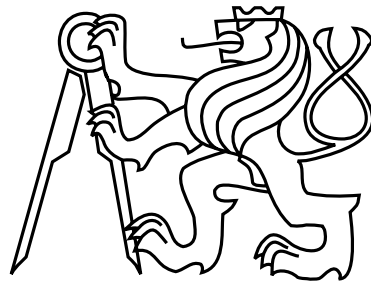


Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Cybernetics



Diploma Thesis

MR Signal Analysis

Erik Saudek

Supervisor: Ing. Daniel Novák, PhD.

Study Program: Electrical Engineering and Information Technology

Specialization: Biomedical Engineering

May 2008

Katedra kybernetiky

Školní rok: 2006/2007

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Erik S a u d e k

Obor: Biomedicínské inženýrství

Název tématu: Zpracování MR signálu

Zásady pro vypracování:

1. Seznamte se s problematikou zobrazování magnetické rezonance.
2. Navrhněte metodiku pro analýzu MR dat.
3. Ověřte navrhnoutou metodu na souboru dat pacientů pomocí základní statistické analýzy.

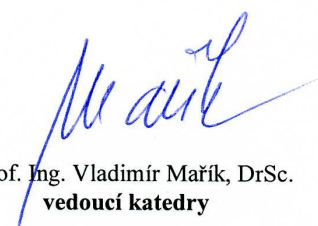
Seznam odborné literatury:

Lerski, R.: *Texture Analysis for Magnetic Resonance Imaging*. Med4publishing.2006.ISBN 8090366007

Vedoucí diplomové práce: Ing. Daniel Novák, Ph.D.

Termín zadání diplomové práce: letní semestr 2006/2007

Termín odevzdání diplomové práce: květen 2008


prof. Ing. Vladimír Mařík, DrSc.
vedoucí katedry




doc. Ing. Boris Šimák, CSc.
děkan

Abstract

Magnetic resonance has proven to be a successful method of in-vivo imaging. Although MRI can help detect various pathologies, its ability to classify the nature of the pathological tissue is limited. Magnetic resonance spectroscopy allows identifying metabolite content of the tissue and estimating the metabolite concentration. Map of metabolite concentration along with the MR image allows proper classification of many pathologies, for example progressive tumorous tissue identification in human brain. Standard methods used to analyze nuclear magnetic resonance spectra such as singular value decomposition or curve fitting algorithms are very time consuming taking several minutes to analyze spectrum from a single voxel. To analyze the spectra from a chemical shift imaging sequence (CSI) in maximal resolution hundreds of spectra need to be processed. The suggested ANN framework proved to be much faster. Networks were trained on the outputs of LCMoDel curve fitting algorithm. Time needed to process a spectrum from a single voxel was reduced to the order of seconds. The total time needed to analyze a CSI in full resolution (hundreds of spectra) was significantly reduced to 5 minutes.

Abstrakt

Magnetická resonance je uznávanou zobrazovací metodou živé tkáně v moderní medicíně. Přestože pomáhá odhalit různé patologické změny, její schopnost určit povahu patologie a přesně ji klasifikovat je značně omezená. Spektroskopie pomocí magnetické resonance umožňuje detekci metabolického obsahu tkáně a určit koncentrace jednotlivých metabolitů. Metabolická mapa společně s MR obrázkem pomáhá přesněji popsat patologickou tkáň, například identifikovat progresivní tumor mozku. Standardní metody analýzy spekter nukleární magnetické resonance, jako například prokládání sadou křivek nebo rozklad na principiální komponenty, jsou výpočetně velmi náročné. Zpracování spektra z jednoho voxelu může trvat až několik minut. Ke zpracování dvojdimensionálního spektroskopického obrázku v plném rozlišení musí být analyzovány stovky spekter. Navržený postup postavený na umělé neuronové síti je daleko rychlejší. Data pro trénink neuronové sítě jsou získávána z výsledku LCMoDelu, algoritmu prokládajícího spektrum sadou křivek. Čas potřebný k analýze jednoho samostatného spektra se snížil a proběhne v řádu vteřin. Celkový čas potřebný k analýze spektroskopického obrázku v plném rozlišení (stovky spekter) byl výrazně snížen, nyní proběhne do 5-ti minut.

