais le filtre Gaussian génère une perte des informations de hautes fréquences.

Dans ce travail, nous proposons d'appliquer une transformation en ondelettes ainsi que de faire l'analyse statistique sur les coefficients obtenus afin d'éviter cette perte d'information et d'augmenter la sensibilité de résolution. En effet, notre analyse statistique est basée sur une stratégie de décision conservatrice, ce qui assure que le nombre de fausses détections reste sous contrôle.

Après une introduction de l'IRMf et une comparaison avec les autres modalités, une caractérisation du bruit des données brutes obtenues ainsi que l'introduction des outils statistiques pour notre analyse vont être présentés. Les différents types de transformations par ondelettes ainsi que la manière avec laquelle elles peuvent être appliquées pour une analyse des images IRMf seront évalués.

En effet, les différentes bases d'ondelettes offrent chacune leurs inconvénients et leurs avantages et par conséquent, une analyse des différents compromis est nécessaire pour déterminer lesquels sont le mieux adaptés à notre problème, c'est-à-dire la détection des activations cérébrales. Pour évaluer les différentes performances, une procédure de test a été développée et nous permet de définir très clairement les bases les plus adaptées à l'analyse des données IRMf. Cette procédure nous permet de comparer des types d'ondelettes différentes, comme celles de Daubechies (complexes et réelles) et les nouvelles ondelettes splines fractionnaires. Nous observons que l'une d'elles —la transformation en ondelettes splines fractionnaires de degré 1.2 (type "dual")—donne des performances nettement plus élevées que les autres ondelettes. On met en évidence une connection théorique intéressante entre cette transformation particulière et le filtrage Gaussian recommandé par SPM.

Les ondelettes séparables sont contraintes à une progression d'échelle dyadique et par

conséquent, pour permettre une progression plus fine, nous proposons l'utilisation de nouvelles base d'ondelettes fractionnaires à 2 et 3 dimensions. Ces dernières, développés par nos soins, utilisent un échantillonnage en quinconce et leur implémentation dans une base de l'espace de Fourier permet un traitement de l'image de manière étonnamment rapide. L'avantage de notre procédé sera finalement démontré à travers différents exemples.

Chapter 1

Introduction

1.1 Motivation

The brain is the most complex organ of a human being. Some authors have even claimed that it is the most complex entity in the universe. For several thousand years humans—philosophers, poets, musicians and more recently scientists—have been trying to disclose the secret of the function of the brain. The question is: what is the relationship between the brain and mental functions, including perception, memory, thought and language ?

Today, we are in the exciting phase of starting to understand how the brain works. A crucial role is played here by imaging techniques such as functional Magnetic Resonance Imaging (fMRI). Several brain areas which are activated by physiological stimuli have already been discovered. Now, more diffuse connections are being investigated. Scientists are starting to study neural correlates of many emotional states. One example, which recently made headlines, is the investigation of the neural basis of romantic love. The activity in the brains of 17 subjects who were deeply in love was scanned using fMRI while they viewed pictures of their partners. It was compared with the activity produced by viewing pictures of three friends of similar age, sex and duration of friendship as their partners [12]. The authors found a unique network of areas which is responsible for evoking this affective state.

Another example is the perception of music. Music plays a crucial role in every culture. It has been discovered that different types of music—or composers—affect different areas of the brain.

Another strong incentive to monitor the function of the brain is the search for better diagnostic tools and treatment planning. There are several neurological diseases such as epilepsy for which the last treatment, when the patient does not respond to medication, is neurosurgery. To minimize the potential risk of such an invasive treatment, neurosurgery relies on a precise delineation of the structural and functional aspects of the brain. fMRI

is becoming the method of choice for such individualized maps of brain functions. Further medical examples where fMRI is playing an increasing role are: the localization of regions with increased metabolism (tumors) for planning radiation therapy, the assessment of the effects of stroke, trauma, or degenerative diseases for guiding therapy and patient follow-up.

One of the practical problems with fMRI is that the signal differences between active areas and less-active areas are very small (1-5%). The detection of such small differences requires the use of signal processing techniques to enhance the signal-to-noise ratio of the data and the development of sophisticated statistical analyses to obtain reliable results. fMRI has typically a better spatial resolution than other functional imaging techniques (PET, SPECT, and MEG). Unfortunately, some of this resolution is lost in standard statistical analysis, which applies a Gaussian blurring filter to reduce the effect of the noise. Here, we investigate an alternative wavelet-based approach which has the capability of preserving the high-frequency content of the image.

1.2   Overview of the thesis

This thesis is organized as follows:

Chapter 2   Brief description of functional Magnetic Resonance Imaging (fMRI). We describe the problem, give an overview of the existing fMRI-data analysis methods and situate our work in this larger context.

Chapter 3   The necessary statistical tools for our analysis. In particular, we give an introduction to hypothesis testing which is a non-standard topic for a signal processing audience. One part of this chapter is also devoted to the characterization of the noise in fMRI data and to the application of the statistical tools to the fMRI problem.

Chapter 4   Foundations of the wavelet transform. We present a brief review of wavelet theory with an emphasis on fractional splines which have the advantage of being tunable. We also consider complex and quincunx wavelet variants, the possibility of separable and nonseparable downsampling, and complex wavelets. Then, we present our implementation in Java of the fractional spline wavelet transform and of the quincunx wavelet transform.

Chapter 5   Design and investigation of a new family of 2D and 3D wavelets appropriate for quincunx sampling. These new wavelets are orthogonal, symmetric and have a continuously-adjustable order parameter. We propose also a fast FFT-based implementation. The linear and nonlinear approximation properties of the separable and nonseparable wavelet transforms are compared.

Chapter 6   Description of a first approach to analyze fMRI data based on orthogonal filterbanks.

Chapter 7   Descriptions and evaluations of the wavelet-based approach for the analysis of fMRI data. We present test data that were created to find out which properties of the wavelet transform are preferable for fMRI. We present also the validation of our algorithm on real data and the comparison of the results with Statistical Parameter Mapping (SPM).

Chapter 8   General conclusion and recommendations for possible extensions.

1.3   Contribution of this thesis

The contribution of this thesis is twofold:

1. Design of new wavelets and efficient implementation.

    → Development of a new family of 2D and 3D quincunx wavelets with fractional order.

    → FFT-based fast algorithm of the quincunx transform. Implementation in Matlab for 2D and 3D and in Java for 2D. This algorithm is almost as efficient as the standard Mallat algorithm.

    → Demonstration that the quincunx wavelets have good energy-compaction properties for linear and nonlinear approximation.

2. fMRI analysis.

    → A first proposal to use orthogonal filterbanks for noise reduction as an alternative to the Gaussian filtering that is used in SPM (the reference technique in the field).

    → Development of wavelet-based algorithms for the statistical analysis of fMRI data.

    → Assessment of quality depending on different aspects of the wavelet transform for a large collection of activation patterns.

        – Fractional splines: evaluation of different types and order parameters. Fast and exact implementation in Matlab for 2D and 3D and in Java for 2D.

        – The proposal to analyze fMRI data using complex wavelets in conjunction with a multivariate test (Hotelling's T^2) which opens new possibilities for multiple comparisons. The motivation is to have a transform that is more nearly shift-invariant.

- – Finer scale progression and less directionality of the wavelets: quincunx in fMRI.
- → Validation of the multiresolution approach on real data.
- → Improvements over the widely used standard method SPM.

Chapter 2

Functional Magnetic Resonance Imaging (fMRI)

2.1 A brief historical view

Before functional Magnetic Resonance Imaging (fMRI), there was only one way to access the functionality of the human brain: One had to study the brains of ill people after they had died and compare them to healthy ones. This way many brain discoveries were made from the examination of injuries or malformations. In 1861, the French surgeon Paul Broca had a patient who could speak only one sound: "tan", after suffering a stroke. Following the death of this patient, Broca examined his brain and discovered a damaged spot on the front left side. He suspected that this area controlled speaking ability [23]. He was right. "Broca's area" was the first recognized link between brain function and its localization.

In 1890 C.W. Roy and C.S. Sherrington, in their paper *On the regulation of blood supply of the brain*, suggested that neural activity was accompanied by a local increase in cerebral blood flow [119]. The possibility of imaging this effect started in 1946, when Felix Bloch at Stanford, studying liquids, and Edward Purcell at Harvard, studying solids, described—independently of each other—the nuclear magnetic resonance phenomenon [19, 112]. They discovered that certain substances are able to absorb and emit radiofrequency electromagnetic radiation, when they are placed in a magnetic field. The absorption and emission was related to the substance being studied. This new technique, Nuclear Magnetic Resonance (NMR), was used for chemical and physical molecular analysis between 1950 and 1970. Bloch and Purcell were awarded the Nobel prize for physics in 1952 for this discovery, but it was not until 1973 that NMR was used to generate images from inside the body. It was in 1971, when Raymond V. Damadian showed that the nuclear magnetic relaxation times of healthy and cancerous tissues differed, that sci-

entists started to consider magnetic resonance for the detection of disease [40]. In 1973, a chemist, Paul Lauterbur, improved the technique significantly. He succeeded in getting from the one-dimensional spatial representation to a complete two-dimensional scan [87]. But the human body is three-dimensional and it was Peter Mansfield who developed the three-dimensional technique [96]. The basis of the current Magnetic Resonance Imaging (MRI) technique was finalized by Anil Kumar, Dieter Welti, and Richard R. Ernst at ETH Zürich [83]. They applied Fourier reconstruction techniques to encode spatial information in the phase and frequency of the NMR signal.

MRI is tuned to image the hydrogen nuclei of water molecules. In functional MRI, Blood Oxygenation Level Dependent (BOLD) effects are measured for mapping brain functions. Until 1990 there was no non-invasive way of measuring the flow of blood in cortical areas. Ogawa and Lee at the AT & T Bell Laboratories and Turner at the National Institutes of Health (NIH), working independently on laboratory animals, discovered that the oxygenation level of blood acts as a contrast agent in MR images [108, 142]. This discovery led to the technique of fMRI as it is in use today [41, 98].

In the past few years, the interest in fMRI has increased continuously, as indicated by the number of papers published in this area (cf. Figure 2.1).
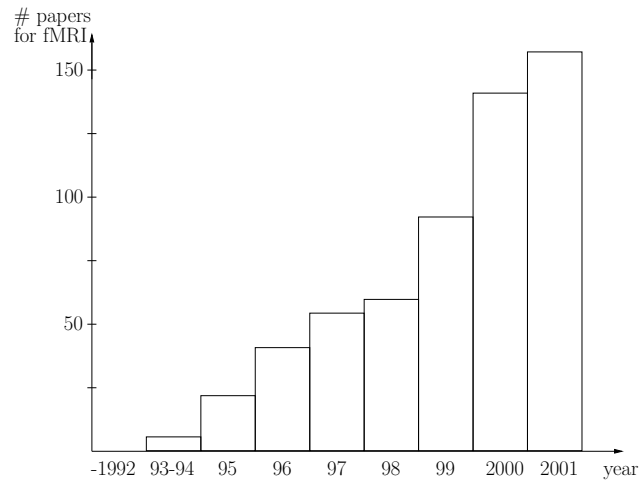
Figure 2.1: Papers published in the last few years related to the subject of functional Magnetic Resonance Imaging (from the INSPEC database only).

2.2 Principle of Magnetic Resonance Imaging

A modern MRI scanner, from which we get our fMRI images, is depicted in Figure 2.2. In this section we describe briefly the underlying principles of fMRI [6, 14, 53, 63, 24, 90, 139, 126].

Figure 2.2: Example of a magnetic resonance scanner with a magnet of 1.5 tesla. It belongs to the Siemens "MAGNETOM Symphony Maestro Class" range [127].

2.2.1 Relaxation phenomena

The MR signal in its classical clinical use is predominantly a signature from the protons of tissue water in humans. The technique of Nuclear Magnetic Resonance (NMR) is based on the fact that atoms of hydrogen H^1, but also elements like sodium (Na^{23}) and phosphorus (P^{31}), have a nuclear spin, which is proportional to the angular momentum of the nucleus (the proportionality constant is the gyromagnetic ratio). In general, atoms which have an odd atomic number or an odd atomic weight have a resulting nuclear spin. Such a nucleus can be seen as a rotating electrically charged object causing a magnetic moment. In human tissue, we deal with a large number of such nuclei. At a macroscopic level these magnetic dipoles form an *ensemble* and the sum of their magnetic moments, the nuclear magnetization, is equal to zero provided there is no exterior magnetic field. When such nuclei are immersed in a static magnetic field $\vec{B}_0 = (0, 0, B_0)$ in the z-direction, the total magnetic moment, the nuclear magnetization $\vec{M}$, is oriented in the direction of the applied field $\vec{B}_0$ (parallel or anti-parallel heading). In MRI a very strong

constant magnetic field (in medical MRI between 0.2 and about 3 tesla) is applied. As a point of reference, the magnetic field of the earth is only about $5 \cdot 10^{-5}$ tesla.

In order to excite the nuclear magnetization in the tissue, a pulse of radiofrequency magnetic field $\vec{B}_{RF}$ of small amplitude and oscillating at a well-defined frequency is superposed, perpendicular to $\vec{B}_0$. In consequence $\vec{M}$ changes the alignment from being parallel to $\vec{B}_0$. The magnetization $\vec{M}$ is then divided into a transverse component $\vec{M}_{xy}$ and a longitudinal component $\vec{M}_z$. The resulting angle between the magnetization vector and the z-axis is called the RF flip angle. After excitation, the vector belonging to $\vec{M}_{xy}$ precesses at a resonance frequency, the *Larmor frequency*, which is nucleus-specific, and decreases exponentially with a time constant T_2^*. This vector induces an oscillating, decaying voltage signal in a receiver coil. The realignment rate of $\vec{M}_z$ is described by a *longitudinal relaxation time T_1*, also called spin lattice relaxation time, which is also nucleus-specific (Figure 2.3). The relaxation time of the vector $\vec{M}_{xy}$, perpendicular to $\vec{B}_0$, is named the *transverse relaxation time T_2* or spin-spin relaxation time. This vector decreases since, in addition to the rotation, the magnetization starts to dephase, because each spin packet rotates at its own Larmor frequency. This process of transverse relaxation is illustrated in Figure 2.4. Note that there is no direct correlation between the decreasing transverse and the increasing longitudinal magnetization, and it always holds that $T_2 \leq T_1$. In contrast, the relation between T_2 and T_2^* is described by $\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}$, where T_2' describes the time decay of dephasing due to field inhomogeneity.
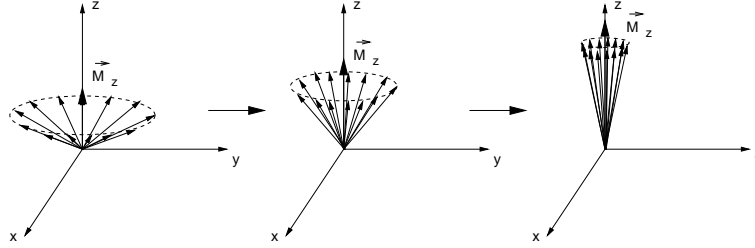
Figure 2.3: The recovery of the longitudinal magnetization component $\vec{M}_z$, which is related to the longitudinal relaxation time T_1.

2.2.2   MR imaging

In order to localize the MR signal, a second magnetic field $\triangle\vec{B}_0$ is added to $\vec{B}_0$. $\triangle\vec{B}_0$ is parallel to $\vec{B}_0$ but inhomogeneous in space. The resonance frequency of the nuclei depends now on $\vec{B}(x,y) = \vec{B}_0 + \triangle\vec{B}_0(x,y)$ and changes within the space. A frequency coding system is achieved when the signal is recorded in the presence of the magnetic field $\vec{B}(x,y)$. A phase coding system results by applying this gradient field before recording the signal, because the arising temporary frequency change involves a spatial variation
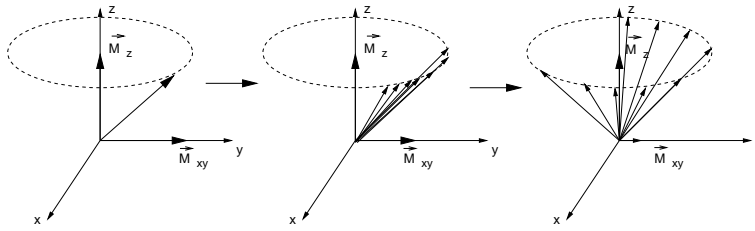
Figure 2.4: The decrease of the transverse magnetization component $\vec{M}_{xy}$ of the spins in a molecule, which is related to the transverse relaxation time T_2.

of the phase. The space coding of two orthogonal axes is usually achieved by applying these two methods.

The imaging effect of the contrast parameters T_1 and T_2 can be suppressed or enhanced in a specific experiment by another set of parameters, such as repetition time (T_R), echo time (T_E), and flip angle (α) [90]. This distinction allows certain structural components to be emphasized. Water for example has very similar T_1 and T_2 times. In cerebrospinal fluid (CSF) T_1 is larger than T_2, and in white matter the difference is even greater. Therefore, T_1-*weighted imaging* is an effective method to obtain images of a good anatomical definition and T_2-*weighted imaging* is a sensitive method for disease detection, since many disease states are characterized by a change of the tissue's T_2 value. Figure 2.5 shows an example of T_1- and T_2-weighted images.

a)   b)
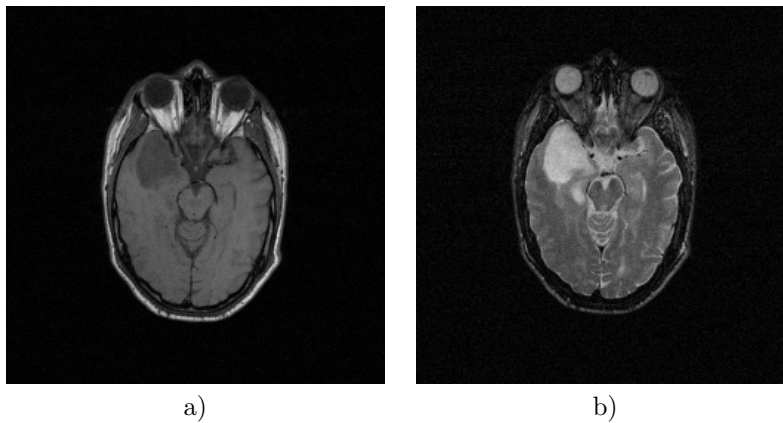
Figure 2.5: a) T_1-weighted head image. b) T_2-weighted head image: the tumor is more visible.

2.2.3 Fast MR imaging

fMRI of a human brain, must be performed rapidly. One reason is that the movements of the patient increase with time; another is that the images of the whole volume of the brain should be acquired within the same activation condition. Furthermore, the patient can perform most of the cognitive tasks of interest only for a few minutes without habituation or fatigue. Two such faster methods are Fast Low Angle SHot (FLASH) and EchoPlanar Imaging (EPI). FLASH uses a small interval between the RF pulses to reduce the flip angle and therefore also T_1 [74], while EPI uses smaller RF pulses that excite only a fraction of the protons, which reduces relaxation time. EPI requires a strong gradient field, which was not available when FLASH was invented. EPI, invented by Mansfield et al. [133], acquires a complete image in less than 100 ms. It is currently a widely deployed method. However, this technique suffers more than others from distortion and signal loss arising from magnetic field inhomogeneities in the brain. Nevertheless the experiments with real data described in this thesis were all performed with EPI, since it is at present the most commonly used technique. Further description of (noise) artifacts of rapid imaging methods can be found in Section 3.3.1.

2.3 BOLD contrast in fMRI

In fMRI experiments, we alternate between an activation state where the patient performs a certain task and a rest state where no task is performed. The task involves changes of neuronal activation during an activation state in the brain. With fMRI, we cannot measure directly the neuronal activities, but their related effects. Ogawa [107] and Turner [142], working with laboratory animals, observed independently in 1990 that MRI image contrast changes around the blood vessels could be obtained by the modification of the oxygenation state of the blood. This effect arises from the fact that deoxyhemoglobin is more paramagnetic than oxyhemoglobin, which itself has almost the same magnetic susceptibility as tissue. Thus deoxyhemoglobin can be seen as nature's own contrast agent. Neuronal activity is related to changes of the oxygenation state because neurons need oxygen to function, since they are cells. This oxygen is carried to the neurons by oxyhemoglobin within the red blood cells, the erythrocytes. The oxygen uptake by the neurons causes a deoxygenation of the oxyhemoglobin, which becomes deoxyhemoglobin. In addition, neurons need glycogen for their metabolic processes during electrical activity, which is also carried by the erythrocytes. The need for oxygen and glucose of the neurons during electrical activity causes an increased blood flow and perfusion in the capillaries. But the oxygen consumption increases only slightly during activation, therefore the relation between oxyhemoglobin and deoxyhemoglobin changes. This relative decrease in the concentration of paramagnetic deoxyhemoglobin indicates activation, which can be measured by the fMRI method. Since the resulting image contrast depends on the oxygen content of the blood, this method is called *Blood Oxygenation Level Dependent*

(BOLD) contrast [110, 53].

There is also a more recent alternative contrast mechanism for visualizing functional activation in the brain that provides direct measurement of cerebral blood flow (CBF) using MRI. This technique, called arterial spin labeled (ASL) perfusion MRI, utilizes electromagnetically labeled arterial blood water as endogenous tracer for measuring CBF [45, 49].

Perfusion contrast has some advantage over the BOLD technique. In particular, it improves the independence of the observations in time under the hypothesis that no activation is present, whereas BOLD-data signals can be correlated. However, perfusion contrast has a poorer temporal resolution and is less sensitive than the BOLD contrast method [4].

2.4 Different imaging modalities

We will now briefly discuss the two other competing functional imaging methods, PET (cf. Figure 2.6) and SPECT, which are older than fMRI.

Figure 2.6: Example of a Positron Emission Tomograph. It belongs to the Siemens "ECAT EXACT" range and is the most widely used PET scanner [127].

2.4.1 Positron Emission Tomography (PET)

The history of PET began in the early 1950s, when the medical imaging possibilities of a particular class of radioactive substances were first realized. By the mid-1980s, PET had become a tool for medical diagnosis and for dynamic studies of human metabolism.

PET imaging begins with the injection of a metabolically active tracer into the patient. The tracer is built by a biological molecule that carries a positron-emitting isotope (known as the radioactive tag), such as ^{11}C (carbon-11), ^{18}F (fluorine-18), ^{15}O (oxygen-15) or ^{13}N (nitrogen-13), which have short decay times. Over a few minutes, the isotope accumulates in an area of the body for which the carrier molecule has an affinity. For example, glucose labeled with ^{11}C accumulates in the brain or in tumors where glucose is used as the primary source of energy. The radioactive nuclei then decay by positron emission. Thus PET provides images of biochemical functions, depending upon the type of molecule that is radioactively tagged. The emitted positron combines with an electron almost instantaneously, and these two particles experience the process of annihilation. The energy associated with the masses of the positron and electron is divided equally between two photons that fly away from one another at a 180° angle. These high-energy γ-rays emerge from the body in opposite directions and are detected by an array of circular γ-ray detectors that surround the body of the patient [24]. These detectors have a series of scintillation crystals, each connected to a photomultiplier tube. The crystals convert the γ-rays to photons, which the photomultiplier tubes convert to electrical signals and amplify. These electrical signals are then processed by the computer to generate images. Figure 2.7 shows such an image.
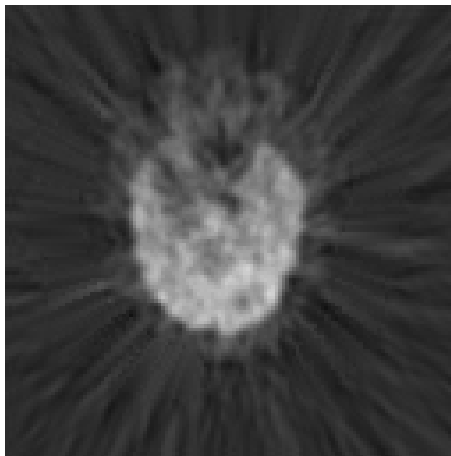
Figure 2.7: PET image from a slice of the human brain.

There are some limitations of PET. Higher order cognitive processes increase regional Cerebral Blood Flow (rCBF) by only about 5%. This small amount can only be detected when multiple subjects are used and this relates to the problem of anatomical variation between the subjects. Another limitation are the spatial and temporal resolutions, which are lower than for fMRI, i.e., 3 mm and 60 s, respectively, compared to $\sim$ 1.5 mm and 100

ms for fMRI. PET scanners are also less available than fMRI scanners, since they must be located near a cyclotron device that produces the short-lived radioisotopes used in the technique. Further advantages of fMRI over PET are lower costs, reduced invasiveness, structural and functional information, and a larger signal-to-noise ratio. On the other hand, PET shows a greater sensitivity in certain aspects, such as soft auditory and overt speech stimulation [124, 79], since the loud noise of the fMRI scanner can interfere with the auditory tasks. (Today, however, good hearing protection is available, so that more and more auditory tasks can also be done with an fMRI scanner.) When performing an experiment with PET and fMRI, the activation zones detected were mostly identical. However the activation zones are slightly larger in the case of fMRI, probably due to the better signal-to-noise ratio.

2.4.2   Single Photon Emission Computed Tomography (SPECT)

Single Photon Emission Computed Tomography (SPECT) is a technique similar to PET. But the radioactive substances used in SPECT (xenon-133, technetium-99, iodine-123) have longer decay times than those used in PET, and emit single instead of double γ-rays. SPECT can provide information about blood flow and the distribution of radioactive substances in the body. Its images have less sensitivity and are less detailed than PET images, but the SPECT technique is less expensive than PET. Also SPECT centers are more accessible than PET centers because they do not have to be located near a cyclotron.

2.5   Description of typical fMRI data

The first step in an fMRI experiment is the set-up of the experimental design by a neurologist. The task succession must be chosen in order to find the relation between the brain activation and a certain motor or cognitive task and to filter out inappropriate actions. For example, a cognitive task meight be listening to a specific sound or to music or viewing a screen displaying different pieces of information or images.

During the fMRI experiment a volunteer lies in an MRI scanner, his head fixed with a helmet-like device to avoid strong movement artifacts. A sequence of images of the volunteer's brain is acquired while he performs the specified tasks. The spatial resolution of the obtained images (slices) is typically 2–3 mm and the thickness of one slice is usually 3–10 mm. An example of an acquired volume is given in Figure 2.8. Using the fast EPI technique, 3–10 images per second can be acquired [63].

The experimental design can be based on a *block paradigm* or an *event-related fMRI*, which we will explain below. All of the real data investigated in this thesis are based on the block paradigm. We received our data through collaborations with the Inselspital Bern and the Hôpital Cantonal Universitaire de Genève (HCUG).
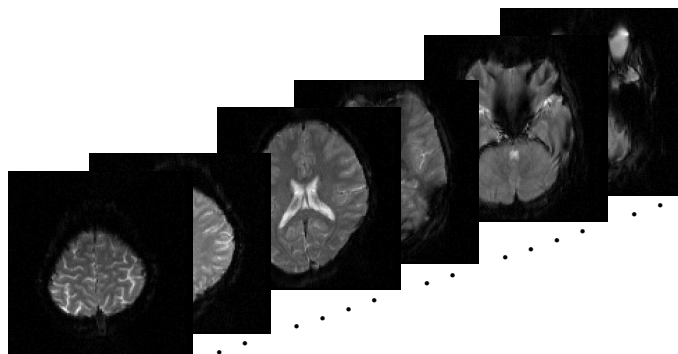
Figure 2.8: Example of some slices of a volume of raw data. The inplane number of pixels is 128×128, the number of slices in a volume is 30. Thus the resolution along the z-axis is lower than in the xy-plane.

2.5.1   Block paradigm

In the case of the block paradigm, fMRI data consist of repetitions of alternating A and B *epochs*. Symbol A stands for the brain volume with activation (the subject performs a certain task) and B for the brain volume in the rest state. An epoch contains a sequence of several repetitions under the same condition. A *cycle* contains one epoch under each condition [17]:

$$\underbrace{\text{A A A A B B B B}}_{cycle} \underbrace{\text{A A A A}}_{epoch} \text{B B} \ldots$$

A time sequence of one voxel of an activated area is shown in Figure 2.9. It is obvious that a lot of noise is present in the data; the underlying signal, which should follow the periodic activation pattern, is hardly recognizable even though the voxel is taken from a strongly activated region (upper limit of the signal-to-noise ratio). The difference image of two slices (A–B) is shown in Figure 2.10. The amount of noise makes it almost impossible to detect any activation with only one realization. Thus the fMRI algorithms are based on averaging over several realizations. Some activation zones are visible by averaging only, see Figure 2.10. However, the limiting factor in averaging is the volunteer's ability to stay that long time in the scanner and to perform identical tasks without moving insight the scanner. Even with few repetitions the images may be corrupted by movement artifacts.

2.5.2   Event-related fMRI

An alternative to the block paradigm is the Event-Related experiment protocol (ER-fMRI), which involves a different stimulus structure. Event-related protocols are charac-
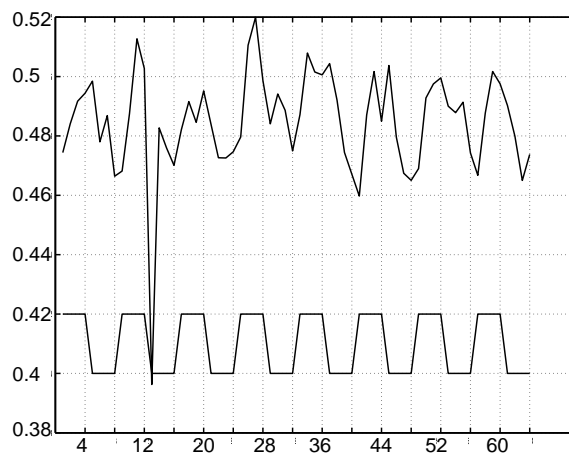
Figure 2.9: A time sequence of an activated voxel. The upper plot is the sequence of the raw data (64 voxels). The lower plot gives the periodic rate of activated (up) state and rest (down) state.

$$A - B \qquad\qquad \overline{A - B}$$
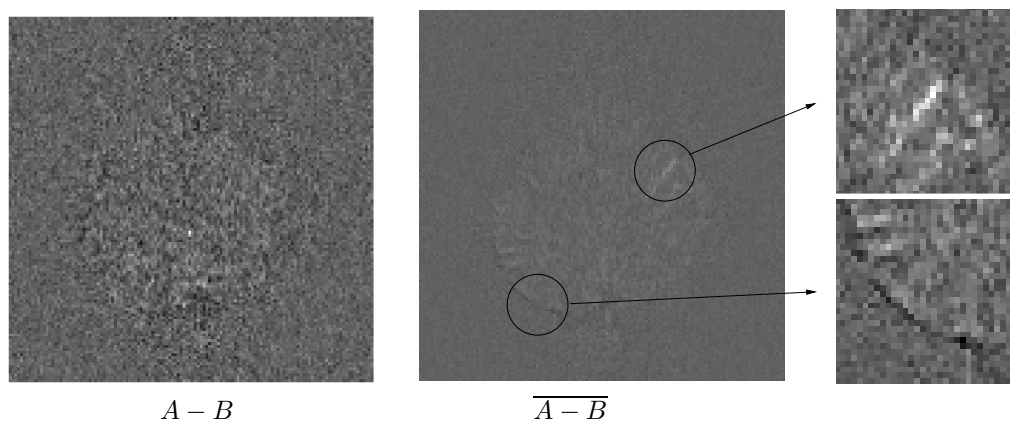
Figure 2.10: Left hand side: average of the difference of one slice of activation and one slice in a rest state. Middle: average of the difference of the activation state and rest state of eight realizations. Right hand side: in the upper image, some activation area is already visible by averaging only; but due to the repetition some movement artifacts could also be interpreted as activation (lower image).

terized by rapid, randomized presentation of stimuli. This is analogous to the technique of *Event-Related Potentials* (ERPs), which uses electric brain potentials that are measured by electrodes on the scalp or implanted in the skull [115]. The disadvantage of ERPs is their bad spatial information, although their time resolution is excellent.

The event-related experiment combines the advantage of time information of ERP and the advantage of good spatial resolution of fMRI. The time information is important for determination of the Hemodynamic Response Function (HRF), which represents the regional BOLD effect signal change after a short stimulus. The time between two stimuli (Inter-Stimulus Interval (ISI)) can be chosen randomly. This has the advantage that the volunteer does not get used to the experiment, which implies that the HRF does not change its shape or decrease in amplitude. This is necessary in order to be able to sum over several realizations [27]. Other advantages, such as the possibility of *post hoc* categorization of an event, are reported in [75, 80, 149]. However, for steady-state analysis, the block paradigm is still the most effective one.

2.6   Preprocessing

We call preprocessing all processes which are prior to the statistical analysis and the possible image transformations of the data.

In the case of the block paradigm, the common practice is to omit the first volume of each block (Figure 2.11) or to shift the paradigm by one or even two volumes. This is because the volumes at the beginning of a block contain a transition signal that would make the steady-state analysis more difficult. Another possibility is to keep this first transition measurement by introducing a hemodynamic function in the model [11, 113]. We restricted ourselves to the first approach as presented in Figure 2.11, which was proposed by our collaborators and enabled us to compare the methods. Also a hypothesis need only be made about the length of the transition signal and not about the shape of the hemodynamic function.

$$A̸ \; A \; A \; A \; B̸ \; B \; B \; B \; A̸ \; A \; A \; A \; B̸ \; B \; B \; B \ldots$$

Figure 2.11: Typical ordering of the fMRI data with suppression of the first volume.

Another very important preprocessing step is motion correction. Getting data with the fewest possible head movements is a requisite. But head motions cannot be completely avoided. A movement of one tenth of a voxel may produce 1–2% signal changes, which is not negligible. In comparison, fMRI BOLD effects are very small: 1–5% signal changes [37, 71]. This calls for the use of accurate image *registration* algorithms [140, 65]. The first step in registration is to determine parameter values for the transformation of the images toward a reference image, the target image, which is often the first acquired

image. The images are transformed by resampling according to the determined geometry, which is evaluated by optimizing some criterion that expresses the goodness of the matching [8]. In order to realign our data, we used mainly the method proposed in SPM [65], see Section 2.8. We also analyzed data realigned with the Automated Image Registration (AIR) method [153, 154]. The third realignment, which we used, is a high-quality realignment procedure developed in our group [140]. Figure 2.12 illustrates the influence of realignment.

a) without realignment   b) with realignment
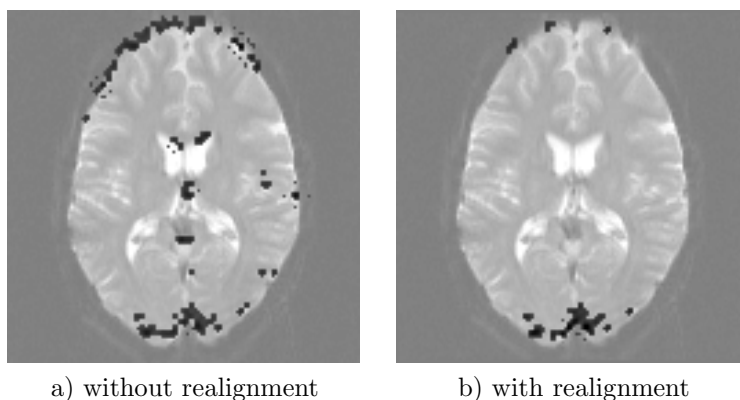
Figure 2.12: Example of the influence of realignment. The activation in the brain was caused by a visual task. a) Detection of activation based on raw data with the algorithm developed in this work. Some of these detected areas are caused by movement artifacts, in particular the one on the border in the upper brain region. b) Detection of activation based on data which were realigned with SPM99. Most of the detected areas which were caused by movement artifacts disappear.

Sometimes it is desirable to compare the analysis of different subjects or even analyze across different subjects. In such cases *spatial normalization* is applied to the data between the subjects. Subjects can even be spatially normalized according to their Euclidian coordinate within a standard space [62]. The most commonly adopted coordinate system within the brain imaging community is that described by Talairach and Tournoux [137]. Matching here is only possible on a coarse scale, since there is not necessarily a one-to-one mapping of the cortical structures between different brains. Consequently, the images are smoothed prior to the matching [8].
No such processing was required in our work since we did not engage in inter-subject studies.

Almost every fMRI scanner acquires the slices of a volume in succession. *Slice timing correction* shifts each voxel's time series so that all voxels in a given volume appear to have been captured at exactly the same time [131].

Spatial filtering with a smoothing kernel is often applied to the data to increase the

signal-to-noise ratio. But this will only work if the activation zone is larger than the size of the smoothing kernel. Smoothing decreases the resolution, whereas the MRI-scanner technique is constantly upgraded to produce higher resolution data.

Global intensity normalization, also known as time normalization, is applied to the data, when the intensities of the voxels drift over time [25, 68]. For this purpose, the mean value of each volume in time is measured, taking only voxels above a certain threshold into account which represents the intracerebral voxels. Then the intensities are adjusted such that each volume has the same mean value. This treatment is more necessary for PET data than for fMRI data, since in the first case the amount of radioactivity in the brain can change during acquisition. Global intensity normalization can affect the detection results when the activation is sufficiently large and very significant such that the mean value is strongly influenced by this activation. It is therefore only useful when there are big changes in global intensity.

The data may include slow, repetitive physiological signals related to the cardiac cycle or to breathing. With a temporal filter, in this case a high-pass filter, these periodic signals can be avoided. Also, scanner-related drifts are suppressed with such a filter.

2.7   An overview of fMRI postprocessing methods

By postprocessing methods, we mean the statistical analysis and possible image transformations; i.e., all treatments that are used to find the activation loci. At this stage, the data have already been preprocessed, e.g., realigned.

2.7.1   Real space methods

An early developed and still often used method is the correlation analysis introduced by Bandettini et al. [11]. The key to this approach is the formation of the correlation coefficient, cc, for each pixel:

$$cc = \frac{\sum_{n=1}^{N} (f_i - \mu_f)(r_i - \mu_r)}{\sqrt{\left(\sum_{n=1}^{N} (f_i - \mu_f)^2\right)\left(\sum_{n=1}^{N} (r_i - \mu_r)^2\right)}}. \tag{2.1}$$

$f_i\,(i = 1 \ldots N)$ is the time-course function at a fixed spatial location. It can be considered to be an N dimensional vector. r_i is a reference waveform or vector. This reference may be an experimental time-course function **f** for some particular pixel or an average of the **f**'s of several experiments, which is then correlated with the time-course function **f** of other pixels. Alternatively, it is a signed sequence (1 1 1 –1 –1 –1 . . .) coding for the paradigm [152]. μ_f and μ_r are the average values of the vectors **f** and **r**, respectively. The value of cc always varies between $+1$ and -1. A threshold value TH between 0 and

+1 is selected and data in each pixel where

$$cc < TH$$

are rejected, i.e., set to zero. A value of 0.7 for TH is a typical threshold [11]. In order to further distinguish between time courses, the application of additional procedures (e.g. amplitude threshold [15] or cluster-size threshold [61]) is advantageous.

Another more powerful statistical approach is the use of a generalized linear model [100] pioneered by Friston et al. [67] which was used for PET studies first and has then been adapted to fMRI; it is called Statistical Parameter Mapping (SPM) and is freely available as a software tool (SPM99). Since SPM is largely used by neurologists and is the *de facto* standard in the field, we devote a whole section to this method (see Section 2.8). We will also compare our results with the results of SPM (cf. Chapter 7).

The analyses discussed so far need prior knowledge about the succession of activation (A) and rest state (B). Among the methods that do not require prior knowledge at all, the earliest are Principal Component Analysis (PCA) [150] and Factor Analysis (FA) [10]. However, these do not work so well for evaluating fMRI data because the total variance of the data is dominated not by the activation changes but by (physiological) noise which is then partitioned into uncorrelated components [13].

Fuzzy Clustering Analysis (FCA) is another technique which is paradigm-independent, fast and reliable [122, 55]. It attempts to find a partition of a dataset X of n feature vectors $(x_1, x_2, \ldots, x_n)$ by producing, for a preselected number of clusters c, c vectors in a feature space $\mathbb{R}^p$, called cluster centroids or prototypes. They represent points around which the data are concentrated. For each datum, a membership vector u_k measures the similarity of the point to each cluster centroid and indicates how well the point is classified. But often, this classification suffers from the high amount of noise present in the data. We believe that a multiresolution approach for classifying could reduce this problem.

For completeness sake, we also mention the existence of other less popular methods that are briefly described and compared to each other in [85]. The conclusion of this author is that a pluralistic empirical strategy coupled formally with substantive prior knowledge of the data set is better than employing a single method only. In [73] different software tools for analyzing fMRI data are described and compared. These software tools are AFNI 2.01 (Analysis of Functional NeuroImages) [38, 36]; SPM96 (Statistical Parameter Mapping) [66]; STIMULATE 5.0 [136]; Yale [128]; MEDIMAX 2.01, conceived by the Infographics Group in 1995; FIASCO (Functional Imaging Analysis Software-Computational Olio) [50]; MEDx 2.0, a commercial multimodality brain imaging processing and analysis software by Sensor Systems; and FIT (Functional Imaging Toolkit) (unpublished). The software BrainVoyager[1], a commercial software-tool conceived by Brain Innovations B.V., offers various possibilities, such as time-course analysis of fMRI. BrainVoyager is easier to handle than SPM96 and is thus being used more and more

---

[1]http://www.brainvoyager.de/

by medical doctors and scientists for analyzing fMRI data. The authors of SPM and BrainVoyager are collaborating on the creation of an interface between these two image analysis environments.

2.7.2 Wavelet methods

The use of the wavelet transform for the analysis of functional Magnetic Resonance Imaging (fMRI) of the brain was pioneered by Ruttimann et al. [120]. Their statistical test consists of a two-stage approach: first all subbands are tested with a χ^2-test globally, to detect the presence of a signal. When the energy in the subband is too low, the subband is discarded. In the second stage, the coefficients of the remaining subbands are thresholded individually via a two-tailed z-test. These authors used orthogonal cubic-spline wavelets.

In [116], Raz et al. perform an analysis of variance (ANOVA) in the wavelet domain by thresholding the wavelet coefficients according to their score in a statistical test. The testing is done per block. At the coarsest level, each block consists of 16 coefficients; at finer levels the block contains all its child coefficients at those levels. They used Daubechies least asymmetric wavelets with 8 filter coefficients, also called "symmlets" [43]. For multiple comparisons correction, they proposed to use the False Discovery Rate (FDR), which is much less conservative than the Bonferroni correction applied by Ruttimann et al.

There have also been some attempts to apply wavelets to the analysis of PET data: Ruttimann et al. [121], Unser et al. [146] and Turkheimer et al. [141]. In [141], the problem of estimation of the spatial dimension is solved by applying the wavelet transform to each scan of the dynamic PET sequence and then performing the kinetic modeling and statistical analysis in the wavelet domain. These authors used a Translation-Invariant Discrete Wavelet Transform (DWT-TI) introduced by [33], which does not imply a unique inverse transform.

In another work [26, 54], the wavelet transform of fMRI time signals is used to remove temporal autocorrelations.

In [102, 101], F. Meyer demonstrates that, with a wavelet-based estimation, a semiparametric generalized linear model of fMRI time-series can remove drifts in time that cannot be adequately represented with low degree polynomials.

In [151], wavelet denoising is performed as an alternative to Gaussian smoothing. The statistical testing (FDR) is done in the space domain. The authors chose to use the symmetric orthonormal spline of degree $\alpha = 5$.

For our part, we have also contributed to this research by extending the statistical testing to the non stationary case [57] and by searching for objective criteria for comparing the performance of various wavelet transforms [59].

2.8 Statistical Parameter Mapping (SPM)

Statistical Parameter Mapping (SPM) is a method that performs the whole analysis process for fMRI as well as for PET and SPECT. A software package is freely available, SPM96 and SPM99 (released in 2000), which is widely used by researchers analyzing functional images of the brain. SPM is also the acronym for the statistical parametric map that is calculated by most fMRI statistical analysis packages.

The main process of the data transformation in SPM is illustrated in Figure 2.13. The first part includes the preprocessing step: realignment of the time series data, normalization of the data to compare different subjects analysis and the Gaussian smoothing for reducing the noise. This part is described briefly in Section 2.8.1. The second part contains the model specification and the parameter estimation that are based on a *general linear model*. The general linear model is defined through a *design matrix*, which can be modelled by the user. The result is a statistical parametric map. This part is described in Section 2.8.2. The third part consists in thresholding the statistical parameter map based on the theory of Gaussian random fields, which is described in Section 3.2.6. We compare our method with the part of SPM that includes smoothing and statistical analysis only.
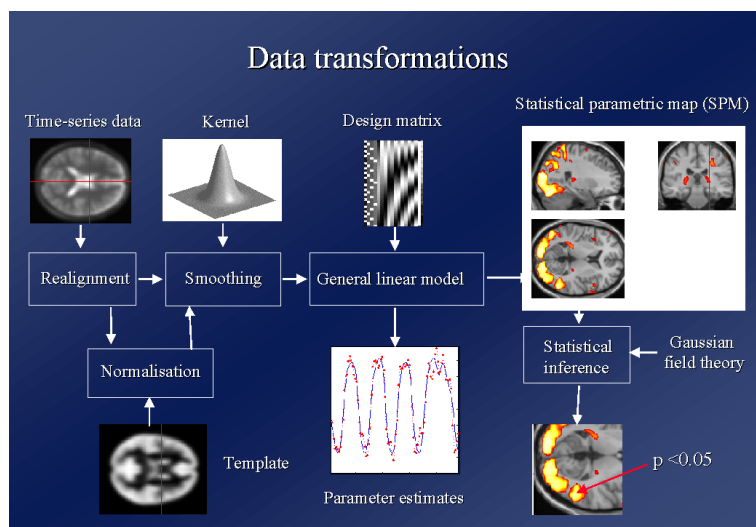

Figure 2.13: An overview of the SPM process, reproduced from K. Friston [64].

2.8.1 Data preprocessing in Statistical Parameter Mapping

Spatial transformation of images

SPM contains an automatic, non landmark-based registration algorithm. By registration, we mean the determination of the geometric transformation (typically rigid body) parameters that map the image onto a reference. To apply the spatial transformation, it is necessary to resample the image. SPM gives different choices of interpolation methods for resampling to the user. The proposed one is the "sinc interpolation", which is the most accurate but also the most time consuming method. For speed considerations a rough approximation using nearest neighbor interpolation is also available. For the transformation method, SPM differentiates between different cases. The first case is the *within modality images co-registration*. It is the case where images of the same subject are transformed. To minimize the difference of two such images, a rigid body transformation, which is defined by six parameters in 3D, is used. The *between modality image co-registration* describes the matching between images of different modalities like PET and fMRI, but of the same subject. This algorithm is based on matching homologous tissue classes together. The tissues are classified as gray matter, white matter or Cerebro-Spinal Fluid (CSF). This method also does not need a preprocessing step or landmark identification, but employs template images of the same modalities and probability images of the different tissue types. First, the affine transformations that map images to the templates are determined by minimizing the sum of squares differences. Then the images are segmented using the probability images and a modified mixture model algorithm. Finally, the image partitions are co-registered using the parameters from the mapping between the images and the templates as a starting estimate.

The *affine spatial normalization* allows for the registration of images from different subjects. Here, prior information about the variability of head sizes is used in the optimization procedure. Also zooms and shears are needed to register heads of different shapes and sizes. A Bayesian framework is used, such that the registration searches for the solution that maximizes the a posteriori probability of it being correct [65]. *Nonlinear spatial normalization* corrects large differences between head shapes, which cannot be resolved by affine transformation alone. These nonlinear warps are modelled by linear combinations of smooth basis functions [9].

Spatial smoothing

To increase the signal-to-noise ratio and to condition the data so that they conform more closely to a Gaussian-field model, SPM uses a Gaussian kernel for spatial smoothing. The amplitude of a Gaussian, j units away from the center, is defined by

$$g[j] = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{j^2}{2s^2}}. \tag{2.2}$$

The standard deviation s is defined by $\text{FWHM}/\sqrt{8ln(2)}$, where FWHM is the *Full Width at Half Maximum* of the Gaussian. The main disadvantage of smoothing the data is the loss of spatial resolution (suppression of high-frequency components).

2.8.2 The general linear model

SPM treats each voxel independently in its statistical analysis. Thus the response variable Y_j is a random variable that represents a voxel value for a certain observation (scan) j in time, where $j = 1 \ldots J$. The predictor variables—also called the explanatory variables—are denoted by x_{jl}, where $l = 1 \ldots L$, $(L < J)$ indexes this variable. x_{jl} can also be called an independent variable. It is used to predict the data in the response variable y_j in terms of a linear combination (linear regression). This variable might consist of information about the task that the volunteer has to perform. In the case of the block paradigm, it might consist simply of the activation and nonactivation statements, like the lower plot in Figure 2.9. The unknown parameters, which have to be worked out and which define the relation between the predictor variables and the response variables, are denoted by β_l and correspond to the L predictor variables. They fit the model such that the residuals ϵ_j are minimized. Ideally, they are independent and identically distributed normal random variables with zero mean and variance σ^2, $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$. Thus the general linear model is expressed by

$$Y_j = x_{j1}\beta_1 + \cdots + x_{jl}\beta_l + \cdots + x_{jL}\beta_L + \epsilon_j. \tag{2.3}$$

The matrix formulation of this model is given by

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{2.4}$$

where $\mathbf{Y}$ is the column vector of observations (of voxel values), $\mathbf{X}$ is the *design matrix* of size $J \times L$, β is the column vector of unknown parameters and ϵ is the column vector of residual terms.

The task now is to find the estimates for β in such a way that the model provides a good fit to the observations $\mathbf{Y}$. This is performed by least squares minimization. The estimates of β are denoted by $\tilde{\beta}$ and the resulting fitted values are denoted by $\tilde{\mathbf{Y}}$, which results in $\tilde{\mathbf{Y}} = \mathbf{X}\tilde{\beta}$. The difference between the observed value $\mathbf{Y}$ and the predicted value $\tilde{\mathbf{Y}}$ is given by the residual error $\mathbf{e}$, thus $\mathbf{e} = \mathbf{Y} - \tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\tilde{\beta}$. The least squares estimates $\tilde{\beta}$ are obtained by minimizing

$$S \ = \ \sum_{j=1}^{J} \mathbf{e}_j^2 \tag{2.5}$$

$$= \ \sum_{j=1}^{J} (Y_j - x_{j1}\beta_1 - \cdots - x_{jL}\beta_L)^2. \tag{2.6}$$

The minimum is reached by $\beta = \tilde{\beta}$, which satisfies

$$\frac{\partial S}{\partial \tilde{\beta}_l} = 2 \sum_{j=1}^{J} (-x_{jl}) \left(Y_j - x_{j1}\tilde{\beta}_1 - \cdots - x_{jL}\tilde{\beta}_L \right) = 0, \tag{2.7}$$

for every $j = 1 \ldots J$. In matrix notation, (2.7) is equivalent to

$$\mathbf{X}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{X})\tilde{\beta} \tag{2.8}$$

and is called the *normal equation*.

When the design matrix $\mathbf{X}$ is of full rank, the least squares estimates, denoted by $\hat{\beta}$, are

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \tag{2.9}$$

The resulting parameter estimates $\hat{\beta}$ are normally distributed with $E\{\hat{\beta}\} = \beta$ and

$$\text{Var}\{\hat{\beta}\} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \tag{2.10}$$

The least squares estimates also correspond to the maximum likelihood estimates if one assumes that the input data are corrupted by additive white Gaussian noise. The residual mean square $\hat{\sigma}^2$ is the estimation of the residual variance σ^2 and is given by

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T \mathbf{e}}{J - p} \sim \sigma^2 \frac{\chi^2_{J-p}}{J - p}, \tag{2.11}$$

where $p = \text{rank}(X)$. Furthermore, $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

The task now is to test the effects, i.e. the predictor variables. We might want to test only for one effect β_l, taking into account the other effects, or we might want to test one effect of one condition (e.g. activation paradigm) against another condition (e.g. control paradigm). Thus SPM introduces a contrast weight vector $\mathbf{c}$, simply called the contrast of a model, to choose the desired test, where each weight corresponds to an effect. In practice, SPM allows to define and test for different contrasts. When we test for one effect the contrast might be $\mathbf{c}^T = [0 \ldots 0\,1\,0 \ldots 0]$, and when we test two effects against each other, the contrast might be $\mathbf{c}^T = [0 \ldots 0\,1\,{-1}\,0 \ldots 0]$.

Thus (2.10) changes to

$$\mathbf{c}^T \hat{\beta} \sim \mathcal{N} \left(\mathbf{c}^T \beta, \sigma^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} \right). \tag{2.12}$$

The corresponding null hypothesis (cf. Chapter 3) is given by $\mathcal{H}_0 : \mathbf{c}^T \hat{\beta} = \mathbf{c}^T \beta$ and the test statistic is given by

$$T = \frac{\mathbf{c}^T \hat{\beta} - \mathbf{c}^T \beta}{\sqrt{\hat{\sigma}^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}}. \tag{2.13}$$

Under the null hypothesis $\mathcal{H}_0$, T follows a Student's t-distribution with $J - p$ degrees of freedom and can therefore be tested with a Student's t-test (cf. Section 3.2).

The majority of the statistical models used for testing are special cases of the general linear model. One example is the two-sample t-test: two groups of different conditions are compared (e.g. activation condition against rest state). Then the response variable $Y1_j$, which stands for the activation state, and $Y2_j$, which stands for the rest state, are assumed to be two independent random variables, with n_1 and n_2 realizations, respectively, where $Y1_j \sim \mathcal{N}(\mu_1, \sigma^2)$ and $Y2_j \sim \mathcal{N}(\mu_2, \sigma^2)$. They can be rearranged into one variable, where $Y_j, j = 1, \ldots, n_1$ represents $Y1_j$ and where $Y_j, j = n_1 + 1, \ldots, n_1 + n_2$ represents $Y2_j$. We assess the null hypothesis:

$$\mathcal{H}_0 : \mu_1 = \mu_2. \tag{2.14}$$

Thus the parameter vector takes the value $\beta = (\mu_1, \mu_2)^T$ which has to be tested. The design matrix X contains two columns of *dummy variables*. The first column represents the membership of the group $Y1_j$, containing n_1 1's and n_2 0's; the second column contains n_1 0's and n_2 1's. The general linear model of (2.3) yields

$$Y_j = x_{j1}\mu_1 + x_{j2}\mu_2 + \epsilon_j \tag{2.15}$$

and the null hypothesis of (2.14) is equal to $\mathcal{H}_0 : \mathbf{c}^T\beta = 0$ with $\vec{c} = (1, -1)^T$, since we test if there is a difference between the samples' mean values. Thus the t-test of (2.13) results in

$$T = \frac{\mu_1 - \mu_2}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \tag{2.16}$$

with $n_1 + n_2 - 2$ degrees of freedom, since $(X^T X) = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}$ and $\mathbf{c}^T(X^T X)^{-1}\mathbf{c} = \frac{1}{n_1} + \frac{1}{n_2}$.

One-way analysis of variance (ANOVA)

Assuming we have different groups of samples or observations, their mean values are $\bar{y}_1, \bar{y}_2, \ldots \bar{y}_c$. There are two components contributing to the variations of those observations. One can measure the variability between the group means or the variation within each group. These two components are then measured by means of an F-test.
The combined estimate of σ^2 from the variation within groups is given by

$$s^2 = \sum_{i=1}^{c} \frac{s_i^2}{c} = \sum_{i=1}^{c} \sum_{j=1}^{J} \frac{(y_{ij} - \bar{y}_i)^2}{c(J-1)}. \tag{2.17}$$

s^2 has $c(J-1)$ degrees of freedom. The observed variance of the group means is given by

$$\sum_{i=1}^{c} \frac{(y_i - \bar{y})^2}{c-1},\tag{2.18}$$

where $\bar{y}$ is the mean value of the group means. Thus

$$s_G^2 = n \sum_{i=1}^{c} \frac{(y_i - \bar{y})^2}{c-1}\tag{2.19}$$

is an unbiased estimate of σ^2 based on $c-1$ degrees of freedom. If the hypothesis $\mathcal{H}_0$, which assumes that there is no difference between the groups, is true, both s^2 and s_G^2 are estimates of σ^2 and thus the ratio $F = s_G^2/s^2$ will follow an F-distribution with $c-1$ and $c(J-1)$ degrees of freedom. When there is a difference between the groups, s_G^2 will be increased by the group differences and thus the F-ratio increases as well. If this ratio is significantly large, the null hypothesis $\mathcal{H}_0$ is rejected and the alternative hypothesis that there is a difference between the groups is accepted [28].
In the case of fMRI data the different groups might represent the voxel acquisitions under different patient conditions.

2.8.3 SPM: further considerations

While the general linear model and ANOVA of SPM are applied to the spatial domain (preprocessed data) directly, it is perfectly conceivable to also apply these techniques in the wavelet domain as is proposed in this thesis.

The major difference in procedure is the way in which one deals with the correction for multiple testing, i.e. the specification of an appropriate threshold on the statistical parameter map provided by the estimation module.

In the case of SPM, the situation is complicated by the use of the Gaussian prefilter which correlates the data and renders a standard Bonferroni correction much too conservative to be of much practical use. SPM gets around the difficulty by applying a threshold for Gaussian "correlated" data derived from the theory of Gaussian random fields (see Section 3.2.6).

2.9 Human brain anatomy

A short introduction to human brain anatomy is given here for reference and for a better understanding of the results obtained with real data in Section 7.4.
An adult human brain weighs about 1.5 kilos and can be divided into five major parts (see also Figure 2.14):

→ Brain stem: regulates autonomic functions like breathing, circulation and digestion. It also controls paradoxical sleep and is responsible for auditory and visual startle reflexes.

→ Thalamus: routes auditory, visual, and tactile sensory input to appropriate regions of the cerebral cortex.

→ Cerebellum: controls voluntary movements, balance and motor coordination.

→ Cerebrum: it is the main mass of the brain, comprising 85% of its weight. The cerebrum is divided into two halves, the left and the right hemispheres. The top 2–4 mm of the cerebrum is the cerebral cortex, the gray matter of the brain. Under the cortex is the white matter, composed of neuronal fibers linking different areas of the cortex and of brain stem and cortex. In humans, the cerebral cortex is strongly convoluted, with ridges (termed *gyri*) and valleys (termed *sulci*).

→ Corpus callosum: links the left and right hemispheres of the cerebrum through a bundle of nerves.
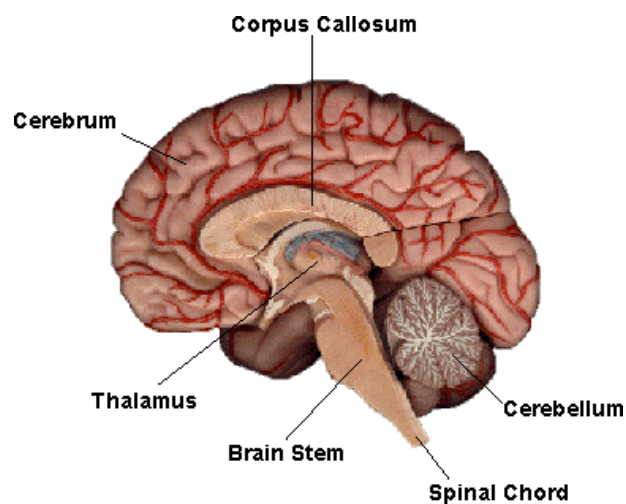
Figure 2.14: The five major parts of the brain [1].

The cerebral cortices of each hemisphere are divided into four lobes (see also Figure 2.15):

→ Frontal lobe, which is the center of consciousness, and deals with planning, judgement, emotional responses, and memory for habitual motor skills (in the motor cortex). It contains also Broca's area, which controls speaking ability (cf. Section 2.1). Through collaboration with Arto Nirkko from the Inselspital Bern, we

have obtained data which result from activation in the motor task. One of the
tasks was finger tapping, which is the most typical task for fMRI analyses since
it causes a relatively strong signal; a typical example of detection result with our
method is shown in Figure 7.46.

→ Parietal lobe, which processes mainly information about movement and body
   awareness (in the sensory cortex, see Figure 2.15).

→ Temporal lobe, which processes auditory information and deals with memory acqui-
   sition and object categorization. It also contains Wernicke's area, which is responsi-
   ble for the ability to understand language. When this area is damaged, the patient
   can speak clearly but the words that are put together make no sense. This is in
   contrast to a damage in Broca's area, in which case the patient can still understand
   language but cannot produce speech. One of our processed data sets corresponds to
   a task where the subject has to listen to bi-syllabic words: see results of activation
   in Figure 7.57.

→ Occipital lobe, which processes visual information. Through different collabora-
   tions, we obtained data which contained activation in the occipital lobe. One task
   was just switching on and off a light source. Activation results with our method
   are shown in Figure 7.56.
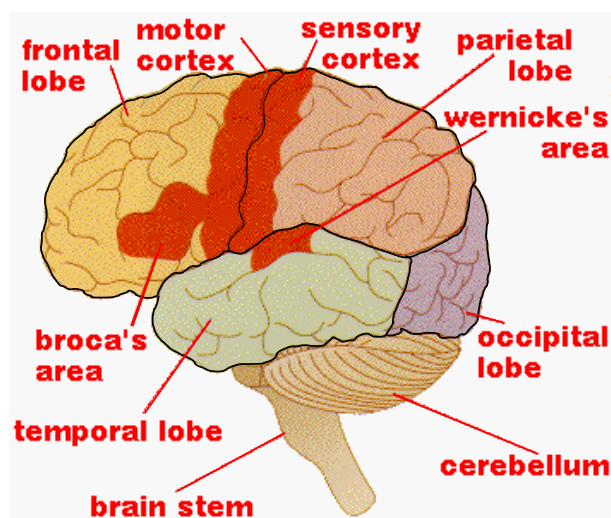
Figure 2.15: The four lobes of the cerebral cortices [1].

Further literature devoted to the understanding of the functional activity of the brain can be found in [132].

Chapter 3

Statistical Processing of fMRI Data

3.1 Statistical tools
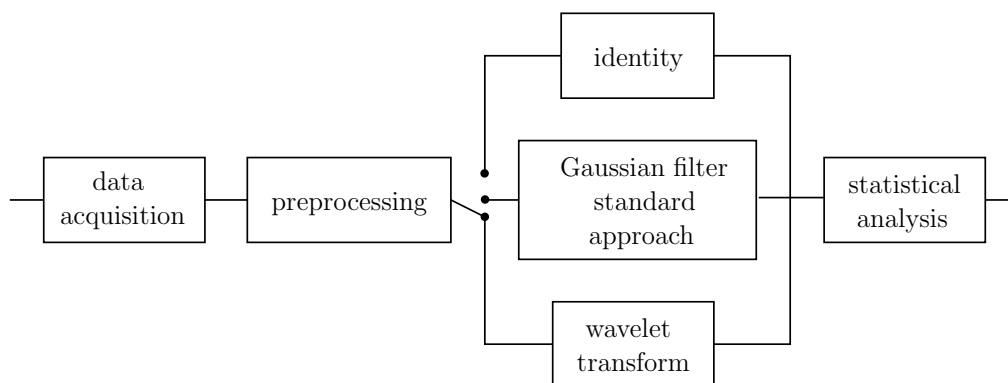
Figure 3.1: Order of the procedures before the statistical analysis.

The statistical tools that are used for the algorithm in this work are described in this section. The application to fMRI can be found in Section 3.3. The block diagram of Figure 3.1 shows where the statistical processing is situated in the analysis of fMRI. In our case, the statistics is applied in the wavelet domain on the wavelet coefficients. In SPM, a Gaussian filter is applied and the detection is done at the pixel level. In principle, the statistics are also applicable to the unprocessed pixels directly, but the SNR is usually too low for this direct method to yield some detection. The statistical

analysis, proposed in this section, is generic and applicable in all three cases.

3.1.1   Random variables

In our statistical process of fMRI-data, we have to deal with hypothesis tests, probability distributions, and thus with random variables. To introduce these topics, we shortly describe the tools which are necessary for their understanding.

A random variable is in our case simply the numerical variable under consideration. For each trial, a real (discrete) signal $x[.]$ is obtained as a realization of the random process $X[.]$. The random variable is represented with a capital letter and its realization with a lower-case letter. An example of a random process is the game of dice, where the random variable represents the outcomes of the dice. Since the outcome is stochastic, only a probability that the random variable X may take a certain value x can be given. The probability function of $X[.]$ is a function $p_X(x)$ which assigns probabilities to all possible outcomes. That is

$$p_X(x) = P(X = x). \tag{3.1}$$

A probability function must satisfy the following properties:

(i) $p_X(x) \geq 0$, for all $x \in A$,

(ii) $\sum_{x \in A} p_X(x) = 1$,

with A the domain of values that X can take. In the continuous case, the random variable is described by the probability density function p_X, defined on the domain A. It satisfies the following properties:

(i) $p_X(x) \geq 0$, for all $x \in A$,

(ii) $\int_{x \in A} p_X(x)dx = 1$,

(iii) for any interval $[a, b] \in A$,

$$P(a \geq X \geq b) = \int_a^b p_X(x)dx.$$

A proposition $x \geq a$—like we make it for a hypothesis test—for the random variable X and a proposition $y \geq b$ for a random variable Y are *independent*, if and only if $P(x \geq a \wedge y \geq b) = P(x \geq a)P(y \geq b)$ for all a and b. Intuitively, X and Y are independent if any information about one variable doesn't tell anything about the other variable; otherwise

$$P(x \geq a \wedge y \geq b) = P(x \geq a | y \geq b)P(y \geq b). \tag{3.2}$$

In general, we have

$$P(x \geq a \vee y \geq b) = P(x \geq a) + P(y \geq b) - P(x \geq a \wedge y \geq b), \tag{3.3}$$

and when X and Y are mutually exclusive, we obtain $P(x \geq a \vee y \geq b) = P(x \geq a) + P(y \geq b)$.

A discrete stochastic process is *stationary* when the random vectors $(X[0], X[1], \ldots, X[n-1])$ and $(X[i], X[i+1], \ldots, X[i+n-1])$ have the same probability density for each $n \in \mathbb{Z}_+$ and each $i \in \mathbb{Z}$. In other words, when the stochastic process is observed through a (time) window of length n, then the observed statistics doesn't depend on the location (i) of the window. All the statistical properties of a stationary stochastic process are time-independent. One example of a stationary random variable is an independent and identically distributed (*i.i.d.*) random variable. For an i.i.d. process, all random variables $X[k]$, $k \in \mathbb{Z}$, have the same probability density and are independent. Thus an i.i.d. process is statistically fully described by only one probability distribution for one random variable, for example $X[0]$.

If a discrete random variable X can take values $x_1, x_2, \ldots, x_n$, with probabilities $p_1, p_2, \ldots, p_n$, then the expected value is defined as

$$E[X] = \sum_{i=1}^{N} x_i p_i. \tag{3.4}$$

In the continuous case, we obtain

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx. \tag{3.5}$$

The variance of X is given for $\mu = E[X]$ by

$$\text{var}[X] = E\left[(X - \mu)^2\right] \tag{3.6}$$

$$= E\left[X^2\right] - (E[X])^2. \tag{3.7}$$

A discrete stochastic process $X[.]$ is *weakly stationary* or *wide-sense stationary*, when both $E[X[i]]$ and $E[X[i+k]X[i]]$ for $k \in \mathbb{Z}$ are independent of i.

In the following section some properties of the expected value are given, for the proof see [28]:

If X is any random variable then

$$E[bX + c] = bE[X] + c, \tag{3.8}$$

when b and c are constants.

If X and Y are any two random variables then

$$E[X + Y] = E[X] + E[Y]. \tag{3.9}$$

If X and Y are independent random variables, then

$$E[XY] = E[X]E[Y] \tag{3.10}$$

and

$$\text{var}[X + Y] = \text{var}[X] + \text{var}[Y]. \tag{3.11}$$

For any random variable X, it holds that $\text{var}[cX] = c^2\text{var}[X]$ and $\text{var}[X + c] = \text{var}[X]$. A stochastic process $X[.]$ is called white noise with variance σ^2, when $X[.]$ is weakly stationary, $E[X[i]] = 0$ and

$$
\begin{aligned}
R_X[k] &= E[X[i+k]X[i]] \\
&= \sigma^2\delta[k],
\end{aligned}
\tag{3.12}
$$

where $R_X[.]$ is the autocorrelation function of $X[.]$.

3.1.2   Multivariate random variables

Suppose we have n p-dimensional observations $\mathbf{X}[1], \mathbf{X}[2], \ldots, \mathbf{X}[n]$. Then each $\mathbf{X}[i] = (X_1[i], \ldots, X_p[i])^T$ denotes a random vector of dimension p, where p random variables $X_1[i], \ldots, X_p[i]$ are defined on a sample space. The mean vector of $\mathbf{X}$, denoted by $\bar{\mathbf{X}}$, is given by:

$$
\begin{aligned}
\bar{\mathbf{X}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}[i] \\
&= \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{pmatrix}.
\end{aligned}
\tag{3.13}
$$

The empirical scatter covariance matrix $\mathbf{S}$ of the n p-variate random vectors $\mathbf{X}[\mathbf{i}]$ is a $p \times p$ matrix

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}[i] - \bar{\mathbf{X}})(\mathbf{X}[i] - \bar{\mathbf{X}})^T \tag{3.14}$$

The expected value of $\mathbf{S}$ is the covariance matrix defined by

$$E[\mathbf{S}] = \begin{pmatrix} \text{var}[X_1] & \text{cov}[X_1, X_2] & \ldots & \text{cov}[X_1, X_p] \\ \text{cov}[X_2, X_1] & \text{var}[X_2] & \ldots & \text{cov}[X_2, X_p] \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}[X_p, X_1] & \text{cov}[X_p, X_2] & \ldots & \text{var}[X_p] \end{pmatrix}. \tag{3.15}$$

Univariate	Multivariate
Normal	Multivariate normal
Variance	Covariance (dispersion) matrix
χ^2-distribution	Wishart's distribution
Student's t-distribution	Hotelling's T^2
Fisher's z-distribution	Ratio of two covariance-type determinants
Maximum likelihood estimation	Maximum likelihood estimation
Analysis of variance (ANOVA)	Multivariate analysis of variance (MANOVA)

Table 3.1: Correspondence between univariate and multivariate statistics [78].

The covariance matrix can as well be expressed by:

$$\mathbf{S} = \overline{\mathbf{X}\mathbf{X}}^T - \bar{\mathbf{X}}\bar{\mathbf{X}}^T. \tag{3.16}$$

Table 3.1 provides a brief summary of some results in multivariate theory that are a generalization of standard notations in the univariate statistical theory [78, 89].

3.1.3 Probability density distributions

We now briefly describe some probability distributions that will be relevant in the context of fMRI.

z-distribution

Let x be a normally distributed random variable with mean μ and standard deviation σ. Then, the normalized variable

$$z = \frac{x - \mu}{\sigma}$$

is Gaussian with mean $= 0$ and standard deviation $= 1$.
It has a so-called z-distribution (normalized Gaussian), which is given by

$$F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2},$$

in one dimension. In several dimensions; i.e., $\mathbf{x} = (x_1, ..., x_D)$, we have

$$F(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det \mathbf{A}}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}}, \tag{3.17}$$

where $\boldsymbol{\mu} = E[\mathbf{x}]$ and $\mathbf{A} = E[(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})]$.
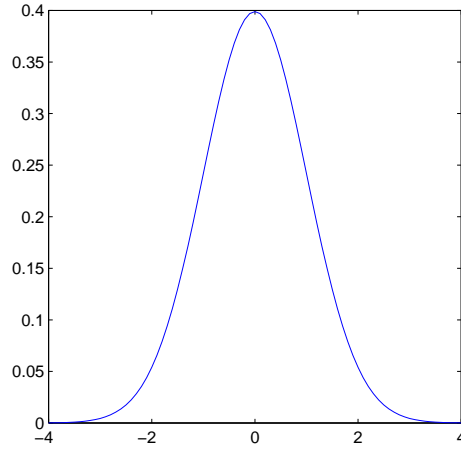
Figure 3.2: Probability density for the normal distribution for $\mu = 0$ and $\sigma = 1$.

χ^2-distribution

Let $x_1, \ldots, x_n$ be n independent normal z-distributed random variables with variance $\sigma^2 = 1$. Then $\xi = \sum_{i=1}^{n} x_i^2$ follows a χ^2-distribution with n degrees of freedom. The χ^2-distribution density is given by:

$$F(\xi) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\xi^{\frac{n}{2}-1}e^{-\frac{\xi}{2}}, \qquad (3.18)$$

where $\Gamma(x) = (x-1)!$ with $x \in \mathbb{N}$ is Euler's Gamma function which generalizes the factorial. Note that $\frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k - \overline{x})^2$ follows a χ^2 distribution with $n-1$ degrees of freedom.

Student t-distribution

If the random variable X is N(0,1), Y is χ^2 with $\nu \in \mathbb{Z}$ degrees of freedom, and X and Y are independent, then the random variable

$$t = \frac{X}{\sqrt{\frac{Y}{\nu}}}$$

is t-distributed with ν degrees of freedom.
The t-distribution density is given by

$$F(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$
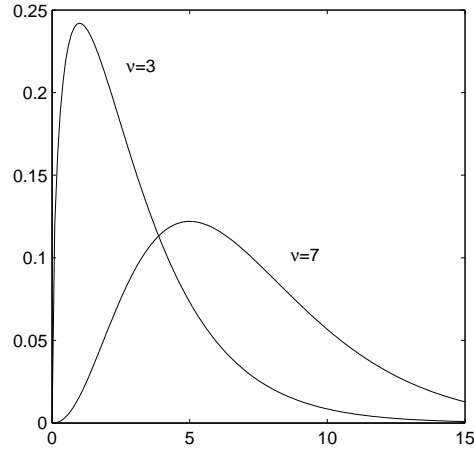
Figure 3.3: Probability density of the χ^2-distribution for several degrees of freedom.

In particular:

$$\sqrt{n-1}\frac{\overline{x}-E[x]}{\sqrt{\sum_{k=1}^{n}(x_k-\overline{x})^2}}$$

follows a t-distribution with $n-1$ degrees of freedom.

F-distribution

If Y_1 and Y_2 are independent χ^2-distributed random variables with ν_1, ν_2 degrees of freedom, then the random variable

$$F = \frac{\frac{Y_1}{\nu_1}}{\frac{Y_2}{\nu_2}}$$

is F-distributed. The F-distribution is given by

$$F(t) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\nu_1^{\frac{\nu_1}{2}}\nu_2^{\frac{\nu_2}{2}}\frac{t_+^{\frac{\nu_1-2}{2}}}{(\nu_1 t + \nu_2)^{\frac{\nu_1+\nu_2}{2}}}.$$

Hotelling's T^2-distribution

If $\mathbf{x}$ and $\mathbf{M}$ are independently distributed as $\mathcal{N}(\mu,\Sigma)$ and $\mathcal{W}_p(\Sigma,m)$, respectively, then

$$m(\mathbf{x}-\mu)^T\mathbf{M}^{-1}(\mathbf{x}-\mu) \sim \mathcal{T}^2(p,m), \tag{3.19}$$
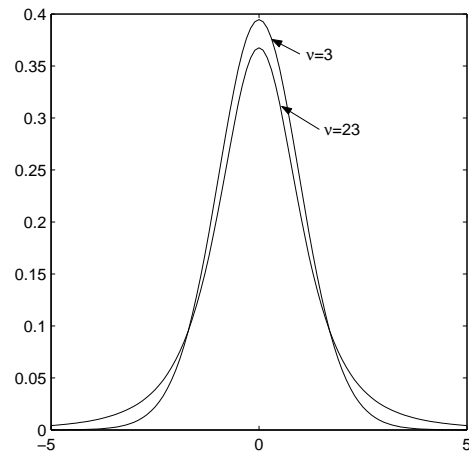
Figure 3.4: Probability density of the Student t-distribution.
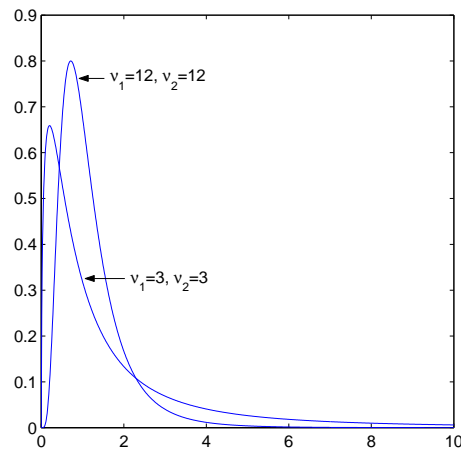
Figure 3.5: Probability density of the F-distribution.

where $\mathcal{W}_p(\Sigma, m)$ is defined as follows:

if $\mathbf{M}$ is a matrix of $(p \times p)$-dimension and can be written as $\mathbf{M} = \mathbf{X}^T\mathbf{X}$, where $\mathbf{X}$ is of size $(m \times p)$ and is $\mathcal{N}(0, \Sigma)$ distributed, then $\mathbf{M}$ is said to have a Wishart distribution with scale matrix Σ and degrees of freedom of m. The Wishart distribution is the multivariate version of the χ^2-distribution (cf. Table 3.1) and is written as $\mathbf{M} \sim \mathcal{W}_p(\Sigma, m)$.

Moreover,

$$
\begin{aligned}
\mathcal{T}^2(p, m) &= m\frac{\chi_p^2}{\chi_{m-p+1}^2} \\
&= \frac{mp}{m-p+1}\mathcal{F}_{p, m-p+1} \quad\quad (3.20)
\end{aligned}
$$

$$(3.21)$$

For derivations and proofs, see [97].

If $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the mean vector and covariance matrix of a sample of size n from $\mathcal{N}_p(\mu, \Sigma)$ than

$$(n-1)(\bar{\mathbf{x}} - \mu)^T\mathbf{S}^{-1}(\bar{\mathbf{x}} - \mu) \sim \mathcal{T}^2(p, n-1). \quad\quad (3.22)$$

Central limit theorem

Let $X_1, X_2, \ldots$ be independent, identically distributed (i.i.d.) random variables with expected value $E[X_i] = \mu$ and variance $\text{var}[X_i] = \sigma^2 < \infty$ for $i \in \mathbb{N}$. Then for every $t < \infty$, it holds that

$$\lim_{n \to \infty} P\left(\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \leq t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}}\mathrm{d}x. \quad\quad (3.23)$$

This theorem states that the distribution of the sum $X_1 + X_2 + \cdots + X_n$ can be approximated by a normal distribution with expected value $E[X_1 + \cdots + X_n] = nE[X_i] = n\mu$ and variance $\text{var}[X_1 + \cdots + X_n] = n\text{var}[X_i] = n\sigma^2$, when n is large.

3.2   Hypothesis testing

In this section, we describe the validation or rejection of a hypothesis, based on a finite number of realizations of some random variables. We consider data that can be described by a probability distribution. In particular, one of its statistical parameter μ is unknown. We propose a particular value μ_0 for μ and we want to be able to decide from a set of measurements whether the hypothesis $\mu = \mu_0$ is trustworthy or must be rejected. A numerical method for testing such a hypothesis or theory is called *hypothesis test* or *significance test*. In any such situation, the experimenter has to decide, if possible, between two rival possibilities. The hypothesis we want to test is called the *null hypothesis*, and is

denoted by H_0. Any other hypothesis is called the *alternative hypothesis*, and is denoted by H_1. Thus, such a hypothesis test could be:

1. Null hypothesis H_0: $\mu = \mu_0$,

2. Alternative hypothesis H_1: $\mu > \mu_0$.

The standard procedure is to summarize the data by a test statistic y that depends on the value μ_0. Under the null hypothesis, the distribution for this statistic is known. Even when H_0 is true, we cannot expect that μ matches μ_0 exactly. Thus we have to think of an expected bound, which should not be exceeded. If y exceeds this expected bound, the null hypothesis is rejected and the alternative hypothesis H_1 is accepted; the decision is made by setting a threshold. Thus, with such a hypothesis test, we can never state that H_0 is true, only that it is trustworthy. On the other hand, we can state that H_0 is wrong, with high evidence whenever we reject the null hypothesis.

Given some probability value α, we compute the threshold t such that the probability that y exceeds t is equal to α, assuming that H_0 is true:

$$P(y \geq t) = \alpha.$$

In other words, α is the probability of observing higher values than t by chance, given that the null hypothesis is true. This probability α is called the *level of significance* of the test.

The p-value is the probability that a test statistic is at least as extreme as the one observed, given that the null hypothesis is true. The smaller the p-value, the stronger the evidence against H_0. If the level of significance of the test statistic is less than 5 %, we say that the result is significant at the 5 % level. This is generally considered as a reasonable evidence that H_0 is not true.

The test on the set of data has to be adapted to its distribution, which transforms the values into p-values.

3.2.1   Two-tailed test

In the section above, the alternative hypothesis to H_0 is defined as H_1: $\mu > \mu_0$. Any such test where the alternative hypothesis tests for values only higher than μ_0 or only lower than μ_0, the test is called *one-tailed* or *one-sided*. But H_1 can also be defined as H_1: $\mu \neq \mu_0$, where the departure of the null hypothesis in both directions is of interest. This means that we are interested in significantly high or low results, which is called appropriately a *two-tailed test* [28].

3.2.2 z-test

Here, a random sample with size n is taken from a normal distribution with unknown mean μ (cf. Section 3.1.3). We want to test if the sample mean is equal to a particular

value μ_0 or not. Thus the null hypothesis is given by

$$H_0 : \mu = \mu_0. \tag{3.24}$$

If the variance σ^2 is constant over the whole set of samples and is known, then $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ follows a normal-distribution, and a z-test may be applied:

$$y = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}. \tag{3.25}$$

If a two-tailed test is appropriate, the level of significance is obtained by calculating the threshold t such that

$$\alpha = P(|y| \geq |t|) \tag{3.26}$$
$$= 2P(y \geq |t|); \tag{3.27}$$

i.e., t is the p-value, related to the significance level α.

3.2.3 Student's t-test

If the variance σ^2 is unknown, we apply a t-test (see Student's t-distribution in Section 3.1.3) with $n-1$ degrees of freedom. The corresponding test statistic with $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is

$$y = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \tag{3.28}$$

where $\mu_0 = E[\bar{x}]$ and where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} |x_i - \bar{x}|^2. \tag{3.29}$$

Two sample t-test

The t-test described above is a one sample t-test. In the case of a two sample t-test, the means of two samples of data are compared. The sizes of the two samples are n_1 and n_2, respectively, with unknown statistical means μ_1 and μ_2, respectively and same standard deviation σ. The first sample mean (measured mean) will be denoted by $\bar{x}_1$ and the second one will be denoted by $\bar{x}_2$. The problem now is to test the null hypothesis

$$H_0 : \mu_1 = \mu_2. \tag{3.30}$$

The alternative hypothesis is expressed in the case of a two-tailed test by

$$H_1 : \mu_1 \neq \mu_2. \tag{3.31}$$

The test statistic is, according to (3.28), obtained by

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \tag{3.32}$$

The unbiased estimate s^2 of σ^2 is given by

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \tag{3.33}$$

where the standard deviation s_1 is obtained by

$$s_1^2 = \sum_{i=1}^{n_1} \frac{(x_1[i] - \bar{x}_1)^2}{n_1 - 1} \tag{3.34}$$

and the standard deviation s_2 is given by

$$s_2^2 = \sum_{i=1}^{n_2} \frac{(x_2[i] - \bar{x}_2)^2}{n_2 - 1}. \tag{3.35}$$

Note, that in the case where $n_1 = n_2$, the estimate of σ^2 is given by

$$s^2 = \frac{s_1^2 + s_2^2}{2}. \tag{3.36}$$

The paired t-test

This test is applied when the experiments are carried out in pairs. Here, the difference between each pair is of interest.

Let us assume we have n pairs $x_1[i]$ and $x_2[i]$ for $i = 1, \ldots, n$, which are independent observations of populations with means μ_1 and μ_2. Then the null hypothesis is described as

$$H_0 = \mu_{1_i} = \mu_{2_i} \qquad \text{for all } i = 1, \ldots, n. \tag{3.37}$$

The difference of the pairs is given by

$$d_i = x_1[i] - x_2[i] \qquad i = 1, \ldots, n, \tag{3.38}$$

where d_i is a sample of size n with mean zero. The average difference is denoted by $\bar{d} = \sum_{i=1}^{n} d_i$ and the standard deviation is denoted by s. Then the test statistic is given by

$$y = \frac{\bar{d}}{\frac{s}{\sqrt{n}}}, \tag{3.39}$$

where the standard deviation is obtained according to (3.29).

3.2.4 F-test

This hypothesis test is mostly used to compare different variance estimates. Let s_1^2 and s_2^2 be two sample variances of a normal distribution with sizes n_1 and n_2. When s_1^2 and s_2^2 are both estimates of the same quantity σ^2 and we take repeated pairs of samples, it can be shown that the ratio $y = \frac{s_1^2}{s_2^2}$ follows an F-distribution. The estimates s_1^2 and s_2^2 must be independent. The distribution depends on two parameters ν_1 and ν_2, which are the degrees of freedom of s_1^2 and s_2^2, respectively. In the case described here, we have $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. In general, a random variable which follows an F-distribution can be obtained from two independent χ^2 random variables, see Section 3.1.3. With an F-test, often the hypothesis that s_1^2 and s_2^2 are both estimates of the same variance σ^2 is tested. Thus, when s_1^2 and s_2^2 are independent estimates of σ_1^2 and σ_2^2, the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ will be tested.

3.2.5 Hotelling's T^2-test

The Hotelling's T^2-test is the multivariate version of the univariate Student's t-test (Section 3.2.3). We have to deal here with multivariate random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_i, \ldots, \mathbf{X}_n$ of dimension p.

Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents the sample from a p-variate distribution with the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)^T \in \mathbb{R}^p$, then we want to test the hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \tag{3.40}$$

where $\boldsymbol{\mu}_0$ is called the hypothetical mean vector. Let $\bar{\mathbf{x}}$ be the sample mean (see equation 3.13) and $\mathbf{S}$ denote the sample covariance matrix (see equation 3.14), respectively. Analogous to the univariate test based on the univariate sample standard distance ($\bar{x} - \mu$ in equation (3.28)), we can base a test of the hypothesis (3.40) on the multivariate standard distance between $\bar{\mathbf{x}}$ and $\boldsymbol{\mu}_0$:

$$D(\bar{\mathbf{x}}, \boldsymbol{\mu}_0) = \left[(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \right]^{\frac{1}{2}}, \tag{3.41}$$

which is well defined, provided the sample covariance matrix $\mathbf{S}$ is positive definite. The corresponding test statistic is:

$$
\begin{aligned}
T^2 &= n D^2(\bar{\mathbf{x}}, \boldsymbol{\mu}_0) \\
&= n (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0). \tag{3.42}
\end{aligned}
$$

The random variable $n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$ follows the T^2-distribution under the assumption that the samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is from a p-variate normal distribution. Further, under the assumption that $\mathbf{S}$ is positive definite and $n > p$, the random variable

$$y = \frac{n}{n-1} \frac{n-p}{p} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \tag{3.43}$$

follows an F-distribution with p and $(n - p)$ degrees of freedom (p degrees of freedom in the numerator and $n - p$ degrees of freedom in the denominator) [7, 60].

3.2.6 Multiple testing

We have different possibilities to set the threshold t for the hypothesis test. Some of them are discussed below.

Bonferroni-correction

The above reasoning of setting the threshold does not apply anymore if we look at several test statistics jointly. In the case of 5 % of the test statistics, H_0 would be rejected by error, when no multiple correction is applied. Thus we have to correct for multiple testing. Since we don't want to have more than $\alpha\%$ of wrong rejections of H_0 globally, we have to divide α by the number of test statistics. This is known as Bonferroni-correction [21] for multiple testing:

$$\alpha' = \frac{\alpha}{N}, \tag{3.44}$$

where N = number of test statistics.
It is almost exact, when no correlation is present, otherwise still correct but perhaps too conservative. It is exact (and not only almost exact) with $(1 - \alpha')^N = 1 - \alpha$, i.e., $\alpha' = 1 - (1 - \alpha)^N \simeq \frac{\alpha}{N}$ (cf. Paragraph "Probability of the maximum"). Bonferroni's correction is the most conservative (=safe) approach to deal with multiple tests.

Binomial law

Another possibility for multiple correction is to determine the number of false rejections of H_0 (n_0) that is expected from our significance level α. The probability to observe n_0 false rejections in a random field with noise only follows a binomial distribution:

$$P(n = n_0) = \binom{N}{n_0} p^{n_0} (1 - p)^{N - n_0}, \tag{3.45}$$

with $p = P(y > t)$ and N = number of test statistics. Equation (3.45) holds when no correlation between the test statistics is present. Since we know how many wrong rejections we should expect, we are able to compare them with the effectively observed ones. If there are significantly more than expected, then we have a significant result within the samples of test statistics. Yet, we have not localized it, because this is a global test only.
In the case of Bonferroni-correction, $n_0 = 0$ and the mean value $E[n]$ of appearance of a false rejections is equal to $Np = \alpha'$ (usually $\alpha' = 0.05$). Consequently the approach is

very conservative.

Probability of the maximum

Another way of dealing with multiple test statistics is to look at the maximum of a vector. We choose the threshold by computing the probability that only the maximum of N random variables is greater or equal to this threshold and without any hypothesis one has:

$$
\begin{aligned}
P(\max(|x_n|) \geq t) &\leq P(|x_1| \geq t) + P(|x_2| \geq t) + \cdots + P(|x_n| \geq t) \\
&\leq N \cdot p
\end{aligned}
\tag{3.46}
$$

where

$$
p = P(|x_0| \geq t).
$$

It follows that $p \geq \frac{\alpha}{N}$; the lower bound is the same result as with Bonferroni. When the data are independent, the lower bound is approximately attained, since $P(\max(|x_n|) \leq t) = (1-p)^N \Rightarrow P(\max(|x_n|) \geq t) = 1 - (1-p)^N \simeq Np$. This means that the approach of the probability of the maximum and the Bonferroni-correction are essentially equivalent. For correlated data, one could in principle find a $p > \frac{\alpha}{N}$, which would yield a lower detection threshold t.

Probability of a Gaussian Random Field

This section describes a method for setting a threshold for "Gaussian correlated" data. This is the approach used in SPM where the data is smoothed with a Gaussian kernel to increase the signal-to-noise ratio (SNR). This correlates the voxels with their spatial neighbors (cf. Section 2.8). For this reason the Bonferroni correction, which treats voxels as if they were independent, is not adequate anymore—it would be too conservative. Thus a different approach must be taken to determine the threshold for the statistical parameter map. In SPM, the Random Field Theory (RFT) is used for this approach. This formulation is made in the continuous domain even though the data is discrete. Specifically, one uses the theory of Gaussian Random Fields (GRF) to find the maximum of a Gaussian-correlated continuous random field. The first step to proceed with RFT is to define how many *resolution elements* (RESELS) are present in the questioned image or volume. This number can be roughly seen as the number of independent observations within the volume and is defined as

$$
\text{RESELS} = \frac{V}{\prod_{i=1}^{D} \text{FWHM}_i},
\tag{3.47}
$$

where FWHM_i represents the smoothness in the volume with the measure of full width of half maximum in the i'th direction of D dimensions.

The next step is to calculate the expected Euler characteristic $E[\chi_t]$ of the statistical parameter map. $E[\chi_t]$ is used to approximate the probability that the maximum value on the map (z_{max}), when no activation is present, is greater than a given threshold t under the null hypothesis $\mathcal{H}_0$. χ_t is a geometrical measure that counts the number of isolated regions on a volume map after thresholding with t, provided the excursion set contains no holes and doesn't touch the boundary of the search volume (approximation for a rather high t) [2, 3, 156]. For such a t, it holds that

$$P(z_{max} \geq t) \approx P(\chi_t \geq 1) \approx 1 - e^{E[\chi_t]} \approx E[\chi_t]. \tag{3.48}$$

Moreover the clusters over such a high threshold t are independent and their number C_t follows approximately a Poisson distribution with mean $E[\chi_t]$:

$$P(C_t = x) = \frac{1}{x!}(E[\chi_t])^x e^{-E[\chi_t]} = \Upsilon(x, E[\chi_t]). \tag{3.49}$$
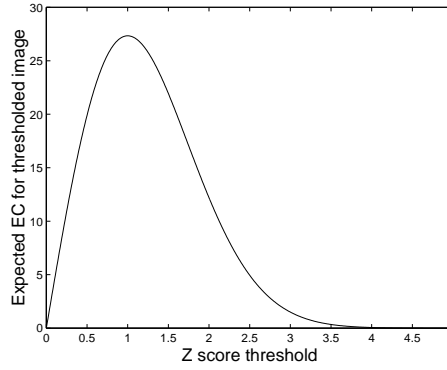
Figure 3.6: Expected Euler characteristic for smoothed image with 256 RESELS, corresponding to equation (3.50).

The expected Euler characteristic is given by:

$$E[\chi_t] = \lambda(V)\sqrt{|\Lambda|}(2\pi)^{-\frac{D+1}{2}} H_{e_D}(t)e^{-\frac{t^2}{2}}, \tag{3.50}$$

where $\lambda(V)$ represents the volume and $H_{e_D}(t)$ is the Hermite polynomial of degree D in t. t represents here the value to test, the local maximum of a potential activation. $\sqrt{|\Lambda|}$ is a smoothness factor which allows to obtain an estimation of the smoothness in the volume, where Λ represents the variance covariance matrix tested in D dimensions. It is shown that ([156])

$$\lambda(V)\sqrt{|\Lambda|} = \text{RESELS}(4\log_e 2)^{\frac{D}{2}}. \tag{3.51}$$

For a volume, (3.50) becomes:

$$\mathrm{E}[\chi_\mathrm{t}] = \lambda(V)\sqrt{|\Lambda|}(2\pi)^{-2}(4t^2 - 2)e^{-\frac{t^2}{2}}. \tag{3.52}$$

Several assumptions have to be made for equations (3.48) and (3.50) to hold: the threshold t must be sufficiently high, the volume V must be large compared to resolution elements in the map, and finally the discrete statistical parameter map must approximate a continuous zero-mean univariate homogeneous smoothed Gaussian Random Field (GRF). In other words, the autocorrelation function of the field must be a Gaussian. Another assumption is the spatial stationarity which, in our view, is a wrong assumption. These assumptions are discussed more exhaustively in [111].

In addition to testing for the intensity of an activation, SPM also tests for the significance of its spatial extent, and for the significance of a set of regions. For the corresponding theory, we refer to [111].

SPM99 takes care of two aspects which can be formulated more precisely and are described in [157]. The first improvement concerns the Euler characteristics, which depends not only on the number of RESELS anymore, but also on the shape of the volume. This precision is especially effective when the volume in which the RESELS are contained is small. In contrast, when the volume is large, the two approaches give similar results. The second improvement concerns the degrees of freedom. In the case of few degrees of freedom, the SPM scores, which are based on z-statistics, are not adequate anymore using the random field theory. Thus the new approach generates thresholds for the t-statistics, from which the z-scores are derived [69, 155, 63].

3.3 fMRI applications

In this section, we briefly discuss fMRI noise and then describe the statistical procedure of our fMRI analysis algorithm.

3.3.1 Noise in fMRI data

Since fMRI-data are extremely noisy, we devote a section to the analysis and modelling of this particular noise.

Origin

The noise in fMRI data originates from several sources. There is the noise of the measurement instrument (e.g., the scanner), in particular, due to hardware imperfections [94], but there is also noise arising from the patient's body, which is the main source of disturbance in an image [51]. Some of the main sources are:

Thermal noise. The most dominant source of thermal noise is the one arising from the patient in the scanner. Thermal vibrations from ions and electrons induce noise in the receiver coil. Since the noise level depends on how many ions and electrons are contributing to the signal, it is important to adjust the size of the receiver coil to the object which is imaged. For this reason, MRI has different coils for different regions of interest, like the head or the spine. The disadvantage is that some small coils have an image intensity which is non-uniform over the region of interest. Other sources of thermal noise are quantization noise in the A/D devices, preamplifier/electronic noise, and finally thermal noise in the RF coils [106, 24].

Motion of the subject. Even when fixing the head of the patient during acquisition, head motion cannot be completely avoided. Such head motions can arise, for example, from respiration. Task-correlated head movements are especially disturbing, since these movement effects can look like task activation. Motion of the order of $\frac{1}{10}$ voxel may produce $1-2\%$ signal changes, which is not negligible. In comparison, fMRI BOLD effects are very small: $1-5\%$ signal changes with a magnetic field of 1.5 T [37, 25]. This calls for the use of accurate image registration algorithms in a preprocessing stage (cf. Section 2.6).

Noise due to physiological activities. The cardiac cycle causes vessel pulsations, cerebrospinal fluid movements, and tissue deformations, which all produce fMRI signal variations [39]. The approaches to mitigate cardiac effects are to record the cardiac cycle, to estimate its influence on the fMRI signal a posteriori, and to finally remove it.
Respiratory variations also induce signal changes in the head. Respiration causes blood pressure changes that may result in slight modifications of the size of the CSF and venous blood spaces in the brain. A special possible respiration effect for some subjects is the unconsciously gating of their breathing to the task, especially when this task is very regular. A subject may hold off breathing until his response to a trial [106, 109, 18].

Low frequency drift. Low frequency drifts (0.0-0.015 Hz) and trends are often reported in fMRI data. It is one of the most poorly understood source of signal variations. Investigations by Smith et al. [130] suggest that scanner instabilities and not motion or physiological noise may be the major cause of the drift.

Spontaneous neural and vascular fluctuations and behavior variations. There are a variety of neuro- and vascular physiological processes that lead to fluctuations in the BOLD signal intensity. In addition to this background processes, there are also variations in the way in which subjects perform tasks from trial to trial [106].

Artifacts in Rapid Imaging Methods. The main disadvantages and artifacts of rapid imaging methods are [106]:

> ▶ Limited spatial resolution.

> ▶ Ghosting (affects mostly EPI) [30, 123, 32].

> ▶ Susceptibility distortions: differences in magnetic susceptibilities between air and watery tissue in the head result in non-uniformities in the magnetic fields, which causes artifacts. In the EPI technique, these artifacts result mainly in geometric distortions [84]. The distortions can be a shift, shear, stretch, or squish of the object; in most cases these effects are local. Fat is an example that can cause artifacts like shifting and therefore rapid imaging techniques use "fat suppression". All the fMRI-data used in this work were acquired under fat suppression.

Characterization

To get a better understanding of the statistical nature of our data, we try to determine the main properties of the noise. This step is crucial since we are aiming at an accurate and rigorous statistical analysis of the results of fMRI experiments.

We make the general hypothesis that the noise follows a Gaussian distribution. This assumption is certainly reasonable when working with averaged finite energy data (due to the central limit theorem, cf. Section 3.1.3), as is the case here. The key question to be answered concerns the covariance structure of the noise (independence, stationarity over space and time).

Time variance over space. We question whether the variance is homogeneously distributed over space, or whether the variance differs significantly within the volume. The hypothesis for the following calculation is that the time sequence of each voxel is independent and normally distributed with variance σ; then, defining $s_k^2 = \sum_{\tau=1}^{n} \frac{(x_k(\tau) - \overline{x}_k)^2}{n-1}$ with $\overline{x}_k = \frac{1}{n} \sum_{\tau=1}^{n} x_k(\tau)$, we know that $\frac{s_k^2}{\sigma^2}$ follows a χ^2 distribution with $n-1$ degrees of freedom (see Section 3.1.3), where k is a 3D integer index locating the voxel in the volume.

Therefore, we may write a confidence interval for this estimate as

$$P(a \leq \underbrace{(n-1)\frac{s_k^2}{\sigma^2}}_{\xi} \leq b) = \alpha, \tag{3.53}$$

where P is the χ^2 cumulative probability, see Section 3.1.3; this means, that $(n-1)\frac{s_k^2}{\sigma^2}$ will fall within the interval $[a, b]$ with probability α.

"a" and "b" are to be computed in such a way as to obtain the smallest interval, depending on the confidence interval probability α. Let us denote by $a(\alpha)$ and $b(\alpha)$ those optimal

values. Then we get an interval for σ^2:

$$(n-1)\frac{s_k^2}{b(\alpha)} \le \sigma^2 \le (n-1)\frac{s_k^2}{a(\alpha)}. \tag{3.54}$$

To get the smallest interval for σ^2, we minimize $\left(\frac{1}{a} - \frac{1}{b}\right)$ under the condition of $P(a \le \xi \le b) = \alpha$. Using Lagrange multipliers, we define the auxiliary function

$$f(a, b, \lambda) = \frac{1}{a} - \frac{1}{b} - \lambda P(a \le \xi \le b). \tag{3.55}$$

Taking the partial derivatives, we get

$$\frac{\partial f(a, b, \lambda)}{\partial a} = -\frac{1}{a^2} + \lambda p(a) = 0 \tag{3.56}$$

$$\frac{\partial f(a, b, \lambda)}{\partial b} = \frac{1}{b^2} - \lambda p(b) \quad = 0, \tag{3.57}$$

since $P(a \le \xi \le b) = \int_a^b p(\xi)d\xi$. From equation (3.56) and (3.57), we obtain the optimality condition:

$$a^2 p(a) = b^2 p(b). \tag{3.58}$$

Below is the code that determines the smallest interval by an iterative process which converges to a solution where (3.53) and (3.58) are satisfied.

```
N := degree of freedom;
a(1) := 0;
for i := 1 to 100 step 1 do
    b(i) = chi2inv(α + chi2cdf(a(i), N), N);
    a(i + 1) = a(i) + 0.1 * (b(i)² * chi2pdf(b(i), N) − a(i)² * chi2pdf(a(i), N));
end,
```

We observe that, on a logarithmic scale, the intervals have a constant size

$$\log(\sigma^2) - \log(s_k^2) \in \underbrace{[\min, \max]}_{\Delta}, \tag{3.59}$$

where

$$\min = \log\left(\frac{n-1}{b(\alpha)}\right),$$

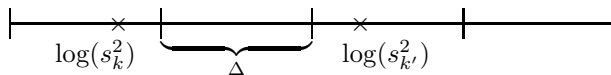$$\max = \log\left(\frac{n-1}{a(\alpha)}\right),$$

Figure 3.7: Equidistant confidence intervals on logarithmic scale. s_k^2 and $s_{k'}^2$ are significantly different.

which suggests a natural quantization step for the representation of the variance data s_k^2. Since we look at the whole image, the value α is corrected for multiple testing with the Bonferroni method.

We thus obtain for each $\log(s_k^2)$ an uncertainty interval Δ, wherein the true $\log(\sigma^2)$ lies with the probability α (chosen to be close to 1). If two intervals are not adjacent, then we can assert that the corresponding estimates for σ^2 are significantly different.

Applying this to one slice of different data-set, we get the result as depicted in Figure 3.8. From experiments such as these, we conclude that the variance is generally not constant over space, contrary to what has often been stated in the literature [120, 52]. The variances, however, appear to be strongly correlated over space—in other words, their variation is usually small within the neighborhood of a pixel.

Stationarity and time-independence. We take the time sequence of the rest state (B) for each voxel and compute the variances of the first and second half of the time sequence. Then, applying an F-test (see Section 3.2.4), we can check if the variances differ significantly, in which case at least one of our hypotheses (stationarity or time-independence) will fail:

$$\rho_k = \frac{\sum_{\tau=0}^{\frac{n}{2}-1} (x_k(\tau) - \xi_0)^2}{\sum_{\tau=0}^{\frac{n}{2}-1} (x_k(\tau + \frac{n}{2}) - \xi_1)^2} \tag{3.60}$$

where

$$\xi_0 = \frac{1}{\frac{n}{2}} \sum_{\tau=0}^{\frac{n}{2}-1} x_k(\tau)$$

$$\xi_1 = \frac{1}{\frac{n}{2}} \sum_{\tau=0}^{\frac{n}{2}-1} x_k(\tau + \frac{n}{2}).$$

We note that most statistical methods developed so far in the fMRI literature make the assumption of stationarity and time-independence.