Acknowledgements

The project was supported by MZOIKEM2005 and by Ministry of Education, Youth and Sport of the Czech Republic with the grant No. MSM6840770012 entitled “Transdisciplinary Research in Biomedical Engineering II”. I would also like to thank all the people that helped me with this thesis: Special thanks to Filip Jírů for technical discussion, remarks and access to SIPRO source code. Thanks to Daniel Novák for technical support and leadership. Thanks to my colleagues Dita Wagnerová, Antonín Škoch, and Monika Dezortová for help with data collection and measurements and for valuable advice and help in understanding the MR principals. Thanks to Jakub Ráfl for the help with \LaTeX text processing and graphical design.

Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v příloženém seznamu.

V Praze dne 16.5.2008.....

Eril Soud
.....
podpis

Contents

| | |
|---|-----------|
| List of Abbreviations | ix |
| 1 Introduction | 1 |
| 1.1 Aims of the Thesis | 2 |
| 2 Theoretical Part | 3 |
| 2.1 Principles of Magnetic Resonance | 3 |
| 2.1.1 Larmor Precession | 3 |
| 2.1.2 NMR Theory | 4 |
| 2.1.3 Magnetization Vector and Bloch Equations | 5 |
| 2.1.4 Spin Echo Sequence | 7 |
| 2.2 NMR Spectroscopy | 8 |
| 2.2.1 Chemical Shift Effect | 9 |
| 2.2.2 J-coupling | 9 |
| 2.2.3 Spectra Acquisition | 10 |
| 2.2.4 NMR Spectra Localization | 11 |
| 2.2.5 PRESS Sequence | 11 |
| 2.2.6 2D PRESS CSI | 12 |
| 2.3 NMR Spectra Processing | 15 |
| 2.3.1 Metabolites of the Human Brain | 15 |
| 2.3.2 Determination of the Signals of Metabolites | 16 |
| 2.3.3 Description of the CULICH Preprocessing | 17 |
| 2.3.4 Description of the LCModel Program | 18 |
| 2.4 Artificial Neural Networks | 20 |
| 2.4.1 Feedforward Artificial Neural Network | 22 |
| 2.4.2 Supervised Learning | 23 |
| 2.4.3 Linear Perceptron | 23 |
| 2.4.4 Multi-Layer Perceptron | 25 |

| | | |
|----------|--|-----------|
| 2.4.5 | Learning through Backpropagation | 26 |
| 2.4.6 | Scaled Conjugate Gradient Backpropagation | 28 |
| 2.5 | Hankel-Lanczos Singular Value Decomposition | 28 |
| 3 | Experimental Part | 30 |
| 3.1 | Siemens Magnetom Trio-3T | 30 |
| 3.2 | Institute for Clinical and Experimental Medicine | 30 |
| 3.3 | MR Examination | 31 |
| 3.4 | MR Image Acquisition | 31 |
| 3.5 | Spectroscopic Image Acquisition | 31 |
| 3.5.1 | Water Suppression Sequence | 32 |
| 3.5.2 | Spectroscopic Data Averaging | 32 |
| 3.5.3 | Delta Frequency | 33 |
| 3.6 | Data Export | 33 |
| 3.7 | Metabolite Concentration Quantification Toolbox | 33 |
| 3.8 | Spectroscopic Data Processing | 34 |
| 3.8.1 | 2D Data Interpolation | 34 |
| 3.8.2 | HLSVD Residual Water Signal Removal | 36 |
| 3.8.3 | Noise Removal | 37 |
| 3.8.4 | Water Referencing | 37 |
| 3.8.5 | NMR Spectrum Calculation | 37 |
| 3.8.6 | Baseline Correction | 37 |
| 3.8.7 | Spectral Band Selection | 39 |
| 3.9 | Training Data | 41 |
| 3.10 | ANN Set-up | 41 |
| 3.10.1 | Training Dataset Reduction | 42 |
| 3.10.2 | Application of the Artificial Neural Network | 42 |
| 3.11 | Data Visualization | 43 |
| 4 | Results | 44 |
| 4.1 | Performance Test Data | 44 |
| 4.2 | Test Criteria | 44 |
| 4.3 | Test Results | 45 |
| 4.3.1 | Patient Diagnosed with a Brain Tumor | 46 |
| 4.3.2 | 7-Month Old Patient with a Suspicion for a Metabolic Disease | 46 |
| 4.3.3 | Cerebellum 2D CSI of a Healthy Adult Volunteer | 46 |

| | |
|--------------------------|-----------|
| 4.4 Discussion | 49 |
| 5 Conclusions | 51 |