We applied this test to several of our data sets. Most of the slices show no detection. However, in some cases the number of detections exceeds the false detection rate of the
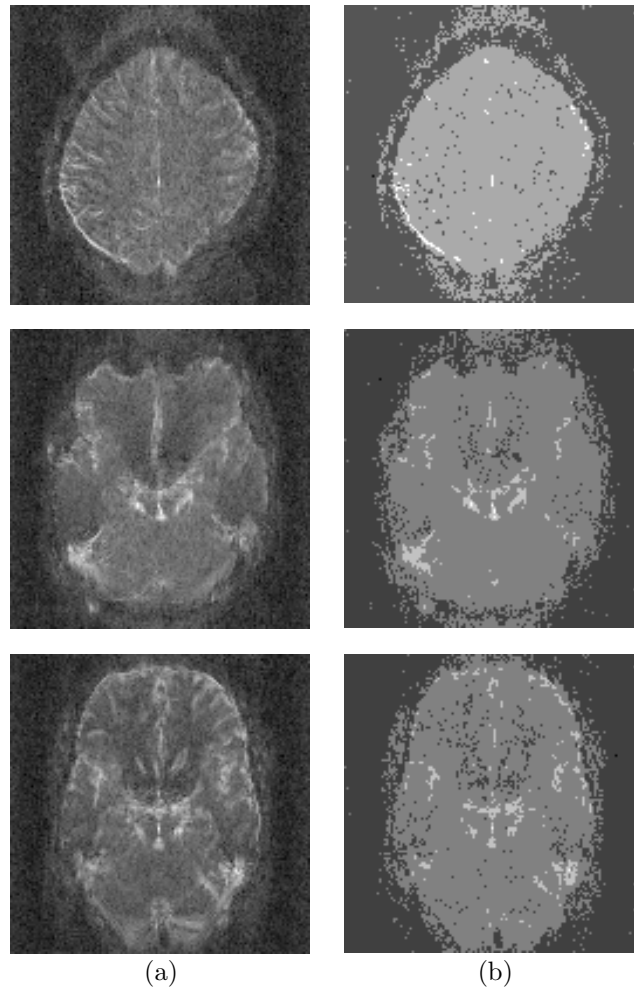
Figure 3.8: Variance estimation for three different slices: a) plot of $\log(s_k)$. b) quantization of (a) so that the classification error is as low as 0.0003% (Bonferroni correction). Obviously the variance is significantly nonuniform.
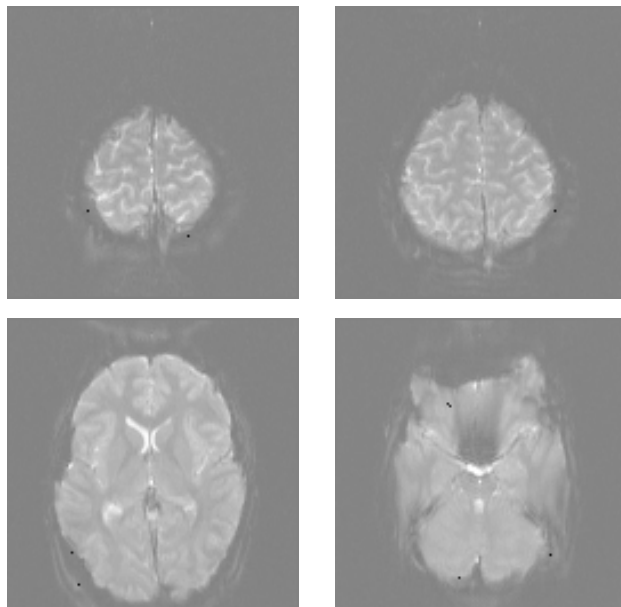
Figure 3.9: Slices of one data set which show some detections. The threshold of the test is Bonferroni-corrected. We believe the false detections arose mostly from motion artifacts.

Bonferroni-correction. Those detected pixels are mostly on the border of the brain or around the eyes. Thus we have some indications that the cases of false detections are due to motion artifacts which could not be prevented by the realignment procedure (cf. Figure 3.9).

3.3.2 Deterministic model

In order to make a statistical inference, we first need to specify a model. Three possible data models at pixel k are being considered, from the simplest to the more complex.

Gaussian white noise assumption

<u>Model 1:</u> In some studies, the noise has been assumed to be white Gaussian [120, 52]. The corresponding model is

$$x_k(\tau) = \mu_k(\tau) + e_k(\tau) \cdot \sigma,$$

where μ_k is the average hemodynamic function, e_k is an i.i.d. normalized noise ($E[e_k] = 0$, $E[e_k^2] = 1$).

The hypothesis is that $\mu_k = \mu_1$ during A and $\mu_k = \mu_2$ during B. Thus, the detection of activation uses a statistical test applied on a pixel by pixel basis to determine whether or not μ_1 and μ_2 are different.

The white noise hypothesis has the advantage of simplicity. However, we found it to be inappropriate for our data (cf. Section 3.3.1).

Time-independent, but space-non-stationary white noise

Model 2: In the second model, we assume to have whiteness (over time), but space-varying (non stationary) Gaussian noise:

$$x_k(\tau) = \mu_k(\tau) + e_k(\tau) \cdot \sigma_k,$$

where the variance σ_k varies over the pixels, but changes only slightly, see Section 3.3.1. We use this model which is consistent with our finding in Section 3.3.1.

General linear model

Model 3: In the third model we include components that also vary over time (hemodynamic function or periodic disturbances). They are represented as a weighted sum of c different functions:

$$x_k(\tau) = \sum_{i=1}^{c} \alpha_{k,i} g_i(\tau) + \sigma_k e_k(\tau),$$

where $g_k(\tau)$ is the explanatory variable and represents an activation function. This approach offers no extension on the noise description. But its deterministic model offers the possibility of refinements, e.g., including baseline drifts. The General Linear Model (GLM) is used among others by SPM, see also Section 2.8.2.

3.3.3 Hypothesis testing for activation

In the case of our algorithm, the sample of the data set to test is the time sequence of a voxel and we want to decide whether this voxel is activated or not. This procedure is repeated for all voxels at locations $k = (i, j, l)$ in the volume. To simplify matters, we will describe the two-dimensional case, where we consider all the pixels $k = (i, j)$ in a slice. The derivation of the samples to test is as follows: spatial noise is reduced either by filtering or wavelet transformation in which case the testing is performed in the wavelet domain. To find the activation, we calculate the pairwise difference between activation and rest states yielding the sequence difference images $x_{D_\tau}(i, j) = x_{A_\tau}(i, j) - x_{B_\tau}(i, j)$, where the symbol A stands for the activation state and the symbol B for the rest state.
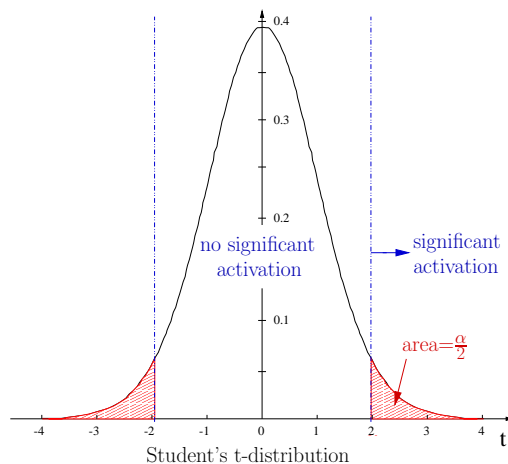
Figure 3.10: Definition of significance level for the two-tailed t-test.

To decide which pixel is activated, we denote the null hypothesis —the pixel under consideration is not activated—by H_0. Under this hypothesis, $\mu_0 = E[\bar{x}_D(i,j)] = 0$ and the distribution for the mean value of each pixel relative to its standard deviation is known. The corresponding test statistic $y(i,j)$ on the distribution has to be adapted to the set of data, which transforms the pixel values into p-values. Which test statistic y is implemented for fMRI data is discussed later in Section 3.3.5. If a test statistic exceeds some expected bound for $\bar{x}_D(i,j)$, the null hypothesis is rejected and the alternative hypothesis H_1 is accepted: we conclude that the pixel (or, eventually, wavelet coefficient) is activated:

1. Null hypothesis H_0: $E[\bar{x}_D(i,j)] = 0$ (no activation).

2. Alternative hypothesis H_1: $E[\bar{x}_D(i,j)] \neq 0$ translates to $|\bar{x}_D(i,j)| \geq t$, where t is our significance threshold as explained in Section 3.2.6.

Given the level of significance (see Section 3.2) $\alpha = 0.05$, we compute the threshold t such that the probability that $|y(i,j)|$ exceeds t is less than or equal to α, assuming that H_0 is true.

$$P(|y(i,j)| \geq t) = \alpha.$$

In the hypothesis test of our algorithm, the alternative hypothesis H_1 is defined as $E[\bar{x}_D(i,j)] \neq 0$, as we saw above. Thus we perform a two-tailed test (see Section 3.2.1). A two-tailed test is chosen, since in the wavelet domain, the high-pass includes possible significant high and low coefficients, which are of interest. The low-pass is tested with

a two-tailed hypothesis test as well. Here, when a significantly low-pass coefficient is detected, one can think of the negative activation, also called deactivation, such as a relative decrease of oxyhemoglobin; A deactivation can happen for two reasons: it may reflect decrease of synaptic activity during the task of interest, or it may reflect an active increase during the rest task [86].

3.3.4 z-test

If the variance σ^2 is constant over the whole set of samples in the image and is known, then $E[\bar{x}_D(i,j)]$ follows a normal-distribution, and a z-test (Section 3.2.2) may be applied as in [120]:

$$z(i,j) = \frac{\bar{x}_D(i,j) - \mu_0}{\frac{\sigma}{\sqrt{n}}}. \tag{3.61}$$

Here a two-tailed test is appropriate too.

3.3.5   **Student's t-test**

If the variance σ^2 changes with (i,j), as was found to be the case with fMRI data in Section 3.3.1, we apply a t-test (see Section 3.2.3) with $n-1$ degrees of freedom. The corresponding test statistic is

$$t(i,j) = \frac{\bar{x}_D(i,j) - \mu_0}{\frac{s_{D(i,j)}}{\sqrt{n}}}, \tag{3.62}$$

where $\bar{x}_D(i,j) = \frac{1}{n} \sum_{\tau=1}^{n} x_{D_\tau}(i,j)$ and where

$$s_D^2(i,j) = \frac{1}{n-1} \sum_\tau |x_{D_\tau}(i,j) - \bar{x}_D(i,j)|^2. \tag{3.63}$$

In our case, $\mu_0 = E[\bar{x}_D(i,j)]$. The t-test described above is a one sample t-test, since we compute the difference of the two samples before testing. In the following section, a method is described, where the number of the samples of the activation state and the rest state are not the same. Also with SPM (see Section 2.8) one can choose between one sample t-test, two sample t-test and paired t-test under the category of "basic models".

Two sample t-test

This case is applied when the number of samples in the activation state and in the rest state are not the same; in other words, when we cannot make the difference between these two states before testing. The two samples size will be denoted by n_A and n_B,

respectively, with the unknown mean value μ_A and μ_B, respectively. The corresponding samples mean are denoted by $\bar{x}_A(i,j)$ and $\bar{x}_B(i,j)$, respectively. The null hypothesis is

$$H_0 : \mu_A = \mu_B. \tag{3.64}$$

Since we choose a two-tailed test, the alternative hypothesis is expressed by

$$H_1 : \mu_A \neq \mu_B. \tag{3.65}$$

The test statistic is, obtained from equation (3.28):

$$t(i,j) = \frac{\bar{x}_A(i,j) - \bar{x}_B(i,j)}{s_{AB}(i,j)\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}. \tag{3.66}$$

The unbiased estimate s_{AB}^2 of σ^2 is given by

$$s_{AB}^2(i,j) = \frac{(n_A - 1)s_A^2(i,j) + (n_B - 1)s_B^2(i,j)}{n_A + n_B - 2}, \tag{3.67}$$

where the standard deviation estimates s_A and s_B are obtained by:

$$s_A^2(i,j) = \sum_{\tau=1}^{n_A} \frac{(x_{A_\tau}(i,j) - \bar{x}_A(i,j))^2}{n_A - 1} \tag{3.68}$$

$$s_B^2(i,j) = \sum_{\tau=1}^{n_B} \frac{(x_{B_\tau}(i,j) - \bar{x}_B(i,j))^2}{n_B - 1}. \tag{3.69}$$

The paired t-test

This test can be applied when there is a natural pairing in the trials (see Section 3.2.3). The n pairs are denoted as $x_{A_\tau}(i,j)$ and $x_{B_\tau}(i,j)$ for $\tau = 1, \ldots, n$ with means $\mu_{A_\tau}(i,j)$ and $\mu_{B_\tau}(i,j)$. Then the null hypothesis is described by

$$H_0 = \mu_{A_\tau}(i,j) = \mu_{B_\tau}(i,j) \qquad \text{for all } \tau = 1, \ldots, n. \tag{3.70}$$

The difference of the pairs is given by

$$x_{D_\tau}(i,j) = x_{A_\tau}(i,j) - x_{B_\tau}(i,j) \qquad \tau = 1, \ldots, n, \tag{3.71}$$

where $x_{D_\tau}(i,j)$ is a sample of size n with mean zero. The average difference is denoted by $\bar{x}_D(i,j)$ and the standard deviation is denoted by $s_D(i,j)$. Then the test statistic is given by

$$t(i,j) = \frac{\bar{x}_D(i,j)}{\frac{s_D(i,j)}{\sqrt{n}}}, \tag{3.72}$$

where the standard deviation is obtained accordingly to equation (3.29).

3.3.6 F-test

Instead of applying a two-tailed t-test to our data, we could also compare the different estimates of $\bar{x}_A^2$ and $\bar{x}_B^2$ and apply an F-test (see Section 3.2.4).

3.3.7   Hotelling's T^2-test

As reported in Section 3.2.5, the Hotelling's T^2-test is the multivariate version of the univariate Student's t-test. In our case, the complex wavelets (see Section 4.4) make the statistics bivariate. Like in the univariate case, to find the activation, we calculate the difference between activation and rest states pairwise to produce the sequence difference images $x_{D_\tau}(i,j) = x_{A_\tau}(i,j) - x_{B_\tau}(i,j)$, where $x_{A_\tau}(i,j)$ and $x_{B_\tau}(i,j) \in \mathbb{C}$. Thus, we have to deal with bivariate random vectors ($p = 2$) $\mathbf{X}[1], \ldots, \mathbf{X}[\tau], \ldots, \mathbf{X}[n]$, where one dimension of the random vectors represents the real part of the wavelet coefficients, while the second dimension represents the imaginary part.

The variables $\mathbf{x}[1], \ldots, \mathbf{x}[\tau], \ldots, \mathbf{x}[n]$ represent the samples from one wavelet location with the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$. Then we want to test the hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}, \tag{3.73}$$

where the hypothetical mean vector $\boldsymbol{\mu_0}$ equals in our case to $(0,0)^T$. The sample mean vector and the sample covariance matrix are denoted by $\bar{\mathbf{x}}_D(i,j)$ and $\mathbf{S}_D(i,j)$, respectively. The corresponding test statistic with $\bar{x}_D(i,j) = \frac{1}{n} \sum_{\tau=1}^{n} x_D(i,j)[\tau]$ is:

$$T^2(i,j) = n\bar{\mathbf{x}}_D^T(i,j)\mathbf{S}_D^{-1}(i,j)\bar{\mathbf{x}}_D(i,j), \tag{3.74}$$

where

$$\mathbf{S}_D = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}, \tag{3.75}$$

with

$$s_{11} = \frac{1}{n-1} \sum_{\tau=1}^{n} \left(x_1(i,j)[\tau] - \bar{x}_1(i,j)\right)^2$$

$$s_{22} = \frac{1}{n-1} \sum_{\tau=1}^{n} \left(x_2(i,j)[\tau] - \bar{x}_2(i,j)\right)^2$$

$$s_{12} = \frac{1}{n-1} \sum_{\tau=1}^{n} \left(x_1(i,j)[\tau] - \bar{x}_1(i,j)\right) \left(x_2(i,j)[\tau] - \bar{x}_2(i,j)\right)$$

$$s_{21} = s_{12}.$$

The hypothesis H_0 is accepted, when

$$\frac{n-2}{(n-1)2}T^2(i,j) \leq f, \qquad (3.76)$$

where f is the $(1-\alpha)$ quantile of the F-distribution with 2 and $(n-2)$ degrees of freedom. Otherwise H_0 is rejected.

3.3.8 Determining the significance level

In our algorithm, we correct for multiple comparisons with Bonferroni (cf. Section 3.2.6):

$$\alpha' = \frac{\alpha}{N},$$

where N = number of wavelet coefficients in the image.

When the noise is white and the wavelet transform is orthogonal, the data remains independent (cf. Section 6.4) and the Bonferroni-correction is in this case exact (i.e. non conservative). When the data is locally correlated, there is no reason to restrict ourselves to orthogonal transformation only. In fact, the transformation may even improve the situation by acting as a whitening operator. The testing remains safe as long as we apply the Bonferroni correction.

When applying the hypothesis test and the correction for multiple comparisons with Bonferroni to the data directly, we usually hardly detect any activation. This is simply because the level of noise in the data set is too high for this kind of straightforward detection. A typical example of such behavior in 2D is shown in Figure 3.11.

To set the threshold, we could also apply the binomial law (see Section 3.2.6). In that case the probability to observe n_0 false detections in a random field without activation follows the binomial equation (3.45) with parameters $p = P(t_k > t)$ and N = number of pixels in the image. This is a global test, but we may also want to know the location of the detection. Since it is most probable that the activation appears in clusters while the false positives appear scattered, one could check for clusters of a given size after the global testing.

The SPM method uses the correction of probability of a Gaussian Random Field (see Section 3.2.6) to test for activation. Its continuous formulation of the problem is not optimal, since in the fMRI problem the data are discrete. The Gaussian random field theory is not adequate when the correlation length is of order of the pixel size. We have evidence that, the threshold value is too conservative (more than Bonferroni), when there is no correlation at all, and that it is too optimistic for moderate correlation, which corresponds to the case of our data.
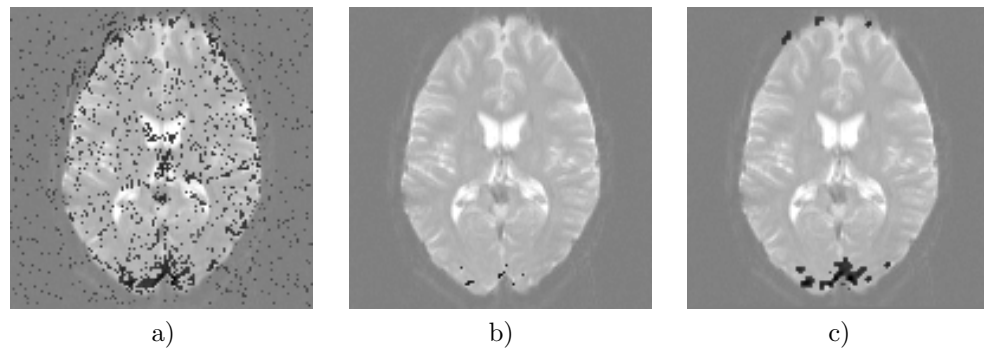
a)   b)   c)

Figure 3.11: Detecting activation of a visual task: a) applying a two-paired t-test to the pixel data without multiple correction to realigned fMRI data. b) Like a) but with Bonferroni-correction. Only few pixels are detected. c) Applying a wavelet transform and using a t-test with Bonferroni-correction. We detect the regions that are clustered in a). The wavelet transform reduces the noise sufficiently so that the Bonferroni-correction does not seem to be penalizing.

Chapter 4

Wavelets

4.1 Introduction

The Fourier basis consists of sines and cosines, which are perfectly localized in the frequency domain, but not in space. In contrast, wavelets decay to zero as $t \to \pm\infty$ while retaining some level of localization in frequency domain. It is thus likely that wavelets are better suited to represent functions that are localized both in space and frequency.

4.1.1 Definition of a multiresolution analysis

We now consider the space of Lebesgue square-integrable functions:

$$L_2(\mathbb{R}) = \left\{ f(x) : x \in \mathbb{R}, \int_{-\infty}^{\infty} |f(x)|^2 dx < +\infty \right\}. \tag{4.1}$$

Note that the L_2 space is considerably larger than the Shannon's conventional space of band-limited functions B_π. We define approximation spaces V_j, which have the following structure:

* $V_j \in L_2$

* $f \in V_j \Leftrightarrow f\left(2^j t\right) \in V_0$.

The subspace of functions that contain signal information down to scale 2^j is denoted by V_j. The first requirement of the MultiResolution Analysis (MRA) is that the information in V_j must also be contained in finer resolution space, V_{j+1}; thus $V_j \subset V_{j+1}$. It follows that V_{j+1} can be decomposed as

$$V_{j+1} = V_j \oplus W_j, \tag{4.2}$$

where W_j is called the detail space (there are several choices for W_j, one of them being the $\perp$ complement of V_j). Iterating equation (4.2) gives

$$V_{j+1} = \oplus_{k=-\infty}^{j} W_k. \tag{4.3}$$

The second requirement for an MRA is that all functions in $L^2(\mathbb{R})$ are included at the finest resolution and only the zero function at the coarsest level. Hence, due to (4.1), we have $\lim_{j \to -\infty} V_j = \{0\}$ and $\lim_{j \to \infty} V_j \to L^2(\mathbb{R})$. Hence, letting $j \to +\infty$ gives

$$L^2(\mathbb{R}) = \oplus_{k=-\infty}^{\infty} W_k. \tag{4.4}$$

The last requirement for an MRA is that there exist a function ϕ such that $\{\phi(t-k)\}_{k \in \mathbb{Z}}$ forms a Riesz basis for V_0.

To summarize, a multiresolution analysis of $L^2(\mathbb{R})$ is a nested sequence of subspaces $\{V_j\}_{j \in \mathbb{Z}}$ such that

i) $\quad \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset L^2(\mathbb{R})$ (4.5)

ii) $\quad \bigcap_j V_j = \{0\}, \qquad \bigcup_j V_j = L^2(\mathbb{R})$ (4.6)

iii) $\quad f(t) \in V_j \qquad \Longleftrightarrow \qquad f(2t) \in V_{j+1}$ (4.7)

iv) $\quad$ There exists a function $\phi(t)$, called the scaling function,

$\qquad$ such that $\{\phi(t-k)\}$ is a Riesz basis of V_0. (4.8)

$\qquad$ Consequently: $f(t) \in V_0 \Longrightarrow f(t-k) \in V_0$

$\phi(t)$ is defined to be a valid scaling function of L_2 if and only if:

i) $\quad A \le \sum_{k \in \mathbb{Z}} \left| \hat{\phi}(\omega + 2\pi k) \right|^2 \le B \quad$ (Riesz basis) (4.9)

ii) $\quad \phi(t) = \sqrt{2} \sum_k h[k] \phi(2t - k) \quad$ (two-scale relation) (4.10)

iii) $\quad \sum_{k \in \mathbb{Z}} \phi(t-k) = 1 \qquad$ (partition of unity) (4.11)

Under slight constraints, these three conditions are equivalent to the four conditions which define an MRA.

(4.11) implies that $\int_{-\infty}^{\infty} \phi(t) \mathrm{d}t = 1$. Thus with the two scale relation (4.10) we obtain

$$\sum_k h[k] = \sqrt{2}. \tag{4.12}$$

If ϕ is orthonormal, then

$$\sum_k |h[k]|^2 = 1 \qquad \text{and} \qquad \sum_k h[k] h^*[k+2l] = 0, \, l \neq 0. \tag{4.13}$$

The detail space W_0 has also a Riesz basis $\{\psi(t-k)\}_{k\in\mathbb{Z}}$, called the wavelet basis. The function $\psi(t)$ is referred to as the *mother wavelet*. Since $W_0 \subset V_1$, $\psi(t)$ can be written as a superposition of the basis functions for V_1:

$$\psi(t) = \sqrt{2} \sum_k g[k]\phi(2t-k), \qquad (4.14)$$

which is called the *wavelet equation*. From the requirement that $\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t = 0$, we obtain

$$\sum_k g[k] = 0. \qquad (4.15)$$

All these relations can be expressed in the Fourier domain. With $\hat{\phi}(\omega) = \int_{-\infty}^{\infty} \phi(t)e^{-j\omega t}\mathrm{d}t$ and $H(\omega) = \frac{1}{\sqrt{2}}\sum_k h[k]e^{-jk\omega}$ and correspondingly $\hat{\psi}(\omega) = \int_{-\infty}^{\infty} \psi(t)e^{-j\omega t}\mathrm{d}t$ and $G(\omega) = \frac{1}{\sqrt{2}}\sum_k g[k]e^{-jk\omega}$, we obtain

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \qquad (4.16)$$

$$\hat{\psi}(\omega) = G\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right). \qquad (4.17)$$

The orthogonality conditions can also be expressed in the frequency domain. It is equivalent to claim that $\{\phi(t-k)\}$ is orthonormal and to have:

$$\sum_{l=-\infty}^{\infty} |\hat{\phi}(\omega+2\pi l)|^2 = 1. \qquad (4.18)$$

To orthogonalize a basis function which is not orthogonal, we divide its Fourier transform by the square root of the left hand side of (4.18):

$$\hat{\phi}_{ortho}(\omega) = \frac{\hat{\phi}(\omega)}{\sqrt{\sum_{l=-\infty}^{\infty} |\hat{\phi}(\omega+2\pi l)|^2}}. \qquad (4.19)$$

It can be verified that for ϕ to be orthogonal, it is necessary and sufficient that

$$|H(\omega)|^2 + |H(\omega+\pi)|^2 = 1. \qquad (4.20)$$

This relation is derived simply by inserting (4.16) into (4.18). Similarly, $\{\psi(t-k)\}$ is orthonormal if

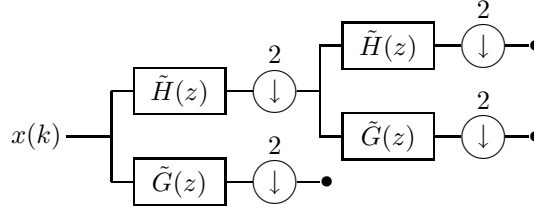$$\sum_{l=-\infty}^{\infty} |\hat{\psi}(\omega+2\pi l)|^2 = 1 \qquad (4.21)$$

Figure 4.1: Analysis part of the 1D discrete wavelet transform (DWT) for two iterations.

or

$$|G(\omega)|^2 + |G(\omega + \pi)|^2 = 1. \tag{4.22}$$

We can also derive the condition that $\{\psi(t - k)\}$ and $\{\phi(t - k)\}$ are orthogonal ($\int_{-\infty}^{\infty} \phi(t)\psi^*(t)dt = 0$),

$$\sum_{l=-\infty}^{\infty} \hat{\phi}(\omega + 2\pi l)\hat{\psi}^*(\omega + 2\pi l) = 0. \tag{4.23}$$

Using (4.16) and (4.17) and rewrite the above equation, we get the condition of orthogonality between $\{\psi(t - k)\}$ and $\{\phi(t - k)\}$ dependent on H and G:

$$H(\omega)G^*(\omega) + H(\omega + \pi)G^*(\omega + \pi) = 0. \tag{4.24}$$

4.1.2   1D filterbank

Wavelet decomposition can be also expressed in terms of subband filtering, a term that is used in signal processing. The connection between the multiresolution analysis described in the previous section and filterbank described here builds the filter $H(\omega)$ and $G(\omega)$. A filterbank is a set of down- and up-sampled filters. The Discrete Wavelet Transform (DWT) can be implemented using an iterated dyadic filterbank. Figure 4.1 represents the analysis part for two iterations and Figure 4.2 represents the synthesis part where $\widetilde{H}$ and $\widetilde{G}$ (resp., H and G) are the transfer functions of the analysis (resp., synthesis) filters. In the orthogonal case, the analysis and synthesis filters are identical up to a time-reversal; the wavelet filter G is simply a modulated version of the low-pass filter H:

$$G(z) = -z^{-1}H^*(-z^{-1}) \tag{4.25}$$

and in the time domain:

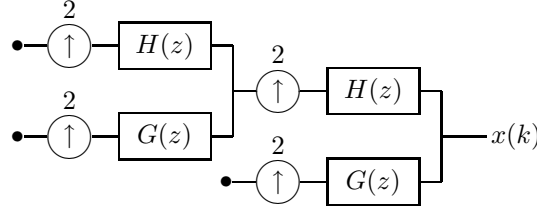$$g[k] = (-1)^k h^*[1 - k]. \tag{4.26}$$

Figure 4.2: Synthesis part of the 1D discrete wavelet transform (IDWT) for two iterations.

Let $x[k]$ denote a one dimensional discrete signal. Then its down-sampled version with 2 is

$$[x]_{\downarrow 2}[k] = x[2k]. \tag{4.27}$$

The Fourier-domain version of (4.27) is

$$[x]_{\downarrow 2}[k] \longleftrightarrow \frac{1}{2}\left[X\left(z^{\frac{1}{2}}\right) + X\left(-z^{\frac{1}{2}}\right)\right].$$

The up-sampling is defined by

$$[x]_{\uparrow 2}[k] = \begin{cases} x[l] & k = 2l, \ k \text{ even} \\ 0 & k \text{ odd} \end{cases}$$

and its effect in the transform domain is as follows:

$$[x]_{\uparrow 2}[k] \longleftrightarrow X\left(z^2\right).$$

If we now chain the down-sampling and up-sampling operators, we obtain

$$[x]_{\downarrow 2 \uparrow 2}[k] = \begin{cases} x[k] & k \text{ even} \\ 0 & k \text{ odd} \end{cases}$$

$$\updownarrow$$

$$\frac{1}{2}\left[X\left(z\right) + X\left(-z\right)\right]. \tag{4.28}$$

The low-pass filter $\widetilde{H}$ reduces the resolution by a factor of 2; the wavelet coefficients correspond to the output of the high-pass filter $\widetilde{G}$. Applying the relation (4.28) to the block diagram in Fig. 4.3, it is easy to derive the conditions for a perfect reconstruction:

$$\begin{cases} \widetilde{H}\left(z\right)H\left(z\right) + \widetilde{G}\left(z\right)G\left(z\right) = 2 & \text{no distortion,} \\ \\ \widetilde{H}\left(-z\right)H\left(z\right) + \widetilde{G}\left(-z\right)G\left(z\right) = 0 & \text{alias cancellation.} \end{cases} \tag{4.29}$$
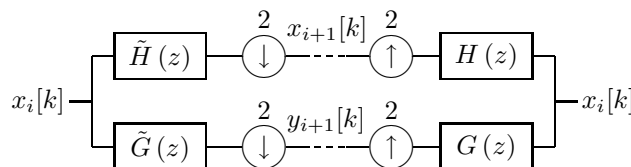
Figure 4.3: Perfect reconstruction filterbank.

4.2 Separable real wavelets

By separable wavelets, we mean here separable filters and separable sampling. In the two-dimensional case, the filters are obtained through the scaling function

$$\phi(x,y) = \phi(x)\phi(y). \tag{4.30}$$

In the two-dimensional case, we obtain this way three basic wavelets,

$$
\begin{aligned}
\psi_1(x,y) &= \phi(x)\psi(y) \\
\psi_2(x,y) &= \psi(x)\phi(y) \\
\psi_3(x,y) &= \psi(x)\psi(y).
\end{aligned}
$$

Sampling in multiple dimensions is represented by a lattice, the set of all vectors generated by $\mathbf{D}\mathbf{k}$, $\mathbf{k} \in \mathbb{Z}^n$, where $\mathbf{D}$ is the matrix characterizing the sampling process. $\mathbf{D}$ is not unique for a given sampling pattern [48]. A separable lattice is a lattice that can be represented by a diagonal matrix. In two dimensions, the dyadic sampling matrix has the expression

$$\mathbf{D} = \left(\begin{array}{cc} 2 & 0 \\ 0 & 2 \end{array} \right).$$

The separable 2D wavelet transform has four outputs: one low-pass and three high-pass branches.

The advantage of using a separable wavelet transform is twofold. First, it is simpler since the 1D approach can be extended to higher dimensions through the use of tensor product basis functions. The second reason is that separability is computationally more efficient, since the transform can be implemented by treating the rows and columns separately.

The tensor product solution has the disadvantage that the filters obtained in this way have a constrained design (nonseparable filters of size $N_1 \times N_2$ would offer $N_1 N_2$ free design variables versus $N_1 + N_2$ in the separable case). The second drawback is that only rectangular divisions of the spectrum are possible. It is usually the case that other divisions may better capture the signal energy [148].

4.2.1 Fractional splines

We describe here the discrete wavelet transform that arises from a multiresolution analysis generated by fractional splines. Fractional splines are a generalization of B-splines to fractional orders. B-splines are constructed from the $(n-1)$-fold convolution of a rectangular pulse β_+^0 [144]:

$$\beta_+^0(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.31}$$

$$\beta_+^n(x) = \underbrace{\beta_+^0 * \beta_+^0 * \cdots * \beta_+^0(x)}_{(n+1) \text{ times}}. \tag{4.32}$$

For n integers, these functions have a central axis of symmetry around $\frac{n+1}{2}$, and can be re-centered to yield the symmetric B-splines:

$$\beta_*^n(x) = \beta_+^n\left(x + \frac{n+1}{2}\right). \tag{4.33}$$

Likewise there are two flavors of fractional splines: the causal ones, denoted by β_+^α, which are a linear combination of the one-sided power functions

$$(x-n)_+^\alpha = \begin{cases} (x-n)^\alpha, & x \geq n \\ 0, & x < n \end{cases} ; \tag{4.34}$$

and the symmetric ones, denoted by β_*^α, which are a linear combination of the radial power functions $|x-n|^\alpha$ if α is any non-even real number, and of $|x-n|^\alpha \log|x-n|$ if α is even. For each version, the degree α of the splines is assumed to be strictly greater than $-\frac{1}{2}$, in order to ensure a valid MRA [135].

These functions have a simple expression in the frequency domain:

$$\hat{\beta}_+^\alpha(\omega) = \left(\frac{1 - e^{-j\omega}}{j\omega}\right)^{\alpha+1} \tag{4.35}$$

and $\hat{\beta}_*^\alpha(\omega) = |\hat{\beta}_+^\alpha(\omega)|$. They satisfy two-scale difference equations whose scaling filters are given, respectively, by

$$B_+^\alpha(z) = \sqrt{2}\left(\frac{1 + z^{-1}}{2}\right)^{\alpha+1} \tag{4.36}$$

for the causal spline, and by

$$B_*^\alpha(z) = \sqrt{2}\left|\frac{1 + z^{-1}}{2}\right|^{\alpha+1} \tag{4.37}$$

for the symmetric one.

The autocorrelation filter of a B-spline of degree α is given by

$$A^{\alpha}(e^{j\omega}) = \sum_n \left| \hat{\beta}_*^{\alpha}(\omega + 2n\pi) \right|^2. \tag{4.38}$$

Letting $B^{\alpha} = B_+^{\alpha}$ or B_*^{α}, the discrete Fractional Spline Wavelet Transform (FSWT) is implemented using the analysis and synthesis filters given in Table 4.1. Note that the

type filter	bspline	ortho	dual
$H(z)$	$B^{\alpha}(z)$	$B^{\alpha}(z)\sqrt{\dfrac{A^{\alpha}(z)}{A^{\alpha}(z^2)}}$	$B^{\alpha}(z^{-1})\dfrac{A^{\alpha}(z)}{A^{\alpha}(z^2)}$
$G(z)$	$-z^{-1}B^{\alpha}(-z^{-1})A^{\alpha}(-z)$	$-z^{-1}B^{\alpha}(-z^{-1})\sqrt{\dfrac{A^{\alpha}(-z)}{A^{\alpha}(z^2)}}$	$\dfrac{-zB^{\alpha}(-z)}{A^{\alpha}(z^2)}$
$\widetilde{H}(z)$	$B^{\alpha}(z^{-1})\dfrac{A^{\alpha}(z)}{A^{\alpha}(z^2)}$	$B^{\alpha}(z^{-1})\sqrt{\dfrac{A^{\alpha}(z)}{A^{\alpha}(z^2)}}$	$B^{\alpha}(z)$
$\widetilde{G}(z)$	$\dfrac{-zB^{\alpha}(-z)}{A^{\alpha}(z^2)}$	$-zB^{\alpha}(-z)\sqrt{\dfrac{A^{\alpha}(-z)}{A^{\alpha}(z^2)}}$	$-z^{-1}B^{\alpha}(-z^{-1})A^{\alpha}(-z)$

Table 4.1: Analysis and synthesis filters for the wavelet transform.

dual type is obtained by exchanging the analysis- and synthesis-part of the bspline case. In Figure 4.4, the low-pass filter $H(z)$ of the symmetric orthogonal fractional spline is presented for different order parameters α. With increasing α, the filter is less local in space but more frequency selective.

Implementation in Fourier

We have worked out a fast and exact implementation of the FSWT in 2D and 3D in the Fourier domain based on [20]. The MATLAB files are freely available on our website[1].

We have also an FFT-based implementation of the FSWT in Java 2.0 [70]. The principal advantage of Java over C++ is portability; indeed, there are now widely available virtual Java machines for most workstations including Mac, PC, and Sun. The final form of the algorithm is a plugin implemented in Java for ImageJ [114], a Java free-software available on the site of the National Institute of Health (NIH). The other final form is an applet, which makes it possible to run the algorithm over the Internet. A screen-shot of this applet is presented in Figure 4.5 and is available on the site http://bigwww.epfl.ch/demo/jwavelet/index.html. The performance of the algorithm is as

---

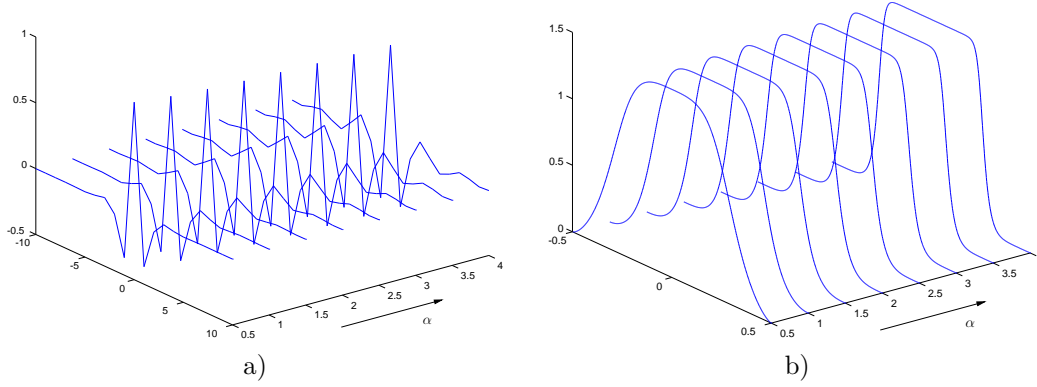[1] http://bigwww.epfl.ch/blu/fractsplinewavelets/

Figure 4.4: a) Impulse response of the orthogonal fractional spline dependent on α. b) Frequency response of the orthogonal fractional spline.

follows (Pentium 500 MHz): an image of size of 512×512, 32-bits takes for one iteration 1.10 s and for nine iterations 1.30 s. The implementation has been conceived to minimize the number and size of FFT's. In the present version, the calculation of the FFT and the IFFT takes up 50 % of the time.

The algorithm is implemented in a separable fashion and the intermediate results are stored in floating point format. The first step is to perform the FFT of the signal (row or column):

$$x_i[k] \overset{\text{FFT}}{\longleftrightarrow} X_i[n] \tag{4.39}$$

where $k \in [0, N-1]$ and $n \in [0, N-1]$. Down-sampling in Fourier domain is achieved by

$$X_i[n]_{\downarrow 2} = \tfrac{1}{2} \left(X_i[n] + X_i[n + \tfrac{N}{2}] \right), \tag{4.40}$$

where $n \in [0, \tfrac{N}{2} - 1]$.

Thus, the two down-sampled filtered components in the Fourier domain ($X_{i+1}[n]$ and $Y_{i+1}[n]$, cf. Figure 4.3) are computed as

$$X_{i+1}[n] = \tfrac{1}{2} \left(X_i[n]\widetilde{H}[n] + X_i[n + \tfrac{N}{2}]\widetilde{H}[n + \tfrac{N}{2}] \right) \tag{4.41}$$

$$Y_{i+1}[n] = \tfrac{1}{2} \left(X_i[n]\widetilde{G}[n] + X_i[n + \tfrac{N}{2}]\widetilde{G}[n + \tfrac{N}{2}] \right), \tag{4.42}$$

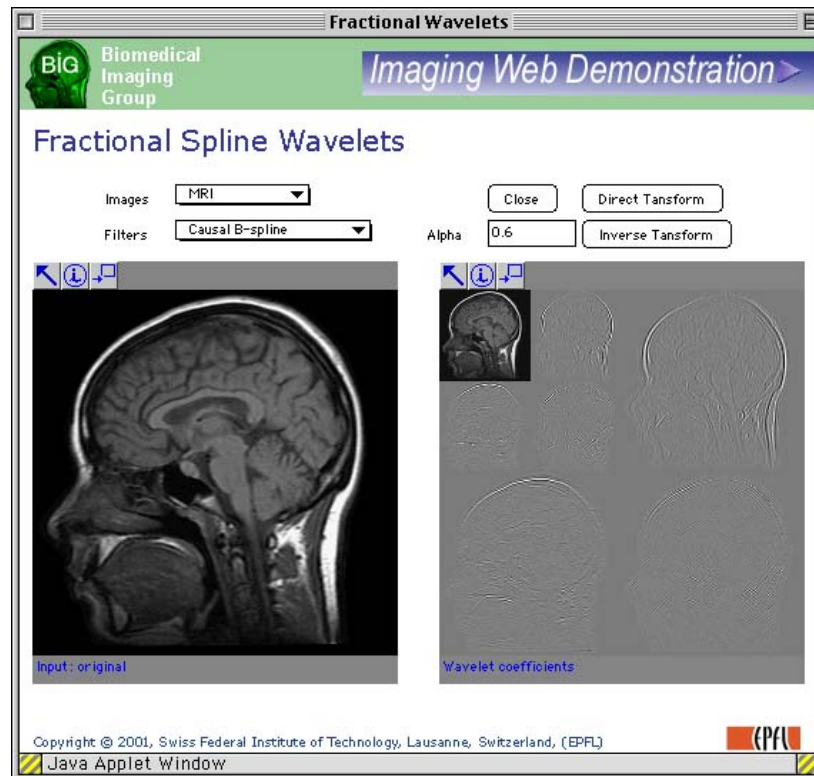where $n \in [0, \tfrac{N}{2} - 1]$. The corresponding $\tfrac{N}{2}$-point signals in Figure 4.3 are obtained by inverse FFT.

Figure 4.5: Applet of the Fourier based implementation of the separable wavelet transform and the fractional splines, available on the site: http://bigwww.epfl.ch/demo/jwavelet/index.html.
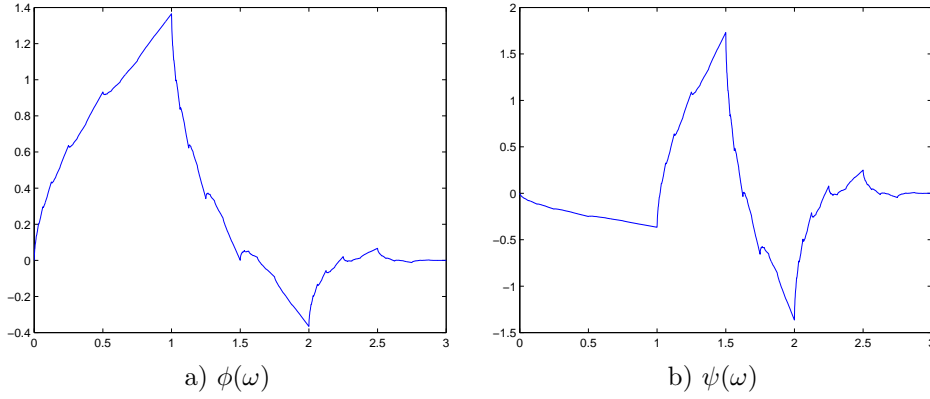
a) $\phi(\omega)$ b) $\psi(\omega)$

Figure 4.6: Approximation of the scaling function $\phi(\omega)$ and the wavelet function $\psi(\omega)$ of order 2, approximated with 10 iterations.

Conversely, the up-sampling is given by

$$X_{i+1}[m]_{\uparrow 2} = \frac{1}{2} \begin{cases} X_{i+1}[n] & m = 0, \ldots, \frac{N}{2} - 1 \\[2mm] X_{i+1}[n] & m = \frac{N}{2}, \ldots, N - 1. \end{cases} \tag{4.43}$$

The reconstruction formula in the Fourier domain is

$$X_i[m] = \tfrac{1}{2} \left(X_{i+1}[m]_{\uparrow 2} H[m] + Y_{i+1}[m]_{\uparrow 2} G[m] \right), \tag{4.44}$$

where $m \in [0, N - 1]$.

In this implementation, the analysis of RGB-images and of a stack of images is possible.

4.2.2 Daubechies wavelets

The Daubechies wavelets are widely used for many applications, also for fMRI. We describe here shortly the real Daubechies wavelets, since we are comparing their performance for the detection of activation patterns with other wavelets. The Daubechies wavelets are orthogonal and compactly supported. They are the shortest, orthogonal wavelets with N vanishing moments ($\int_{-\infty}^{\infty} t^l \psi(t) dt = 0$ for $l = 0, 1, \ldots, n - 1$). Their refinement filter is given $H(z) = (1 + z)^N Q(z)$, where $Q(z)$ is the shortest polynomial such that the orthogonality relation (4.13) is satisfied and $\deg Q \leq N - 1$. The exact formulation can be found in [43].

4.3   Nonseparable wavelets: quincunx

The quincunx transform will be described in more detail in Chapter 5. Quincunx wavelets
have some advantageous features that make them well suited for our application (Sec-
tion 7.2.2). The quincunx wavelet algorithm has been also implemented in Java [77]. A
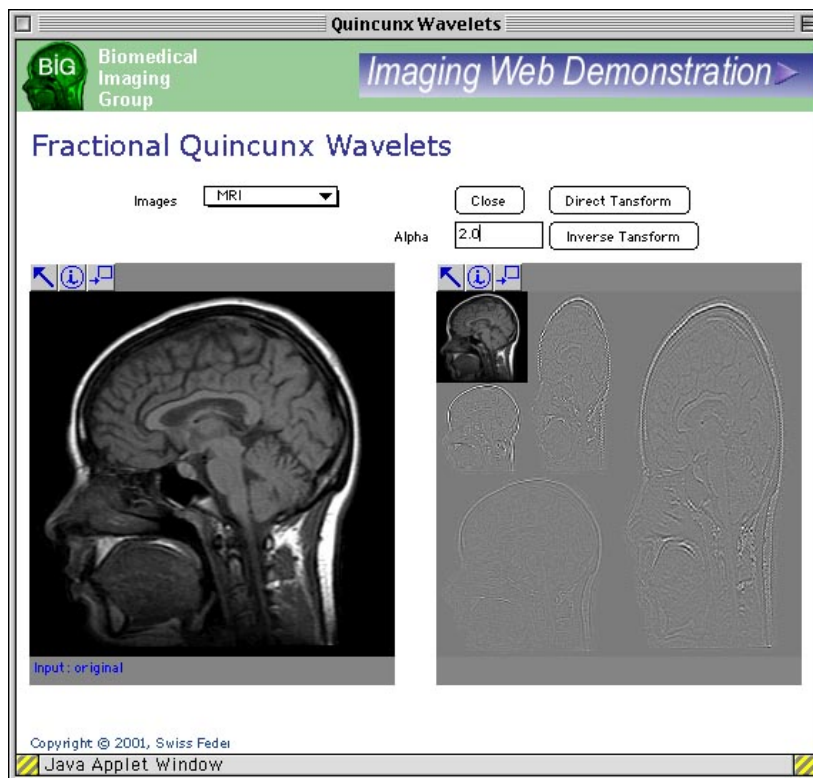screen-shot of the applet is presented in Figure 4.7.

Figure 4.7: Applet of the Fourier based implementation of the quincunx wavelet transform,
available on the site: http://bigwww.epfl.ch/demo/jquincunx/index.html.

4.4   Complex wavelets

Recent wavelet development has primarily focused on real-valued wavelet bases. However,
complex wavelet bases offer a number of potential advantageous properties [159]:
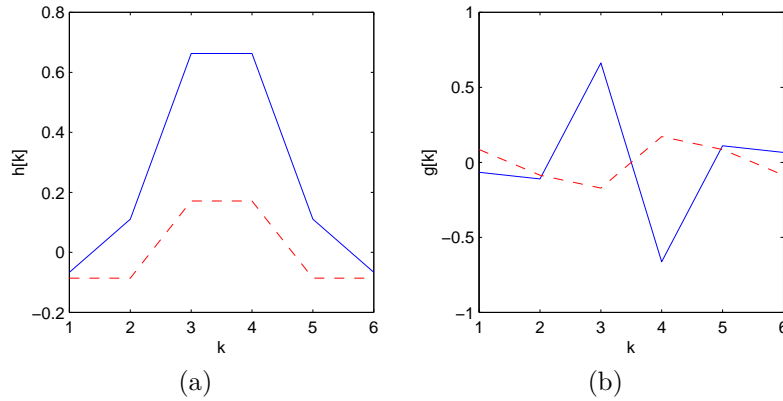
Figure 4.8: Complex Daubechies filters. The solid line represents the real part, the dashed line represents the imaginary part. a) $h[k]$ for complex Daubechies of order 3. b) $g[k]$ for complex Daubechies of order 3.

– Daubechies has shown that all compactly supported, real orthogonal wavelet bases and their associated conjugate quadrature filter (CQF) filterbanks are neither symmetric nor antisymmetric, except for the Haar transform [43, 42]. Lawton [88] and Lina *et al.* [93] showed that symmetry could be regained by using complex scaling functions. This led then to the definition of the Complex Daubechies wavelets, which are symmetric, orthogonal and of minimum length. These wavelets have been used in image coding applications in [16] and showed better compression performance than asymmetric wavelets.

– Another advantage is the near shift-invariance of the complex wavelet transform without significant increase of the complexity of the algorithm [118, 29]. The price to pay is that the complex wavelet transform is redundant by a factor of two when the input signal is real.

We also used the complex symmetric Daubechies wavelets for our fMRI application and implemented the filters like described in [91, 92]. We implemented the order $n = 3, 5, 7, 9$. The first order wavelet is the Haar wavelet. Note that the amplitude responses for the complex real Daubechies filters are the same, see Figure 7.20. The filter coefficients $h[k]$ and $g[k]$ for the order $n = 3$ are illustrated in Figure 4.8.

4.5 Applications

The wavelet transform is a powerful tool for many applications. Its space-frequency localization property and the existence of fast algorithms make it attractive for the anal-

ysis of non-stationary signals. The space-frequency localization property motivates new algorithms for example for image denoising applications, where the wavelet transform is able to preserve boundaries better than, for example, the Fourier transform [5, 72].

Another advantage of the wavelet transform as compared to the Fourier transform, is that the analyzing wavelet can be specifically tailored to the application. The Windowed Fourier Transform (WFT) circumvents the lack of time localization of the Fourier transform, however, it remains the problem of the fixed size of each shifted window, whereas for the wavelet transform we have different scales.

But also the scale specific non-redundant information of the wavelet coefficient makes the wavelet transform attractive. For detection problems like fMRI or microcalcifications in mammography, where the size of the feature to detect is not known a priori, the multiresolution properties of the wavelet transform give access to signal information at different resolution levels. This allows for the separation of signal components of various sizes including high frequency noise. Another example than fMRI of successful application is the detection of microcalcifications [158, 35].

The wavelet transform also de-correlates the signal, which results in a concentration of power into a smaller number of coefficients. This makes the wavelet transform especially attractive for the compression of signals or images. Wavelets have already been incorporated into the new JPEG2000 algorithm; they tend to be quite flexible and yield subjectively better results than the standard DCT-based JPEG, at especially high compression ratios.

Here, we will apply the wavelet transform to fMRI data.

Chapter 5

Orthogonal Quincunx Wavelets with Fractional Orders

In this chapter, we present a new family of 2D and 3D orthogonal wavelets which uses quincunx sampling. The orthogonal refinement filters have a simple analytical expression in the Fourier domain as a function of the order λ, which may be non-integer. We can also prove that they yield wavelet bases of $L_2(\mathbb{R}^2)$ for any $\lambda > 0$. The wavelets are fractional in the sense that the approximation error at a given scale a decays like $O(a^\lambda)$; they also essentially behave like fractional derivative operators. To make our construction practical, we propose an FFT-based implementation that turns out to be surprisingly fast. In fact, our method is almost as efficient as the standard Mallat algorithm for separable wavelets.

5.1 Introduction

The great majority of wavelet bases that are currently used for image processing are separable. There are two primary reasons for this. The first is convenience because wavelet theory is most developed in 1D and that these results are directly transposable to higher dimensions through the use of tensor product basis functions. The second is efficiency because a separable transform can be implemented by successive 1D processing of the rows and columns of the image. The downside, however, is that separable transforms tend to privilege the vertical and horizontal directions. They also produce a so-called "diagonal" wavelet component, which does not have a straightforward directional interpretation.

Non separable wavelets, by contrast, offer more freedom and can be better tuned to the characteristics of images [82, 103]. Their less attractive side is that they require more computations. The quincunx wavelets are especially interesting because they can

be designed to be nearly isotropic [56]. In contrast with the separable case, there is a single wavelet and the scale reduction is more progressive: a factor $\sqrt{2}$ instead of 2. The preferred technique for designing quincunx wavelets with good isotropy properties is to use the McClellan transform to map 1D biorthogonal designs to the multidimensional case [99]. Since this approach requires the filters to be symmetric, it has only been applied to the biorthogonal case because of the strong incentive to produce filters that are compactly supported [31, 81, 125, 138].

In this chapter, we construct a new family of quincunx wavelets that are orthogonal and have a fractional order of approximation. The idea of fractional orders was introduced recently in the context of spline wavelets for extending the family to non-integer degrees [145]. The main advantage of having a continuously-varying order parameter—not just integer steps as in the traditional wavelet families—is flexibility. It allows for a continuous adjustment of the key parameters of the transform; e.g., regularity and localization of the basis functions. The price that we are paying for these new features—orthogonality with symmetry as well as fractional orders—is that the filters can no longer be compactly supported. We will make up for this handicap by proposing a fast FFT-based implementation which is almost as efficient as Mallat's algorithm for separable wavelets [95].

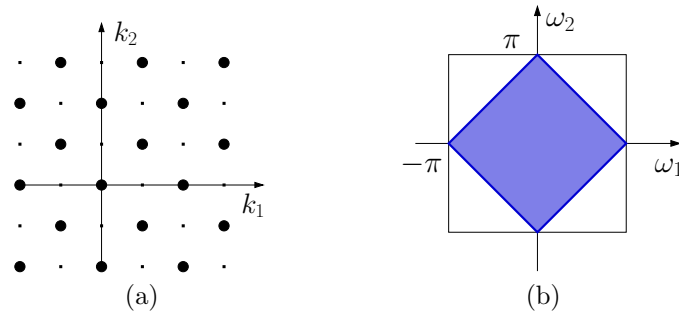## 5.2   Quincunx sampling and filterbanks

Figure 5.1: The quincunx lattice (a) and its corresponding bandwidth (b).

First, we recall some basic results on quincunx sampling and perfect reconstruction filterbanks [148]. The quincunx sampling lattice is shown in Fig. 5.1. Let $x[\vec{k}]$ with $\vec{k} = (k_1, k_2) \in \mathbb{Z}^2$ denote the discrete signal on the initial grid. The 2D Fourier transform of $x[\vec{k}]$ is denoted by $X(\vec{\omega}) = \sum_{\vec{k} \in \mathbb{Z}^2} x[\vec{k}] e^{-j\langle \vec{\omega}, \vec{k} \rangle}$ with $\vec{\omega} = (\omega_1, \omega_2)$. Then, its quincunx

sampled version is

$$[x]_{\downarrow \mathbf{D}}[\vec{k}] = x[\mathbf{D}\vec{k}] \text{ where } \mathbf{D} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{5.1}$$

Our down-sampling matrix $\mathbf{D}$ is such that $\mathbf{D^2} = 2\mathbf{I}$. The Fourier-domain version of (5.1) is

$$[x]_{\downarrow \mathbf{D}}[\vec{k}] \longleftrightarrow \frac{1}{2} \left[X\left(\mathbf{D}^{-T}\vec{\omega} \right) + X\left(\mathbf{D}^{-T}\vec{\omega} + \vec{\pi} \right) \right],$$

where $\vec{\pi} = (\pi, \pi)$.
The upsampling is defined by

$$[x]_{\uparrow \mathbf{D}}[\vec{k}] = \begin{cases} x[\mathbf{D^{-1}}\vec{k}] & \text{if } k_1 + k_2 \text{ even} \\ 0 & \text{else where} \end{cases}$$

and its effect in the transform domain is as follows:

$$[x]_{\uparrow \mathbf{D}}[\vec{k}] \longleftrightarrow X\left(\mathbf{D}^T\vec{\omega} \right).$$

If we now chain the down-sampling and up-sampling operators, we get

$$[x]_{\downarrow \mathbf{D} \uparrow \mathbf{D}}[\vec{k}] = \begin{cases} x[\vec{k}] & \text{if } k_1 + k_2 \text{ even} \\ 0 & \text{else where} \end{cases}$$

$$\updownarrow$$

$$\frac{1}{2} \left[X\left(\vec{\omega} \right) + X\left(\vec{\omega} + \vec{\pi} \right) \right]. \tag{5.2}$$

Since quincunx sampling reduces the number of image samples by a factor of two, the corresponding reconstruction filterbank has two channels (cf. Fig. 5.2). The low-pass filter $\widetilde{H}$ reduces the resolution by a factor of $\sqrt{2}$; the wavelet coefficients correspond to the output of the high-pass filter $\widetilde{G}$.
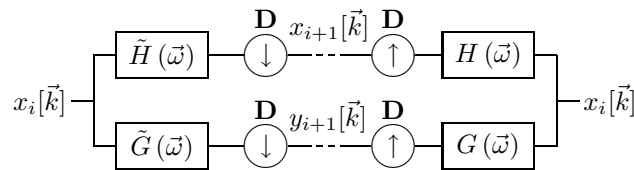

Figure 5.2: Perfect reconstruction filterbank on a quincunx lattice.

Applying the relation (5.2) to the blockdiagram in Fig. 5.2, it is easy to derive the conditions for a perfect reconstruction:

$$\begin{cases} \widetilde{H}(\vec{\omega})\,H(\vec{\omega}) + \widetilde{G}(\vec{\omega})\,G(\vec{\omega}) = 2 \\[2mm] \widetilde{H}(\vec{\omega}+\vec{\pi})\,H(\vec{\omega}) + \widetilde{G}(\vec{\omega}+\vec{\pi})\,G(\vec{\omega}) = 0, \end{cases} \qquad (5.3)$$

where H and G (resp., $\widetilde{H}$ and $\widetilde{G}$) are the transfer functions of the synthesis (resp., analysis) filters. In the orthogonal case, the analysis and synthesis filters are identical up to a central symmetry; the wavelet filter G is simply a modulated version of the low-pass filter H.

5.3   Fractional quincunx filters

To generate quincunx filters, we will use the standard approach which is to apply the diamond McClellan transform to map a one dimensional design onto the quincunx structure.

5.3.1   A new 1D wavelet family

As starting point for our construction, we introduce a new one-dimensional family of orthogonal filters:

$$\begin{aligned} H_\lambda(z) &= \frac{\sqrt{2}(z+2+z^{-1})^{\frac{\lambda}{2}}}{\sqrt{(z+2+z^{-1})^\lambda + (-z+2-z^{-1})^\lambda}} \\[2mm] &= \frac{\sqrt{2}(2+2\cos\omega)^{\frac{\lambda}{2}}}{\sqrt{(2+2\cos\omega)^\lambda + (2-2\cos\omega)^\lambda}}, \end{aligned} \qquad (5.4)$$

which is indexed by the continuously-varying order parameter λ.

These filters are symmetric and are designed to have zeros of order λ at $z = -1$; the numerator is a fractional power of $(z+2+z^{-1})$ (the simplest symmetric refinement filter) and the denominator is the appropriate orthonormalization factor. By varying λ, we can adjust the frequency response as shown in Fig. 5.3. As λ increases, $H_\lambda(z)$ converges to the ideal half-band low-pass filter. Also note that these filters are maximally flat at the origin; they essentially behave like $H_\lambda(\omega)/\sqrt{2} = 1 + O(\omega^\lambda)$ as $\omega \to 0$. Their frequency response is similar to the Daubechies' filters with two important differences: (1) the filters are symmetric, and (2) the order is not restricted to integer values.

We can prove mathematically that these filters will generate valid 1D fractional wavelet bases of L_2 similar to the fractional splines presented in [145]. The order property (here fractional) is essential because it determines the rate of decay of the approximation error as a function of the scale. It also conditions the behavior of the corresponding wavelet which will act like a fractional derivative of order λ; in other words, it will kill all polynomials of degree $n \le \lceil \lambda - 1 \rceil$.
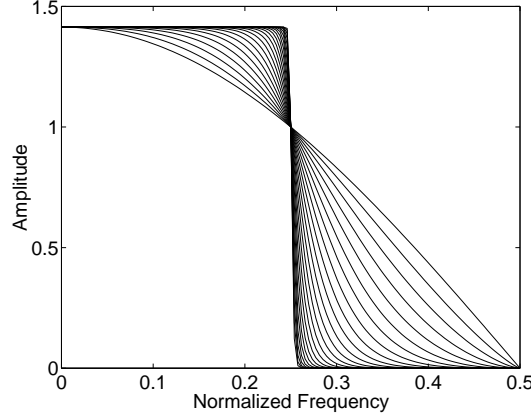
Figure 5.3: Frequency responses of the orthogonal refinement filters for $\lambda = 1, \ldots, 100$.

5.3.2 Corresponding 2D wavelet family

Applying the diamond McClellan transform to the filter above is straightforward; it amounts to replacing $\cos \omega$ by $\frac{1}{2}(\cos \omega_1 + \cos \omega_2)$ in (5.4). Thus, our quincunx refinement filter is given by

$$H_\lambda(\omega_1, \omega_2) = \frac{\sqrt{2}(2 + \cos \omega_1 + \cos \omega_2)^{\frac{\lambda}{2}}}{\sqrt{(2 + \cos \omega_1 + \cos \omega_2)^\lambda + (2 - \cos \omega_1 - \cos \omega_2)^\lambda}}. \tag{5.5}$$

This filter is guaranteed to be orthogonal because the McClellan transform has the property of preserving biorthogonality. Also, by construction, the λth order zero at $\omega = \pi$ gets mapped into a corresponding zero at $(\omega_1, \omega_2) = (\pi, \pi)$; this is precisely the condition that is required to get a two dimensional wavelet transform of order λ.

The orthogonal wavelet filter is obtained by modulation

$$G_\lambda(\omega_1, \omega_2) = e^{j\omega_1} H_\lambda(-\omega_1 - \pi, -\omega_2 - \pi). \tag{5.6}$$

The corresponding orthogonal scaling function $\varphi_\lambda(\vec{x})$ is defined implicitly as the solution of the quincunx two-scale relation:

$$\varphi_\lambda(\vec{x}) = \sqrt{2} \sum_{\vec{k} \in \mathbb{Z}^2} h_\lambda[\vec{k}] \varphi_\lambda\left(\mathbf{D}\vec{x} - \vec{k}\right).$$

Since the refinement filter is orthogonal with respect to the quincunx lattice, it follows that $\varphi_\lambda(\vec{x}) \in L_2(\mathbb{R}^2)$ and that it is orthogonal to its integer translates. Moreover, for $\lambda > 0$, it will satisfy the partition of unity condition, which comes as a direct consequence

of the vanishing of the filter at $(\omega_1, \omega_2) = (\pi, \pi)$. Thus, we have the guarantee that our scheme will yield orthogonal wavelet bases of $L_2(\mathbb{R}^2)$. The underlying orthogonal quincunx wavelet is simply

$$\psi_\lambda\left(\vec{x}\right) = \sqrt{2} \sum_{\vec{k} \in \mathbb{Z}^2} g_\lambda[\vec{k}] \varphi_\lambda\left(\mathbf{D}\vec{x} - \vec{k}\right).$$
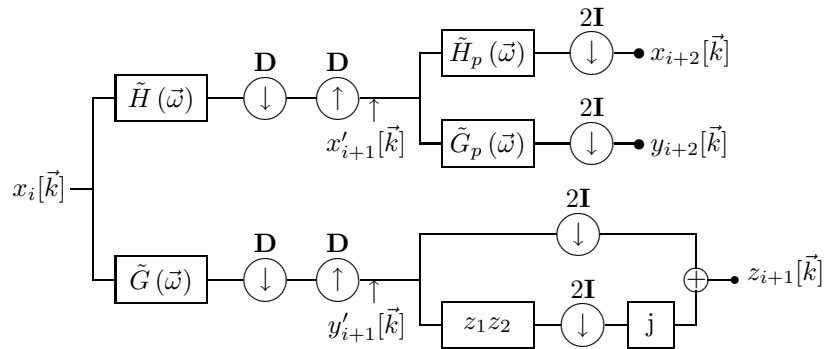
5.4   Implementation in Fourier domain

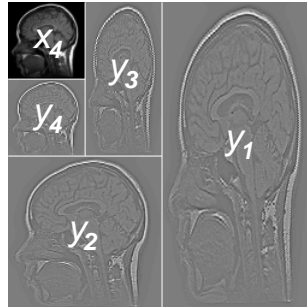Figure 5.4: Analysis part of the 2D QWT for two iterations.

Figure 5.5: Formation of quincunx coefficients for 4 iterations.

The major objection that can be made to our construction is that the filters are not FIR and that it may be difficult and costly to implement the transform in practice. We will see here that we can turn the situation around and obtain a very simple and efficient algorithm that is based on the FFT, following the idea of [117]. Working in the frequency domain is also very convenient because of the way in which we have specified our filters
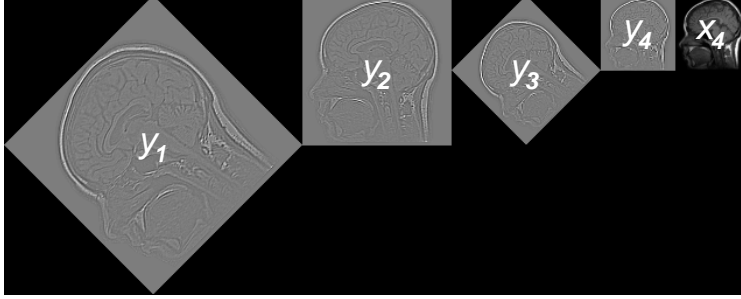
Figure 5.6: Other possible formation of quincunx coefficients for 4 iterations.

(cf. Eqs. (5.5) and (5.6)). A first such implementation has been done by [104]. Our algorithm is an other variation, which minimizes the number and size of FFT's and seems to be faster.

Here, we will only describe the decomposition part of our algorithm which corresponds to the block diagram presented in Fig. 5.4 where we have pooled together two levels of decomposition. The initialization step is to evaluate the FFT of the initial input image $x[\vec{k}]$ and to precompute the corresponding sampled frequency responses of the analysis filters $\widetilde{H}[\vec{n}]$ and $\widetilde{G}[\vec{n}]$ using (5.5) and (5.6).

Assuming that the current image size is $N \times N$, the input variable, following a 2D FFT, is

$$X_i[\vec{n}] = \sum_{\vec{k}} x_i[\vec{k}] e^{-j\frac{2\pi \langle \vec{k}, \vec{n} \rangle}{N}} \qquad \text{for } n_1, n_2 = 0 \ldots N - 1.$$

Globally, at the end of the process, the output variables are the quincunx wavelet coefficients $y_1[\vec{k}]$, $y_2[\vec{k}], ..., y_J[\vec{k}]$ and $x_J[\vec{k}]$ (see Fig. 5.5). Their Fourier transforms for the odd iteration are derived from the auxiliary $N \times N$ signals:

$$
\begin{aligned}
X'_{i+1}[\vec{n}] &= \sum_{\vec{k}} x'_{i+1}[\vec{k}] e^{-j\frac{2\pi \langle \vec{k}, \vec{n} \rangle}{N}}, \\
Y'_{i+1}[\vec{n}] &= \sum_{\vec{k}} y'_{i+1}[\vec{k}] e^{-j\frac{2\pi \langle \vec{k}, \vec{n} \rangle}{N}} \qquad n_1, n_2 \in [0, N-1]^2.
\end{aligned}
\qquad (5.7)
$$

Down- and upsampling with $\mathbf{D}$ in the first iteration step (see Figure 5.4) introduces zeros in the space domain and preserves size ($Y'_{i+1}[\vec{n}]$); however, it implies some symmetry/redundancy in frequency domain. Only half of the coefficients need to be computed which saves operations. This reduced signal $Y'_{i+1}[\vec{k}]$ and its corresponding low-pass signal

is obtained by

$$
\begin{aligned}
Y'_{i+1}\left[m_1, n_2\right] &= \tfrac{1}{2}\left(\widetilde{G}\left[m_1, n_2\right] X_i\left[m_1, n_2\right]\right. \\
&\quad \left. +\widetilde{G}\left[m_1+\left(\tfrac{N}{2}\right), n_2+\left(\tfrac{N}{2}\right)\right] X_i\left[m_1+\left(\tfrac{N}{2}\right), n_2+\left(\tfrac{N}{2}\right)\right]\right) \\
X'_{i+1}\left[m_1, n_2\right] &= \tfrac{1}{2}\left(\widetilde{H}\left[m_1, n_2\right] X_i\left[m_1, n_2\right]\right. \\
&\quad \left. +\widetilde{H}\left[m_1+\left(\tfrac{N}{2}\right), n_2+\left(\tfrac{N}{2}\right)\right] X_i\left[m_1+\left(\tfrac{N}{2}\right), n_2+\left(\tfrac{N}{2}\right)\right]\right),
\end{aligned}
\tag{5.8}
$$

where $m_1 \in [0, \tfrac{N}{2}-1]$, $n_2 \in [0, N-1]$.

To generate the signal $y_{i+1}[\vec{k}]$ of (5.7) in the way that is depicted in Figure 5.5 with every second row shifted by one pixel, we separate the image in even ($y_{i+1,\text{even}}$) and odd ($y_{i+1,\text{odd}}$) rows already in the Fourier domain, using the auxiliary variable $Z[\vec{m}]$:

$$
\begin{aligned}
Z[\vec{m}] &= Y'_{i+1}[\vec{m}]+Y'_{i+1}\left[\vec{m}+\left(0, \tfrac{N}{2}\right)\right] \\
&\quad +j\left(Y'_{i+1}[\vec{m}]-Y'_{i+1}\left[\vec{m}+\left(0, \tfrac{N}{2}\right)\right]\right) e^{j\frac{2\pi(m_1+m_2)}{N}} \\
&\quad\quad \updownarrow \\
z_{i+1}[\vec{k}] &= y_{i+1,\text{even}}[\vec{k}]+j y_{i+1,\text{odd}}[\vec{k}].
\end{aligned}
\tag{5.9}
$$

The sum in the real part $(Y'_{i+1}[\vec{m}]+Y'_{i+1}[\vec{m}+\left(0, \tfrac{N}{2}\right)])$ represents downsampling by two, keeping all the even rows, whereas the sum in the imaginary part represents the odd rows.

In the space domain, we alternate the rows $y_{i+1}[k_1, 2k_2+1] = \text{Re}\left\{z[\vec{k}]\right\}$ and $y_{i+1}[k_1, 2k_2] = \text{Im}\left\{z[\vec{k}]\right\}$. Since $z[\vec{k}]$ is four times smaller than $y'_{i+1}[\vec{k}]$, we save computations with the reduced-size IFFT.

Instead of rotating the frequency variables after each iteration, we calculate a rotated version of the filters $\widetilde{H}[\vec{z}] \rightarrow \widetilde{H}_p[\vec{z}]$ and $\widetilde{G}[\vec{z}] \rightarrow \widetilde{G}_p[\vec{z}]$, which we apply at all even iterations. In this way, we also save two rotations per iteration in the frequency domain.

The Fourier transforms of the output for the even iterations are:

$$
Y_{i+2}[\vec{m}] = \sum_{\vec{k}} y_{i+2}[\vec{k}]e^{-j\frac{2\pi\langle\vec{k},\vec{m}\rangle}{\frac{N}{2}}} \quad \text{for } m_1, m_2 = 0\ldots\tfrac{N}{2}-1.
$$

Its source is:

$$
\begin{aligned}
X_{i+2}[\vec{m}] &= \tfrac{1}{2}\left(\widetilde{H}_p[\vec{m}] X'_{i+1}[\vec{m}]+\widetilde{H}_p\left[\vec{m}+\left(0, \tfrac{N}{2}\right)\right] X'_{i+1}\left[\vec{m}+\left(0, \tfrac{N}{2}\right)\right]\right) \\
Y_{i+2}[\vec{m}] &= \tfrac{1}{2}\left(\widetilde{G}_p[\vec{m}] X'_{i+1}[\vec{m}]+\widetilde{G}_p\left[\vec{m}+\left(0, \tfrac{N}{2}\right)\right] X'_{i+1}\left[\vec{m}+\left(0, \tfrac{N}{2}\right)\right]\right),
\end{aligned}
\tag{5.10}
$$

where $\widetilde{H}_p\left[\vec{m}\right] = \widetilde{H}\left[\mathbf{D}\vec{m} \bmod (N, N)\right]$ and $\widetilde{G}_p\left[\vec{m}\right] = \widetilde{G}\left[\mathbf{D}\vec{m} \bmod (N, N)\right]$. The process is then iterated until one reaches the final resolution. When the last iteration is even, we lower the computation costs with the FFT by utilizing its imaginary part: $z[\vec{k}] = \sum_{\vec{m}} \left(X_{i+2}[\vec{m}] + jY_{i+2}[\vec{m}]\right) e^{j\frac{2\pi\langle\vec{m},\vec{k}\rangle}{\frac{N}{2}}}$, where $x_{i+2}[\vec{k}] = \mathrm{Re}\left\{z[\vec{k}]\right\}$ and $y_{i+2}[\vec{k}] = \mathrm{Im}\left\{z[\vec{k}]\right\}$.

Obviously, as one gets coarser, the Fourier transforms of the filters need not be recalculated; they are simply obtained by down-sampling the previous arrays.

The synthesis algorithm operates according to the same principles and corresponds to the flow graph transpose of the decomposition algorithm using up-sampling instead. For instance, the synthesis counterpart of (5.10) is:

$$X'_{i+2}\left[m_1, m_2 + \left(\tfrac{N}{2}\right)\right] = X_{i+2}\left[m_1, m_2\right]$$
$$Y'_{i+2}\left[m_1, m_2 + \left(\tfrac{N}{2}\right)\right] = Y_{i+2}\left[m_1, m_2\right]$$
$$X_{i+1}\left[m_1, n_2\right] = X'_{i+2}[m_1, n_2]H_p[m_1, n_2] + Y'_{i+2}[m_1, n_2]G_p[m_1, n_2].$$

5.5 Experiments

5.5.1 Benchmark and testing

We have implemented the algorithm in Matlab and report computation times below $0.85\,sec$ for four iterations and for a 256×256 image on an aging Sun Ultra 30 workstation; the decomposition is essentially perfect with an (RMS) reconstruction error below 10^{-12}. The method is generic and works for any set of filters that can be specified in the frequency domain.

Two examples of fractional quincunx wavelet decompositions with $\lambda = \sqrt{2}$ and $\lambda = \pi$ are shown in Fig. 5.7. Note how the residual image details are more visible for the lower value of λ. The larger λ reduces the energy of the wavelet coefficient, but this also comes at the expense of some ringing. Thus, it is convenient to have an adjustable parameter to search for the best tradeoff.

An advantage of the present approach is that the filters for small λ are nearly isotropic; this is the reason why the wavelet details in Fig. 5.7 do not present any preferential orientation. The degree of isotropy of the various filters can be visualized in Figure 5.8. The shape of the contour-plots of the low-pass filter $H(\vec{\omega})$ confirms that the degree of isotropy is the best for small values of λ. At the other extreme, when $\lambda \to \infty$, $H(\vec{\omega})$ tends to the diamond-shaped ideal filter.

Another nice feature of the algorithm is that the computational cost remains the same irrespective of the value of λ.
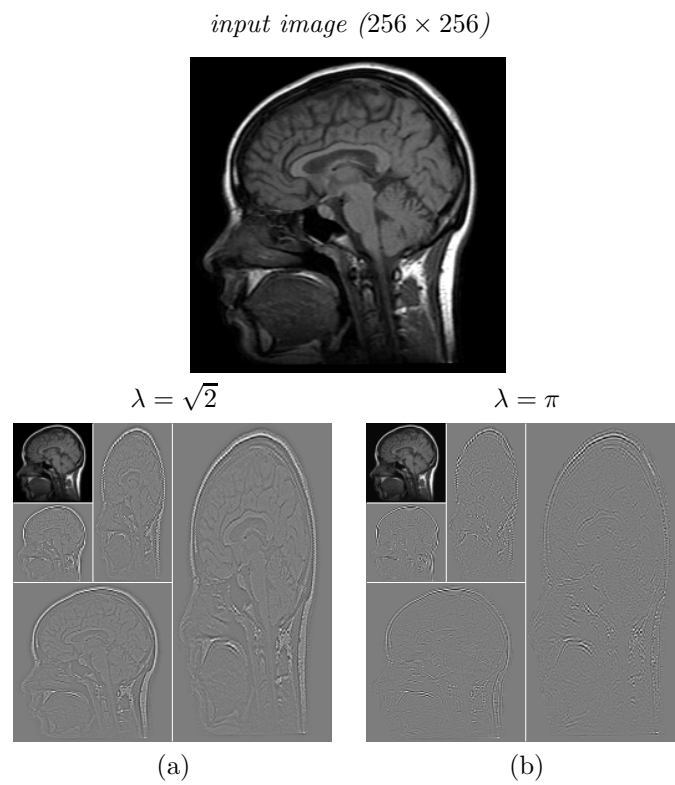
Figure 5.7: Quincunx wavelet transforms with four iterations: (a) $\lambda = \sqrt{2}$, (b) $\lambda = \pi$.

$\lambda = 1$ $\lambda = \sqrt{2}$

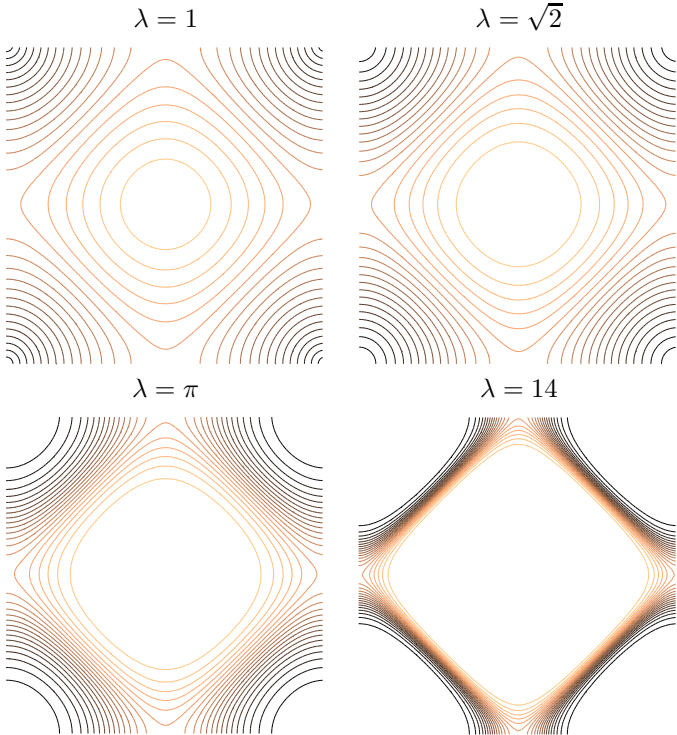$\lambda = \pi$ $\lambda = 14$

Figure 5.8: Contour-plot of the low-pass filter $H(\vec{\omega})$ for different order parameters.

5.5.2   Dependence of the order parameter

The usefulness of a tunable order parameter is demonstrated in the following experiment: we apply the quincunx transform to an image ("cameraman") and reconstruct the image with only 15 % of the largest coefficients and measure then the SNR depending on the order parameter. The graph 5.9 shows the SNR dependence on λ; the optimum is achieved with $\lambda = 2.1$.
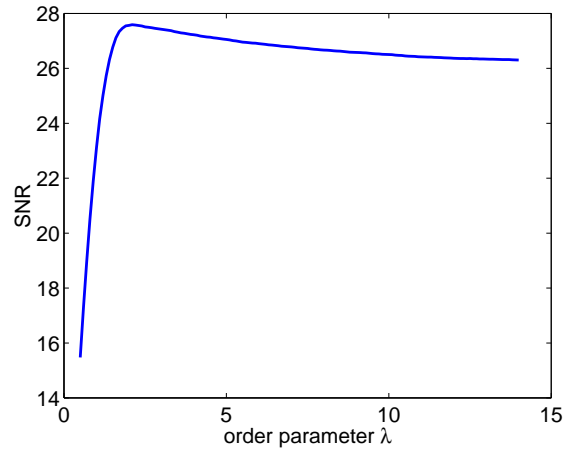

Figure 5.9: Relation between the order parameter λ and the SNR of the reconstructed image with only 10000 of the largest coefficients. The reconstruction with $\lambda = 2.1$ reaches the highest SNR.

5.5.3   Approximation properties

The main differences between the quincunx (`quin`) and the conventional algorithm (`sep`) is the finer scale progression and the nonseparability. To test the impact that this may have on compression capability, we compared the approximation of `quin` and `sep`. Since the wavelet transform is orthogonal, the approximation error (distortion) is equal to $D^2 = \|x - \hat{x}\|^2 = \|y - \hat{y}\|^2$, where y are the wavelet coefficients, x is the input, $\hat{x}$ and $\hat{y}$ is the estimated image. Also D^2 in the space domain is equivalent to the sum of squares of discarded wavelet coefficients [143].

Linear approximation

In classical rate-distortion theory, the coefficients are grouped into channels and coded independently. In the orthogonal case, D^2 is equivalent to the difference between the

original of "cameraman"

$\lambda = 2.1$ $\lambda = 14$
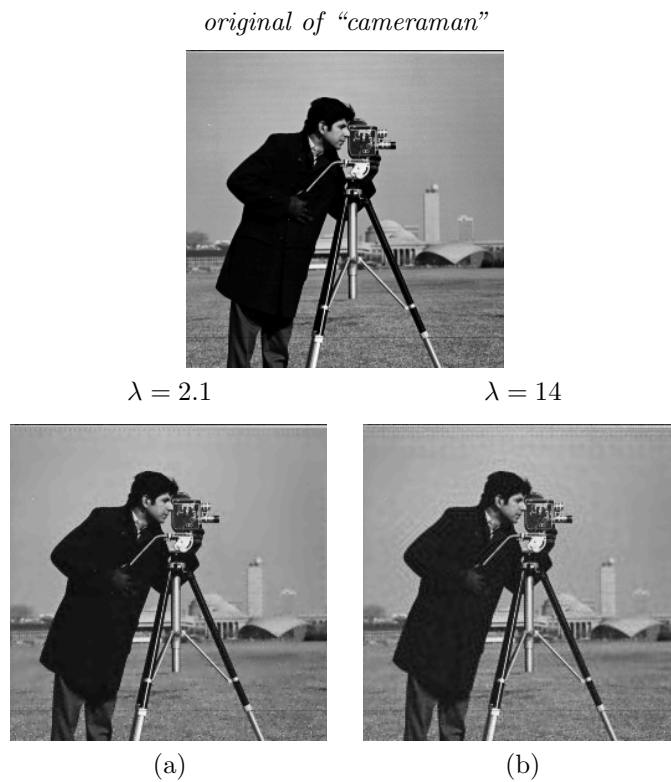

(a)   (b)

Figure 5.10: Reconstruction of the "cameraman" with the 10000 largest coefficients with an order parameter a) $\lambda = 2.1$ and b) $\lambda = 14$.

signal's energy and the energy of the reconstructed signal: $\|x - \hat{x}\|^2 = \|x\|^2 - \|\hat{x}\|^2 = \|x\|^2 - \left(\|\hat{x}_J\|^2 + \sum_{j=1}^{J} \|\hat{y}_j\|^2 \right)$. The distortion across J channels with variance σ_i is

$$D = J \cdot C \cdot 2^{-2\bar{R}} \cdot \rho^2,$$

where C is a constant, $\bar{R}$ is the mean rate and ρ is the geometric mean of the subband variances:

$$\rho = \left(\prod_{i=1}^{J} \sigma_i^2 \right)^{\frac{1}{J}}.$$

When ρ is smaller, the distortion is smaller as well. What this means qualitatively is that the wavelet transform which has the larger spread in the variances will achieve the better coding gain [148]. The linear approximation subband coding gain for sample-by-sample quantization (PCM) is described by:

$$G_{PCM} = \frac{D_{PCM}}{D_{SBC}} = \frac{\frac{1}{J} \sum_{i=1}^{J} \sigma_i^2}{\left(\prod_{i=1}^{J} \sigma_i^2 \right)^{\frac{1}{J}}}.$$

In Figure 5.12 a) we compare the energy packing properties of `quin` and `sep` for linear approximation. "Energy packing" refers to the property that the more the first coefficients contain energy, the better the DWT yields compression. We decomposed an image $I(x, y)$ for the maximal possible iteration J. The image $I(x, y)$ is depicted in Figure 5.11; its decompositions for `quin` and `sep` are shown on the right hand side. We start to sum up the energy of the subbands with the lowest resolution. Each step of the stairs represents a subband. The first subbands of `quin` report higher energy than `sep`, as shown in the Graph in Figure 5.12. Its coding gain for `quin` is $G_{PCM} = 77.9$ and for `sep` $G_{PCM} = 75.6$. For the majority of images we considered, `quin` yielded better energy packing than `sep`. However, we did also encounter one notable counter example, "Lena", (cf. Fig. 5.13), which seems to be better suited for a separable treatment.

Since the branches are orthogonal, the transformation that provides the maximum energy compaction in the low-pass channel is also the one that results in the minimum approximation error [143]. Since most images have a power spectrum that is roughly rotationally-invariant and decreases with higher frequencies, separable systems are usually not best suited for isolating a low-pass channel containing most energy and having high-pass channels with low energy. In contrast, a quincunx low-pass filter will retain more of the original signal's energy [148].

Nonlinear approximation

A more recent trend in wavelet theory is the study of *nonlinear approximation*. In this case we don't take the "n-first" coefficients, but the "n-largest" coefficients to approximate a signal with $n < N$ coefficients. This yields better energy packing, since in the
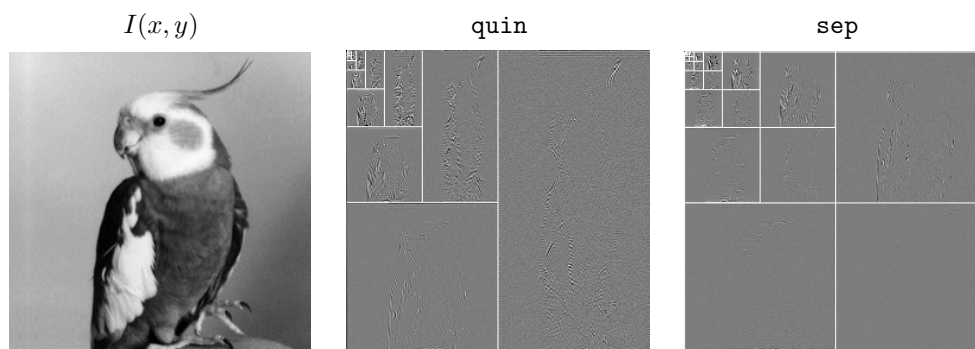
Figure 5.11: $I(x, y)$ has the size 256×256. Left hand side: $I(x, y)$ is decomposed for the maximal possible iteration J.
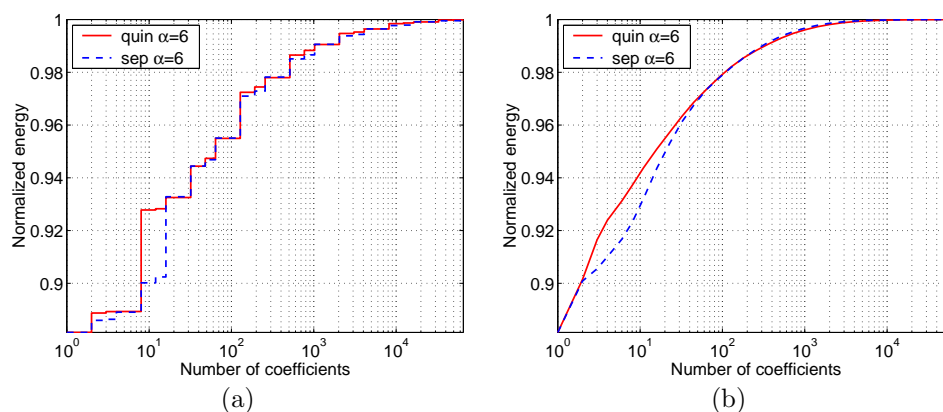

Figure 5.12: Comparison of energy-compaction property between `quin` and `sep` of image $I(x, y)$ (Figure 5.11): a) linear approximation depending on number of subbands (in log), b) nonlinear approximation depending of the n largest coefficients (in log). In both cases `quin` reports better energy packing.
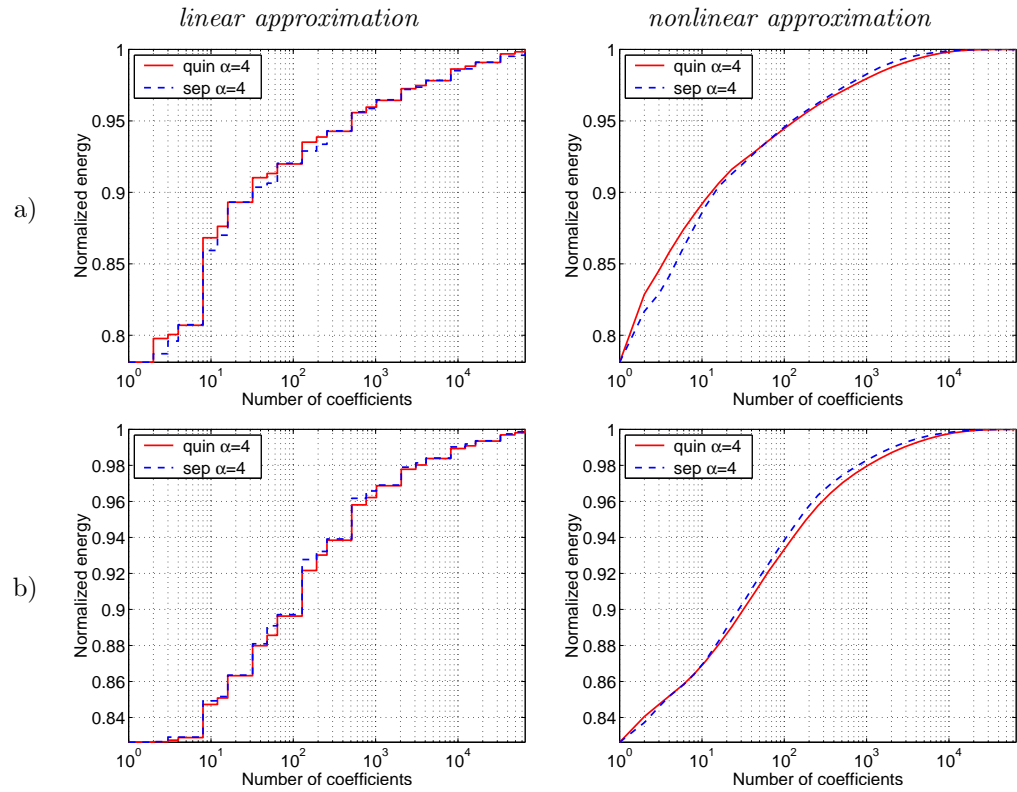
Figure 5.13: a) represents the approximation curves of the "cameraman". It is again an example, where `quin` is better than `sep`. b) represents the curves of "Lena". This is an example where both, the linear approximation and the nonlinear approximation are better for `sep` than for `quin`.
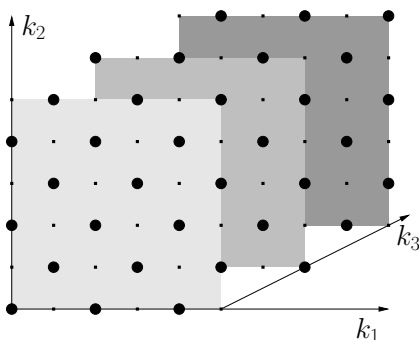
Figure 5.14: 3D-FCO (face-centered orthorhombic) lattice, corresponding to the sampling matrix (5.11).

wavelet domain the "n-first" coefficients are not necessarily the largest one, especially along the position-indices [47]. The distortion is described by [46]:

$$D^2 = \|y - y_{N(T)}\|^2 = \sum_{|y|<T} |y[n]|^2.$$

Moreover, it can be shown that

$$D \propto C \cdot (N(T))^{-\frac{\gamma}{2}},$$

when the smoothness of y is measured by its inclusion in some critical Besov space $B_q^\gamma(L^q(I))$ with $\frac{1}{q} = \frac{\gamma}{2} + \frac{1}{2}$, roughly when y is a function with γ derivatives in $L^q(I)$.
For the nonlinear approximation, `quin` also yields a better approximation than `sep` for a small n in many cases. Figure 5.12 b) represents the energy depending on the n largest coefficients (in log). Further examples of (non-) linear examples are shown in Figure 5.13.

5.6 Extension to 3D

The extension of quincunx sampling to 3D is straight forward. The quincunx sampling lattice for 3D is shown in Fig. 5.14. Let $x[\vec{k}]$ denote the discrete signal on the initial grid. Then, its quincunx sampled version is

$$[x]_{\downarrow\mathbf{D}}[\vec{k}] = x[\mathbf{D}\vec{k}] \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & -1 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{5.11}$$

Our down-sampling matrix $\mathbf{D}$ is such that $\mathbf{D^3} = 2\mathbf{I}$ and $|\det D| = 2$. The Fourier-domain version of this formula is similar to the case in 2D:

$$[x]_{\downarrow \mathbf{D}}[\vec{k}] \longleftrightarrow \frac{1}{2} \left[X \left(\mathbf{D}^{-T} \vec{\omega} \right) + X \left(\mathbf{D}^{-T} \vec{\omega} + \vec{\pi} \right) \right],$$

where $\vec{\pi} = (\pi, \pi, \pi)$.

The implementation for 3D is as follows:

The output variables are the discrete Fourier transforms of the wavelet coefficients

$$Y_{i+1}\left[\vec{n}\right] = \sum_{\vec{k}} y_{i+1}[\vec{k}] e^{-j \frac{2\pi \langle \vec{k}, \vec{n} \rangle}{N}} \quad \text{for } n_1, n_2, n_3 = 0 \ldots N - 1$$

$$Y_{i+2}\left[\vec{n}\right] = \sum_{\vec{k}} y_{i+2}[\vec{k}] e^{-j \frac{2\pi \langle \vec{k}, \vec{n} \rangle}{N}} \quad \text{for } n_1, n_2, n_3 = 0 \ldots N - 1$$

$$Y_{i+3}\left[\vec{m}\right] = \sum_{\vec{k}} y_{i+3}[\vec{k}] e^{-j \frac{2\pi \langle \vec{k}, \vec{m} \rangle}{\frac{N}{2}}} \quad \text{for } m_1, m_2, m_3 = 0 \ldots \frac{N}{2} - 1$$

The coefficients themselves are recovered by inverse FFT. The Fourier transforms after the first level of filtering are given by

$$X_{i+1}\left[\vec{n}\right] = \frac{1}{2} \left(\widetilde{H}\left[\vec{n}\right] X_i \left[\vec{n}\right] + \widetilde{H} \left[\vec{n} + \left(\frac{N}{2}, \frac{N}{2}, \frac{N}{2}\right)\right] X_i \left[\vec{n} + \left(\frac{N}{2}, \frac{N}{2}, \frac{N}{2}\right)\right] \right)$$

$$Y_{i+1}\left[\vec{n}\right] = \frac{1}{2} \left(\widetilde{G}\left[\vec{n}\right] X_i \left[\vec{n}\right] + \widetilde{G} \left[\vec{n} + \left(\frac{N}{2}, \frac{N}{2}, \frac{N}{2}\right)\right] X_i \left[\vec{n} + \left(\frac{N}{2}, \frac{N}{2}, \frac{N}{2}\right)\right] \right).$$

After the second level of filtering, we have:

$$X_{i+2}\left[\vec{n}\right] = \frac{1}{2} \left(\widetilde{H}_p\left[\vec{n}\right] X_{i+1} \left[\vec{n}\right] + \widetilde{H}_p \left[\vec{n} + \left(0, \frac{N}{2}, 0\right)\right] X_{i+1} \left[\vec{n} + \left(0, \frac{N}{2}, 0\right)\right] \right)$$

$$Y_{i+2}\left[\vec{n}\right] = \frac{1}{2} \left(\widetilde{G}_p\left[\vec{n}\right] X_{i+1} \left[\vec{n}\right] + \widetilde{G}_p \left[\vec{n} + \left(0, \frac{N}{2}, 0\right)\right] X_{i+1} \left[\vec{n} + \left(0, \frac{N}{2}, 0\right)\right] \right).$$

Note that these are computed at the resolution of the input. The size reduction only takes place during the third step:

$$X_{i+3}\left[\vec{m}\right] = \frac{1}{2} \left(\widetilde{H}_{pp}\left[\vec{m}\right] X_{i+2} \left[\vec{m}\right] + \widetilde{H}_{pp} \left[\vec{m} + \left(\frac{N}{2}, 0, 0\right)\right] X_{i+2} \left[\vec{m} + \left(\frac{N}{2}, 0, 0\right)\right] \right)$$

$$Y_{i+3}\left[\vec{m}\right] = \frac{1}{2} \left(\widetilde{G}_{pp}\left[\vec{m}\right] X_{i+2} \left[\vec{m}\right] + \widetilde{G}_{pp} \left[\vec{m} + \left(\frac{N}{2}, 0, 0\right)\right] X_{i+2} \left[\vec{m} + \left(\frac{N}{2}, 0, 0\right)\right] \right),$$

where $\widetilde{H}_p\left[\vec{m}\right] = \widetilde{H}\left[\mathbf{D}\vec{m} \bmod\left(N,N,N\right)\right]$ and $\widetilde{H}_{pp}\left[\vec{m}\right] = \widetilde{H}\left[\mathbf{D}^2\vec{m} \bmod\left(N,N,N\right)\right]$. Analogously, we have that: $\widetilde{G}_p\left[\vec{m}\right] = \widetilde{G}\left[\mathbf{D}\vec{m} \bmod\left(N,N,N\right)\right]$ and $\widetilde{G}_{pp}\left[\vec{m}\right] = \widetilde{G}\left[\mathbf{D}^2\vec{m} \bmod\left(N,N,N\right)\right]$.

It is still possible to optimize this implementation, especially for the calculation of FFT's. Also the redundancy for the second iteration may be removed. For reason of calculation time, the computation of H_p, H_{pp}, G_p and G_{pp} as well the corresponding analysis filters are implemented in C.

5.6.1 Approximation properties in 3D

We compared the compression capability for `quin` and `sep` for 3D-images, similar to what is described for 2D in Section 5.5.3. An example is the volume $V(x,y,z)$ depicted in Figure 5.15 (human head). In Figure 5.16 are the graphs which represents the percentage of the energy of $V(x,y,z)$ (y-axis) corresponding to the number of coefficients, which approximate $V(x,y,z)$ (x-axis). Another example of this issue is shown in Figure 5.17.

(a)   (b)

Figure 5.15: Volume $V(x,y,z)$: a) surface of $V(x,y,z)$ b) one slice of $V(x,y,z)$.

5.7 Conclusion

We have introduced a new family of orthogonal wavelet transforms for quincunx lattices. A key feature is the continuously-varying order parameter λ which can be used to adjust the bandpass characteristics as well as the localization of the basis functions.

We have also demonstrated that these wavelet transforms could be computed quite efficiently using FFTs. This should help dispel the commonly held belief that non separable wavelet decompositions are computationally much more demanding than the separable ones.
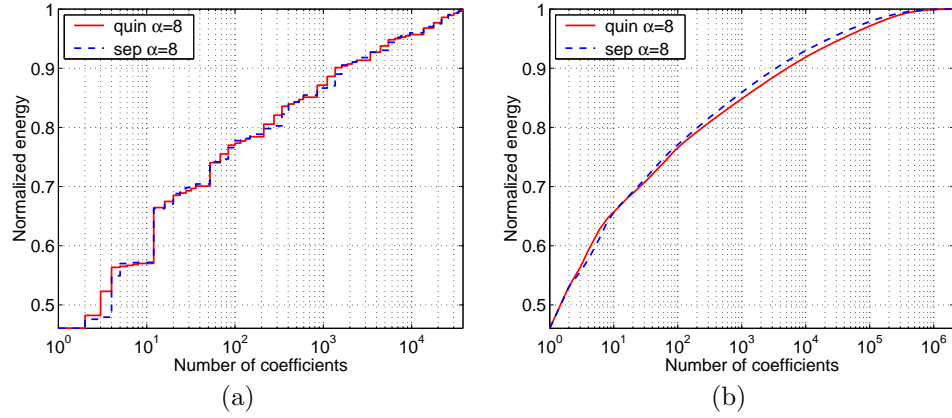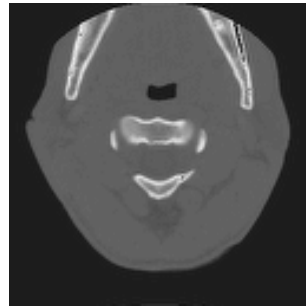
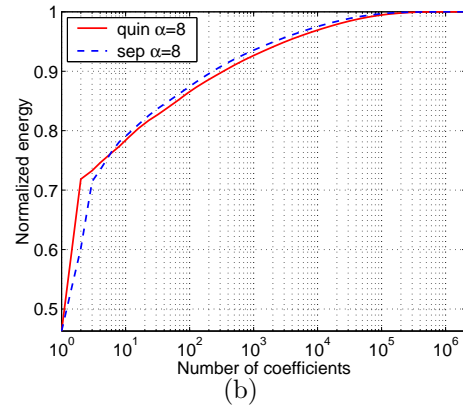Figure 5.16: Comparison of energy-compaction property between `quin` and `sep` of a volume $V(x, y, z)$ (Figure 5.15): a) linear approximation depending on number of subbands (in log), b) nonlinear approximation depending of the n largest coefficients (in log).

Figure 5.17: a) One slice of a spine. b) Corresponding nonlinear approximation. `quin` reports better energy packing with few coefficients.

Because of their nice properties and their ease of implementation, these wavelets should present a good alternative to the separable ones that are being used in a variety of image processing applications (data compression, filtering, texture analysis etc.).

Chapter 6

Analysis of fMRI Data using Orthogonal Filterbanks

In this chapter, we propose to analyze the data using an orthogonal filterbank as an alternative to the traditional approach which uses Gaussian spatial filtering to reduce measurement noise. This procedure is intended to simplify and eventually improve the statistical analysis. The system is designed to concentrate the signal into a fewer number of components thereby improving the signal-to-noise ratio. Thanks to the orthogonality property, we can test the filtered components independently on a voxel-by-voxel basis; this testing procedure is optimal for i.i.d. measurement noise. The number of components to test is also reduced because of down-sampling. This offers a straightforward approach to increase the sensitivity of the analysis (allowing a lower detection threshold) while applying the standard Bonferroni correction for multiple statistical tests. We present experimental results to illustrate the procedure. In addition, we discuss filter design issues. In particular, we introduce a family of orthogonal filters which are such that any integer reduction m can be implemented as a succession of elementary reductions m_1 to m_p where $m = m_1 \ldots m_p$ is a prime number factorization of m.

6.1   Introduction

Our first approach to simplify the statistical analysis, compared to the method of SPM, is to use a filterbank with downsampling for spatial noise reduction as an alternative to the Gaussian filtering of SPM. The general idea is that smoothing suppresses high-frequency components so that the data may as well be represented with less samples. In addition, we try to preserve the independence of the noise as much as possible in order to simplify subsequent statistical analyses.

Our solution is to apply an orthonormal filter, which has the property of transforming white noise into white noise. Here, we assume independent (over time and space), but (non-stationary) Gaussian noise. The corresponding data model is given in Section 3.3.2. We will show that the independence assumption in this model remains approximately valid, under the assumption that the local standard deviation, σ_i^2, changes only slightly over space.

The orthonormal filterbank is applied to the pairwise difference between activation and rest states $I_{D_\tau}(x, y) = I_{A_\tau}(x, y) - I_{B_\tau}(x, y)$. The result is a data set, which is reduced in the x- and y-coordinate according to the down-sampling factor.

Subsequently, a statistical testing is applied to the reduced images. A hypothesis test is performed as described in Section 3.3.5. The reduction of the image in the down-sampling process is of advantage for our multiple correction method, the Bonferroni-correction, since this approach is less conservative for fewer pixels. The simplified statistical analysis also makes the whole method more transparent.

For visualizing the results, the detected pixels are superposed to the mean fMRI-slice in the background. The detected pixels have a meaningful p-value, marked in gray scales (black large p-value, white small p-value). It is also customary in the field, to use colored maps (yellow large p-value, red small p-value).

6.2   Orthonormal filterbank

The downsampled filter is shown in Figure 6.1; $h_m[k]$ is symmetric, orthonormal, and the downsampling factor is m.

This sequence of operations can be expressed as a discrete inner product

$$
\begin{aligned}
y[l] \quad &= \quad \sum_k x[k] h_m[ml - k] \\
&:= \quad < x[.], h_m[ml - .] > .
\end{aligned}
\tag{6.1}
$$

The definition of orthogonality is:

$$
< h_m[.], h_m[. - ml] >= \delta_l \quad \Leftrightarrow \quad \sum_{l=0}^{m-1} |H_m(e^{j\omega - j\frac{2\pi l}{m}})|^2 = m,
$$

where $H_m(z) = \sum_k h_m[k] z^{-k}$ is the z-transform of h_m.

If the object to filter is of dimension p, we build the corresponding p-D filter by tensor product.

For our application, the requirements on the filters are:

i) symmetry: needed in order to avoid phase distortion when localizing activated voxels.

ii) approximation order λ (which implies the reproduction of polynomials of degree $\leq n = \lambda - 1$). This property is beneficial for a better approximation of signals with most energy in the low frequency part.

iii) all possible integer decimation factors.

iv) consistent family with respect to chaining (for the simplicity of the design $\rightarrow$ modular library of transforms):
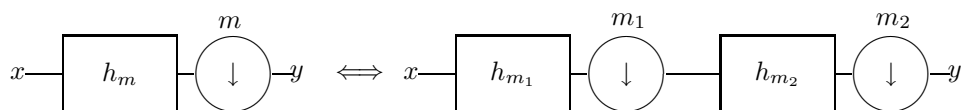

Figure 6.1: Signal flow diagram, where $m = m_1 m_2$.

We will use the orthonormal fractional spline filters of degree α, which are given by

$$H_m(e^{j\omega}) = \left| \frac{\sin(m\frac{\omega}{2})}{m \sin(\frac{\omega}{2})} \right|^{\alpha+1} \sqrt{\frac{A(e^{j\omega})}{A(e^{jm\omega})}} \tag{6.2}$$

$$A(e^{j\omega}) = \sum_{k \in \mathbb{Z}} \left| \frac{\sin(\frac{\omega}{2})}{(\frac{\omega}{2} + k\pi)} \right|^{2(\alpha+1)} \tag{6.3}$$

Proposition 6.1 *The spline family satisfies the requirements i)—iv).*

Proof. ii) This property is valid, where $\lambda = \lfloor \alpha + 1 \rfloor$ in (6.2) because splines satisfy the Strang-Fix conditions [134].
iv) We start from the right-hand signal flow diagram in Figure 6.1. In the first step, we exchange the decimation factor m_1 with the filter h_{m_2} and apply a *noble identity* [148] to get the system in Figure 6.2. The equivalent filter $G(z)$ is:
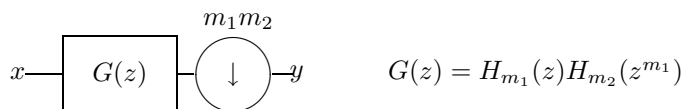

$$G(z) = H_{m_1}(z) H_{m_2}(z^{m_1})$$

Figure 6.2: Signal flow diagram after the exchange.

$$
\begin{aligned}
G(e^{j\omega}) &= \left| \frac{\sin(m_1 \frac{\omega}{2})}{m_1 \sin(\frac{\omega}{2})} \right|^{\alpha+1} \sqrt{\frac{A(e^{j\omega})}{A(e^{jm_1\omega})}} \left| \frac{\sin(m_1 m_2 \frac{\omega}{2})}{m_2 \sin(m_1 \frac{\omega}{2})} \right|^{\alpha+1} \sqrt{\frac{A(e^{jm_1\omega})}{A(e^{jm_1 m_2\omega})}} \\
&= \left| \frac{\sin(m_1 \frac{\omega}{2})}{m_1 \sin(\frac{\omega}{2})} \frac{\sin(m_1 m_2 \frac{\omega}{2})}{m_2 \sin(m_1 \frac{\omega}{2})} \right|^{\alpha+1} \sqrt{\frac{A(e^{j\omega})}{A(e^{jm_1\omega})} \frac{A(e^{jm_1\omega})}{A(e^{jm_1 m_2\omega})}}.
\end{aligned}
\tag{6.4}
$$

After cancellation of common factors, we get:

$$
\begin{aligned}
G(e^{j\omega}) &= \left| \frac{\sin(m_1 m_2 \frac{\omega}{2})}{m_1 m_2 \sin(\frac{\omega}{2})} \right|^{\alpha+1} \sqrt{\frac{A(e^{j\omega})}{A(e^{jm_1 m_2\omega})}} \\
&= H_{m_1 \cdot m_2}(e^{j\omega}).
\end{aligned}
\tag{6.5}
$$

□

The proof that an orthogonal filter preserves the white noise is given in Section 6.4. We also show that the "model 2" given in Section 3.3.2 remains approximately valid: If the variance varies only slightly over the size of the filter, then the noise is still reasonably independent.

6.3   Comparison of Gaussian versus orthogonal filter

Given an orthogonal filter with a decimation factor of m, we want to determine an "equivalent" Gaussian filter which has exactly the same noise reduction power. The notations are as follows:

i) $x[.]$: input signal, white noise with $\sigma^2 = 1$ and $\mu = 0$ ($\mathcal{N}(0,1)$).

ii) $g[.]$: filter coefficients.

iii) $y[.]$: output signal after filtering.

In general: $y[n] = \sum_k g[n-k]x[k]$ with $\sum_n g[n] = 1$. The variance is:

$$
\begin{aligned}
E\left[y^2[n]\right] &= \sum_n \sum_k g[k]x[n-k] \sum_l g[l]x[n-l] \\
&= \sum_k \sum_l g[k]g[l] \sum_n x[n-k]x[n-l] \\
&= \sum_k \sum_l g[k]g[l]\delta[k-l] \\
&= \sum_k g^2[k]
\end{aligned}
\tag{6.6}
$$

- orthonormal filter:

$$\sum_k g[nM - k]g[n'M - k] = \frac{1}{M}\delta[n - n'] \qquad (6.7)$$

$$\Rightarrow \sum_k g^2[k] = \frac{1}{M} \qquad (6.8)$$

Equation (6.8) follows from equation (6.7) because of the Strang-Fix condition of order 1.

- Gaussian filter is normalized such that: $\sum_n g[n] = 1$ with $g[n] = Ce^{\frac{-n^2}{2\sigma^2}}$. The constant C is determined as follows:

$$\begin{aligned}
\sum_n e^{\frac{-n^2}{2\sigma^2}} &\simeq \int e^{\frac{-x^2}{2\sigma^2}}\, dx \\
&\cong \sigma\sqrt{2\pi} \qquad (6.9)
\end{aligned}$$

$$\Rightarrow C \simeq \frac{1}{\sigma\sqrt{2\pi}}$$

$$\sum_n g^2[n] = C^2 \sum_n e^{\frac{-n^2}{\sigma^2}} \simeq C^2 \int e^{\frac{-x^2}{\sigma^2}}\, dx = C^2\sigma\sqrt{\pi} = \frac{1}{2\sigma\sqrt{\pi}}$$

- identification of the energies: $\sigma = \frac{M}{2\sqrt{\pi}}$

$M = 2 : \sigma = 0.564$
$M = 4 : \sigma = 1.128$
$M = 8 : \sigma = 2.257$

SPM recommends a Full Width of Half Maximum (FWHM) between 2 and 3. Since FWHM $= 2\sigma\sqrt{2\log 2}$ and $M = \frac{\sqrt{\pi}}{\sqrt{2log(2)}}$FWHM, the range for the scaling factor is $3.0 \leq M \leq 4.5$ and we obtain for $M = 4 :$ FWHM $= 2.66$.

6.4 Decorrelation property

We prove here that an orthogonal filter transforms decorrelated noise into noise that is essentially decorrelated as well. In particular, this means that it preserves white noise.

Theorem 6.4.1 *Assume that the signal $x[n]$ can be expressed as the sum of a deterministic part, $\mu[n]$, and a non-stationary uncorrelated part $b[n]$. Let $y[n]$ be the filtered version of $x[n]$, i.e., $y[n] = h[n] * x[n]$, where $h[n]$ is orthonormal. If the fluctuations of the standard deviation of $b[n]$ are sufficiently small, then $y[n] = h[n] * \mu[n] + z[n]$, where $E(z[n]z[n']) \cong \sigma^2 \delta[n - n']$.*

Proof.

input signal: $x[n] = \mu[n] + \sigma[n]e[n]$, where $E\left[e[n]\right] = 0$ and $E\left[e[n]e[k]\right] = \delta[n - k]$

output signal: $y[n] = \sum_k h[nM - k]x[k]$

$$y[n] = \underbrace{\sum_k h[nM - k]\mu[k]}_{\hat{y}[n]} + \sum_k h[nM - k]\sigma[k]e[k]$$

$$z[n] = y[n] - \hat{y}[n]$$

$$E[z[n]] = 0$$

$$
\begin{aligned}
E\left[z[n]z[n']\right] &= E\left[\sum_k h[nM - k]\sigma[k]e[k] \sum_{k'} h[n'M - k']\sigma[k']e[k']\right] \\
&= \sum_k \sum_{k'} h[nM - k]h[n'M - k']\sigma[k]\sigma[k'] \underbrace{E\left[e[k]e[k']\right]}_{\delta[k - k']} \\
&= \sum_k h[nM - k]h[n'M - k]\sigma^2[k]
\end{aligned}
$$

First case: uniform variance: $\sigma[k] = \sigma$:

$$E\left[z[n]z[n']\right] = \sigma^2 \underbrace{\sum_k h[nM - k]h[n'M - k]}_{\frac{1}{M}\delta[n - n'], \text{ if } h[n] \text{ is orthonormal}}$$

Second case: close to uniform variance: $\sigma[k] = \sigma(1 + \varepsilon[k])$, $\varepsilon[k] \ll 1$

$$
\begin{aligned}
E\left[z[n]z[n']\right] &= \sum_k h[nM - k]h[n'M - k]\sigma^2(1 + \varepsilon[k])^2 \\
&= \sigma^2 \delta[n - n'] + \sigma^2 \underbrace{\sum_k h[nM - k]h[n'M - k](2\varepsilon[k] + \varepsilon[k]^2)}_{r}
\end{aligned}
$$

We bound the second term as follows:

$$
\begin{aligned}
r &\leq \sum_k |h[nM - k]h[n'M - k]|(\sup 2|\varepsilon[k]| + \sup \varepsilon[k]^2) \\
&\leq \underbrace{\sqrt{\sum_k h[nM - k]^2 \sum_k h[n'M - k]^2}}_{=\frac{1}{M}}(\sup 2|\varepsilon[k]| + \sup \varepsilon[k]^2) \quad \text{(Cauchy-Schwartz)} \\
&\leq \frac{1}{M}\left(\sup 2|\varepsilon[k]| + \sup \varepsilon[k]^2\right).
\end{aligned}
$$

So that

$$
E\left[z[n]z[n']\right] = \sigma^2(\delta[n - n'] + \rho[n, n']),
$$

where

$$
|\rho[n, n']| \leq \frac{2}{M}\sup|\varepsilon[k]| + \sup \varepsilon[k]^2 \ll 1.
$$

Thus, in both cases, $z[n]$ remains essentially uncorrelated. $\square$

This theorem implies that the noise remains uncorrelated in every homogeneous zones ($\sigma = $ constant). However, the correlation property is not valid anymore on the border of zones, like the border of the brain against the background, over a band that has the width of the "efficient" length of the filter G.

6.5 Results

To illustrate the technique, we present some examples of real data analysis (cf. Figures 6.3 to 6.4). The data set is obtained from HUG and has the specification described in Table 7.4. The task was watching a moving checker board of colors. After orthogonal filtering and applying the Bonferroni-correction, we detect activity in the occipital lobe (cf. Section 2.9). We also applied SPM [156] with a σ corresponding to the smoothing factor of the orthogonal filter (Figure 6.3). The difference between our detection and SPM is not negligible. The activation of the right hemisphere of the brain seems to be comparable. However, in the left hemisphere, we detect an additional activation on the border (cf. Figure 6.4). Applying a Gaussian filter of the same width as in SPM and performing a t-test with Bonferroni-correction, we could confirm this detection. It is unclear why SPM did not detect this pattern.

6.6 Conclusion

We have proposed a first approach that simplifies the statistical analysis of the data, as compared to SPM. The idea is to continue "orthogonal" smoothing with down-sampling
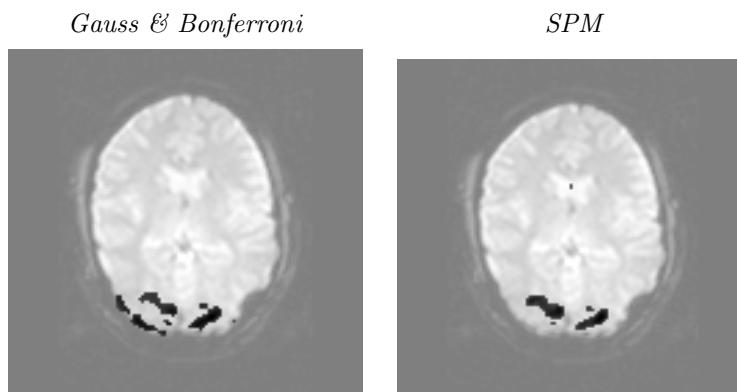
Gauss & Bonferroni   *SPM*

Figure 6.3: On the left hand side, the image is filtered with a Gaussian filter with $\sigma = 1.128$. Subsequently a t-test is performed with Bonferroni-correction. The significance level is $\alpha = 0.05$. Since the data are correlated, the Bonferroni-correction is too conservative here. On the right-hand side, we applied SPM, with a Gaussian filter of $\sigma = 1.128$, i.e., a FWHM corresponding to our scaling factor of 4. The detected pixels have different gray-scales, because they are represented here as p-values.
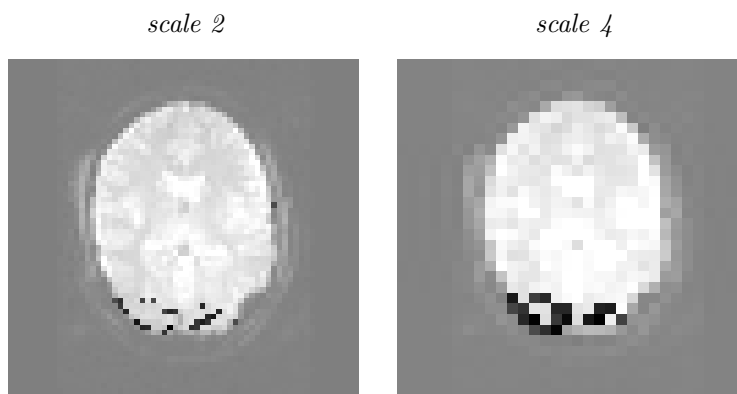
scale 2   *scale 4*

Figure 6.4: On the left hand side, the image is obtained by filtering and downsampling with the orthogonal spline filter with scale 2 and order 3. The significance level is $\alpha = 0.05$ (Bonferroni-corrected). On the right hand side, we filtered and downsampled with scale 4 (same order).

which preserves the independence of the noise in the transformed domain. Additional advantages of our decimated orthogonal filterbank approach are speed and straightforward control of resolution. Here, we use decimation as an alternative to SPM's not-so-straightforward notion of RESEL.

However, the disadvantage of both methods is the loss of resolution. In our final algorithm we eliminate this problem by keeping the high-pass of the filterbank which amounts to using wavelets (see Chapter 7).

Chapter 7

Analysis of fMRI Data using Wavelets

7.1 Introduction

The use of the wavelet transform for the analysis of functional Magnetic Resonance Imaging (fMRI) of the brain was pioneered by Ruttimann *et al.* [120]. The advantages of this method over the more conventional spatial approaches are twofold. First, the wavelet transform has a compression effect because it tends to pack the signal of interest into a relatively few number of large coefficients. On the other hand, the noise gets diluted evenly in the wavelet domain (provided the wavelet is close to orthogonal). This results in an increase of the signal-to-noise ratio for those coefficients where the signal is concentrated and subsequently improves the detection rate. Moreover, the wavelet transform is a one-to-one representation without any redundancies, which permits to apply a conservative decision strategy to keep the false-detection rate well under check (Bonferroni correction for multiple testing). This also means that the procedure remains valid—albeit sub-optimal—even when the basic hypothesis of noise independence breaks down.

The method of Ruttimann *et al.* was restricted to stationary noise. We propose here to relax this constraint by replacing their z-score calculation in the wavelet domain by a series of t-tests. This algorithm is presented in [58].

A key point when applying the wavelet transform is the appropriate choice of the wavelet, the order of the transform and the iteration depth. Different basis functions offer different compromises. The iteration depth controls the accuracy and the sensitivity. The order, depending on the basis functions, has a great influence on the number of detected coefficients and consequently on the sensitivity and specificity. Ruttimann made use of the Battle-Lemarié wavelets. Brammer, in a more recent study [22], applied Daubechies

wavelets. Both studies didn't justify their choice of basis functions with respect to fMRI. Moreover, they did not compare different filters nor discuss the effect of varying the transform parameters (order and iteration depth). These choices are usually dealt with in an ad hoc fashion because it is difficult to assess objectively the quality of the detection since the "true" location of the activation is not known a priori.

Our contribution in this area is to provide a test procedure, which assesses the quality objectively [59] and has shown to make the findings coincide with reasonable activation loci. Thanks to this paradigm, we are able to identify the properties of the basis functions that are most beneficial for fMRI analysis. The properties to be considered are:

- orthogonality versus bi-orthogonality

- frequency selectivity vs. spatial localization

- symmetry vs. causality

- FIR vs. IIR

- scale factor: is a dyadic scale progression (powers of two) adequate ?

- directionality: does the separability of the standard wavelet transform privilege horizontal or vertical patterns ?

- does the shift-variance of the wavelet transform degrade the detection accuracy ?

- does redundancy benefit the analysis ?

To explore these various dimensions, we consider a large class of wavelets, the newly-designed fractional spline wavelet transforms (FSWT) [145]. These are indexed by a *continuously-varying* degree parameter α. This offers the potential for a much finer tuning of the transform between frequency selectivity and spatial localization. And as we will see, the possibility of having a fractional degree is advantageous.

To address the question of the scale progression, we compare the standard dyadic transform with the Quincunx-transform presented in Section 5, which offers a scale progression that is half of that of the separable transform. The quincunx transform also offers the possibility of applying more isotropic filters.

To answer the question if the shift-variance lowers the detection rate, we implemented a more shift-invariant transform. Since wavelet frames would require too conservative a statistics, we chose the complex wavelet transform, which is a newly-developed promising transform. In order to adjust the statistical part to the complex wavelet transform, we implemented a multivariate Hotelling's T^2 test.

A summary of our detection results depending on the different wavelets is presented in Table 7.2, at the end of Section 7.3.

7.2 Description of the algorithm

The overview of the whole system is shown in Figure 7.2. The input is a series of time measurements (slices or volumes) falling within the activation (A) or rest (B) states (standard block paradigm); these image sequences are denoted by $I_{A_\tau}(x, y)$ and $I_{B_\tau}(x, y)$ respectively, where τ is a time index. After the preprocessing step (Section 7.2.1), the Discrete Wavelet Transform (DWT) is applied. Statistical testing is then performed in the wavelet domain considering each coefficient individually (Section 7.2.3). We only keep the detected coefficients on the mean difference image. The resulting image is called $I_{est,W}$, on which we apply the Inverse Discrete Wavelet Transform (IDWT). The signal is then reconstructed, yielding I_{est} and superimposed onto to the corresponding slice of the brain for visualization purposes. The signal reconstruction includes a threshold which is proportional to the noise level in the data. Typical slices of a brain are shown in Figure 7.1. These correspond to the examples that will be used to illustrate each step of the algorithm.
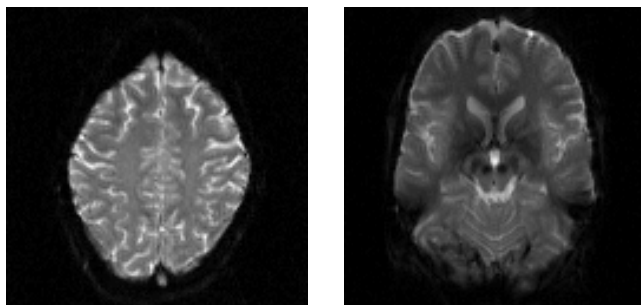

Figure 7.1: Transversal slices taken from the same volume that are used to illustrate the different steps of the algorithm. A light source was switched on (A) and off (B) while the images of the brain of a volunteer looking at the light source were acquired. For such a task, the activation is known to be located in the occipital lobe (cf. Section 2.9).

7.2.1 Preprocessing

During the preprocessing step, the data are aligned with a reference volume using a 6 parameters 3D rigid-body transformation [65, 68] (refer to Section 2.6). Note that the motion correction is usually small with respect to the spatial resolution. An example of realignment for the data set of Figure 7.1 with "Statistical Parameter Mapping" (SPM) (Section 2.8.1) is given in Figure 7.3: the degree of translation and rotation of the data set is given, dependent on the time sequence. In Figure 7.4 the activation of two slices of this data set before and after realignment is depicted. Drifts in time are also removed using the standard procedure SPM99 (further description of SPM in Section 2.8). Then,
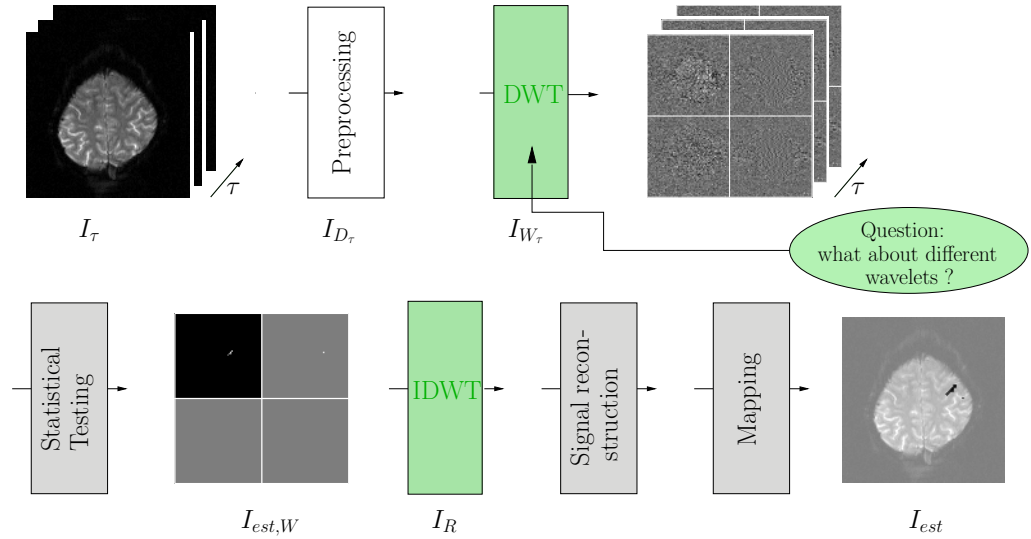
Figure 7.2: Block diagram of the algorithm.

the pairwise difference between activation and rest states are computed to produce the sequence of difference images $I_{D_\tau}(x,y) = I_{A_\tau}(x,y) - I_{B_\tau}(x,y)$.

7.2.2   Discrete wavelet transform

The application of the Discrete Wavelet Transform (DWT) to $I_{D_\tau}(x,y)$ yields the decomposed fMRI-data $I_{W_\tau}(x,y)$, where "W" stands for "wavelet domain", as described below. The DWT is implemented through an iterated filterbank [95] (refer to Chapter 4). Note that, due to the linearity of the transform, the Gaussian hypothesis for the noise is preserved in the wavelet domain. The advantage of a wavelet transform for analyzing fMRI data is that not only the low-pass contents is processed for the detection of the activation, but also the band-pass components which contribute to increase the resolution of the detected regions. In our test procedure, the DWT can be seen as a module where different wavelets are considered for evaluating their advantages to detect activation patterns. We have chosen the fractional splines, which provide a large variety of filters and Daubechies wavelets as real separable wavelets. We also considered nonseparable quincunx wavelets and finally the Complex Daubechies Wavelets. These different wavelets are defined through their order, which we denote by $\alpha + 1$ for fractional splines, where α is the degree, λ for quincunx and by their order of approximation n for Daubechies. Note that for fractional splines the order of approximation is $\alpha + 1$. As further parameter, we consider the iteration depth J of the wavelet transform.

$\rightarrow$ **Fractional splines.** Even though, in full generality, the synthesis and analysis
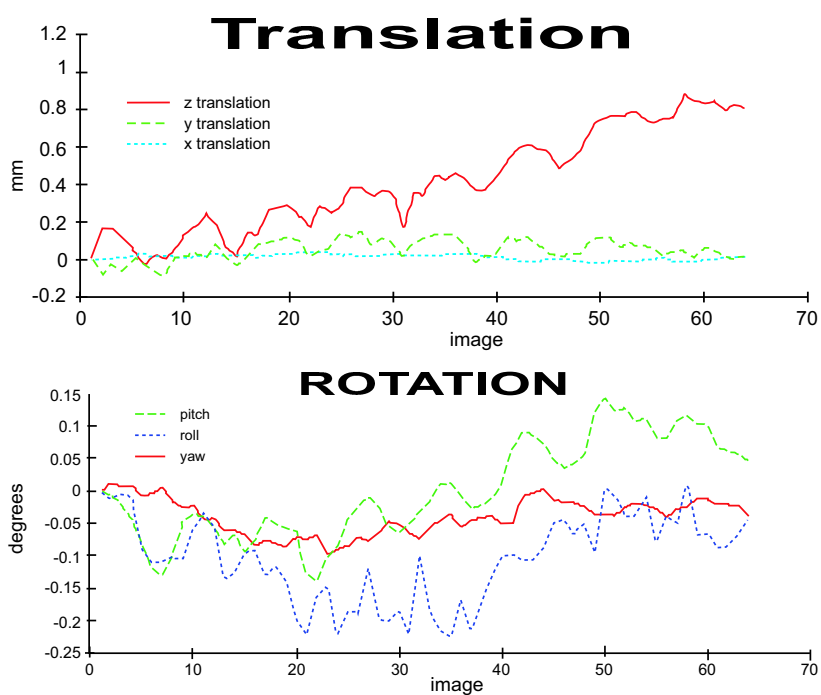
Figure 7.3: Illustration of a realignment example of a temporal sequence of 64 images with SPM. The translation and rotation parameters with respect to the first image.

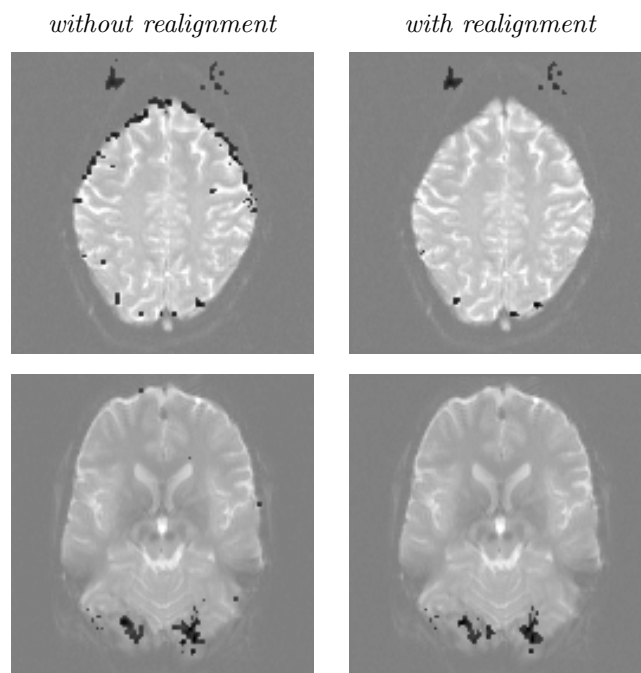*without realignment*   *with realignment*

Figure 7.4: Example of detected regions (I_{est}) before realignment and after realignment of the data set described in Figure 7.1. The corresponding realignment parameter are given in Figure 7.3. The activation pattern outside of the brain in the upper images originates from a ghosting-artifact, see references in Section 3.3.1.

filters are bi-orthonormal, they are mostly chosen orthonormal in the literature, i.e., $\tilde{H}(z) = H(z^{-1})$ and $\tilde{G}(z) = G(z^{-1})$; this is in particular the case for the Daubechies filters [43]. For such filters, which are automatically half-band, there remains only one tradeoff between the sharpness of the frequency cutoff and the spatial localization. Because we want to optimize our detection scheme with respect to orthonormal and bi-orthonormal transforms, we use as first choice the new class of scaling functions and wavelets that provide this flexibility: the fractional splines [145] (Section 4.2.1).

$\rightarrow$ **Complex wavelets.** The discrete real wavelet transform is quite sensitive to a shift of the input signal due to down-sampling. This effect is significantly reduced when considering the coefficient magnitudes of the Complex Wavelet Transform (CWT). An additional advantage of the CWT for 2 dimensions is that it offers six angle-selective subbands rather than three in the case of real DWT. Furthermore, the computational cost of the CWT is the same as two real DWT [118, 29].

$\rightarrow$ **Quincunx wavelets.** The use of such wavelets is interesting because of some advantages it offers over separable wavelets (e.g. fractional splines and Daubechies wavelets):

i) The scale reduction of the separable scheme progresses by powers of 2. Quincunx has a finer scale progression—by powers of $\sqrt{2}$—and can thus detect more accurately very fine/capillary activation patterns or can better provide the adequate iteration depth for noise reduction.

ii) The separable wavelets have a strong directionality. Activation patterns which are expanded in horizontal or vertical directions are better detected than diagonal expanded patterns. Quincunx has the advantage of providing an analysis that is more isotropic.

7.2.3 Activation detection by statistical testing on the wavelet coefficients

Because of the noise present in the data, it is almost impossible to detect any activation without resorting to statistical analysis. This analysis requires a model of noise [57]; here, we assume that $I_{D_\tau}(x, y)$ follows a spatial Gaussian distribution. Thanks to the wavelet transform, simple statistics can be applied, which makes the analysis fast and straightforward. Thus, a statistical testing on each coefficient of $I_{W_\tau}(x, y)$ will be applied to determine whether it is activated or not (refer to Section 3.3.5). In the case of real wavelets, a Student t-test is applied as described in Section 3.3.3. At this stage, all non-significant wavelet coefficients are set to zero, keeping only those that have been labeled as "activated". This yields the W-estimated activation $I_{est,W}(x, y)$ (WEA). An example of $I_{est,W}(x, y)$ for the data of Figure 7.1 is presented in Figure 7.5. In the

case of complex wavelets, Hotelling's T^2 test has to be applied (see Section 3.3.7), since the complex wavelets turn the statistics into a bivariate statistic, taking the real part and the complex part of the wavelet coefficients into account. The resulting W-estimated activation WEA is a complex valued image or volume: $I_{\text{est,W}}(x,y) \in \mathbb{C}$. The new method of testing with Hotelling's T^2 opens up new possibilities for testing multiple coefficients in a multivariate sense, also in the case of real wavelets. Several coefficients can be grouped together, such as neighboring coefficients or related coefficients from different band-pass channels.
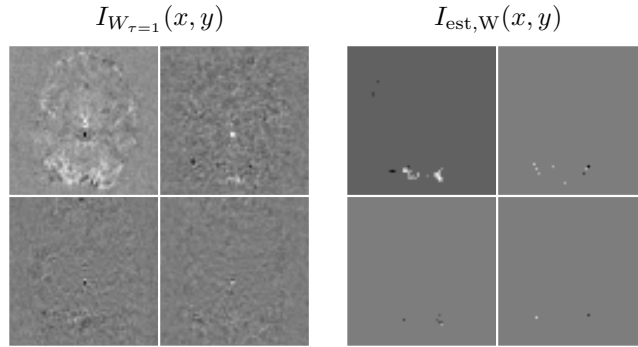
$$I_{W_{\tau=1}}(x,y) \qquad\qquad I_{\text{est,W}}(x,y)$$

Figure 7.5: Example of intermediate results, before applying the statistics $(I_{W_{\tau=1}}(x,y))$ and after the statistics $(I_{\text{est,W}})$. The slices are taken from the same data set than the one of Figure 7.4.

7.2.4   Signal reconstruction

Once the wavelet coefficients of $I_{\text{W}}(x,y)$ have been flagged either as significantly activated or not, the estimation of the activation pattern is achieved by zeroing out all non significant samples and performing an inverse wavelet transform on $I_{\text{est,W}}(x,y)$. Since the generality of the developed algorithm allows also synthesis filters which are of infinite length, a decision has to be made as to what constitutes a valid activation pattern. Specifically, a threshold is applied to the reconstructed activation pattern that takes into account the amount of noise: when this noise is very low, the quantization level of the initial data is simply chosen; if the noise in every realization of I_{D} has a non-negligible standard deviation σ, the threshold level is set at $level = \sigma/\sqrt{n}$; i.e., the noise level of the average image $I_{\text{D}} = \frac{1}{n}\sum_{\tau} I_{\text{D}_{\tau}}$. Without this threshold, the best wavelets seem to be the shortest one. With threshold, the detection quality gain increases strongly and longer wavelets yield good results.

We provide a test procedure to validate this value, which is presented in Figure 7.6: Given a data set with an activation pattern (real data or synthetic data) and a noise level of σ, detection is performed, which yields the estimated activation pattern in the wavelet

domain $I_{est,W}(x,y)$ and the estimated activation pattern $I_{est}(x,y)$ after applying the threshold *level* to the output of the synthesis part. A data set is then reconstructed by adding a noise of level σ ($\mathcal{N}(0,\sigma)$) to $I_{est}(x,y)$. Detection is performed again yielding an estimated activation pattern $\tilde{I}_{est,W}(x,y)$. When the value of *level* is chosen adequately, $I_{est,W}(x,y)$ and $\tilde{I}_{est,W}(x,y)$ should match. A distance map to measure the matching error between $I_{est,W}(x,y)$ and $\tilde{I}_{est,W}(x,y)$ is evaluated :

$$d_{est,W} = \begin{cases} 0, & (x,y) \text{ is in the same activation state in } I_{est,W} \text{ and in } \tilde{I}_{est,W} \\ 1, & \text{otherwise} \end{cases}$$

This process is repeated r-times by adding each time a different realization of noise to $I_{est}(x,y)$. The final measure is given by:

$$d(I_{est,W}, \tilde{I}_{est,W}) = \sum_{x,y} \frac{1}{r} \sum_{i=1}^{r} d_{est,W}^{(i)}(x,y) \tag{7.1}$$
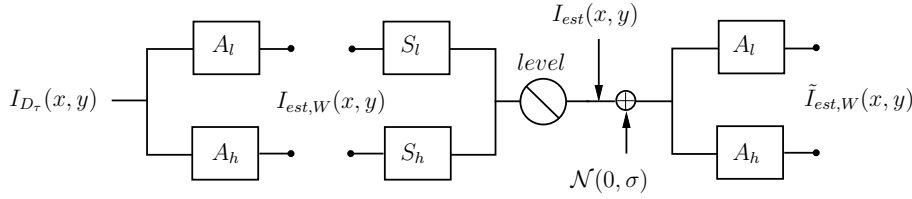

Figure 7.6: Scheme of a test procedure to ensure the consistency of the threshold *level*. A_l and A_h represent the analysis part of the wavelet transform and the statistical testing of the wavelet coefficients. The output of these branches is $I_{est,W}(x,y)$. S_l and S_h represent the synthesis part of the wavelet transform, which is applied to $I_{est,W}(x,y)$. The output of the synthesis part is then thresholded above "*level*" which yields $I_{est}(x,y)$. Adding noise with variance σ^2 to $I_{est}(x,y)$ creates a new data set. The thresholding procedure is consistent whenever, on the average over all possible realizations of the resynthesized noise, $I_{est,W}(x,y)$ and $\tilde{I}_{est,W}(x,y)$ match. In practice, we average over 130 realizations.

7.3 Which properties of the wavelet transform are preferable for fMRI?

In order to investigate the influence of the structure parameters of the algorithm —its (bi-) orthogonality, symmetry / causality, its order of approximation, its iteration depth J etc.— an error measure must first be defined between the true activation and that which
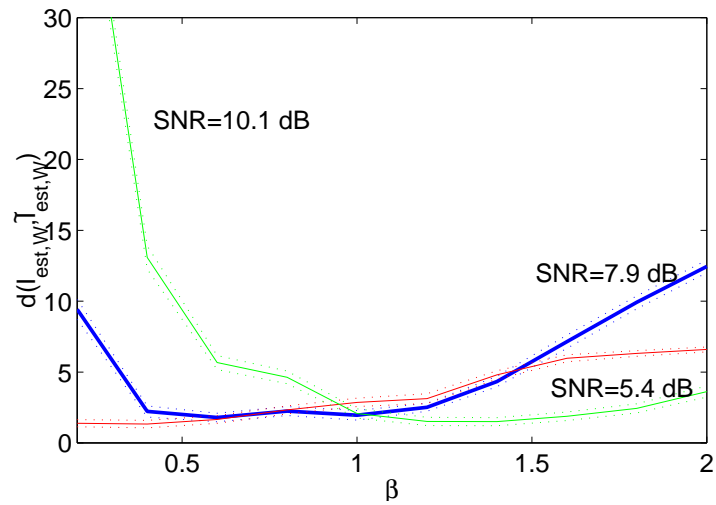
Figure 7.7: Example of $d(I_{est,W}, \tilde{I}_{est,W})$ of 130 realizations for a test image I_{est}, where 10 activation patterns of size 9 to 14 pixels are scattered in the image. Each SNR is given for $\bar{I}_D(x,y)$. The dotted lines represent the confidence interval for $d(I_{est,W}, \tilde{I}_{est,W})$ of 95 %. The threshold *level* is normalized to $level = \beta \frac{\sigma}{\sqrt{n}}$. We observe that $d(I_{est,W}, \tilde{I}_{est,W})$ is flat around its minimum, over a wide range of values of β. Moreover, $\beta = 1$ usually achieves or is close to achieving this minimum, which indicates that our choice for *level*, $\frac{\sigma}{\sqrt{n}}$, makes our algorithm nearly consistent.

is detected with the proposed method (see Section 7.3.2). Our strategy is to generate synthetic data (see Section 7.3.1) for which the activation is known a priori. By processing such data, the detection of the activation pattern can be evaluated quantitatively. Such an error measure allows to compare different structure parameters, especially the different wavelets, in an efficient way. Not only does it give information on the performance of the different wavelets, but it also allows to identify the most significant structure parameters; i.e. those that are likely to influence the quality of the detection.

7.3.1 Test data

Various test data $I(x, y)$ were used, with known activated area A, defined as $\{(x, y)|I(x, y) \neq 0\}$. The complementary region A^c is defined by $\{(x, y)|I(x, y) = 0\}$. As primary test pattern, we used thresholded ellipse-shaped Gaussian activation patterns, which are suitable for testing the possible lack of rotation invariance of our detection schemes. The test-data were generated by adding $n = 20$ realizations of white Gaussian noise. We changed the scale and ellipticity of the pattern, its orientation and position, and the variance of the noise. An example of such an activation pattern is presented in Figure 7.8. In addition to elliptic shapes, we also tested activations with more high frequency contents.
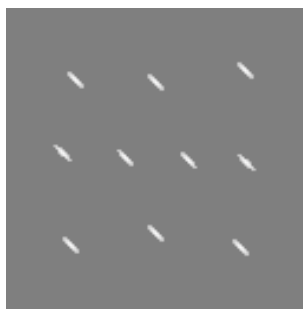
Figure 7.8: Example of test activation pattern. 10 ellipses are randomly placed in an image which contains 128×128 pixels. The intensity of the ellipses follows a Gaussian distribution and is thresholded at half-maximum amplitude. The image is quantized to 8 bit. The size of the ellipses are chosen in the range of possible real activations.

7.3.2 Error measures with prior knowledge of the activation pattern

Two kinds of errors were considered:

- the number of false detections, i.e., errors of type I ($E1$), relative to the number of

activated pixels:

$$E1 = \frac{\text{Card}\big(\{(x,y) : I(x,y) = 0 \text{ and } I_{\text{est}}(x,y) \neq 0\}\big)}{\text{Card}(A)} \qquad (7.2)$$

- the number of missed detections, i.e., errors of type II ($E2$), relative to the number of activated pixels:

$$E2 = \frac{\text{Card}\big(\{(x,y) : I(x,y) \neq 0 \text{ and } I_{\text{est}}(x,y) = 0\}\big)}{\text{Card}(A)}$$

Thus the relation $(1 - E1)$ is the percentage of pixels correctly detected as non-activated, which determines the specificity. In theory, this value can be negative, but in practice, negative values rarely occur. The relation $(1 - E2)$ is the percentage of pixels which were properly detected, and determines the sensitivity. The error types $E1$ and $E2$ are illustrated in Figure 7.9. Our global error criterion to minimize is, by definition, the total error, which is the sum of the two types of errors: $E = E1 + E2$. Note that even though this quantity provides a fair control on the localization of the activation zone, it cannot give clues on the amplitude mismatch. Obviously, the activation pattern needs to be known to compute E. A related measure, first applied to the analysis of
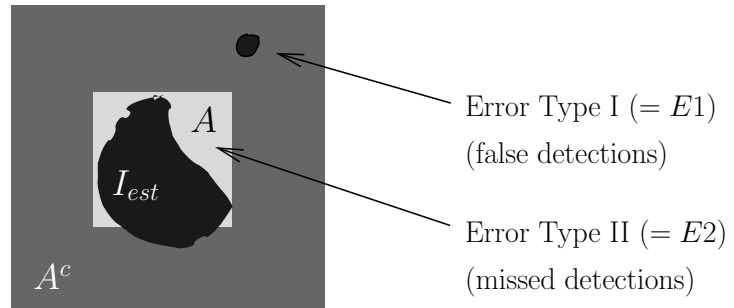

Figure 7.9: Example of a schematic detection of a quadratic pattern. The global error criterion is defined as $E = E1 + E2$.

fMRI processing techniques by [34], is the Receiver Operator Characteristic (ROC). With ROC, the relationship between the true-positive ratio (proportion of correctly detected activations to all true activations) and the false-positive ratio (proportion of pixels that were incorrectly recognized as active in all pixels without true activations) is measured, depending on a parameter that controls the sensitivity of the detection procedure [129]. In our case this parameter would be the threshold in the wavelet t-domain (which we set to the value that ensures less than 5% of false detections). Compared to the measures

$E1$ and $E2$, the true-positive ratio is equal to $(1 - E2)$ and the false-positive ratio is proportional to $E1$. It is the denominator of the definition of these measures that makes the difference: compare (7.2) with:

$$\text{false positive ratio} = E1 \frac{\text{Card}(A)}{\text{Card}(A^c)} \tag{7.3}$$

Thus, the ROC-curve, compared to our measures is:

$$\text{ROC}(\gamma) = \begin{pmatrix} 1 - E2(\gamma) \\ E1(\gamma)\frac{\text{Card}(A)}{\text{Card}(A^c)} \end{pmatrix} \quad \text{where } \gamma \in [0\%, 100\%]. \tag{7.4}$$

We believe that it is more reasonable to express the false detections ($E1$) relative to A, rather than A^c because then: (i) the quantity $E1$ and $E2$ are comparable (ii) $E1$ does not depend on the image size and (iii) a larger activation zone is likely to cause more false activations. In Figure 7.10 such a ROC curve is illustrated in terms of $E1$ and $E2$. Minimizing E amounts to maximizing $1 - E$, i.e., the distance between the ROC-curve and the locus of random detection, which represents $1 - E = 1 - E2 - E1 = 0$. When the curve lies below the locus of random detection, the minimal error is the minimal distance of the ROC-curve to this diagonal. A field where these error measures are intensively
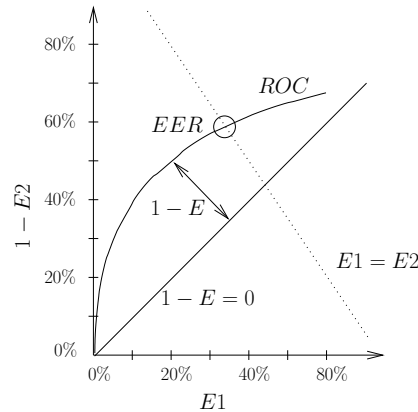
Figure 7.10: Comparison of Receiver Operator Characteristic (ROC), the Equal Error Rate (EER) and our error measure (E).

studied is biometrics, the field of automatic identity verification methods, such as voice, face, eyes or fingerprint identification. They cope with the False Rejection Rate (FRR), which is equal to our $E2$ and with False Acceptance Rate (FAR) which is equal to the false positive ratio. In this field, the Equal Error Rate (EER), given by $E_1 = E_2$ on the ROC-curve, is often used as a performance measure; see the position compared to the

other measures in Figure 7.10. Our global error criterion E is related to their Total Error Rate (TER), the sum of FA and FR, divided by their total number of tested units.

7.3.3   Other measures with unknown activation pattern

We now present three other quantities which, unlike the error measure E defined above, are applied directly on the noisy data. For them, the "true" activation pattern does not have to be known a priori. They are presented as following:

i) Wavelet count C: the number of wavelet coefficients that are detected as activated in our statistical procedure. It is thus defined by $C = \mathrm{Card}\big(I_{\mathrm{est,W}} \neq 0\big)$.

ii) Information measure: the average information content of the detected wavelet coefficients is defined by

$$I = -\frac{1}{\mathrm{Card}\big(I_{\mathrm{est,W}}\big)} \sum_{(x,y)\in I_{\mathrm{est,W}}} \log_2 p\big(t(x,y)\big),$$

where $t(x,y)$ is defined by (3.28) and $p(t)$ is Student's distribution density.

iii) Kullback-Leibler information: the map $t(x,y)$ follows the experimental density (histogram) $p_{\mathrm{data}}(t)$, while the statistical model—in the absence of activation—is described by Student's t-distribution, $p(t)$. The similarity between these densities is computed using Kullback-Leibler information:

$$I(p_{\mathrm{data}}, p) = \int_{-\infty}^{\infty} \log_2 \left\{ \frac{p_{\mathrm{data}}(t)}{p(t)} \right\} p_{\mathrm{data}}(t)\, dt \tag{7.5}$$

Even though this is not a true distance between probability densities, the Kullback-Leibler information has the important property that $I(p_1, p_2)$ is always strictly positive, except when $p_1 = p_2$, where $I(p_1, p_2)$ vanishes.

The drawback of the information and the Kullback-Leibler measure is the small and mostly non-significant variation when an activation is present in the data. This is because activation is only marginally perturbating these global information measures, and this drawback gets even more prominent as the size of the activation is smaller.

7.3.4   Fractional spline wavelets

With fractional splines, we have a variety of different filters which can be applied to the data. Orthogonality versus bi-orthogonality; symmetry versus causality; frequency selectivity versus spatial localization can be explored. In the bi-orthogonal case, strong frequency selectivity on the analysis side versus strong frequency selectivity on the synthesis side can be tested.

Here the three types of analysis-synthesis systems for the Fractional Spline Wavelet Transform (FSWT) are recalled (further description of FSWT in 4.2.1), whose filters are exemplified in Figure 7.11. Its main characteristics and typical performance with different noise levels is described in the following section:

- the `dual` type: it is characterized on the analysis side by a strongly selective scaling filter which ensures a higher pixel-wise signal-to-noise ratio in the wavelet domain. This generally leads to a high wavelet count C and thus fewer missed detections in the low-pass. The filter is all the more selective as α is large. However, on the reconstruction side, the badly localized scaling filter (which even worsens when α increases) is responsible for a higher number of false detections. Thus a compromise on α must be found. Because it is more robust to noise, this type is expected to be the best performer when the level of noise is high.

- the `bspline` type: it is characterized by a strongly localized synthesis scaling filter, which ensures fewer false detections. This behavior is enhanced as α increases. However, on the analysis side, the badly selective scaling filter (which gets even more prominent as α increases) is responsible for a bad signal-to-noise ratio, and thus misses more detections. Because it has a better localization property and because it is more sensitive to noise, this type is expected to be the best performer when the level of noise is low.

- the `ortho` type is characterized by the same space-frequency behavior on the analysis and synthesis sides. By adjusting α, the frequency selectivity and the spatial localization of the filter can be controlled as summarized below:

$$
\alpha \left\{ \begin{array}{l} \text{large } \alpha \left\{ \begin{array}{l} \text{good frequency selectivity} \\ \text{bad spatial localization} \end{array} \right. \\ \text{small } \alpha \left\{ \begin{array}{l} \text{bad frequency selectivity} \\ \text{good spatial localization.} \end{array} \right. \end{array} \right.
$$

We expect the performance of the `ortho` type of transform to be optimal for intermediary noise levels.

More specifically, the following trends have been found in simulations on test data:

→ **Symmetry.** In most cases, the causal FSWT has fewer errors than its symmetric counterpart. The error E and the number of detected coefficients for the causal fractional splines are depicted in Figure 7.13; the corresponding graphs for the symmetric wavelets are shown in Figure 7.14. Both Figures are similar, the causal splines showing a smaller error E than the symmetric splines. This may be due to a better space localization for the causal splines (the integer degree causal splines are even compactly supported), whereas their symmetric counterparts coincide with the regular integer splines only for odd degrees.
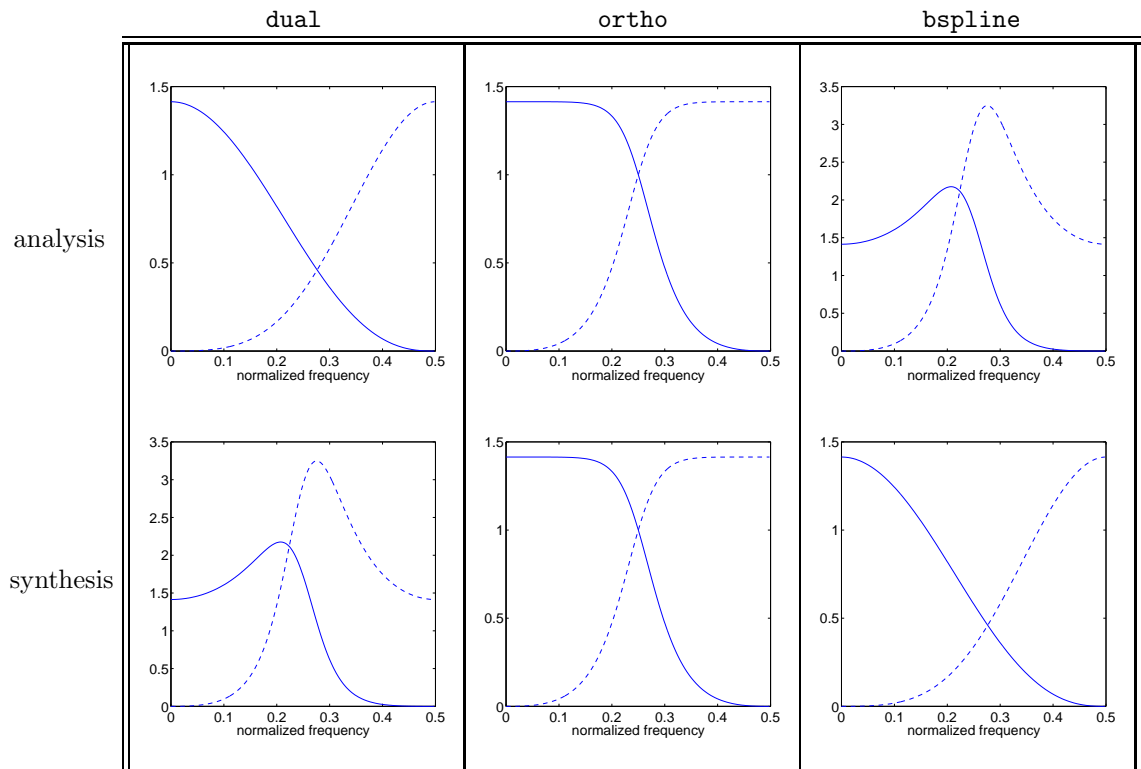
Figure 7.11: Example of frequency response of the low-pass (solid line) and high-pass (dashed line) filters for the three types of FSWT. Note, that the filter for `dual` in the analysis part is the same filter for `bspline` in the synthesis part.
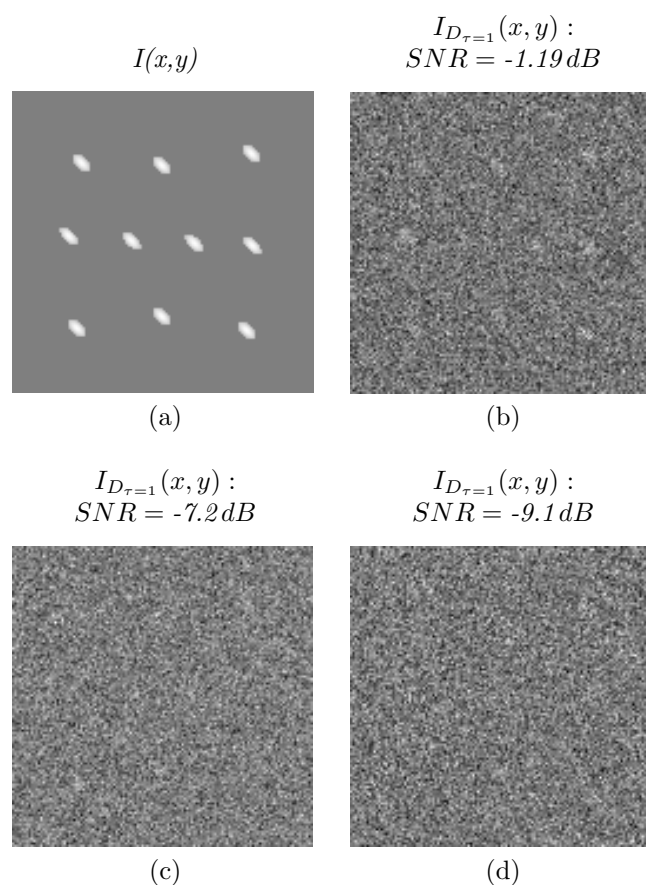
Figure 7.12: Example of test data. Used for the illustrations in this Section. a) Template image $I(x,y)$. b) Noise is added to the template image $I(x,y)$, such that the average image $\bar{I}_D(x,y)$ has an SNR of 11.8 dB. c) The average image over 20 realizations $\bar{I}_D(x,y)$ has an SNR of 5.8 dB. d) The average image $\bar{I}_D(x,y)$ has an SNR of 3.9 dB.

$\rightarrow$ **Type.** The `ortho` and `dual` splines in general outperform the `bspline` type
FSWT, due to less missed detections. In Figure 7.15 the qualitative results of
detected activation patterns are shown for the different types, based on the test
data of Figure 7.12 c). Mostly, the type `dual` detects best. In this example, `ortho`
and `dual` show similar results. In Figure 7.13, the quantitative error is illustrated
for this example of test data. These curves are drawn for one realization of the
noise only; taking into account an infinite number of such realizations is likely to
give a "thickness" to these curves, making them smooth. For smaller patterns the
minimum of `dual` gets sharper and closer to $\alpha = 1.2$. As a matter of fact, the
developed algorithm was always tested with large noise levels, which are typically
observed in practical data. But we surmise that with lower levels, the `bspline`
FSWT might be more competitive because false detections then become a more
important issue than missed detections. To summarize:

$$
\begin{array}{lll}
\text{dual} & \longrightarrow & \text{large } E1, \text{ small } E2 \\
\text{ortho} & \longrightarrow & E1 \sim E2 \\
\text{bspline} & \longrightarrow & \text{small } E1, \text{ large } E2
\end{array}
$$

$\rightarrow$ **Degree.** The optimal spline degree varies widely in the `ortho` type. In the case
of `+dual` FSWT, the optimal α lies between $1.2 - 1.6$, or at least it builds a local
minimum. It is obviously a good compromise between a selective scaling filter at
the analysis side and a localized scaling filter on the synthesis side. Here, having
a fractional order parameter is of great use. In the case of `*dual` the best α varies
slightly more, and $\alpha = 1.6$ is a suitable choice. Comparing the α of `dual` with
the Gaussian filter applied by SPM and its notion of RESELS, we obtained an
interestingly close range for α, see equation (7.21) and (7.22). Thus the `dual` type
optimizes strongly the control of $E2$ errors, whereas the $E1$ error increases when
α exceeds $\alpha = 2$. On the contrary, when α tends to zero (where $\alpha = 0$ represents
the Haar wavelet for `+dual`), the $E1$ error is as low as for the `ortho`, since then the
filters of `+dual` and of `+ortho` get similar. The optimal spline degree for `bspline`
lies between $\alpha = 0.2$ and $\alpha = 0.6$. The degree is low, since the selectivity of the
scaling filter is worse with high α.

$\rightarrow$ **Iteration depth.** Less surprisingly, when the noise is higher, resp. when a smaller
voxel size is chosen to increase the resolution, more iterations are needed to detect
activated pixels in the wavelet domain. Indeed, we observe that, for lower noise
levels the optimal iteration depth is invariably 1, whereas when the variance of the
noise gets large (i.e., for the pattern in Figure 7.12 a) when the SNR $< -5\,$dB for
I_{D_τ}), this optimum is 2. This effect is counterbalanced by the size of the activation:
a smaller size can hardly be detected with more than one iteration. For this reason,
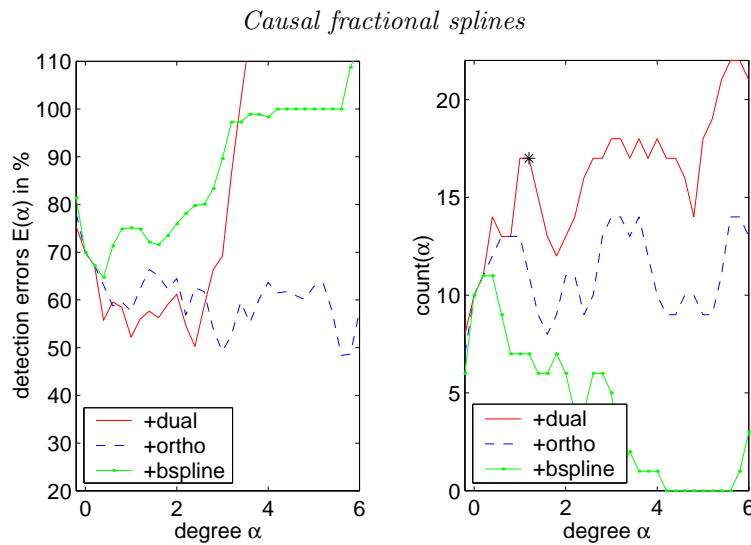values of J_{opt} exceeding 2 were never encountered.

Figure 7.13: Presentation of the errors and the number of detections in the wavelet domain (`count`) corresponding to the activation pattern $I(x, y)$ in Figure 7.12 c) (SNR $= -7.2$ dB). `count` increases with α for the type `+dual`, since the scaling filter gets more low-pass for a high α and has a local maximum for $\alpha = 1.2$ (star). In contrast, for `+bspline`, the number of counts decreases as α increases, since the filter on the analysis side gets less low-pass (and thus filters less the noise) with increasing degree parameter. `+ortho` is a compromise between the `+dual` type and the `+bspline` type and doesn't seem to depend strongly on the degree α. For the bi-orthogonal filters, the error E has a minimum for some α and increases then for higher α's. In the case of `dual`, this increase of E results from false detections, whereas the low `count` in the case of `bspline` yields errors of missed detections.
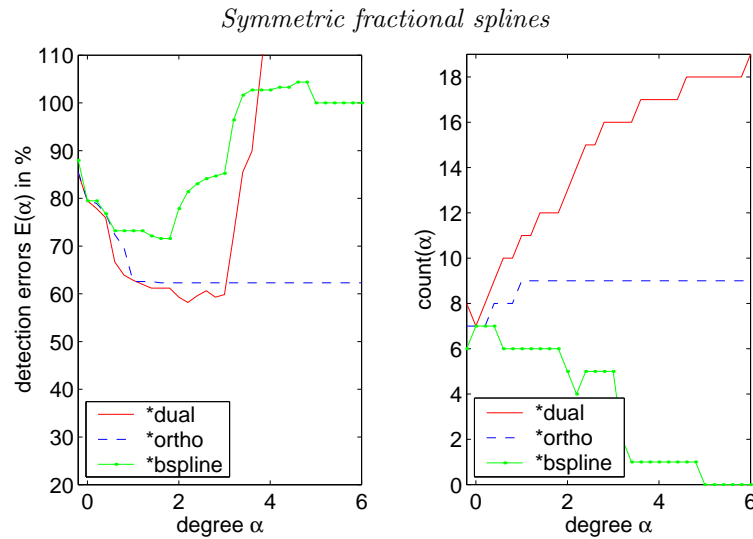
Figure 7.14: Presentation of the errors for the symmetric fractional wavelets based on the data set of Figure 7.12 c). The error E and the number of detections `count` behave very similarly as for causal fractional splines.

$\rightarrow$ **Wavelet count C** As already mentioned, `dual` causes the highest `count` with increasing α and `bspline` the lowest. The number of detections in the high-pass alone is exactly opposed, `bspline` shows more detections than `dual`. This comes from the high frequency behavior of the dual spline (used on the analysis-side of the `bspline` type transform). `+dual` often has a local peak for $\alpha = 1.2$, which is also the case in our example in Figure 7.13 and is marked with a star. This peak coincides with the minimum of the error E.

$\rightarrow$ **Correlation E–C.** For the `bspline` and `ortho` FSWT, the error criterion E is always tightly (inverse) correlated with the number of detected wavelet coefficients, i.e., the wavelet count C. This is extremely useful because, unlike the false+missed detection error count E, the wavelet count does not require any prior knowledge of the activation pattern. This suggest a direct application of the developed FSWT optimization algorithm to real data: instead of minimizing E, a maximization of the wavelet count for the structure parameters could be done. However, this would theoretically require a correction of the threshold in the wavelet domain for multiple testing.

By analyzing the situation in more details, we found that, for all types of FSWT (i.e., including `dual`), the error type II is strongly correlated with C; i.e., the

+bspline   +ortho   +dual

*bspline   *ortho   *dual
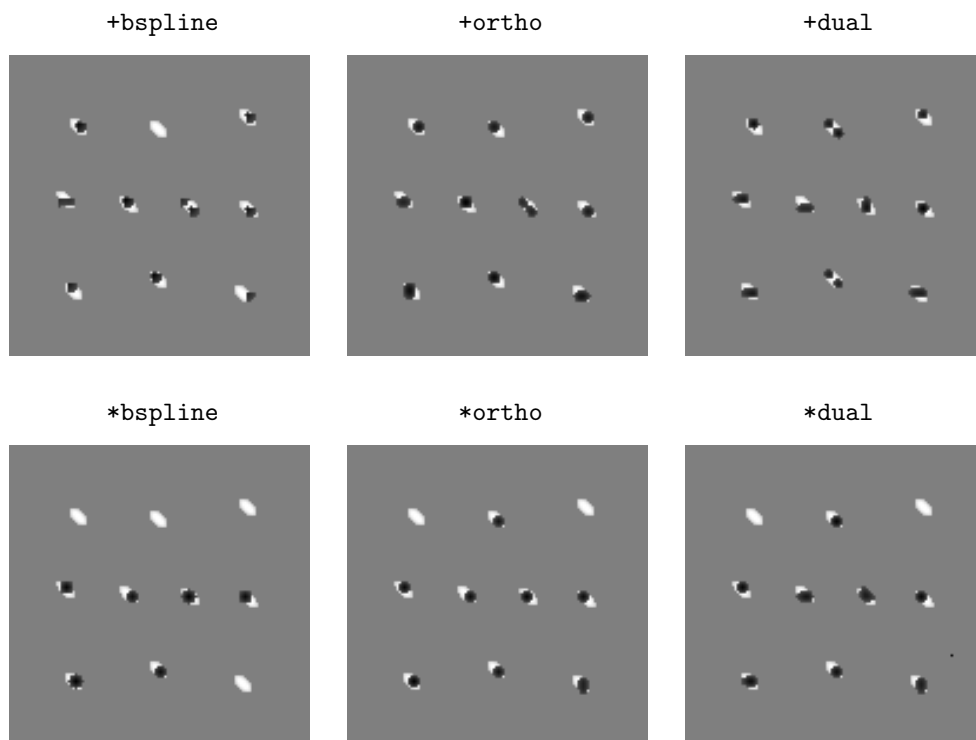
Figure 7.15: These are the detections with the FSWT corresponding to the graphs of Figures (7.13) and (7.14), when α is chosen optimal. The detected pixels are shown in black. The corresponding template image $I(x, y)$ is shown in Figure 7.12 a), the realization of one difference image is given in Figure 7.12 c). We observe that the causal splines perform qualitatively better than the symmetric one: they miss less ellipses.
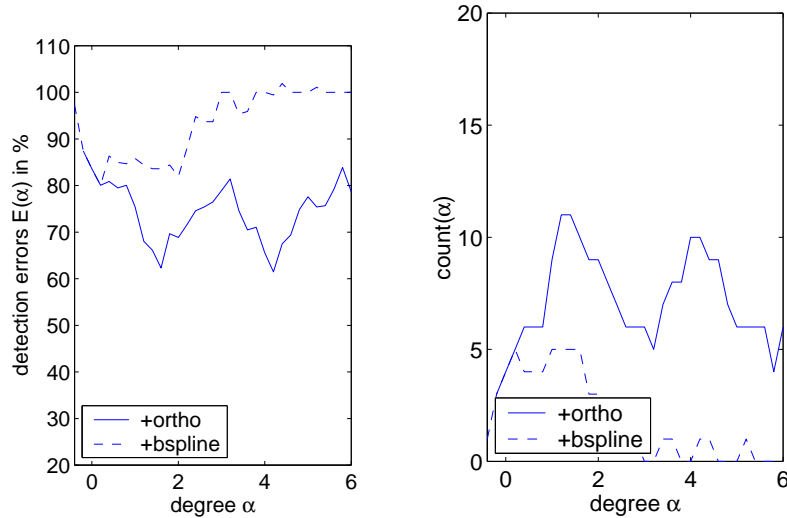
Figure 7.16: Variations of the percentage of detection errors E and the wavelet count C in function of the spline degree α for the causal `ortho` and `bspline` types. The corresponding test data is given in Figure 7.12 d) and, due to the high noise level (-9.1dB for I_{D_τ}), the analysis needs 2 iterations. Note the sharp minima of E, especially for `ortho`, and the inverse correlation between E and C. Their qualitative result for detection is shown in Figure 7.18.

product $C \times E2$ remains approximately constant. As far as the `bspline` and `ortho` types are concerned, it is thus because, the missed detections count is always so much larger than the false detections $E1$ for high noise that we finally have $E \propto 1/C$. The higher the noise, the better the correlation. The relation between E and C is evidenced in Figure 7.16. For this specific case, the optimal value of α is 4.2 in the causal `ortho` type and 0.2 in the causal `bspline`; both are reached at a depth of two iterations. The correlation coefficient between C and $1/E$ is close to 0.94, out of a maximum of 1 for perfectly proportional functions as presented in Figure 7.17. A bad SNR does not favor `bspline`: indeed, 80% of mainly missed detections are due to its lack of denoising efficiency; this is to be compared to the 60% detection errors of `ortho`, which performs best among all the other wavelets. Corresponding detection results are shown in Figure 7.18. Here, for $\alpha = 1.6$ and 2 iteration steps, the causal `dual` FSWT is only slightly worse (by 1%) than `ortho`.

$\rightarrow$ **High-pass detections.** Most of the band-pass coefficients of the wavelet transform are usually left undetected when high noise is present, which questions the utility of testing for them. However, for activations that have a high frequency content, some of these coefficients can be detected even with a low SNR. The gain over low-pass detection alone is then clear as exemplified in Figure 7.19.
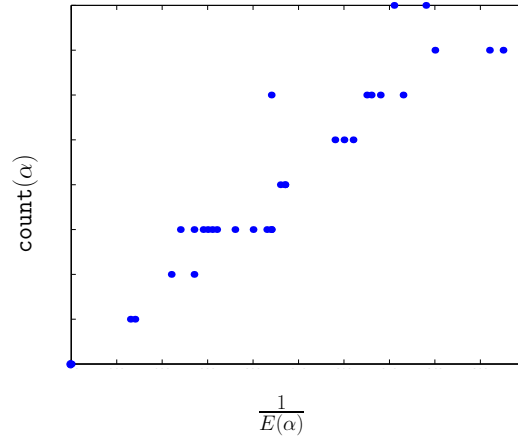
Figure 7.17: Scatterplot between the wavelet count $C(\alpha)$ and the inverse of the detection error percentage, $\frac{1}{E(\alpha)}$ for `ortho`. Same data as in Figure 7.16. The correlation coefficient $\frac{\sum C(\alpha)\frac{1}{E(\alpha)}}{\sqrt{\sum C(\alpha)^2}\sqrt{\sum \frac{1}{E(\alpha)^2}}}$ is 0.94.
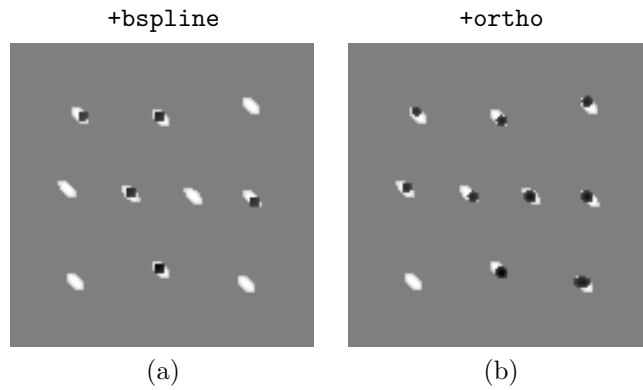
Figure 7.18: Detections corresponding to the data of Figure 7.12 using the optimal value of α for `+bspline` and for `+ortho`. Note that the high level of noise (-9.1dB) make useless and less efficient to detect high-pass activation. In such a case, the method performs better by using only the low-pass.
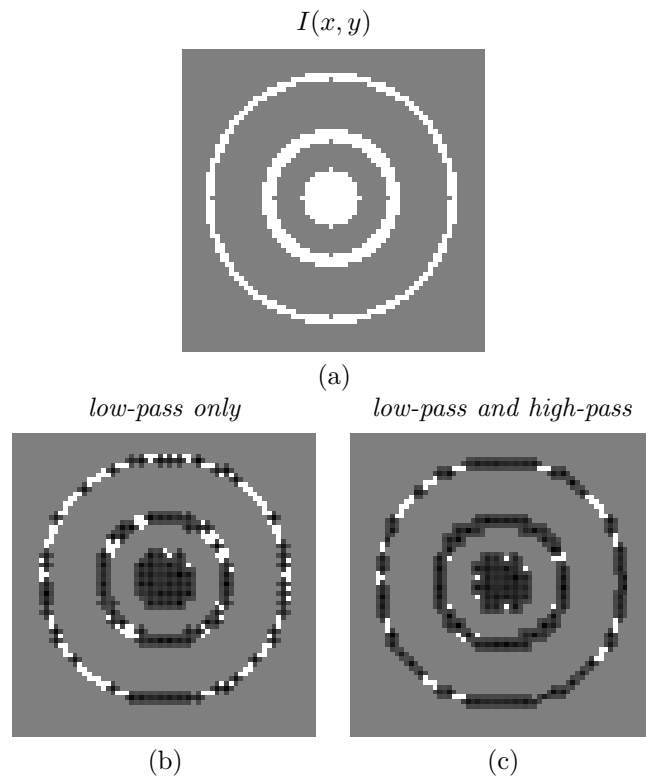
Figure 7.19: Detection of the high frequency activation $I(x, y)$. b) Processing the low-pass wavelet coefficients only. c) Processing both low-pass and high-pass wavelet coefficients.

→ **Translation invariance.** Similarly to other real wavelets due to the down-sampling operation, the fractional spline wavelets are not translation invariant. In addition the causal fractional splines shift their center by $\frac{\alpha+1}{2}$.

→ **Rotation invariance.** Since the FSWT is a separable transform, it should not exhibit any rotational invariance. However, the `dual` shows the best isotropy and thus should yield the best rotation invariance.

To make a global choice over all FSWT, we propose to take `+dual` with $\alpha = 1.2 - 1.6$ and an iteration step of 1 or 2, depending on the amount of noise. When the noise level is very high, the method performs better by using only the low-pass. `+bspline` performs well for the degree parameter $\alpha = 0.2 - 0.6$ when low noise is present. It controls the error of false detections strongly.

7.3.5 Complex wavelets

The structure of the complex wavelets is described in Section 4.4. We recall here that the CWT has the advantage of being closer to shift invariance than its real counterpart while keeping the computational cost low.

The frequency response of the real Daubechies and of the complex Daubechies with low-pass and with high-pass filters are presented in Figure 7.20 for the order parameter $n = 3$ and $n = 7$. The filters are orthogonal and compactly supported (in the complex case $[-n, n+1]$, $n \in \mathbb{Z}$). In addition, the complex Daubechies filters are symmetric.
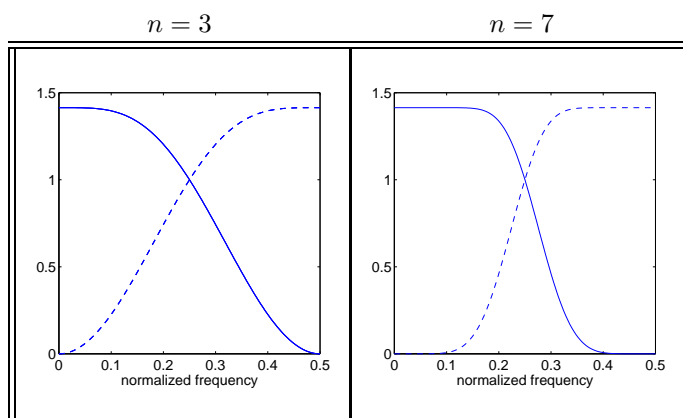
Figure 7.20: Example of frequency response of the low-pass (solid line) and high-pass (dashed line) filters for the complex Daubechies wavelets (`complex`), which are the same for the real Daubechies.

→ **Shift invariance.** Figure 7.21 illustrates that the complex filters are more shift invariant than the real Daubechies filters. Each plot shows the energy of the 1D
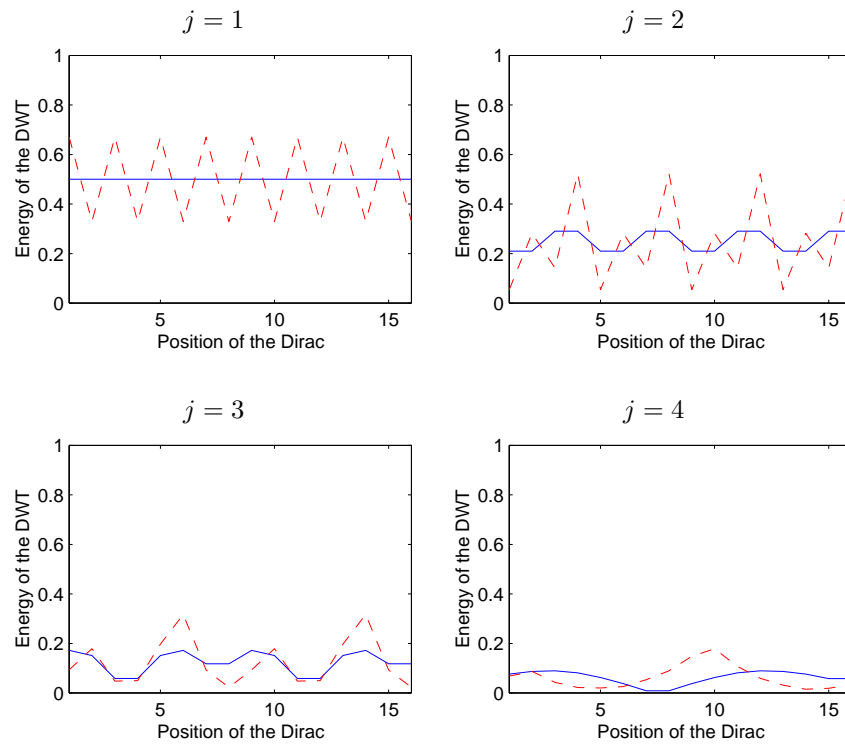
Figure 7.21: Comparison of shift invariance between the Daubechies wavelets (dashed line) and the complex Daubechies wavelets (solid line). The graphs show the band-pass energies of the DWT of a Dirac pulse of varying positions. The first four iterations $j = 1, \ldots, 4$ are shown here. The complex Daubechies is more nearly shift-invariant than its real counterpart.

wavelet coefficients along 16 different input signals. For the first iteration, the Complex Daubechies stay invariant. For higher iterations, their shift invariance decreases.

→ **High noise.** Another experiment, presented in Figure 7.24, shows the better shift-invariance for the complex Daubechies by detecting ellipses which are at equally spaced fixed positions (Figure 7.22). The shift-variance is more prominent as the noise is larger (here $-9.3\,\text{dB}$). The error $E(n) = E1(n) + E2(n)$ is measured, which depends on the order n of the wavelets. In Figure 7.24 this error is illustrated quantitatively for all n for two different set of data, which differ only by a shift of the ellipses. Obviously the results for `daub` change depending on the shift of the ellipses. In the case of the complex Daubechies (`complex`) the results remain very similar. However, the minimal error $E(n)$ is only slightly better for the complex case compared to the real Daubechies filter. Typically, for a very high noise like in this example, the complex Daubechies do not perform better, although `complex` detects more ellipses than `daub`, as it can be seen in Figure 7.23.

→ **Medium level of noise.** With lower amount of noise, $5\text{dB} \leq \text{SNR} \leq 10\text{dB}$ the shift variance of the real wavelets loses its influence on the results of detection. A remarkable result with lower amount of noise, same realization than in Figure 7.22 b), is the performance of `complex` for detection compared to `daub`. The complex Daubechies outperform the real Daubechies by far, as shown in Figure 7.25 b). But also compared to other wavelets like the fractional spline wavelets, presented in Section 7.3.4, the complex Daubechies show good results, as presented in Figure 7.25. The larger the activation pattern, the better the result of `complex`.

→ **Low level of noise.** Very low noise, $\text{SNR} > 10\,\text{dB}$ for $\bar{I}_D$ yields increased $E1$ errors and thus, the performance of `complex` decreases. However it still gives results comparable to `daub`. In Figure 7.26 an example is presented where the test data is illustrated in Figure 7.12 b). In this example, the complex Daubechies and the dual fractional spline wavelets give the best result, see Figure 7.26 and table in Figure 7.1. Independently of the noise level, the number of detected wavelet coefficients `count` is always much higher than in the case of `daub`, as can be seen in Figure 7.27 a). `count` for `complex` is often higher than for `dual`, when the degree α of `dual` is less than 2. These many detections yield a low $E2$ error. Thus `complex` tends to detect more pixels rightly, but at the cost of more false detections than in the case of real Daubechies (compare Figure 7.27). Figure 7.28 shows these detection results qualitatively. The detection `count` in the high-pass are also relatively frequent. `complex` detects more in the high-pass than `daub` and also `dual`.

$$I(x,y)$$

(a)

$$I_{D_{\tau=1}}(x,y):$$
$$SNR = \text{-}4.9\,dB$$

$$I_{D_{\tau=1}}(x,y):$$
$$SNR = \text{-}9.3\,dB$$

(b)

(c)
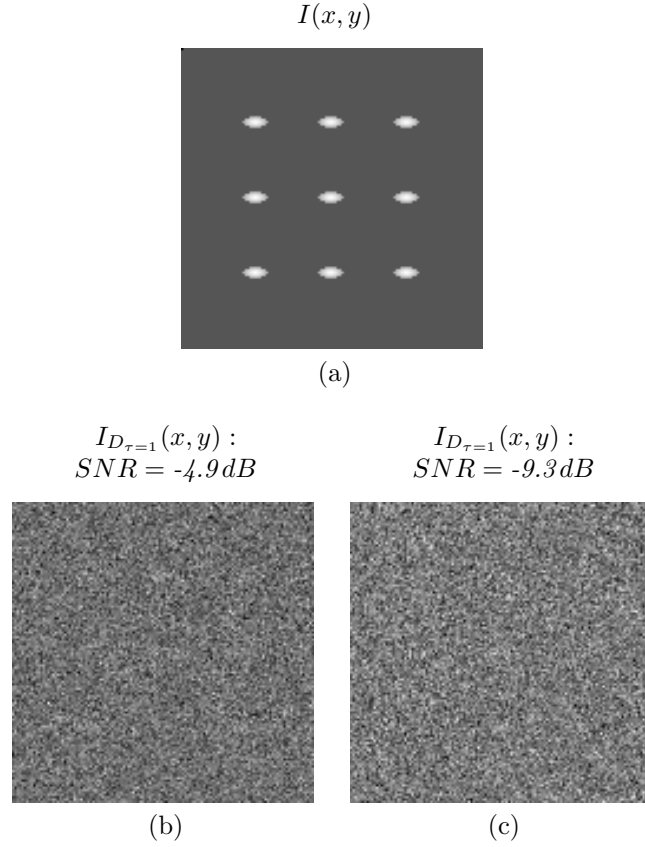
Figure 7.22: To compare the shift-variance, these data and its shifted version of ellipses at equally spaced positions are used for the illustrations in this Section. a) Template image $I(x,y)$. b) Noise is added to the template image $I(x,y)$, so that the average image $\bar{I}_D(x,y)$ has an SNR of 8.2. c) The average image $\bar{I}_D(x,y)$ has an SNR of 3.7 dB.

type	min $E(\alpha)$ resp. min $E(n)$	argmin α resp. argmin n
complex	18.0	7
+dual	18.0	1
+ortho	22.1	1
daub	26.8	4
+bspline	32.0	0.6

Table 7.1: Quantitative results corresponding to Figure 7.26.

.
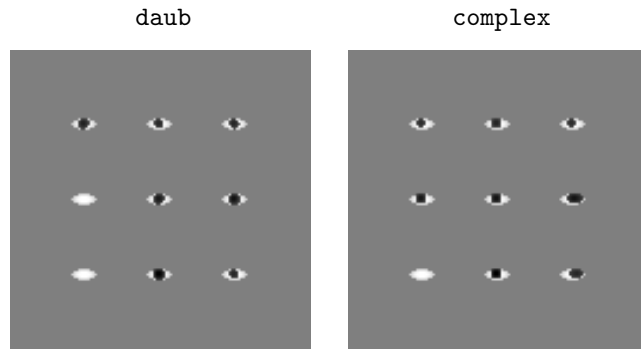
Figure 7.23: Detection of the pattern of Figure 7.22 c) using real and complex Daubechies. Here the noise level is very high: -9.3dB. The complex Daubechies (`complex`) detect more ellipses than the real Daubechies (`daub`).

a) Standard activation pattern
(Figure 7.22)

b) Shifted activation pattern
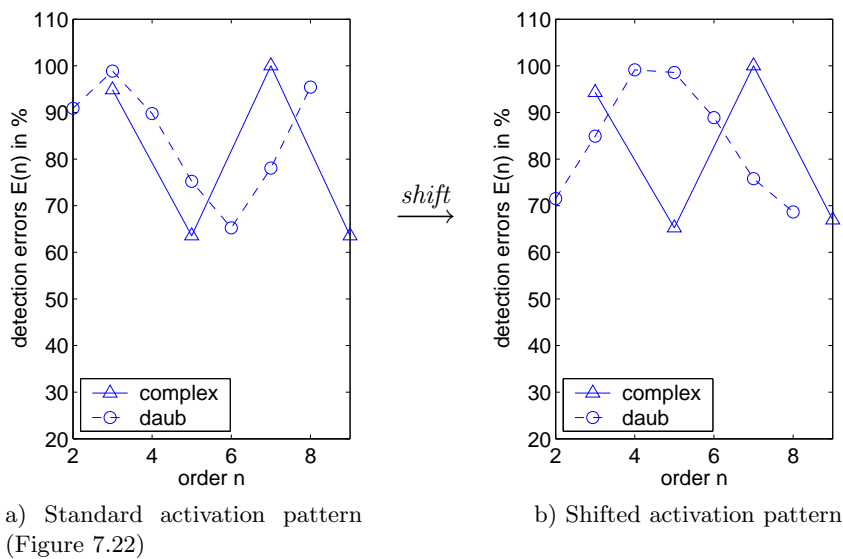
Figure 7.24: Detection error percentage of the pattern corresponding to $I(x, y)$ presented in Figure 7.23. Note the periodic waveform of the error E depending on the order of Daubechies wavelets. The results of a) and b) differ by a relative shift of the activation pattern for one pixel: $E(n)$ moves for `daub` with the order n, but $E(n)$ for the case of `complex` remains similar. Thus `complex` is more shift invariant than `daub`.
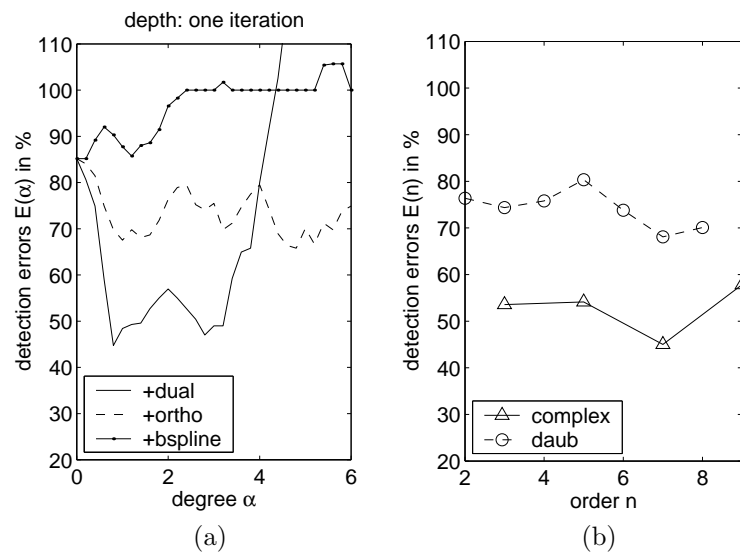
Figure 7.25: Detection error percentage of the pattern of Figure 7.22 with the realization of noise of Figure 7.22 b). For lower noise, the causal wavelets show less dependency on the shift (a) and (b)). A typical result is the good detection performance for `complex` for middle amount of noise like in this case. The detection quality of `complex` with respect to the error E is comparable to the one of `dual` and outperforms `daub` by far.
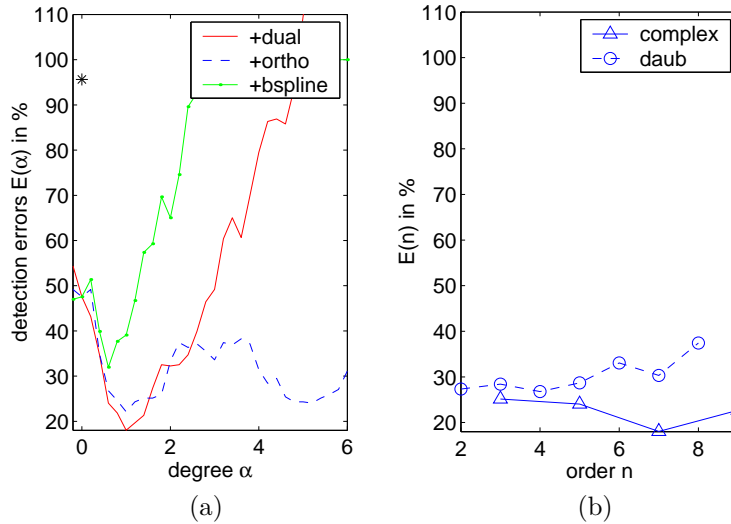
Figure 7.26: Performance of `complex` compared to `daub` and the FSWT when low noise is present and the activation pattern are placed randomly. `complex` and `daub` give the best detection results. The detection error when testing is performed on the raw data is marked with a star ($E=E2$).

The iteration step never exceed 2 iterations for small activation patterns, which is the case for images processed in Section 7.3.4.

7.3.6 Quincunx: comparison of separable with nonseparable transform

The Quincunx wavelet transform is extensively described in Chapter 5. The advantages of the quincunx wavelets over the separable ones are described in Section 7.2.2. We use a new family of quincunx wavelets, see Section 5.3, which are orthogonal, symmetric and have a fractional order of approximation. The advantage of a continuously varying order parameter λ is the possibility of a continuous adjustment of frequency selectivity and localization, similarly to the FSWT, see Section 7.3.4. To generate the quincunx filters, we applied the diamond McClellan transform to map a one dimensional design $H_\lambda(\omega)$ onto the quincunx structure. It is the separable transform based on this one dimensional filter $H_\lambda(\omega)$ with which we compare the nonseparable transform; because $H_\lambda(\omega_1)H_\lambda(\omega_2)$ has similar behavior as the quincunx filter. This filter is also quite close to the symmetric
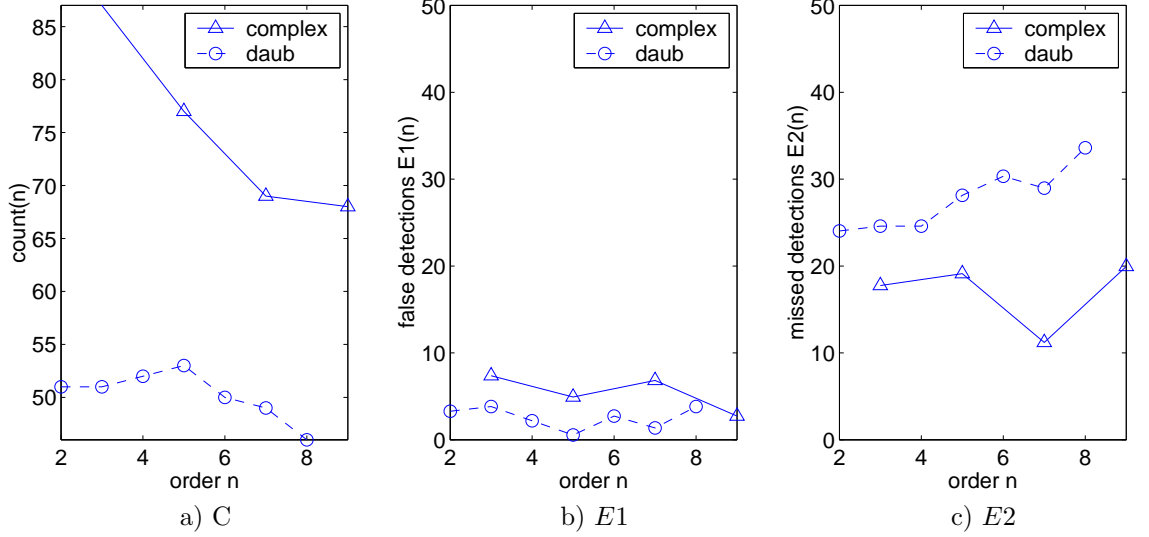
Figure 7.27: Illustration of wavelet count C, $E1$ and $E2$ corresponding to the detection with the same data like in Figure 7.26. $\mathtt{count}(\alpha)$ is always higher for $\mathtt{complex}$ than for $\mathtt{daub}$ (a). For this reason, the missed detections $E2$ are usually much smaller for $\mathtt{complex}$ (c). On the contrary the false detections $E1$ are higher for $\mathtt{complex}$ (b).

orthogonal fractional spline filter B^{α}_{*ortho} with $\alpha = \lambda - 1$ as shown below:

$$
\begin{aligned}
H_\lambda(\omega) &= \frac{\sqrt{2}(2 + 2\cos\omega)^{\frac{\lambda}{2}}}{\sqrt{(2 + 2\cos\omega)^\lambda + (2 - 2\cos\omega)^\lambda}} \\
&= \frac{\sqrt{2}\left|\cos\frac{\omega}{2}\right|^\lambda}{\sqrt{\left|\cos\frac{\omega}{2}\right|^{2\lambda} + \left|\sin\frac{\omega}{2}\right|^{2\lambda}}}
\end{aligned}
\tag{7.6}
$$

whereas

$$
\begin{aligned}
B^{\alpha}_{*ortho}(\omega) &= \sqrt{2}\left|\frac{1 + e^{-j\omega}}{2}\right|^{\alpha+1}\sqrt{\frac{A^\alpha(\omega)}{A^\alpha(2\omega)}} \\
&= \sqrt{2}\left|\cos\frac{\omega}{2}\right|^{\alpha+1}\sqrt{\frac{A^\alpha(\omega)}{A^\alpha(2\omega)}}.
\end{aligned}
\tag{7.7}
$$

Call that the autocorrelation filter $A(\omega)$ satisfies ([135]) the two-scale relation:

$$
A(2\omega) = \frac{1}{2}|B^{\alpha}_*(\omega)|^2 A(\omega) + \frac{1}{2}|B^{\alpha}_*(\omega + \pi)|^2 A(\omega + \pi).
\tag{7.8}
$$

Figure 7.28: Detection of the activation pattern of Figure 7.12 b) with SNR $= -1.19\,$dB. The corresponding error and degree parameter α is given in the table of Figure 7.1. `complex` and `+dual` detect obviously the best: there are not many visible missed detections.
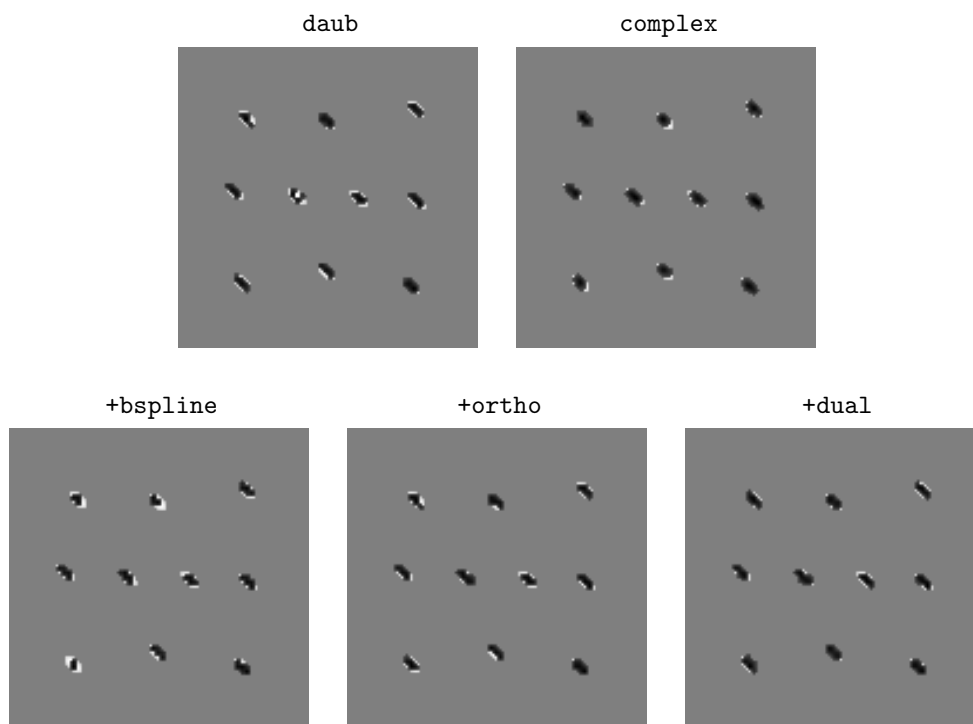
A way to compute $A(\omega)$ is to iterate the following process (ensured to converge if $\phi \in \mathbb{L}^2$ and if ϕ is a Riesz basis):

$$A_{i+1}(2\omega) = \frac{1}{2}|H(\omega)|^2 A_i(\omega) + \frac{1}{2}|H(\omega + \pi)|^2 A_i(\omega + \pi),$$

starting with $A_0(\omega) = 1$. The convergence speed of this scheme is directly given by the Sobolev regularity of ϕ: higher regularity, faster convergence. The filter

$$\sqrt{\frac{A_i(\omega)}{A_{i+1}(2\omega)}} B_*^\alpha(\omega)$$

automatically satisfies the orthogonality relation and the first iteration gives (7.6) while the infinite iteration gives (7.7) if we identify $\lambda = \alpha + 1$. For $\alpha = 0$ and $\lambda = 1$ the filters are exactly the same. For $0 < \alpha \lesssim 2$, their difference increases and for a higher degree parameter $(\alpha \gtrsim 2)$, the filters match again better, which illustrates Figure 7.29.

The frequency response of $H_\lambda(\omega)$, called `sep` is depicted in Figure 7.30 for the order $\lambda = 2$ and $\lambda = 6$. The nonseparable filter for the quincunx transform is called `quin` and is depicted in Figure 7.31.

In fMRI, the activation zones can be very small, compared to the size of the volumes or images taken from the brain. This is illustrated in Figure 7.32. One limitation of the separable wavelets is the scale progression by powers of 2 only. Thus the number of coefficients of the low-pass which represents mainly the activation zones is reduced by 4 with every iteration, as presented in Figure 7.33 upper row. To detect more accurately small patterns, a finer scale progression would be preferable.

The second drawback of the separable wavelets is the strong directionality of the filters. Thus the separable wavelets detect more easily horizontal and vertical expanded patterns, whereas the activation zones can have any directionality. To reduce this problem, nonseparable wavelets are more appropriate.

The quincunx transform is a solution to these issues: the scale progresses by powers of $\sqrt{2}$ and provides thus a bandpass of finer resolution than in the separable case; moreover, the filtering is less directional, which was made possible because of the non-separability of the filters. We developed a fast FFT-based implementation for two and three dimensions, refer to Chapter 5.

$\rightarrow$ **Detection accuracy.** We devised an experiment with a single pattern, scattered over the size of the image, to which a noise of SNR = 2.5 dB for $I_{D_\tau}(x, y)$, respectively SNR = 15.5 dB for $\bar{I}_D(x, y)$ has been added. The corresponding test image $I(x, y)$ is shown in Figure 7.34. Both algorithms, the quincunx transform with its fractional filter `quin` and the separable algorithm with the corresponding separable filters were applied to $I(x, y)$. This is a typical example where `quin` detects more accurately, even though the patterns have a rather horizontal extension, which should favor separable wavelets. A qualitative detection is illustrated in Figure
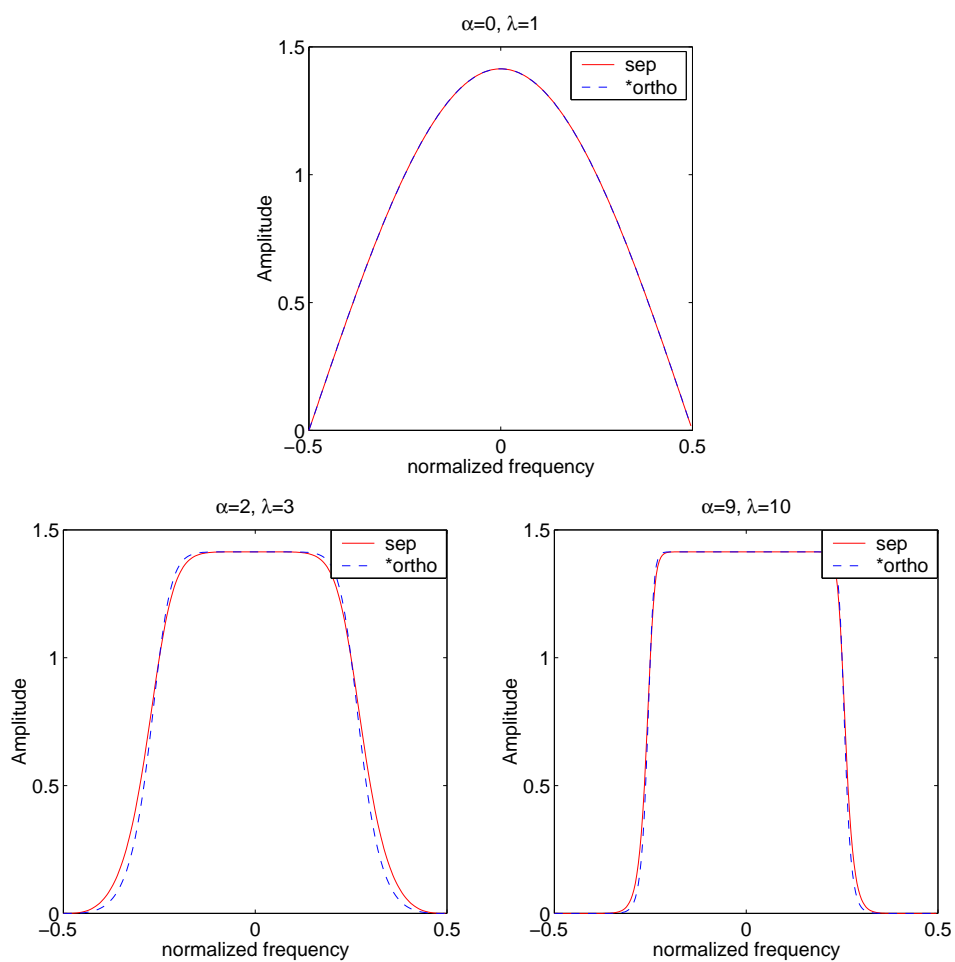
Figure 7.29: Frequency response of the filter $H_\lambda(\omega)$ (`sep`) and $B^\alpha_{*ortho}(\omega)$ (`*ortho`), where $\lambda = \alpha + 1$. For large α, i.e., $\alpha \gtrsim 2$ the filters are close.
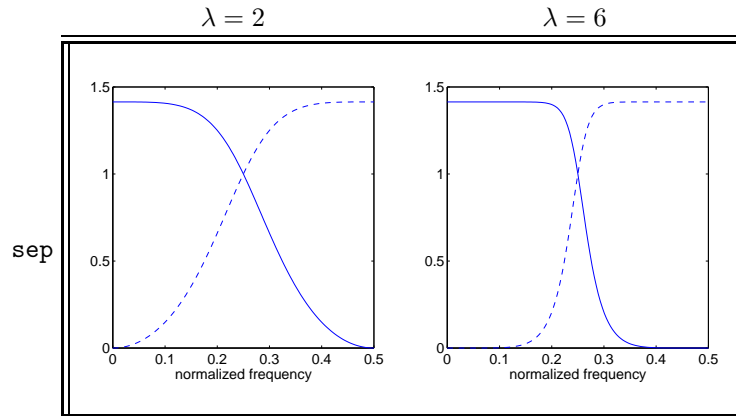
Figure 7.30: Example of frequency response of the low-pass (solid line) and high-pass (dashed line) filters for the one dimensional design $H_\lambda(\omega)$ (`sep`).
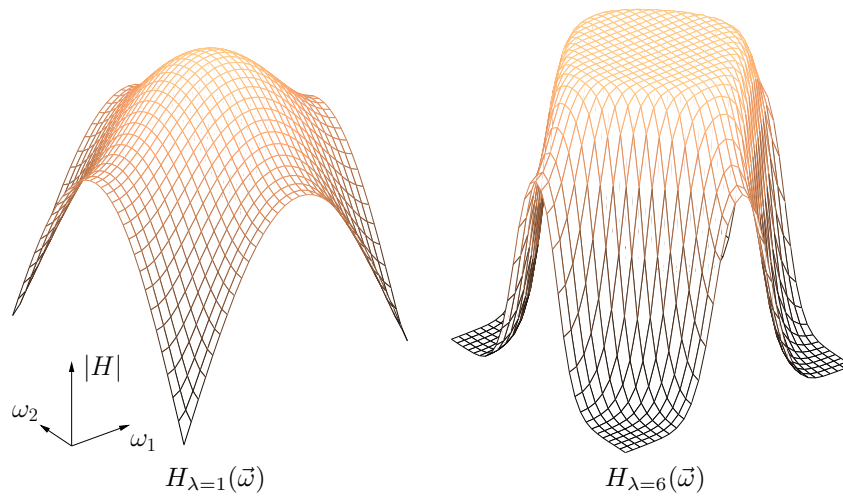
Figure 7.31: Example of frequency response of the quincunx low-pass filter.
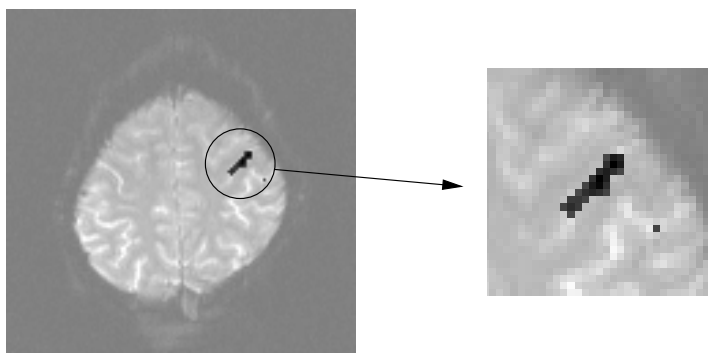
Figure 7.32: Example of detection of activation in the primary motor area. This activation zone is small compared to the size of the image.

sep
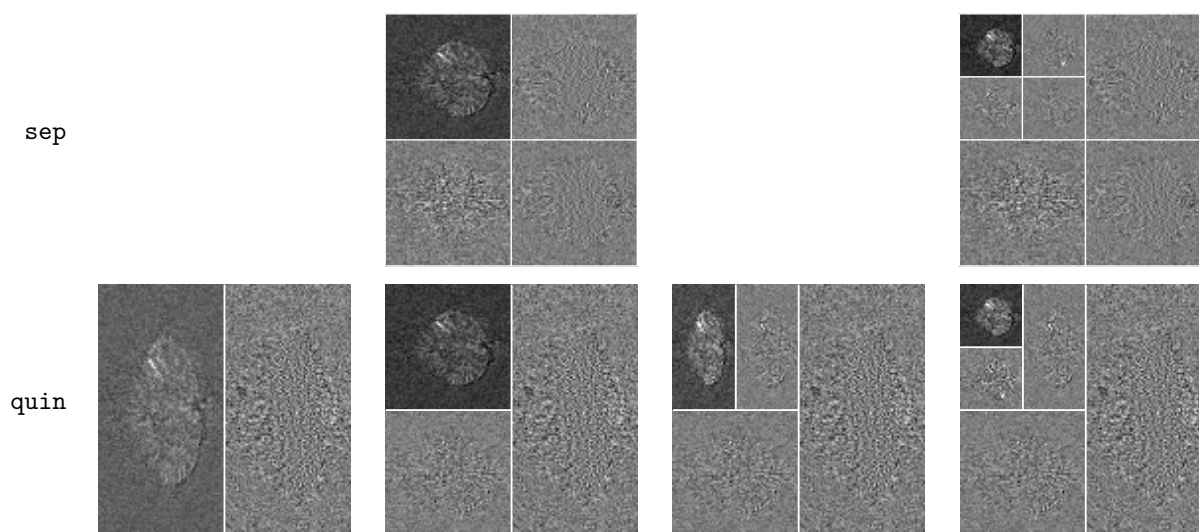
quin

Figure 7.33: Wavelet coefficients for the separable (upper row) and the quincunx transform (lower row). From left to right, the number of iterations increases. Every second iteration, the size of the low-pass (upper left corner) can be compared to one iteration of the separable case.
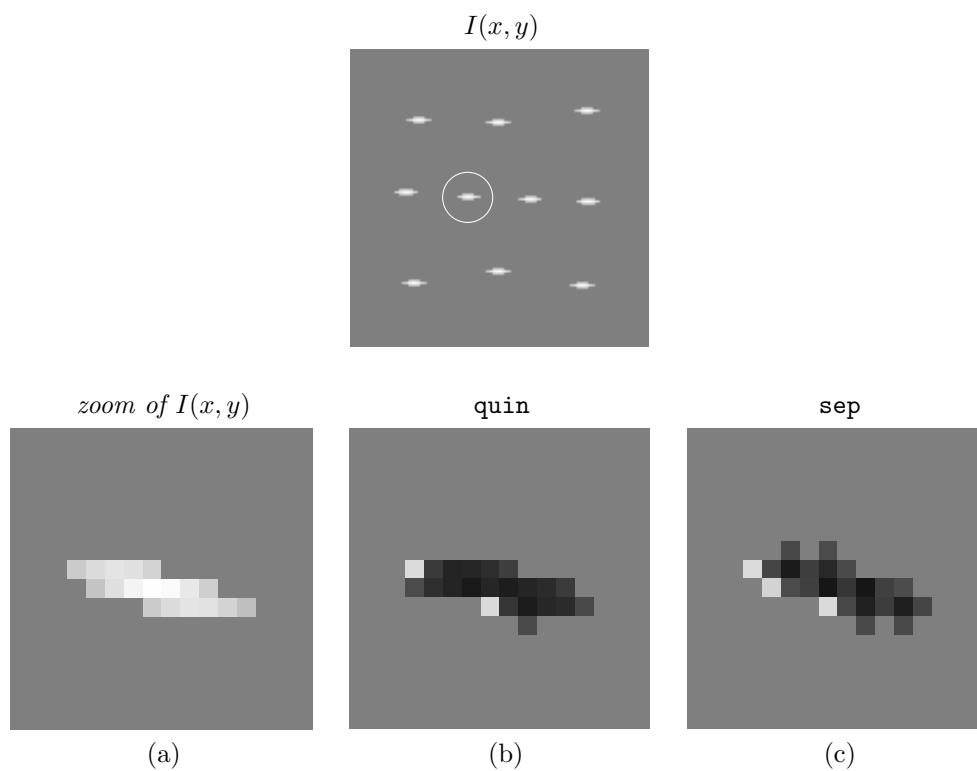
Figure 7.34: $I(x, y)$ contains several patterns to detect. The level of noise is SNR = 2.5dB for I_{D_τ}. For each method the results are presented for the optimal order parameter λ, which causes the smallest error E. The white pixels represent the pattern. In b) and c) the black pixels represent the detection. b) Representation of the detection of one of those patterns for `quin`. The errors are relatively few compared to the small size of the pattern. With this low level of noise, one iteration is sufficient. c) Representation of the qualitative result for the separable algorithm for the same zoomed pattern.

7.34. The quantitative results of this example are given in Figure 7.35. Especially for a larger order parameter λ, there is a big gap between the error E of `quin` and `sep`. We have indications that the false detections $E1$ are much smaller for `quin`, because of the finer scale progression. Even though the threshold *level* seems to be adequate, the false detections $E1$ increases, when the iteration depth is chosen to be higher. On the contrary, when we keep the iteration depth on a finer resolution compared to the separable case, the noise is less reduced and thus in some cases more $E2$-errors appear in the quincunx case (Figure 7.35). The advantage of the quincunx transform is the more noticeable when the optimal iteration depth J is in between the possible iterations of the separable algorithm, i.e., when J is odd. An example of such a situation is given in Figure 7.36, where `sep` performs best with two iterations and causes an error $E = 58\%$, while `quin` performs best with three iterations and causes an error $E = 41\%$ with $\lambda = 2$. Thus the gap between both methods is almost 20%.

Another example, closer to real data is given in Figure 7.37. $I(x, y)$ is extracted
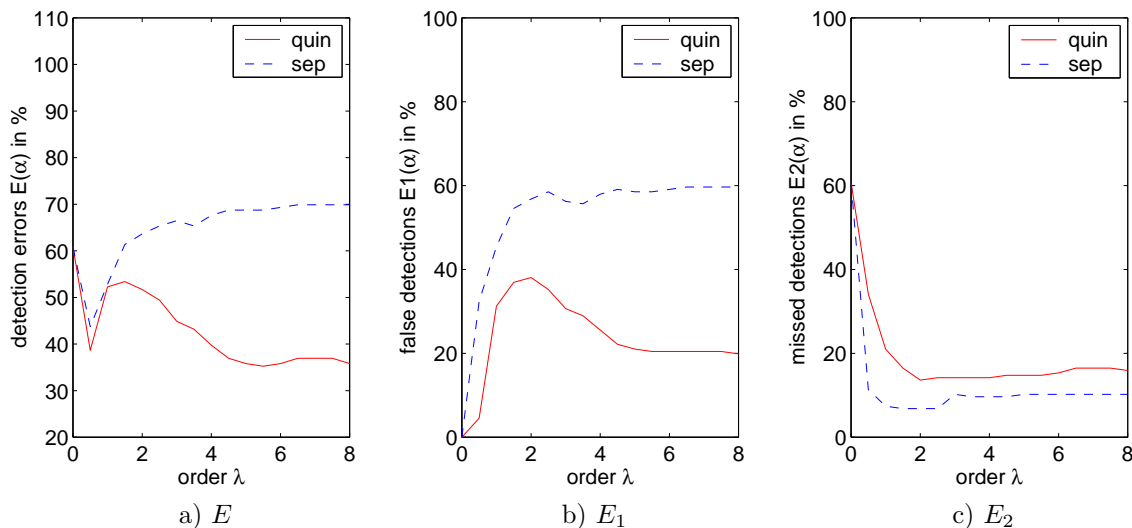
Figure 7.35: Error of detection of $I(x, y)$ of Figure 7.34. Comparison of `quin` (solid line) and `sep` (dashed line). For both methods the optimal iteration depth is one.

from a slice of real data and contains small shapes of different intensities. Like for the other examples, no mask is applied, which would enhance the detection quality. Again, it is evident that `quin` detects more accurately. Thus the finer scale progression of the quincunx transform offers more flexibility and leads to better results.

$\rightarrow$ `count.` It can be observed that `quin` typically detects more coefficients than `sep`.

a) E b) E_1 c) E_2
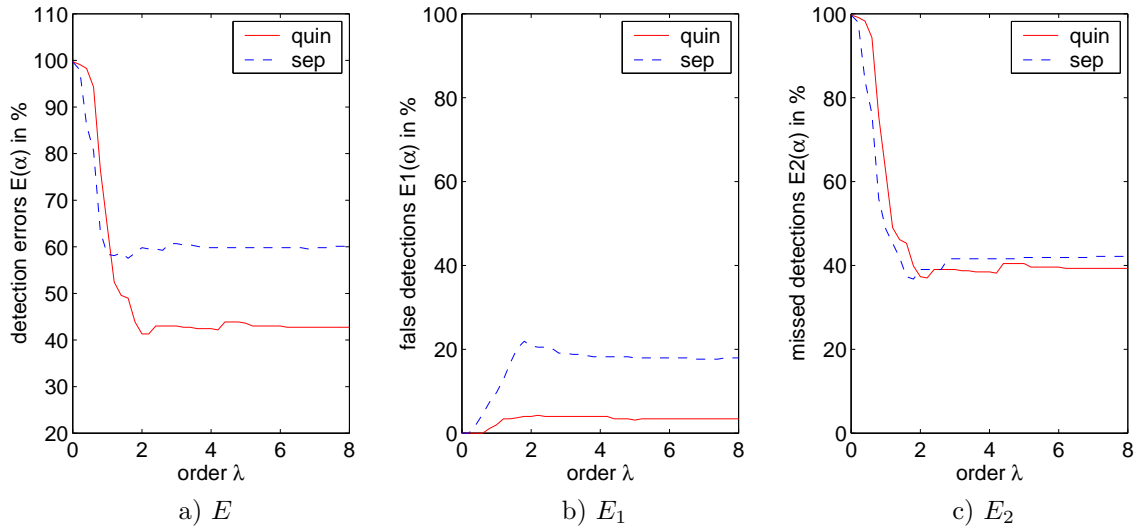
Figure 7.36: Detection error percentage of the data of Figure 7.22 b). `quin` detects with less errors since it can take advantage of its odd iteration depth ($J = 3$). The better result is mainly due to less false detections $E1$.
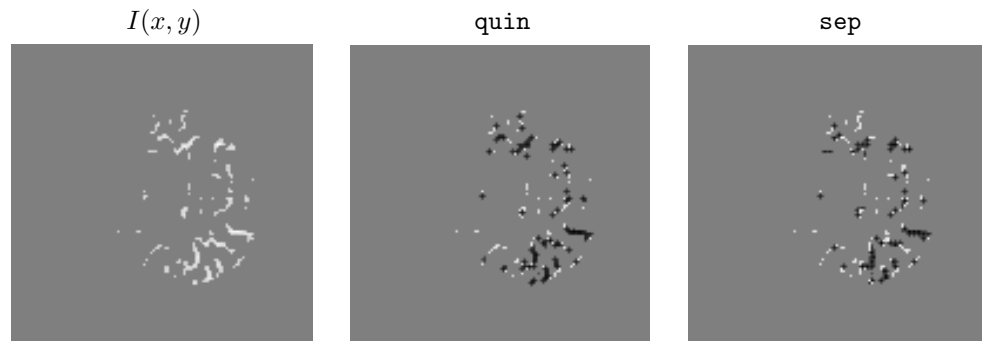
Figure 7.37: $I(x, y)$ is extracted from a slice of real data. The level of noise added to $I(x, y)$ results in an SNR = 14.8dB for $\bar{I}_D$. The order parameter λ for `quin` and `sep` is chosen such that the error E is minimal.

This is reasonable insofar as the optimum iteration depth for `quin` is lower compared to the separable case and thus the low-pass is of larger size. The graph of the numbers of detected coefficients corresponding to Figure 7.37 is presented in Figure 7.38. The number of detections in the high-pass depends highly on the order λ. It could be observed with the experiments of ellipses that the lower the order the more detections are obtained in the high-pass. This is due to the decrease
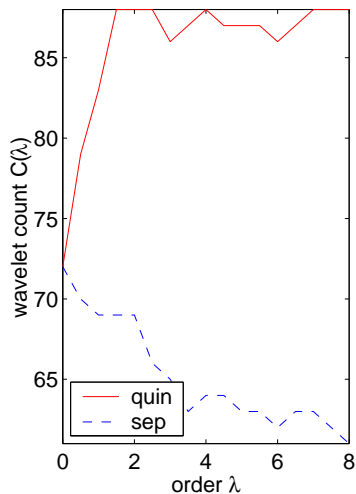
Figure 7.38: `quin` detects more coefficients, also because of the higher resolution of the low-pass for one iteration.

of the frequency selectivity for a lower λ, where detections of rather not so high frequencies contained in the band-pass of low order parameters are more frequent.

$\rightarrow$ **Rotation invariance.** To see whether the non-separability of the quincunx transform leads to near rotation invariance, several test data with ellipses were created which differ by rotation of the ellipses by the angle $\theta = [0°, 15°, ..., 90°]$. Two such test images are shown in Figure 7.39. For each set of data, the minimal error $E(\lambda)$

$\theta = 0°$ $\qquad\qquad\qquad$ $\theta = 30°$
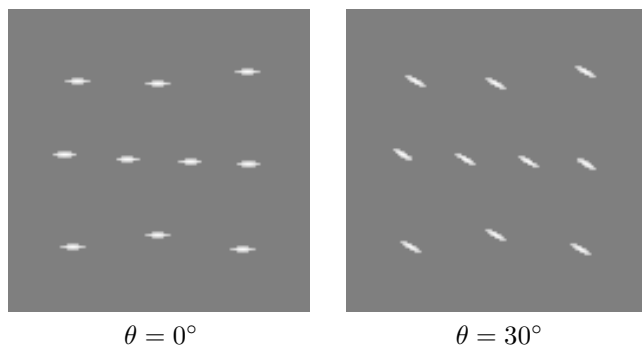
Figure 7.39: Two examples of test images with rotated ellipses.

is calculated for the separable case `sep` and for `quin`. In Figure 7.40 the error $E2$

and E are plotted depending on the angle θ. Obviously, the separable algorithm has more difficulty with detection when the angle approaches to $\theta = 45°$ (Figure 7.40 a)). On the contrary, the nonseparable method doesn't change much when the angle θ varies, but detects slightly better diagonal expanded patterns, as expected. Also the error E changes slightly less in the case of `quin` and again `quin` detects better (see Figure 7.40 b)).

a) E_2 \qquad\qquad\qquad\qquad b) E
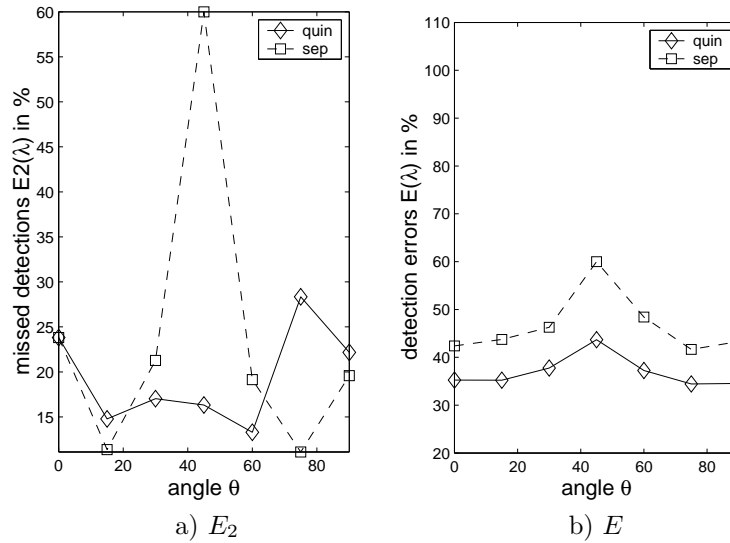
Figure 7.40: Comparison of `quin` and `sep` regarding directionality. Detection errors of rotated ellipses like in Figure 7.39, with the same level of noise added as in Figure 7.34. a) The missed detections $E2(\theta)$ is less influenced by θ in the case of `quin`. b) $E(\theta)$ is less dependent on the angle of the ellipses (θ) in the case of `quin`.

To conclude, it is especially the finer scale progression that brings advantages: depending on the data set, the detection gain can be 20 %. However, compared to the fractional splines, the quincunx transform presented in this work is not always superior, since its nature is comparable with `*ortho`, which doesn't perform as well as `+dual`, but provides a good compromise between a selective analysis filter and a localizing synthesis filter and is thus a more general approach. This calls for a "dual" version of the symmetric filters to be used in the quincunx transform.

7.3.7   Results

The results and observations of the whole series of experiments are summarized in Table 7.2.

	+bspline	+ortho	+dual	quin	complex	daub	sep	*dual
min E	$--$	$+++$	$+++$	$++$ low noise: $+++$	$++$	$++$	$+$	$+++$
min $E2$ (sensitivity)	$--$	$++$	$++++$	$-$	big shape: $+++$ otherwise: $-$	$-$	$-$	$+++$
min $E1$ (specificity)	$+++$	$++$	$\alpha \to 0:$ $++$ otherwise: $--$	$+++$	$++$	$++$	$++$	$--$
invariance of α optimal	$++$	$--$	$+++$	$++$	$-$	$-$	$+$	$++$
argmin(E)	0.2-0.6	vary	1.2-1.6	2/5.8	3/5	vary	vary	1.2-2
translation invariance	$++$	$--$	$++$	$++$	$++$	$--$	$++$	$++$
rotation invariance	?	?	?	$+++$	?	?	$--$	?
detection when low noise present	$++$	$\alpha \to 0:$ $++$	$\alpha \to 0:$ $++$	$++$	$-$	$--$	$--$	$--$
detection when very high noise present	$--$	$+++$	$+++$	$++$	$--$	$-$	$++$	$+++$
`count`	$--$	$++$	$+++$	$++$	$+++$	$++$	$+$	$+++$
`count` in highpass	$+++$	$++$	$--$	λ small: $++$ otherwise: $--$	$++$	$--$	$++$	$--$

Table 7.2: Summary of different methods. Notation: $--$ bad, $-$ not optimal, $+$ good, $++$ very good, $+++$ excellent, $++++$ the very best. The symmetric fractional bspline (`*bspline`) behaves close to the causal bspline, especially for higher α and is thus omitted in the table. The filters `*ortho` are close to `sep` and are also omitted for this reason.

7.3.8   Discussion

A recent work [44] (2001) uses an algorithm which follows in general terms the one proposed by Ruttimann and compares different wavelet decomposition schemes. A template with only circular activation patterns is used. The conclusion of the authors is that the Gabor transform performs the best, since it best fits the circular activation areas. Obviously, their conclusions are only valid for such patterns. To carefully prevent this problem, we used different non symmetric activation shapes [59] (2000). We also avoided using different pattern sizes in the same image, since in this case the large patterns dominate the error measure and no information is given for small activation patterns. The signal reconstruction method of [44] seems to cause ringing and thus the false detection errors get more prominent for wavelets as the length of the synthesis filters increase. This explains why, in their experiments, wavelet transforms with very short synthesis filters perform the best. To avoid this effect we applied a threshold on the reconstructed signal which enhances the detection rate substantially, since this restriction of applying short wavelets only is not necessary anymore. In addition, compared to [44], we validate our findings with real data.

7.4   Real data

In this section, we validate the results of the synthetic data (Section 7.3) with real data. For this purpose, we considered several real data sets from different subjects and for different tasks. For example, tapping with the left or right hand fingers was one of these tasks. Other tasks were moving the shoulder, looking at a light source switching on and off or listening to bi-syllabic words.
The orientation of the images is shown in Figure 7.41.
We also compare our method with SPM (Section 2.8).
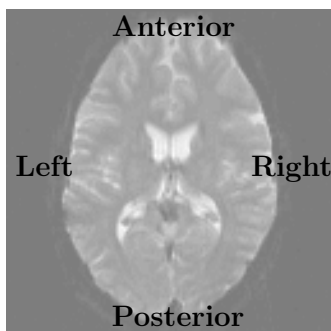
Figure 7.41: Orientation of the brain in the images along this section.

7.4.1   Image acquisition

We obtained the data sets from different sources.

A. Inselspital of Bern The data sets acquired by Arto Nirkko from the Inselspital of Bern have the following characteristics ([105]): The fMR-images were acquired on a standard clinical 1.5 T whole body scanner (Siemens Magnetom Vision). Blood Oxygenation Level Dependent (BOLD) (see Section 2.3) fMRI was achieved with an Echo-Planar Imaging sequence (EPI) (see Section 2.2.3) with an Echo Time (TE) of 62 ms and a Repetition Time (TR) of 6 s between whole brain data sets. A 128 pixel matrix and a 200 mm field-of-view resulted in an inplane pixel resolution of $1.56\,\mathrm{mm} \times 1.56\,\mathrm{mm}$. Thirty 2-D slices of 5 mm thickness and 4 mm spacing between slice centers were acquired. Since distortions in the phase-encoding direction cannot be avoided completely, the phase-encoding direction was selected in the antero-posterior direction, thus preserving brain symmetry. In total, 68 measurements were acquired per experiment. The volunteer had to alternate between activation (A) and rest (B) periods according to auditory cues. Cycles of A and B performance (each with four data sets, equaling 24 s) were repeated eight times. The total of 68

data sets per experiment was obtained in 6.8 min. The first four sets were discarded in order to analyze only the data obtained with steady-state MRI signals and task performance. The summary is given in Table 7.3.

Paradigm length A/B	4/4
Voxelsize	$1.56 \times 1.56 \times 4$ mm
# Voxels per volume	$128 \times 128 \times 30$
TR	6 s
# volumes	68

Table 7.3: Parameters of acquisition of the data acquired by the Inselspital in Bern.

B. University Hospital of Geneva (HUG) The data sets acquired by Francois Lazeyras from the University Hospital of Geneva have the following characteristics: The fMR-images were acquired on a 1.5 T whole body scanner (Eclipse, Marconi/Philips). Blood oxygenation level dependent (BOLD) fMRI was achieved with a gradient echo, echo-planar imaging sequence (GRE, EPI) with an echo time (TE) of 40 ms and a repetition time (TR) of 1.4 s between whole brain data sets. The delay between two successive acquisitions were 2 s. A 128 pixel matrix and a 250 mm field-of-view resulted in an inplane pixel resolution of 1.95 mm × 1.95 mm. Thirteen 2-D slices of 5 mm without gap were acquired. A flip angle of 80 degrees was used. Cycles of A and B performance (with 10 data sets for A and 20 data sets for B) were repeated five times. In total, 150 acquisitions were made.

Paradigm length A/B	10/20
Voxelsize	$1.95 \times 1.95 \times 5$ mm
# Voxels per volume	$128 \times 128 \times 13$
TR	1.4 s
# volumes	150

Table 7.4: Parameters of acquisition of the data acquired by HUG.

C. Wellcome Department of Imaging Neuroscience, University College London These are publically available data. The experiments conducted by Geriant Rees under the direction of Karl Friston and the Functional Imaging Laboratory (FIL) method group [76] have the following characteristics: The images were acquired with an echo-planar imaging sequence (EPI) on a modified 2 T Siemens MAGNETOM Vision system. Each acquisition consisted of 64 contiguous slices

with $64 \times 64 \times 64$ voxels of $3\,\text{mm} \times 3\,\text{mm} \times 3\,\text{mm}$. The acquisition of the whole volume took 6.05 s with a scan to scan repetition time (TR) of 7 s.
96 acquisitions were made in blocks of 6 volumes, giving 16 blocks of 42 s. The condition for successive blocks alternated between rest and auditory stimulation, starting with the rest-condition. The auditory stimulation was bi-syllabic words presented binaurally at a rate of 60 per minute.

Paradigm length A/B	6/6
Voxelsize	$3 \times 3 \times 3$ mm
# Voxels per volume	$64 \times 64 \times 64$
TR	7 s
# volumes	96

Table 7.5: Parameters of acquisition of the data conducted by the Wellcome Department of Imaging Neuroscience, University College London.

7.4.2   Relation between our algorithm and SPM

To enhance the signal-to-noise ratio (SNR), SPM filters the images with a Gaussian kernel (see Section 2.8.1). Our approach consists of performing a wavelet transform, where the low-pass contains the SNR-enhanced information. In this section, we show that these approaches are related (if we choose a symmetric dual fractional spline wavelet transform), with the difference that we keep the high-pass content and downsample after filtering. It is the statistics of the two methods that exhibit the main difference. While SPM defines the threshold based on the theory of Gaussian random fields (Section 3.2.6), our method uses a t-test (Section 3.3.5), respectively a Hotelling's T^2 test (Section 3.3.7), and defines the threshold by a simple Bonferroni-correction (Section 3.2.6). The approach with the wavelet transform requires signal reconstruction after the statistical part. Note that despite this difference, the Gaussian random field statistics are roughly comparable with a downsampling over the size of a RESEL to cancel the correlation in the space domain and with a subsequent Bonferroni-correction.

Relation between the degree parameter α of the fractional spline *dual and the FWHM of the Gaussian filter

We compare here the Gaussian filter with the symmetric dual fractional spline filter *dual rather than +dual, since *dual is a symmetric filter like the Gaussian filter.
We have the following equivalence for *dual when α is large (which allows to consider

only frequencies ω close to 0):

$$
\begin{aligned}
B_*^\alpha \left(e^{j\omega}\right) &= \sqrt{2} \left| \frac{1 + e^{-j\omega}}{2} \right|^{\alpha+1} \\
&= \sqrt{2} \left| \cos \frac{\omega}{2} \right|^{\alpha+1} \\
&= \sqrt{2} e^{(\alpha+1)\log|\cos\frac{\omega}{2}|} \\
&\approx \sqrt{2} e^{(\alpha+1)\log\left(1-\frac{\omega^2}{8}\right)} \\
&\approx \sqrt{2} e^{-(\alpha+1)\frac{\omega^2}{8}}.
\end{aligned}
\tag{7.9}
$$

Due to the Noble identity, the J iterations of the filter $B_*^\alpha(e^{j\omega})$ take the expression:

$$
B_{*J}^\alpha(e^{j\omega}) = B_*^\alpha(e^{j\omega})B_*^\alpha(e^{2j\omega})\dots B_*^\alpha(e^{2^{J-1}j\omega})
\tag{7.10}
$$

According to (7.9), (7.10) can be approximated by

$$
\begin{aligned}
B_{*J}^\alpha(e^{j\omega}) &\approx 2^{\frac{J}{2}} e^{-\frac{\alpha+1}{8}\left(\omega^2 + 4\omega^2 + \dots + 4^{J-1}\omega^2\right)} \\
&\approx 2^{\frac{J}{2}} e^{-\frac{\alpha+1}{8}\omega^2 \frac{4^J-1}{3}}.
\end{aligned}
\tag{7.11}
$$

The Fourier transform of the Gaussian kernel is given by:

$$
\frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_g^2}} \overset{\mathcal{F}}{\longleftrightarrow} e^{-\frac{\sigma_g^2 \omega^2}{2}}.
\tag{7.12}
$$

From equation (7.11) and (7.12) it follows that we can approximate the `*dual` filter with a Gaussian filter of width:

$$
\sigma_g \approx \frac{\sqrt{\alpha+1}}{2}\sqrt{\frac{4^J-1}{3}}
\tag{7.13}
$$

The relation between the Full width of half maximum (FWHM) and σ_g can be obtained by:

$$
e^{-\frac{\frac{\text{FWHM}}{2}x^2}{2\sigma_g^2}} = \frac{1}{2}
$$

and is thus given by

$$
\text{FWHM} = 2\sigma_g \sqrt{2\ln(2)}.
\tag{7.14}
$$

From equations (7.13) and (7.14) we thus get:

$$
\text{FWHM} \approx \sqrt{2\ln(2)}\sqrt{\alpha+1}\sqrt{\frac{4^J-1}{3}}.
\tag{7.15}
$$
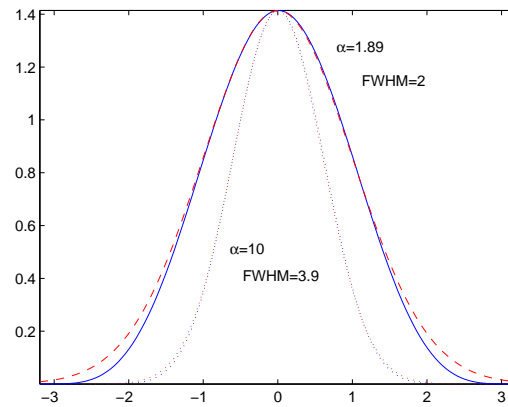
Figure 7.42: Comparison of the Gaussian filter (dashed line) with the `*dual` (solid line) in the Fourier domain. For a larger α the approximation is closer (dotted line for Gaussian filter and for `*dual`).
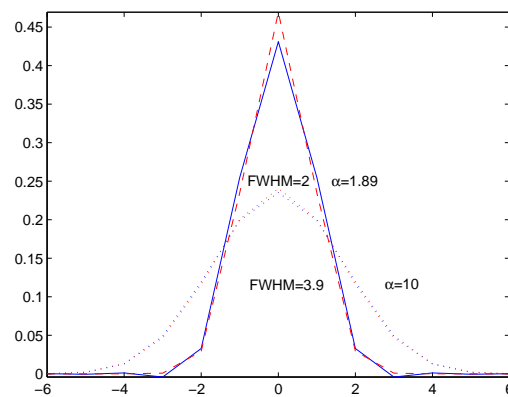
Figure 7.43: Comparison of the Gaussian filter (dashed line) with the `*dual` (solid line) in the time domain. For a larger α the approximation is closer (dotted line for Gaussian filter and for `*dual`).

 The equivalence between the Gaussian filter and the `*dual` filter, when related parameters are chosen, using equation (7.15), is presented in Figure 7.42 in the Fourier domain and in Figure 7.43 in the time domain. It can be noticed that for an increasing degree parameter α this equivalence is getting more exact. When applying both filters (the Gaussian filter and the `*dual` filter) with comparable parameters to real data, we obtain the same activation pattern, even after downsampling and applying a t-test with Bonferroni-correction (see Figure 7.44).
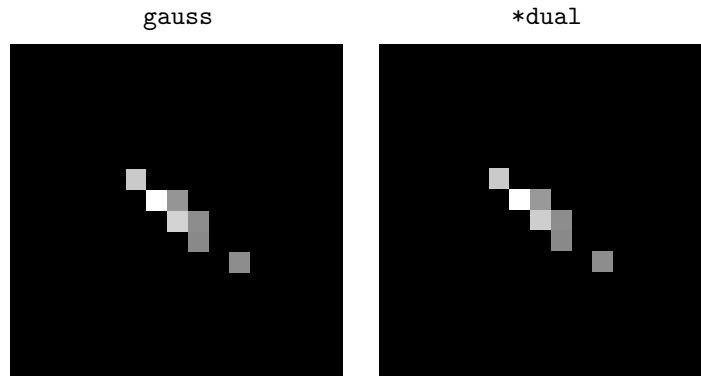
Figure 7.44: Zoom of the detection in the wavelet domain after applying the Gaussian filter or the `*dual` filter. The result is the same for the Gaussian filter and the `*dual` filter.

   The notion of RESELS (Section 3.2.6) of SPM is related to our number of iterations J through the equation

$$\text{RESELS} = \frac{V}{2^{JD}}, \tag{7.16}$$

where 2^{JD} is the low resolution wavelet sub-sampling factor. On the other hand, SPM defines the number of RESELS by the expression:

$$\text{RESELS} \quad \approx \quad \frac{V}{\prod_{i=1}^{D} \text{FWHM}_i}, \tag{7.17}$$

where V represents the volume of the dataset. Thus, assuming the same FWHM in each dimension, we have

$$\text{FWHM} = 2^{J}. \tag{7.18}$$

   With these equations we can derive a bound depending on the iteration depth J for α and compare it to the best α found with the simulations in Section 7.3. When we

a) FWHM=2× pixels   b) FWHM=4× pixels
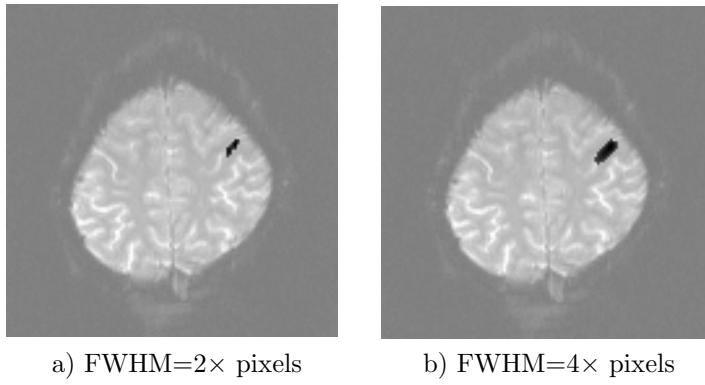
Figure 7.45: Results for two different resolutions using the standard possibility recommended by SPM.

a)   b)
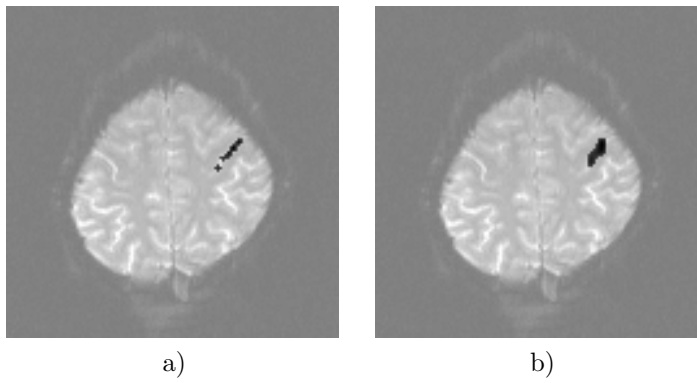
Figure 7.46: Detection results using the downsampled low-pass of the symmetric fractional spline (*dual): a) J=1 and $\alpha = 1.89$; b) J=2 and $\alpha = 1.31$.

replace equation (7.15) in (7.18), we obtain

$$\text{RESELS} = \frac{V}{\left(\sqrt{2 \ln(2)} \sqrt{\alpha + 1} \sqrt{\frac{4^J - 1}{3}} \right)^D}. \tag{7.19}$$

Finally, by equating (7.19) and (7.16) we find that

$$1 + \alpha \quad = \quad \frac{4^J}{2 \ln(2) \frac{4^J - 1}{3}}. \tag{7.20}$$

For the first iteration ($J = 1$) we get from equation (7.20)

$$\alpha \quad = \quad \frac{2}{\ln(2)} - 1$$
$$\simeq \quad 1.885. \tag{7.21}$$

Asymptotically, when $J \to \infty$

$$\alpha \quad = \quad \frac{3}{2 \ln(2)} - 1$$
$$\simeq \quad 1.164. \tag{7.22}$$

Thus α doesn't change much with the iteration and lies in the range of

$$1.164 \lesssim \alpha \lesssim 1.885, \tag{7.23}$$

which is very close to the "optimal" range found with simulations in Section 7.3: $1.2 \lesssim \alpha \lesssim 1.6$.

We now present some experimental results that support these calculations. The data set that is used for this purpose is acquired with the method described in Table 7.3. The task that the volunteer had to perform was tapping with the left hand fingers. The volunteer was a member of our group. In Figure 7.45, SPM is applied to real data using the parameters obtained above. The results should be comparable to those of Figure 7.46. Indeed, the activation zones are similar, but there are visible differences, most probably due the fact that the statistical analysis performed in the two methods are very different, even though the filters are similar.

7.4.3   Control of $E1-$error

Obviously, we cannot measure the false detection errors $E1$ nor the missed detection errors $E2$ as for the synthetic data, since the underlying true activation is unknown. One possible approach is to control the false detection errors $E1$ with null-data; i.e., data

which doesn't contain any activation. In this case we know that all detected activations are false.

For this experiment, our data set has the parameters of Table 7.3. The data have been realigned. We analyzed the data with the 2D algorithm slice per slice and counted the wavelet detections C. The number of detections, i.e., maximally one coefficient per 26 slices, lies in the range of false detections of the Bonferroni-correction, where we can expect 5 false detections per 100 slices.

	+bspline $\alpha = 0.6$		+dual $\alpha = 1.2$		complex $n = 3$		daub $n = 3$	
	$J = 1$	$J = 2$	$J = 1$	$J = 2$	$J = 1$	$J = 2$	$J = 1$	$J = 2$
C	0	0	0	0	1	1	1	1
	quin $\lambda = 2$		sep $\lambda = 2$					
	$J = 1$	$J = 2$	$J = 1$	$J = 2$				
C	1	0	1	1				

Table 7.6: Table of the number of detections in the whole volume ($128 \times 128 \times 26$) of realigned null-data, i.e., data without activation. The number of detected coefficients lies in the range of Bonferroni correction: 1 coefficient per 26 slices.

7.4.4 Algorithm validation with real data

The results are presented with the real data described in Figure 7.1. Two examples will be shown, "example 1", which contains a weak signal and "example 2", which contains a relatively strong signal.

With real data we cannot compare the different types of wavelets with respect to the error E, but a possibility to assess the results is to compare them to the result of a pixel-based analysis with a t-test and the Bonferroni-correction over a small window. The pixel-based Bonferroni-correction depends on the number of tested pixels and therefore we detect more activated pixels when we choose a window instead of testing the whole image. The window can be defined a priori, since in our cases the locations of activation are known before hand based on the experience of the neurologist.

Fractional splines

With synthetic data, we observed that the error measure E and the $1/$count quantity correlate for +dual only for a low degree parameter α, and for all α in the case of +bspline and of +ortho. Thus another possibility to evaluate the results is to observe the wavelet count $C(\alpha)$.

In this section we also investigate the usefulness of the high-pass in the analysis.

$\rightarrow$ **High SNR.** We observe many high-pass detections, already for one iteration depth $J = 1$. For some α of `+bspline` there are even more detections in the high-pass than in the low-pass. In Table 7.7 the `counts` are given for various degree parameters for `+bspline`. For $\alpha = 0.6$, $C(\alpha)$ is maximal, which supports the proposition in Section 7.3.4 to choose a degree parameter between $\alpha = 0.2 - 0.6$. According to our findings with synthetic data comparing the result to `+dual`, `+bspline` is competitive for low noise data, since it has good localization properties. It typically has lower $E1-$errors and has a constant short range of best degree parameter α. Due to the good detection-rate in the low-pass band, `+dual` is still preferable.

The symmetric fractional splines behave similarly to the causal fractional splines.

	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$	$\alpha = 1.2$
`count` in low-pass	34	36	34	32	33	31
`count` in high-pass	33	34	39	35	34	27
all `count`	67	70	73	67	67	58

Table 7.7: Detection with `+bspline`. The type `bspline` controls the $E1$-error (false detections). We saw in the experiments that the number of errors E correlates inversely with `count`. We find the highest `count` for $\alpha = 0.6$. This reinforces the findings with the synthetic data, where we found the optimal α to be in the range of $\alpha = 0.2 - 0.6$. Notice that many coefficients are detected in the high-pass band.

They detect slightly less coefficients than the causal splines, which could be seen also with synthetic data.

$\rightarrow$ **High-pass detections.** The gain over low-pass detection alone and adding the high-pass is illustrated for "example 1" in Figure 7.47 with `+bspline`. When adding the high-pass, some low-pass detections are missed, since the threshold increases, but more high-pass detections occur. In this example it seems to be worthy to add the high-pass since some additional regions are detected whereas only few pixels are lost compared to the "low-pass only" case. Also the number of detections in the high-pass indicates their importance: for the chosen parameters, more high-pass detections are made than low-pass detections. This tends to indicate that, with a low-pass filter alone, the SPM method might fail to detect high frequency activations that are likely to be detected by our wavelet method. Also noticeable are the finer contours. In Figure 7.48 the negative activations are illustrated (its meaning is discussed in Section 3.3.3). With the high-pass we clearly detect more pixels without losing pixels compared to the detection with the low-pass only. Comparing against detecting with a pixel-based Bonferroni over a window, the detections due

to the high-pass seem to be correct. Thus the high-pass is able to detect negative activated regions, also when they bound positive activated regions, as is the case in this example, see Figure 7.48 a).

Bonferroni over window

a)

+bspline   +bspline

b) low-pass only   c) low- and high-pass
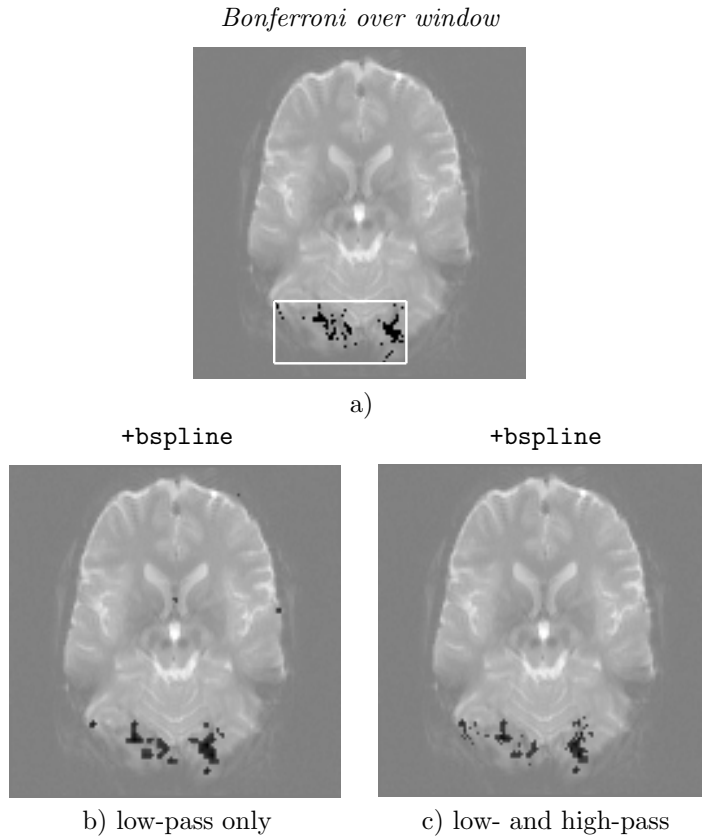
Figure 7.47: Detected activation in "example 1". a) pixel-based Bonferroni over only the window. b) **+bspline** with low-pass only and a degree parameter of $\alpha = 0.6$. c) **+bspline** low- and high-pass with $\alpha = 0.6$.

Another advantage of the high-pass is illustrated in Figures 7.49 and 7.50. The detected region gets clumpy with increasing iteration depth J when detecting only with the low-pass, and eventually disappears completely (Figure 7.49). When adding the high-pass for detection, the shape of the detected region gets better preserved and eventually, when $J \to \infty$, remains constant, since the detections are made with the high-pass only for higher iterations.

Bonferroni over window

a)

negative activation   *negative activation*

b) low-pass only   c) low- and high-pass
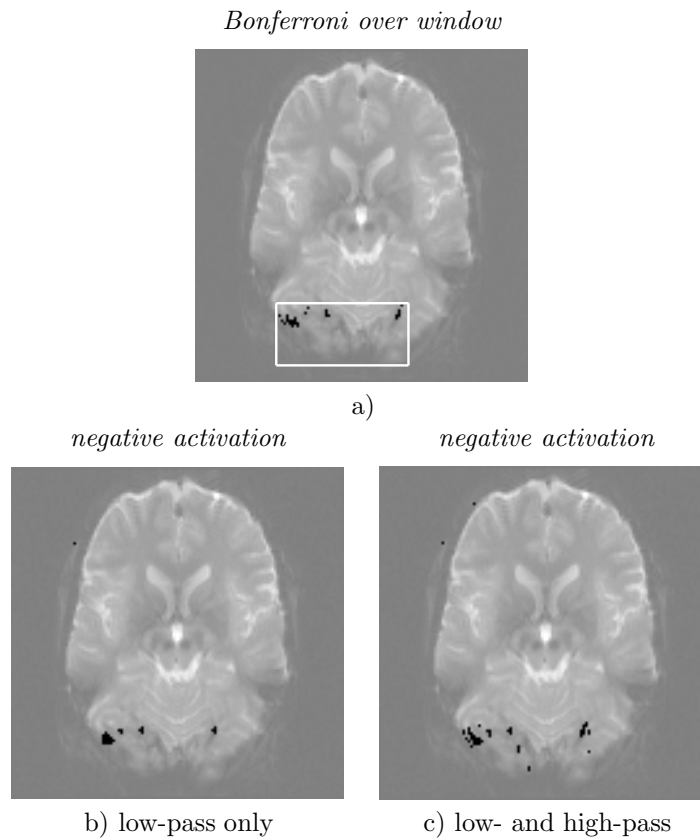
Figure 7.48: Detected negative activation of "example 1". a) Detection with testing for negative activation only and Bonferroni-correction over window. b) Negative activation (in black) detected with +bspline with low-pass only and a degree parameter of $\alpha = 0.6$ c) Negative activation (in black) detected with +bspline low- and high-pass also with $\alpha = 0.6$.

1 iteration *2 iterations*

3 iterations *4 iterations*

5 iterations *6 iterations*
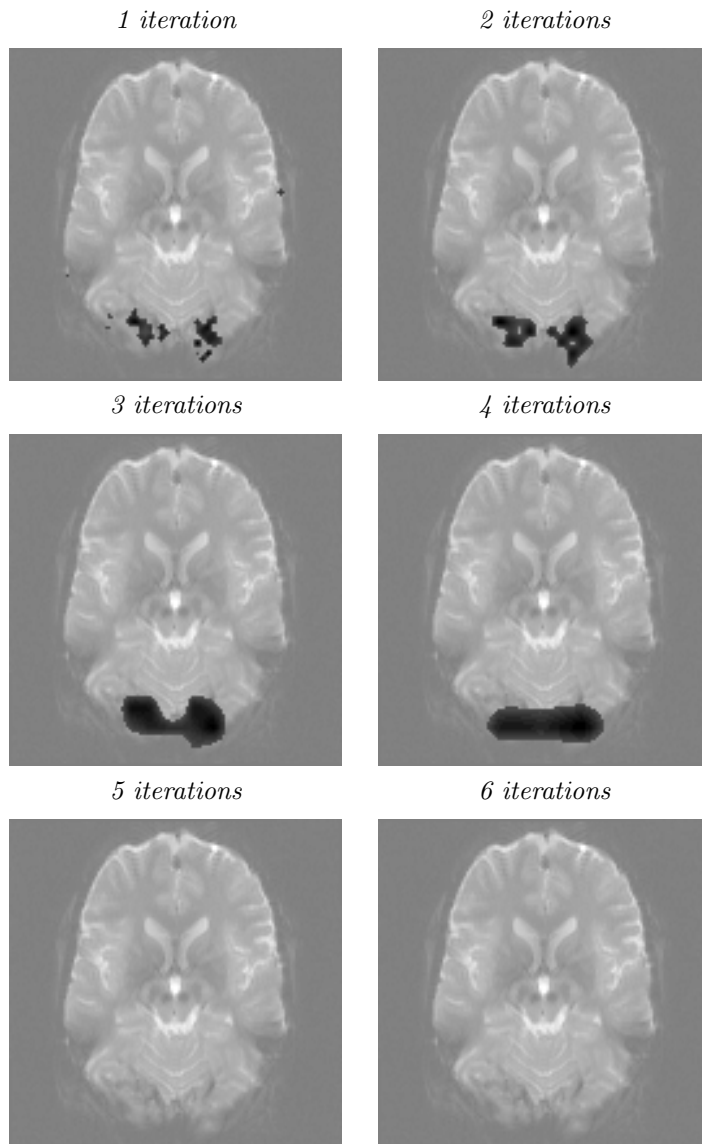
Figure 7.49: Result of our method for `+dual` with $\alpha = 1.2$ only with low-pass. Increasing iterations, from the upper left image down to the lower right image. The detected region becomes clumpy with increasing iteration, and eventually disappears.
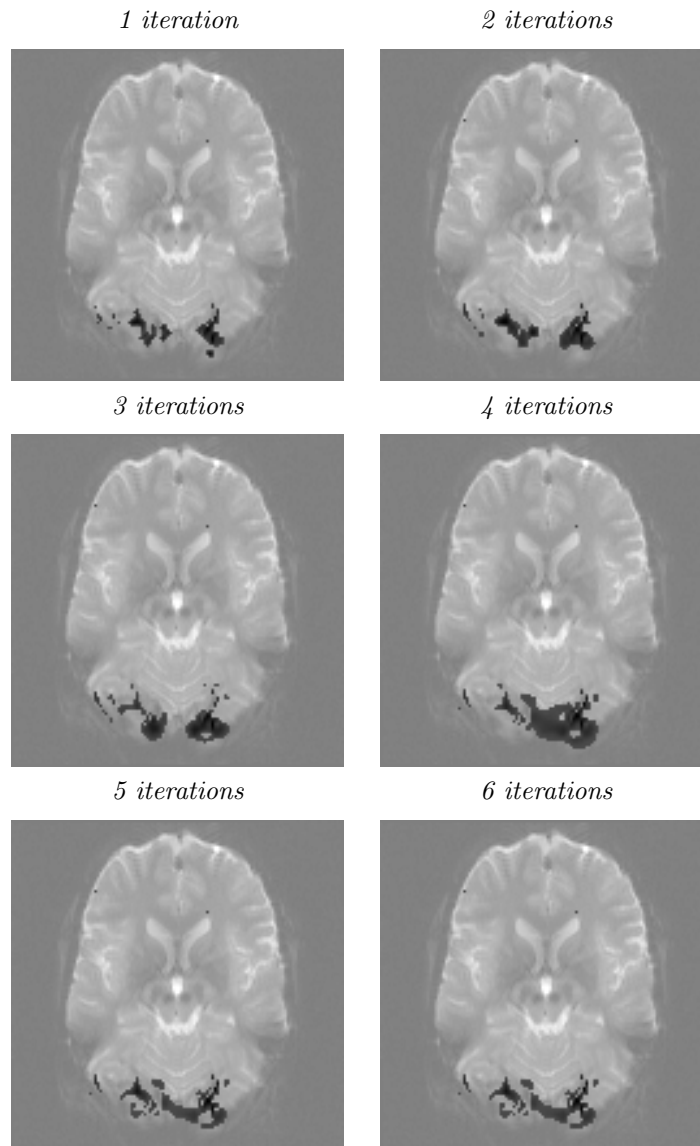
Figure 7.50: Result of method for +dual with $\alpha = 1.2$ with low- and high-pass. Increasing iterations, from the upper left image to the lower right image.
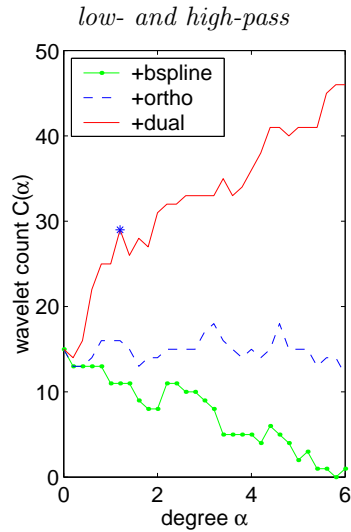
Figure 7.51: `count` of the causal fractional splines for "example 2". The star marks the degree parameter for $\alpha = 1.2$. The characteristics of these curves represent what we found also with synthetic data.

$\rightarrow$ **Low SNR.** When the signal is weak or when the noise is large, only few detections are observed in the high-pass, which questions the utility of the high-pass in this case. "Example 2" is an example of data with a weak signal. It can be observed that the number of detected coefficients of the different types differ all the more as the SNR decreases. `dual` detects in this case much more than `ortho` and `bspline`.

The qualitative detections only with statistics, but without Bonferroni corrections (Figure 7.52 b)) give a reference of what should at least be detected in the region of interest, which is hard to see, since 5% of the pixels are falsely detected. With a pixel-based Bonferroni-correction there is almost no detection at all (see Figure 7.52 a)). Comparing the connectivity of the activation pattern without Bonferroni-correction (i.e. the windows in Figure 7.52 a)), we detect most accurately with `+dual` with a degree parameter $\alpha = 1.6$ and applying the low-pass only. With a lower degree parameter or with the high-pass, one connection of the activation pattern gets lost, even with one iteration. In order to exploit the high-pass efficiently, two iterations ($J = 2$) seem necessary (Figure 7.52 d)).

To conclude, `+dual` performs the best, especially when high noise is present.
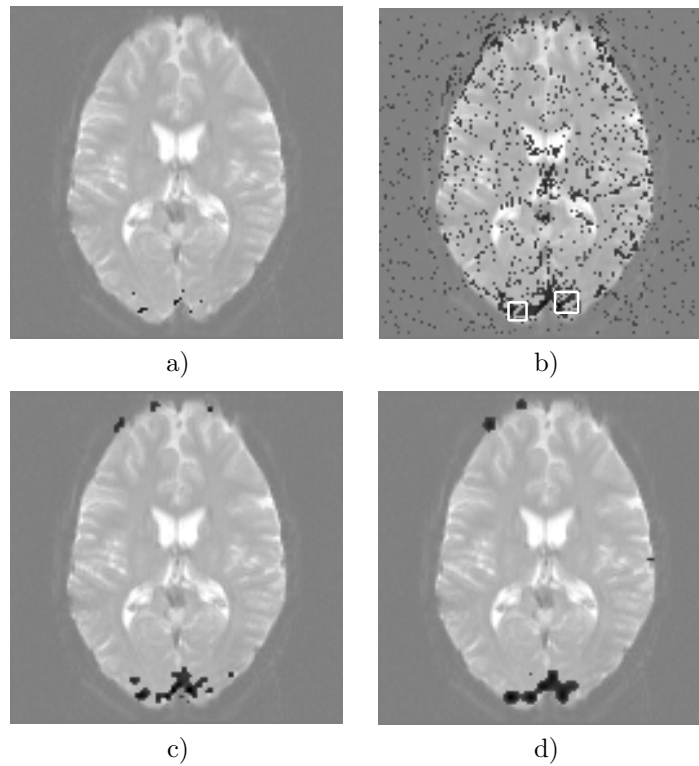
Figure 7.52: a) Pixel-based detection with Bonferroni-correction. b) Pixel-based detection without multiple corrections. The windows show two cases: one case where there is a connection, another case where no connection is detected with an uncorrected test, this means that probably there is effectively no connection or at least with little chance. c) For one iteration, `+dual` with degree parameter $\alpha = 1.6$ and with low-pass only gives a result that matches highly with the detections observed in b). d) For two iterations, `+dual` and $\alpha = 1.2$ with high-pass gives a comparable result, but misses isolated activation regions.

iteration	$J = 1$			$J = 2$	$J = 3$
shift	1	2	3	mean of all shifts	mean of all shifts
$d(I_{\mathrm{est}}^0, I_{\mathrm{est}}^{\mathrm{shift}})$ for `complex`	0	0	0	139.8	176.2
$d(I_{\mathrm{est}}^0, I_{\mathrm{est}}^{\mathrm{shift}})$ for `daub`	91	93	96	136.9	231.2

Table 7.8: Comparing shift-invariance ($d(I_{\mathrm{est}}^0, I_{\mathrm{est}}^{\mathrm{shift}})$, see equation (7.24)) between real Daubechies (`daub`) and complex Daubechies (`complex`), depending on the iteration depth J and on the shift, where "1" stands for a shift of one pixel along the row, "2" stands for a shift of one pixel along the column and "3" stands for a diagonal shift of one pixel. For the first iteration, the complex Daubechies is completely shift invariant, whereas the Daubechies detection changes with the shift. For more iterations, we show only the mean value of all shifts, since we have more than three shifts in those cases.

Complex daubechies

→ **Shift invariance.** Real wavelets are shift-variant and thus are likely to give different results when the activation is shifted. We show in this section the influence of shift-variance on an example and validate the shift-invariance of the complex Daubechies. Thus we analyze "example 1" with all possible three shifts (for one iteration)—one shift along the row, one shift along the column and one shift along the diagonal—and compare it to the first analysis with no shift (marked with 0). In Figure 7.53 the result of `complex` and the influence of shift-variance by `daub` are presented. A distance map to measure the matching error between I_{est}^0 (the non-shifted data set) and $I_{\mathrm{est}}^{\mathrm{shift}}$ (shifted data set) is evaluated:

$$d_{\mathrm{est}} = \begin{cases} 0, & (x, y) \text{ is in the same activation state in } I_{\mathrm{est}}^0 \text{ and in } I_{\mathrm{est}}^{\mathrm{shift}}, \\ 1, & \text{otherwise.} \end{cases}$$

The final measure is given by

$$d(I_{\mathrm{est}}^0, I_{\mathrm{est}}^{\mathrm{shift}}) = \sum_{x,y} d_{\mathrm{est}}(x, y) \tag{7.24}$$

and is illustrated in Table 7.8 for the order parameter 3 and for different iterations. For the first iteration, the complex Daubechies (`complex`) show complete shift-invariance, whereas real Daubechies (`daub`) distance measure is larger than zero. For more iterations, `complex` is not shift-invariant anymore. For higher noise levels ("example 2") the visual difference between the detection of shifted images is even larger for `daub`, whereas `complex` remains shift-invariant for the first iteration.

→ **count.** With synthetic data, the complex Daubechies (`complex`) detects more coefficients than the real Daubechies `daub` (Section 7.3.5). With real data, we

a) `complex`   b) `daub`   c) `daub` of shifted image
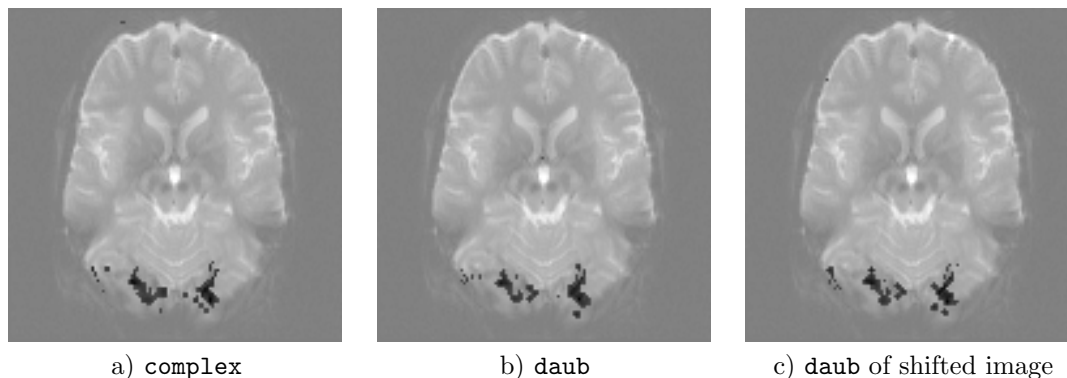
Figure 7.53: Detection map of "example 1" for `complex` and for `daub`. Comparing b) and c) shows the influence of a shift-variant detection on the detection pattern.

	low-pass	high-pass	low- and high-pass
`complex`	61	56	117
`daub`	33	22	55

Table 7.9: Data "example 1": `count` for complex Daubechies and real Daubechies for the order parameter $n = 3$ and one iteration.

observe that `complex` detects more in the low- and in the high-pass, as shown in Table 7.9 for the data set "example 1". The qualitative detections are illustrated in Figure 7.53.

$\rightarrow$ **High SNR.** The qualitative results resemble the one found for `+dual`, compare Figures 7.50 and 7.53.

$\rightarrow$ **Low SNR.** When the noise is relatively high, for the first iteration `+daub` seems to perform visually better than `complex`, since one connectivity is missing in the case of the complex Daubechies. For the second iteration, the results are very similar.

Quincunx

$\rightarrow$ **High SNR.** For low noise, the best order parameter λ is difficult to determine. However, $\lambda = 2$ gave good results with synthetic data. The resulting activation regions are very close to the one found with `+dual` and `+bspline`. The contours are finer, when only one iteration is used corresponding to less subsampling than in the separable case. Thus with quincunx, we also seem to gain some more detected coefficients compared to `sep`.
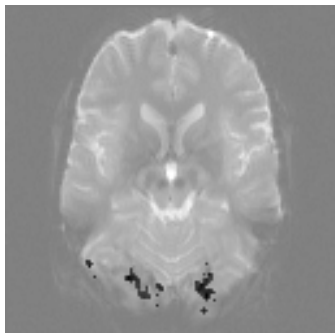
Figure 7.54: Detection with quincunx for one iteration and the order parameter $\lambda = 2$.

$\rightarrow$ **Low SNR.** When high noise is present, a large λ allows to detect more coefficients in the low- and in the high-pass. Then, $\lambda = 6$ is a good choice when the SNR is low. For $\lambda = 6$, `quin` detects some coefficients more for the first and the second iteration than `sep`.

Comparison with SPM

We also analyzed "example 2" and "example 1" with SPM, the standard fMRI-method, and, for control, with the two-sample t-test. We preprocessed the data with a Gaussian kernel of the size of $2\times$ the resolution and $3\times$ the resolution, i.e., the lower and upper bound of pre-smoothing, which is recommended by SPM [147]. Since these two extremal analyses give a good indication of what happens in between these parameters, we have not tried to find the optimal size of the Gaussian kernel. When we choose a large Gaussian kernel, the large activations are spreading and close activation zones melt together. In contrast, very small activation zones disappear. For even larger Gaussian kernels, all activation zones disappear.

When specifying the design-matrix, we choose for the type of response the fixed response (Box-Car), in order to be close to our previous analysis. We did not convolve with a hemodynamic response function (HRF) and did not add temporal derivatives. To estimate the specified model, we did not take advantage of the choice of removing global effects, nor applied a high- or low-pass filter along the time dimension.

Figure 7.55 displays the results of SPM for "example 1". We show here the results of the analysis of the two-sample t-test, since the standard fMRI-method gives even less detections. Comparing these results with the activation map of the method `+bspline` in Figure 7.47, we notice that the results of SPM, which are in general close to ours, are much poorer than the one of `+bspline`, which controls the false detections. One explanation could be that the signal which contains substantial high-frequencies is suppressed by the

Gaussian filter. The detected activation with Bonferroni of "example 1" and the results

a) 2× resolution   b) 3× resolution
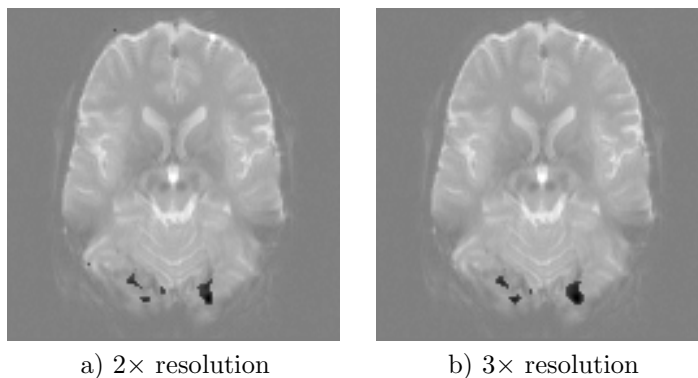
Figure 7.55: Result of "example 1" with SPM using the paired t-test method.

of Quincunx and Complex Daubechies are presented in Figure 7.56.

a) Pixel-based Bonferroni   b) Quincunx   c) Complex Daubechies
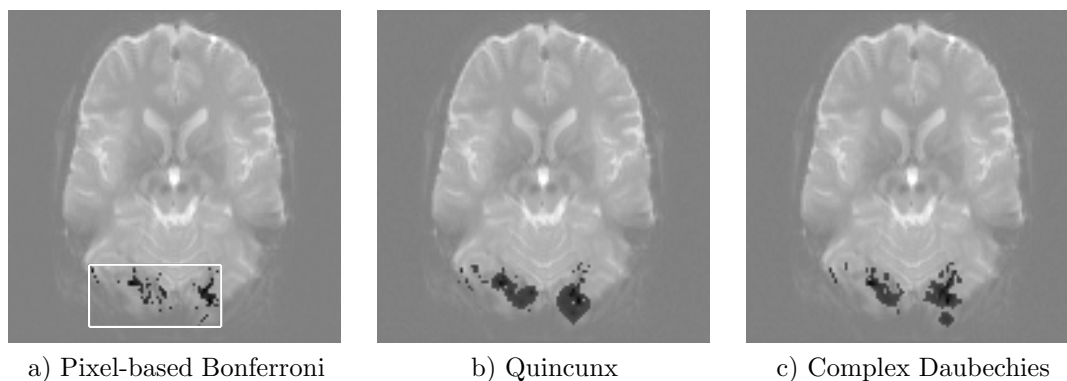
Figure 7.56: a) Result of pixel-based Bonferroni over the window. b) Quincunx: result with best match with respect to the detections of a). c) Complex Daubechies: result with good match with respect to a).

7.4.5   Validation of our 3D algorithm

The analyses which were done in two dimensions can also be done in three dimensions. We have implemented a 3D algorithm for the different types of wavelets—fractional splines, complex wavelets, and quincunx. This algorithm can be applied to volumes

which have isotropic voxels, which is not the case of many datasets. The described data set is acquired by the Wellcome Department of Imaging Neuroscience, University College London, as described in Table 7.5. The patient's task was to listen to bi-syllabic words. In Figure 7.57, we compare the results obtained for `+dual` with the results obtained by SPM. In general, the results are similar. Nevertheless, it can be observed that some activation details are not detected with SPM, but are detected with the method `+dual`. To decide whether this activation is false or right, we applied pixel-based Bonferroni over a small window. When detected with this method, we can be sure that this activation is true; when not detected it is still possible that this activation is true, but we cannot ensure it. An example is the window of the left side of the brain in Figure 7.57. With Bonferroni, we detect a line-like structure in the observation window; with `+dual`, the detected pattern reaches almost the border; with SPM smoothed with $2\times$ the resolution, the activation is shorter and with a smoothing of $3\times$ the resolution this extension is missing. We seem to gain something with the method of `+dual`. However a little pattern detected with SPM, but not with `+dual` at the right side of the brain, could not be ascertained by pixel-based Bonferroni, which doesn't mean that there is no activation, but more probably that this activation is weak. We tested the different methods over 18 windows in this data set. Some results are very close to each other. However, by closer comparisons, `+dual`, complex Daubechies with two iterations, followed by quincunx with four iterations yielded good results with the fewest missed detections compared to pixel-based Bonferroni over a window.

7.5 Conclusion

A critical issue for detection methods that only use the low-pass information is the fact that not the whole region of a potentially low signal is detected with little smoothing. By increasing the smoothing factor, a larger region of low signal gets better detected, but small regions, possibly of strong signal, disappear. Therefore, a multiscale approach is more appropriate.

Adding the high-pass information of the wavelet-transform requires the use of a higher threshold to correct for multiple testing. However we could show that the gain of detection quality by keeping the signal details is in general worthwhile. Especially it avoids the effect of clumpy regions or regions melting together when more smoothing is necessary to increase the SNR by keeping smaller activation regions.

Not all wavelet bases lead to fully satisfactory detection results. Orthogonality—which is usually used for detection—is not a necessary property; for instance, we found that bi-orthogonal wavelets, in particular the causal dual fractional spline, led to especially good results. Experiments with synthetic, real data, and theoretical considerations each showed that a small range of the order parameter—$\alpha = 1.2 \ldots 1.6$—yields the best results.

Another property, which improves the detection substantially, is the finer scale progres-
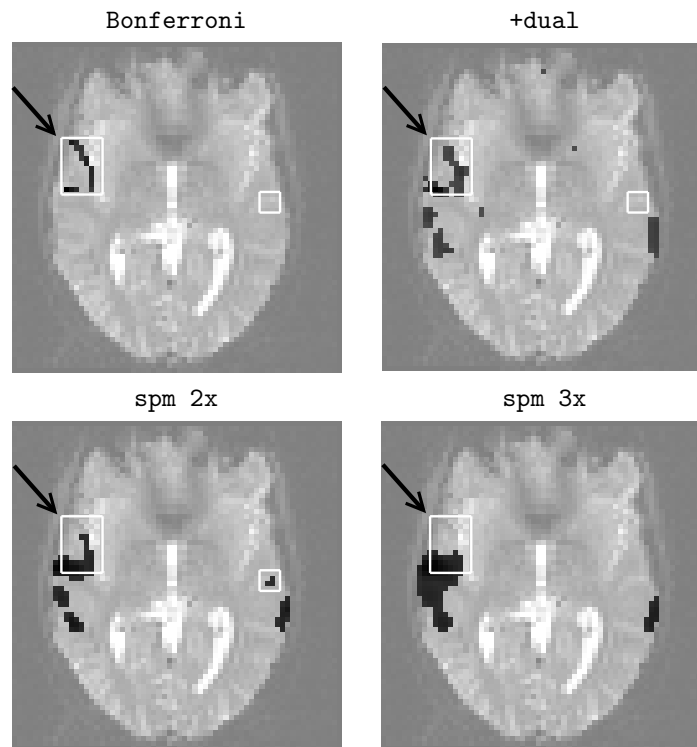
Figure 7.57: Analysis of a 3D data set. The detections in question are tested with pixel-based Bonferroni over a window. The method +dual seems to be closer to the detections of Bonferroni over a window than SPM.

sion of the quincunx wavelet transform. It potentially leads to better results; in one case study, we have observed as much as 20% improvement.

The redundant approach of the complex wavelet transform leads to better shift-invariance and yields good results for fMRI as well. Here instead of a t-test, Hotelling's T^2-test is used.

Chapter 8

Conclusion and Outlook

In this work, we applied recent multiresolution techniques to the fMRI problem. We showed that the wavelet approach was promising and that the fact of retaining the finer scale coefficients was generally beneficial.

The contributions of the thesis can be summarized as follows. First, we introduced a new family of orthogonal wavelet transforms for quincunx lattices. A key feature is the continuously-varying order parameter λ which can be used to adjust the band-pass characteristics as well as the localization of the basis functions. We showed that these quincunx wavelets have better approximation properties than comparable separable wavelets. We also demonstrated that these transforms can be computed efficiently using FFTs. This should help dispel the commonly held belief that nonseparable wavelet decompositions are computationally much more demanding than separable ones. Because of their nice properties and their ease of implementation, these wavelets present a good alternative to the separable ones that are used in a variety of image processing applications (data compression, filtering, texture analysis and so on).

Second, we investigated a wavelet-based approach for the analysis of fMRI. In particular, we identified, using various test data, the wavelet properties that are important for the present application. For instance, we found that orthogonality—which is usually used for detection—is not a crucial property. Our best results were obtained with bi-orthogonal wavelets, especially with short and selective analysis filters such as B-spline filters (corresponding to the "dual" fractional spline wavelet transform). Experiments with synthetic, real data, and theoretical considerations all indicated that a small range of the order parameter—$\alpha = 1.2 \ldots 1.6$—yields the best results.

Another property, which can improve the detection, is the finer scale progression of the quincunx wavelet transform. It leads to either similar or better results.

The redundant approach of the complex wavelet transform gives better shift-invariance, a property which improves the results for fMRI as well.

Analyzing the noise of real data, we could show that a non-stationary model is more appropriate for fMRI than the white-noise model which is often considered. In other words, it seems to be more appropriate to use a t-test rather than a z-test. We further proposed to apply Hotelling's T^2-test, which yielded good results when applied to the complex wavelet transform.

Outlook

The main possible extensions of this work are as follows:

- Our 3D algorithm of analyzing fMRI data could be adjusted to compute also appropriately data with voxels of anisotropic size.

- Bi-orthogonality and causality proved to be properties which enhance the detection results significantly. Since the quincunx properties yielded good results as well, it would be appropriate to combine these properties by designing dual quincunx wavelets. An extension to complex dual quincunx wavelets could be considered as well.

- The statistical analysis could be enhanced by treating the low-pass and the band-pass individually with an adaptive threshold.

Acknowledgments

I would like to thank my thesis advisor Prof. Michael Unser who offered me the possibility to carry out this work and for his scientific support. He also gave me the opportunity to take part on several international conferences. I also thank Thierry Blu for his helpful scientific support and advises. He made great efforts for the review of the draft of this thesis. I'm grateful to Dimitri van de Ville, Michael Sühling and Muthuvel Arigovindan who also reviewed the draft very carefully. Thanks to our system administrator Daniel Sage for his availability and to Philippe Thévenaz for offering his precise knowledge.

My sincere thanks go also to Prof. Hannes Bleuler, Prof. Wulfram Gerstner, Dr. François Lazeyras, Dr. Elena Loli Piccolomini and Dr. Fabiana Zama for having accepted to participate as jury of this thesis.

I wish to thank Dr. Arto Nirkko from the Inselspital Bern for providing us with data and for the very informative discussions contributing a lot to the fMRI part. Many thanks to Dr. François Lazeyras from the University Hospital of Geneva for providing us also with data and for being always open for questions.

I would like to acknowledge the kind visit and helpful discussions with Dr. Jean Baptiste Poline.

J'aimerais remercier tous mes collègues de l'institut d'optique appliquée (IOA), devenu depuis peu BIO-E, pour la très bonne atmosphère et pour leur support. Notamment merci de m'avoir fait découvrir le valais.

Je tiens aussi à remercier toutes les personnes de l'EPFL qui m'ont soutenue moralement et financièrement, spécialement lors de la phase finale de mon travail. Merci pour votre aide compétente, c'étais très agréable et motivant de se sentir encadrée.

Besonderer Dank gilt Prof. Peter Ryser für seine wertvollen Ratschläge und seine Ermutigungen.

Merci à tous mes amis, ainsi que ceux que j'ai rencontré lors de mon séjour à Lausanne, pour les bons moments que nous avons passé ensemble. J'aimerais particulièrement remercier Caroline Jacq pour la bonne équipe de jogging, avec qui on a fait un bon nombre de kilomètres.

Meinen herzlichsten Dank auch meinen Eltern und meiner Schwester Andrea, auf

deren vielfaeltige Unterstützung ich immer zaehlen konnte. Diese Arbeit ist ihnen gewidmet.

Abbreviations and Acronyms

ANOVA	**An**alysis **of** **Va**riance
ASL	**A**rterial **S**pin **L**abelled perfusion MRI
BOLD	**B**lood **O**xygenation **L**evel **D**ependent
CBF	**C**erebral **B**lood **F**low
CQF	**C**onjugate **Q**uadrature **F**ilter
CSF	**C**erebro**S**pinal **F**luid
CWT	**C**omplex **W**avelet **T**ransform
DWT	**D**iscrete **W**avelet **T**ransform
EER	**E**qual **E**rror **R**ate
EPI	**E**cho**p**lanar **I**maging
ER-fMRI	**E**xperiment **P**rotocol-**fMRI**
ERPs	**E**vent-**R**elated **P**otential**s**
FA	**F**actor **A**nalysis
FAR	**F**alse **A**cceptance **R**ate
FCA	**F**uzzy **C**lustering **A**nalysis

FCO	**F**ace-**C**entered **O**rthorombic
FDR	**F**alse **D**iscovery **R**ate
FFT	**F**ast **F**ourier **T**ransform
FIL	**F**unctional **I**maging **L**aboratory
FIR	**F**inite **I**mpulse **R**esponse
FLASH	**F**ast **L**ow **A**ngle **Sh**ot
fMRI	**f**unctional **M**agnetic **R**esonance **I**maging
FRR	**F**alse **R**ejection **R**ate
FSWT	**F**ractional **S**pline **W**avelet **T**ransform
FWHM	**F**ull **W**idth of **H**alf **M**aximum
GRE	**GR**adient **E**cho
GRF	**G**aussian **R**andom **F**ield
HCUG	**H**ôpital **C**antonal **U**niversitaire **G**enève
HRF	**H**emodynamic **R**esponse **F**unction
IDWT	**I**nverse **D**iscrete **W**avelet **T**ransform
IFFT	**I**nverse **F**ast **F**ourier **T**ransform
IIR	**I**nfinite **I**mpulse **R**esponse
ISI	**I**nter-**S**timulus **I**nterval
JPEG	**J**oint **P**hotographic **E**xperts **G**roup
MANOVA	**M**ultivariate **An**alysis **o**f **Va**riance

MRA	**M**ulti**r**esolution **A**nalysis
MRI	**M**agnetic **R**esonance **I**maging
NIH	**N**ational **I**nstitute of **H**ealth
NMR	**N**uclear **M**agnetic **R**esonance
rCBF	**r**egional **C**erebral **B**lood **F**low
RESELS	**Res**olution **El**ement**s**
RF	**R**adio-**f**requency
RFT	**R**andom **F**ield **T**heory
RMS	**R**oot **M**ean **S**quare
ROC	**R**eceiver **O**perator **C**haracteristic
PCA	**P**rincipal **C**omponent **A**nalysis
PCM	**P**ulse **C**ode **M**odulation
PET	**P**ositron **E**mission **T**omography
SBC	**S**ub**B**and **C**oding
SNR	**S**ignal-to-**N**oise **R**atio
SPECT	**S**ingle **P**hoton **E**mission **C**omputed **T**omography
SPM	**S**tatistical **P**arameter **M**apping
TE	**E**cho **T**ime
TER	**T**otal **E**rror **R**ate
TR	**R**epetition **T**ime
WEA	**W**avelet **E**stimated **A**ctivation

WFT   **W**indowed **F**ourier **T**ransform

Bibliography

[1] "Architecture of the brain", http://www.arts.uwa.edu.au/LingWWW/-LIN102_99/Notes/brain.html.

[2] R. J. Adler, *The Geometry of Random Fields*. Wiley, New York, 1981.

[3] R. J. Adler and A. M. Hasofer, "Level crossings for random fields", *Annals of Probability*, vol. 4, pp. 1–12, 1976.

[4] G. K. Aguirre, J. A. Detre, E. Zarahn and D. C. Alsop, "Experimental design and the relative sensitivity of BOLD and perfusion fMRI", *NeuroImage*, vol. 15, pp. 488–500, 2002.

[5] A. Aldroubi. *Wavelets in Medicine and Biology.* CRC Press, ch. 1. The Wavelet Transform: A Surfing Guide, pp. 3–36, 1996.

[6] M. Alonso and E. Finn, *Quanten-Physik*. Addison-Wesley, 1988.

[7] S. F. Arnold, *Mathematical Statistics*. Prentice-Hall, 1990.

[8] J. Ashburner and K. Friston, *Human Brain Function*. Academic Press, 1997.

[9] J. Ashburner and K. J. Friston, "Nonlinear spatial normalization using basis functions", *Human Brain Mapping*, vol. 7, pp. 254–266, 1999.

[10] W. Backfrieder, R. Baumgartner and M. Samal, "Quantification of intensity variations in fMRI using rotated principal components", *Phys. Med. Biol.*, vol. 41, pp. 1425–1438, 1996.

[11] P. Bandettini, A. Jesmanowicz, E. Wong and J. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain", *Magnetic Resonance in Medicine*, vol. 30, pp. 161–173, 1993.

[12] A. Bartels and S. Zeki, "The neural basis of romantic love", *NeuroReport*, vol. 11, pp. 3829–3834, 2000.

[13] M. Barth. *Multiple Gradient-Echo Functional Magnetic Resonance Imaging of the Human Brain*. PhD thesis, Technische Universität Wien, Wien, 1999.

[14] R. Baumgartner. *Post-processing Strategies in Functional MRI of the Human Brain*. PhD thesis, Technische Universität Wien, Wien, 1998.

[15] R. Beisteiner, M. Erdler, C. Teichtmeister, M. Diemling, E. Moser, V. Edward and L. Deecke, "Magnetoencephalography (MEG) may help to improve functional MRI brain mapping", *Eur. J. Neurosci.*, vol. 9, pp. 101–107, 1997.

[16] B. Belzer, J. Lina and J. Villasenor, "Complex, linear-phase filters for efficient image coding", *IEEE Trans. Signal Processing*, vol. 43, pp. 2425–2427, 1995.

[17] B. Bernal, "Understanding the auditory-fMRI", http://www.mch.com/clinical/-radiology/fmri/understandingauditoryfmri3.htm.

[18] B. Biswal, E. DeYoe and J. Hyde, "Reduction of physiological fluctuations in fMRI using digital filters", *Magnetic Resonance in Medicine*, vol. 35, pp. 107–113, 1996.

[19] R. Bloch, W. Hansen and M. Packard, "Nuclear induction", *Phys. Rev.*, vol. 69, p. 127, 1946.

[20] T. Blu and M. Unser. "The fractional spline wavelet transform: Definition and implementation". In *Proc. ICASSP*, Istanbul, Turkey, vol. 1, pp. 512–515, June 2000.

[21] C. E. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilità", *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, vol. 8, pp. 3–62, 1936.

[22] M. J. Brammer, "Multidimensional wavelet analysis of functional magnetic resonance images", *Human Brain Mapping*, vol. 6, pp. 378–382, 1998.

[23] P. Broca, "Remarques sur le siège de la faculté du langage articulé; suivies d'une observation d'aphémie", *Bull. Soc. Anat. Paris*, vol. 6, pp. 330–357, 1861.

[24] T. Budinger, F. Wehrli, et al., *Mathematics and Physics of Emerging Biomedical Imaging*. National Academic Press, 1996.

[25] E. Bullmore, M. Brammer, S. Rabe-Hesketh, V. Curtis, R. Morris, S. Williams, T. Sharma and P. McGuire, "Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI", *Human Brain Mapping*, vol. 7, pp. 38–48, 1999.

[26] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter and M. Brammer, "Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains.", *Human Brain Mapping*, vol. 12, pp. 61–78, 2001.

[27] M. Burock, R. Buckner, M. Woldorff, B. Rosen and A. Dale, "Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI", *NeuroReport*, vol. 9, pp. 3735–3739, 1998.

[28] C. Chatfield, *Statistics for Technology.* Chapman & Hall, 1983.

[29] H. Choi, J. Romberg, R. Baraniuk and N. Kingsbury. "Hidden Markov tree modeling of complex wavelet transforms". In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2000*, vol. 1, pp. 133–136, 2000.

[30] S. Clare. *Functional MRI: Methods and Applications.* PhD thesis, University of Nottingham, 1997.

[31] A. Cohen and I. Daubechies, "Nonseparable bidimensional wavelet bases", *Rev. Mat. Iberoamericana*, vol. 9, pp. 51–137, 1993.

[32] M. Cohen, "Echo-planar imaging (EPI) and functional MRI", http://www.uib.no/-med/avd/miapr/arvid/MOD3/Diffusjon/EPI_fMRI_Cohen.pdf, 1998.

[33] P. Coifman and D. Donoho. "Translation-invariant denoising". In *Wavelets and Statistics; Lecture notes in statistics*, A. Antoniadis and G. Oppenheim, Eds., vol. 103. Springer, 1995, pp. 125–150.

[34] R. T. Constable and P. Skudlarski, "An ROC approach for evaluating functional brain MR imaging and postprocessing protocols.", *Magnetic Resonance in Medicine*, vol. 34, pp. 57–64, 1995.

[35] C. Costamagna, M. Feilner and M. Unser, "Détection de microcalcifications en mammographie", 2002. Projet de semestre.

[36] R. Cox, "Software for analysis and visualization of functional magnetic resonance neuroimages", *Computers and Biomedical Research*, vol. 29, pp. 162–173, 1996.

[37] R. Cox. "A two day workshop on functional MRI". In *Proc. Colloque de Physique C2*, June 1996.

[38] R. Cox and J. Hyde, "Software tools for analysis and visualization of fMRI data", *NMR in Biomedicine*, vol. 10, pp. 171–178, 1997.

[39] M. Dagli, J. Ingeholm and J. Haxby, "Localization of cardiac-induced signal change in fMRI", *NeuroImage*, vol. 9, pp. 407–415, 1999.

[40] R. Damadian, "Tumor detection by nuclear magnetic resonance", *Science*, vol. 171, pp. 1151–1153, 1971.

[41] R. Damadian. "The history of MRI". In *High Care '97 International Congress Bochum*, Bochum, Germany, January 1997.

[42] I. Daubechies, "Orthonormal bases of compactly supported wavelets", *Comm. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.

[43] I. Daubechies, *Ten Lectures on Wavelets*. Siam, 1992.

[44] M. Desco, J. A. Hernandez, A. Santos and M. Brammer, "Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation", *Human Brain Mapping*, vol. 14, pp. 16–27, 2001.

[45] J. A. Detre, "Perfusion fMRI", http://www.lde.upenn.edu/brainlang/Detre.html.

[46] R. A. DeVore, B. Jawerth and B. J. Lucier, "Image compression through wavelet transform coding", *IEEE Trans. Inform. Th.*, vol. 38, no. 2, pp. 719–746, 1992.

[47] D. L. Donoho, M. Vetterli, R. A. DeVore and I. Daubechies, "Data compression and harmonic analysis", *IEEE Trans. Inform. Th.*, vol. 44, no. 6, pp. 2435–2475, 1998.

[48] E. Dubois, "The sampling and reconstruction of time-varying imagery wih application in video systems", *Proc. of the IEEE*, vol. 73, no. 4, pp. 502–522, 1985.

[49] V. Ducry, "Introduction to functional magnetic resonance imaging", 2002.

[50] W. Eddy, M. Fitzgerald, C. Genovese, A. Mockus and D. Noll. "Functional imaging analysis software—computational Olio". In *Proceedings in Computational Statistics*, Physica-Verlag, 1996.

[51] W. A. Edelstein, G. H. Glover, C. J. Hardy and R. W. Redington, "The intrinsic signal-to-noise ratio in NMR imaging", *Magnetic Resonance in Medicine*, vol. 3, pp. 604–618, 1986.

[52] D. Ekatodramis. *A Hypothesis-Testing Approach to Functional Inference of Brain Activity from Magnetic Resonance Image Series*. PhD thesis, Swiss Federal Institute of Technology Zurich, Zürich, 1998.

[53] J. Ellermann and K. Ugurbil, "Bilder aus dem Gehirn, funktionelle Bildgebung mit NMR", *Physik in unserer Zeit*, vol. 1, pp. 17–27, 1996.

[54] M. Fadili and E. Bullmore, "Wavelet-generalised least squares: a new BLU estimator of regression models with long-memory errors", *NeuroImage*, vol. 15, pp. 217–232, 2001.

[55] M. Fadili, S. Ruan, D. Bloyet and B. Mazoyer, "A multistep unsupervised fuzzy clustering analysis of fMRI time series", *Human Brain Mapping*, vol. 10, pp. 160–178, 2000.

[56] J. C. Feauveau, "Analyse multirésolution avec un facteur de résolution $\sqrt{2}$", *Journal de Traitement du Signal*, vol. 7, no. 2, pp. 117–128, 1990.

[57] M. Feilner, T. Blu and M. Unser. "Statistical analysis of fMRI data using orthogonal filterbanks". In *Proc. SPIE, Wavelet Applications in Signal and Image Processing VII, Denver, CO, USA*, July 19-23 1999.

[58] M. Feilner, T. Blu and M. Unser. "Analysis of fMRI data using spline wavelets". In *Proc. of the Tenth European Signal Processing Conference (EUSIPCO'00), Tampere, CO, Finland*, September 4-8 2000.

[59] M. Feilner, T. Blu and M. Unser. "Optimizing wavelets for the analysis of fMRI data". In *Proc. SPIE, Wavelet Applications in Signal and Image Processing VIII, San Diego, CA, USA*, July 31-August 4 2000.

[60] B. Flury, *A First Course in Multivariate Statistics.* Springer, 1997.

[61] S. Forman, J. Cohen, M. Fitzgerald, W. Eddy, M. Mintun and D. Noll, "Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold", *Magnetic Resonance in Medicine*, vol. 33, pp. 636–647, 1995.

[62] P. T. Fox, "Spatial normalization origins: Objectives, applications, and alternatives", *Human Brain Mapping*, vol. 3, pp. 161–164, 1995.

[63] R. Frackowiak, K. Friston, C. Frith, R. Dolan and J. Mazziotta, *Human Brain Function.* Academic Press, 1997.

[64] K. Friston, "Statistical parametric mapping", http://www.fil.ion.ucl.ac.uk/spm/course/notes01/intro/img3.htm, 2002.

[65] K. Friston, J. Ashburner, J. Poline, C. Frith, J. Heather and R. Frackowiak, "Spatial registration and normalization of images", *Human Brain Mapping*, vol. 2, pp. 165–189, 1995.

[66] K. Friston, C. Frith, P. Liddle and R. Frackowiak, "Comparing functional (PET) images: The assessment of significant change.", *Cereb. Blood Flow Metab.*, vol. 11, pp. 690–699, 1991.

[67] K. Friston, P. Jezzard and R. Turner, "Analysis of fMRI time series", *Human Brain Mapping*, vol. 2, pp. 69–78, 1994.

[68] K. Friston, S. Williams, R. Howard, R. Frackowiak and R. Turner, "Movement-related effects in fMRI time series.", *Magnetic Resonance in Medicine*, vol. 35, pp. 346–355, 1996.

[69] K. Friston, K. Worsley, R. Frackowiak, J. Mazziotta and A. Evans, "Assessing the significance of focal activations using their spatial extent", *Human Brain Mapping*, vol. 1, pp. 214–220, 1994.

[70] G. Gaillard, M. Feilner and M. Unser. "Nouvel algorithme de transformation en ondelettes". Master's thesis, EPFL, 2001.

[71] W. Ganong, *Review of medical physiology*. Prentice-Hall, 1995.

[72] C. Glauser, M. Feilner and M. Unser, "Réduction de bruit dans des images biomédicales par des algorithmes pyramidaux et par des ondelettes", 2000. Projet de semestre.

[73] S. Gold, B. Christian, S. Arndt, G. Zeien, T. Cizadlo, D. Johnson, M. Flaum and N. Andreasen, "Functional MRI statistical software packages: A comparative analysis", *Human Brain Mapping*, vol. 6, pp. 73–84, 1998.

[74] A. Haase, J. Frahm, D. Matthaei, W. Hänicke and k D. Merboldt, "FLASH imaging: Rapid NMR imaging using low flip angle pulses", *Magn. Reson.*, vol. 67, pp. 258–266, 1986.

[75] R. Henson. "Event-related fMRI: Introduction, statistical modelling, design optimisation and example". In *Japanese Journal of Cognitive Neuroscience*, in press.

[76] A. Holmes, "SPM: Data sets", http://www.fil.ion.ucl.ac.uk/spm/data/-#fMRI_MoAEpilot, 2002.

[77] A. Jebali, M. Feilner and M. Unser, "Implémentation en Fourier de la transformée en quinconce à filtres fractionnaires", 2001. Projet de semestre.

[78] M. Kendall, *Multivariate Analysis*. Griffin, 1975.

[79] A. Kertesz, *Localization and neuroimaging in neuropsychology*. Academic Press, 1994.

[80] A. Kleinschmidt, C. Buchel, S. Zeki and R. Frackowiak. "Human brain activity during spontaneously reversing perception of ambiguous figures". In *Proc. R. Soc. London B. Biol. Sci.*, vol. 1, pp. 2427–2433, 1998.

[81] J. Kovačević and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathbb{R}^n$", *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 533–555, 1992.

[82] J. Kovačević and M. Vetterli, "Nonseparable two- and three-dimensional wavelets", *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1269–1273, 1995.

[83] A. Kumar, D. Welti and R. Ernst, "NMR Fourier zeugmatography", *J. Magn. Reson.*, pp. 18–69, 1975.

[84] J. Kybic, P. Thévenaz and A. N. M. Unser, "Unwarping of unidirectionally distorted EPI images", *IEEE Transaction on Medical Imaging*, vol. 19, pp. 80–93, 2000.

[85] N. Lange, S. Strother, J. Anderson, F. Nielson, A. Holmes, T. Kolenda, R. Savoy and L. Hansen, "Plurality and resemblance in fMRI data analysis", *NeuroImage*, vol. 10, pp. 282–303, 1999.

[86] P. Laurienti, J. Burdette, M. Wallace, Y.-F. Yen, A. Field and B. Stein, "Deactivation of sensory-specific cortex by cross-modal stimuli", *Journal of Cogn. Neuroscience*, vol. 14, pp. 1–10, 2002.

[87] P. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance", *Nature*, vol. 242, pp. 190–191, 1973.

[88] W. Lawton, "Applications of complex valued wavelet transforms to subband decomposition", *IEEE Trans. Signal Processing*, vol. 41, pp. 3566–3568, 1993.

[89] E. L. Lehmann, *Testing Statistical Hypotheses*. Springer, 1997.

[90] Z. Liang and P. Lauterbur, *Principles of Magnetic Resonance Imaging*. IEEE Press, 2000.

[91] J. Lina, "Complex Daubechies wavelets", *Applied and Computational Harmonic Analysis*, vol. 2, pp. 219–229, 1995.

[92] J. Lina. "Complex Daubechies wavelets: Filters design and applications". In *Proc. ISAAC Conference, Univ. of Delaware*, 1997.

[93] J. Lina and M. Mayrand. "Parameterizations for complex Daubechies wavelets". In *Proc. SPIE Wavelet Applicat. Conf., Orlando, FL, USA*, 1994.

[94] A. Macovski, "Noise in MRI", *Magnetic Resonance in Medicine*, vol. 36, pp. 494–497, 1996.

[95] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.

[96] P. Mansfield, "Multi-planar image formation using NMR spin echoes", *J. Phys. C: Solid State Physics*, vol. 10, pp. 55–58, 1977.

[97] K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis*. Academic Press, 1979.

[98] J. Mattson and M. Simon, *The pioneers of NMR and magnetic resonance in medicine: the story of MRI*. Bar-Ilan University Press, 1996.

[99] J. H. McClellan. "The design of two-dimensional digital filters by transformations". In *Proc. 7th Annual Princeton Conf. Inf. Sciences and Systems*, Princeton (USA), pp. 247–251, 1973.

[100] G. McCullagh and J. Nelder, *Generalized Linear Model*. Chapman and Hill, 1989.

[101] F. G. Meyer. "Estimation of a semi parametric model of fMRI data". In *Proc. SPIE, Wavelet Applications in Signal and Image Processing IX*, vol. 4478, 2001.

[102] F. G. Meyer. "Wavelet based estimation of a semi parametric generalized linear model of fMRI time-series". In *Proc. ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake, UT, USA, pp. 3681–3684, 2001.

[103] A. Mojsilović, M. Popović, S. Marković and M. Krstić, "Characterization of visually similar diffuse diseases from B-scan liver images using nonseparable wavelet transform", *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 541–549, 1998.

[104] F. Nicolier, O. Laligant and F. Truchetet. "B-spline quincunx wavelet transform and implementation in Fourier domain". In *Proc. SPIE*, Boston, Massachussets, USA, vol. 3522, pp. 223–234, november 1998.

[105] A. C. Nirkko, C. Ozdoba, S. M. Redmond, M. Bürki, G. Schroth, C. W. Hess and M. Wiesendanger, "Different ipsilateral representations for distal and proximal movements in the sensorimotor cortex: Activation and deactivation patterns", *NeuroImage*, vol. 13, pp. 825–835, 2001.

[106] D. Noll, "Noise and artifacts in fMRI", http://www.bme.umich.edu/ dnoll/-BME499/noise.pdf, 2001.

[107] S. Ogawa, T. Lee, A. Kay and D. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation", *Proc. Natl. Acad. Sci. USA*, vol. 87, pp. 9868–9872, 1990.

[108] S. Ogawa, T. Lee, A. Nayak and P. Glynn, "Oxygenation-sensitive contrast in magnetic-resonance image of rodent brain at high-magnetic fields.", *Magnetic Resonance in Medicine*, vol. 14, pp. 68–78, 1990.

[109] S. Ogawa, R. Menon, S. Kim and K. Ugurbil, "On the characteristic of functional magnetic resonance imaging of the brain", *Annu. Rev. Biophys.*, vol. 27, pp. 447–474, 1998.

[110] S. Ogawa, R. Menon, D. Tank, S. Kim, H. Merkle, J. Ellerman and K. Ugurbil, "Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging", *Biophysical J.*, vol. 64, pp. 803–812, 1993.

[111] J.-P. Poline, A. Holmes, K. Worsley and K. Friston, *Human Brain Function*. Academic Press, 1997.

[112] E. Purcell, H. Torrey and R. Pound, "Resonance absorption by nuclear magnetic moments in a solid", *Phys. Rev.*, vol. 69, pp. 37–38, 1946.

[113] S. Rabe-Hesketh, E. Bullmore and M. Brammer, "The analysis of functional magnetic resonance images", *Statistical Methods in Medical Research*, vol. 6, pp. 215–237, 1997.

[114] W. Rasband, "ImageJ web site", http://rsb.info.nih.gov/ij/.

[115] J. Raz and B. Turetski, "Event-related potentials", *Encyclopedia of Biostatistics*, vol. 2, pp. 1407–1409, 1998.

[116] J. Raz and B. Turetsky. "Wavelet ANOVA and fMRI". In *Proc. SPIE, Wavelet Applications in Signal and Image Processing VII*, Denver, CO, USA, vol. 3813, pp. 561–570, July 1999.

[117] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms", *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 569–586, 1992.

[118] J. Romberg, H. Choi, R. Baraniuk and N. Kingsbury. "Multiscale classification using complex wavelets and hidden Markov tree models". In *IEEE International Conference on Image Processing, 2000*, vol. 2, pp. 371–374, 2000.

[119] C. Roy and C. Sherrington, "On the regulation of blood supply of the brain", *Journal of Physiology*, vol. 11, pp. 85–108, 1890.

[120] U. Ruttimann, M. Unser, R. Rawlings, D. Rio, N. Ramsey, V. Mattay, D. Hommer, J. Frank and D. Weinberger, "Statistical analysis of functional MRI data in the wavelet domain", *IEEE Trans. Medical Imaging*, vol. 17, no. 2, pp. 142–154, 1998.

[121] U. Ruttimann, M. Unser and D. Rio. "Statistical analysis of image differences by wavelet decomposition". In *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Engineering Advances: New Opportunities for Biomedical Engineers*, vol. 1, pp. 28–29, 1994.

[122] G. Scarth and M. McIntyre. "Detection of novelty in functional images using fuzzy clustering". In *Proc. ISMRM*, Nice, vol. 1, p. 238, 1995.

[123] V. Schmithorst, B. Dardzinski and S. Holland, "Simultaneous correction of ghost and geometric distortion artifacts in EPI using a multiecho reference scan", *IEEE Trans. on Medical Imaging*, vol. 20, no. 36, pp. 535–539, 2001.

[124] D. Seminowics, "PET and fMRI neuroimaging", http://www.uoguelph.ch/ dseminow/, 2001.

[125] J. Shapiro, "Adaptive McClellan transformations for quincunx filter banks", *IEEE Trans. on Signal Processing*, vol. 42, no. 3, pp. 642–648, 1994.

[126] K. Shung, M. Smith and B. Tsu, *Principles of Medical Imaging*. Academic Press, 1992.

[127] Siemens, *MAGNETOM Symphony Maestro Class.* http://www.siemens-medical.com/, 2002.

[128] P. Skudlarski, "fMRI analysis package", http://mri.med.yale.edu/individual/-pawel/fMRIpackage.html, 1998.

[129] P. Skudlarski, R. Constable and J. Gore, "ROC analysis of statistical methods used in functional MRI: Individual subjects", *NeuroImage*, vol. 9, no. 3, pp. 311–329, 1999.

[130] A. Smith, B. Lewis, U. Ruttimann, F. Ye, T. Sinnwell, Y. Yang, J. Duyn and J. Frank, "Investigation of low frequency drift in fMRI signal", *NeuroImage*, vol. 9, pp. 526–533, 1999.

[131] S. Smith, "Introduction to fMRI analysis and fMRI image preprocessing", http://www.fmrib.ox.ac.uk/ steve/lectures/intro+preproc/s_0100_title.htm, 2002.

[132] S. Springer and G. Deutsch, *Left Brain, Right Brain, Perspective from Cognitive Neuroscience.* Freeman, 1998.

[133] M. Stehling, R. Turner and P. Mansfield, "Echo-planar imaging-magnetic-resonance-imaging in a fraction of a second", *Science*, vol. 254, pp. 43–50, 1991.

[134] G. Strang and G. Fix. "A Fourier analysis of the finite element variational method". In *Constructive Aspect of Functional Analysis*. Edizioni Cremonese, Rome, 1971, pp. 796–830.

[135] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1997.

[136] J. Strupp, "Stimulate: A GUI based fMRI analysis software package.", *NeuroImage*, vol. 3, p. 607, 1996.

[137] J. Talairach and P. Tournoux, *Referentially Oriented Cerebral MRI Anatomy.* Thieme, 1993.

[138] D. B. H. Tay and N. G. Kingsbury, "Flexible design of multidimensional perfect reconstruction FIR 2-band filters using transformations of variables", *IEEE Trans. on Image Processing*, vol. 2, no. 4, pp. 466–480, 1993.

[139] C. Teichtmeister. *Methodische Untersuchungen zum Funktionellen Magnetresonanzimaging bei 1.5 Tesla.* PhD thesis, Technische Universität Wien, Wien, 1996.

[140] P. Thévenaz, U. Ruttimann and M. Unser, "A pyramid approach to subpixel registration based on intensity", *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.

[141] F. Turkheimer, R. Banati, D. Visvikis, J. Aston, R. Gunn and V. Cunningham, "Modeling dynamic PET-SPECT studies in the wavelet domain", *Journal of Cerebral Blood Flow and Metabolism*, vol. 20, pp. 879–893, 2000.

[142] R. Turner, D. L. Bihan, C. Moonen, D. Despres and J. Frank, "Echo-planar time course MRI of cat brain deoxygenation changes", *Magn. Reson. Med.*, vol. 22, pp. 159–166, 1991.

[143] M. Unser, "On the optimality of ideal filters for pyramid and wavelet signal approximation", *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3591–3596, 1993.

[144] M. Unser, "Splines: A perfect fit for signal and image processing", *IEEE Signal Processing Magazine*, vol. 16, pp. 22–38, 1999.

[145] M. Unser and T. Blu, "Fractional splines and wavelets", *SIAM Review*, vol. 42, pp. 43–67, 2000.

[146] M. Unser, P. Thévenaz, C. Lee and U. Ruttimann, "Registration and statistical analysis of PET images using the wavelet transform", *IEEE Engineering in Medecine and Biology Magazine*, vol. 14, pp. 603–611, 1995.

[147] D. Veltmann, C. Hutton, J. Ashborner and R. Henson, "SPM99 manual", http://www.fil.ion.ucl.ac.uk/spm/course/manual/spatial.htm, 2001.

[148] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding.* Prentice-Hall, 1995.

[149] A. Wagner, D. Schacter, M. Rotte, W. Koustal, A. Maril, A. Dale, B. Rosen and R. Buckner, "Building memories: remembering and forgetting of verbal experiences as predicted by brain activity", *Science*, vol. 21, pp. 188–191, 1998.

[150] J. Weaver, A. Saykin, R. Burr, H. Riordan and A. Maerlender. "Principal component analysis of functional MRI of memory". In *Proceedings of the Society for Magnetic Resonance, Second Annual Meeting, SMR*, p. 808, 1994.

[151] A. Wink. "Denoising in SPM: comparing wavelets and Gaussian filters". not published yet, 2002.

[152] P. B. E. Wong, R. Hinks, R. Tikofsky and J. Hyde, "Time course EPI of human brain function during task activation", *Magnetic Resonance in Medicine*, vol. 25, pp. 390–397, 1992.

[153] R. Woods, S. Grafton, C. Holmes, S. Cherry and J. Mazziotta, "Automated image registration: I. general methods and intrasubject, intramodality validation", *Journal of Computer Assisted Tomography*, vol. 22, pp. 141–154, 1998.

[154] R. Woods, S. Grafton, J. W. N. Sicotte and J. Mazziotta, "Automated image registration: II. intersubject validation of linear and nonlinear models", *Journal of Computer Assisted Tomography*, vol. 22, pp. 155–165, 1998.

[155] K. Worsley, "Local maxima and the expected Euler characteristic of excursion sets of ξ^2, t and F fields.", *Journal of Applied Probability*, vol. 26, pp. 13–42, 1994.

[156] K. Worsley, A. Evens, S. Marrett and P. Neelin, "A three dimensional statistical analysis for CBF activation studies in human brain.", *J. Cerebral Blood Flow and Metabolism*, vol. 12, pp. 900–918, 1992.

[157] K. Worsley, S. Marrett, P. Neelin, A. Vandal, K. Friston and A. Evans, "A unified statistical approach for determining significant signals in images of cerebral activation", *Human Brain Mapping*, vol. 4, pp. 58–73, 1996.

[158] S. Yu, S. Brown, Y. Xue and L. Guan. "Enhancement and identification of microcalcifications in mammogram images using wavelets". In *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, pp. 1166–1171, 1996.

[159] X. Zhang, M. Desai and Y. Peng, "Orthogonal complex filter banks and wavelets: Some properties and design", *IEEE Transactions on Signal processing*, vol. 47, no. 4, pp. 1039–1048, 1999.

Curriculum vitæ

Manuela Patricia FEILNER

BIG/BIO-E/STI
Swiss Federal Institute of Technology
CH-1015 Lausanne
Switzerland
Phone (home): (+41) 1 915 2991
Email: Manuela.Feilner@epfl.ch

Born on 7th May 1973 in Zürich, Switzerland.

Education

1998–2002	*Ph.D student and research/teaching assistant* **Swiss Federal Institute of Technology (EPFL), Lausanne** • Subject: *Statistical Wavelet Analysis of Functional Images of the Brain*
1993 Oct.–1998	*M.Sc. in Electrical Engineering* **Swiss Federal Institute of Technology (ETHZ), Zürich** • Diplomawork: *Frequenzsynthesizer und Open-Loop Modulator für DECT (1.9 GHz)*, Fachgruppe für Höchstfrequenzgruppe, 1997/1998 • 2nd Semesterwork: *Wavelet-basierte Konturanalyse*, Fachgruppe Bildwissenschaften, 1997 • 1nd Semesterwork: *Charakterisierung von Magnetfeldsensoren*, Institut für Quantenelektronik, 1996/97
1992	*Matura humanistic (Typus A)*

Technical Skills

Operating systems:	Unix, Linux, Mac OS, Windows 2000
Programming Languages:	Matlab, C, Oberon, Java, Assembler
Application Software:	LaTeX, AVS, Mathematica, Maple, CAD, H-Spice, Web design and data basis, Microsoft office

Language

German	Mother tongue
French	Good knowledge
English	Good knowledge

Publications

Journal papers

[1] M. Feilner, M. Unser, "An orthogonal family of quincunx wavelets with continuously-adjustable order," IEEE Transaction on Image Processing, submitted.

Conference papers

[1] M. Feilner, M. Jacob, M. Unser, "Orthogonal Quincunx Wavelets with Fractional Orders," Proceedings of the 2001 IEEE International Conference on Image Processing (ICIP'01), Thessaloniki, Greece, October 7-10, 2001, vol. I, pp. 606-609.

[2] M. Feilner, T. Blu, M. Unser, "Analysis of fMRI Data Using Spline Wavelets," Proceedings of the Tenth European Signal Processing Conference (EUSIPCO'00), Tampere, Finland, September 4-8, 2000, vol. IV, pp. 2013-2016.

[3] M. Feilner, T. Blu, M. Unser, "Optimizing Wavelets for the Analysis of fMRI Data," Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing VIII, San Diego CA, USA, July 31-August 4, 2000, vol. 4119, pp. 626-637.

[4] M. Feilner, T. Blu, M. Unser, "Statistical Analysis of fMRI Data Using Orthogonal Filterbanks," Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing VII, Denver CO, USA, July 19-23, 1999, vol. 3813, pp. 551-560.