List of Abbreviations

| | |
|-------|--|
| ADP | Adenosine Di-Phosphate |
| ANN | Artificial Neural Network |
| ATP | Adenosine Tri-Phosphate |
| Cho | Choline |
| Cr | Creatine |
| CSF | Cerebral-Spinal Fluid |
| CSI | Chemical Shift Imaging |
| DFFT | Discrete Fast Fourier Transform |
| DFT | Discrete Fourier Transform |
| FFT | Fast Fourier Transform |
| FID | Free Induction Decay |
| FT | Fourier Transform |
| HLSVD | Hankel-Lanczos Singular Value Decomposition |
| IKEM | Institute for Clinical and Experimental Medicine, Prague |
| MR | Magnetic Resonance |
| MRI | Magnetic Resonance Imaging |
| NAA | N-Acetyl Aspartate |
| NMR | Nuclear Magnetic Resonance |
| NN | Neural Network |
| OVS | Outer Volume Suppression |
| PCr | Phosphocreatine |
| PRESS | Point Resolved Spectroscopy |

| | |
|-------|----------------------------------|
| RF | Radio Frequency |
| SCG | Scaled Conjugate Gradient |
| SE | Spin Echo |
| SI | Spectral Imaging |
| SNR | Signal to Noise Ratio |
| STEAM | Stimulated Echo Acquisition Mode |
| SVD | Singular Value Decomposition |
| SVS | Single Voxel Spectroscopy |
| TE | Echo Time |
| TR | Repetition Time |

Chapter 1

Introduction

Accurate quantification of cerebral metabolites by in vivo proton magnetic resonance spectroscopy is essential to the study of many brain disorders. Proton nuclear magnetic resonance (NMR) spectroscopy remains one of the few methods by which the chemical content of brain tissue may be probed non-invasively. Several procedures have been applied in biomedical NMR spectroscopy to quantify areas of metabolite resonances from frequency domain spectra [1]. With methods such as the lineshape fitting analysis and the linear combination of model spectra, noisy ^1H NMR spectra or overlapping peaks in spectra can be assessed. However, the reliable use of the lineshape fitting analysis methods needs spectroscopic expertise, so that a fully automated analyzer by these methods is rather difficult to develop. All these procedures involve relatively long computational times.

It has recently been shown that ANN analysis offers some important advantages for biomedical NMR data analysis [2], [3]. However, these techniques were limited to the use of simulated artificial inputs with mathematically estimated outputs for training. The proposed study shows the ability of an artificial neural network (ANN) to estimate concentrations of 3 main metabolites from human brain spectrum measured. In this work, absolute spectra of Free Induction Decay (FID) signals were used as input vectors. Matching metabolite concentrations acquired from the LCModel [4] curve fitting algorithm are set as targets.

Curve fitting algorithms, used to analyze NMR spectra in frequency domain, calculate the best possible combination of base signals (simulated or previously measured in-vitro) to fit the spectrum of the measured FID signal. Currently, the most commonly used quantification methods are based on curve fitting using a non linear least square optimization, such as Levenberg-Marquardt algorithm. Once the network is trained the processing of a new (unknown) input is reduced to a set of matrix multiplications and can be done

within seconds. The developed ANN gives comparable results in much shorter time than widely used curve fitting algorithms.

All the experimental data for the purpose of this thesis was measured at the Institute for Clinical and Experimental Medicine in Prague.

1.1 Aims of the Thesis

The major aim of this work is to develop an algorithm which would approximate the results of an application based on the LCModel NMR spectra processing algorithm in shorter time and with the output error as low as possible. The core of the developed algorithm is realized by an artificial neural network. Therefore, practically, the aim is to train an artificial neural network which would be able to estimate metabolite concentrations from in vivo NMR spectra from human brain. This goal is approached in several steps. First an algorithm needs to be developed which reads raw data from a NMR scanner. Then a preprocessing algorithm is designed which prepares the data to be fed to an ANN network. The same data is processed by commercial software and results are used as targets for the purpose of ANN training. In the next step an artificial neural network is trained and optimized. Finally test data is preprocessed according to the developed procedure and the results are compared to the results of the commercial software. The intention is to duplicate the results of the commercial software in shorter computational time.

Chapter 2

Theoretical Part

In the theoretical part basic concepts of nuclear magnetic resonance, signal acquisition and processing principles are described. Furthermore the theory of artificial neural networks is introduced.

2.1 Principles of Magnetic Resonance

2.1.1 Larmor Precession

All atom nuclei consist of particles called protons and neutrons. These elementary kernel particles have a characteristic inner feature accruing from quantum mechanics, so called spin or magnetic spin momentum \mathbf{S} . Protons and neutrons in the atom nucleus also possess magnetic moment, therefore, they do interact with magnetic field. The overall nuclear magnetic spin momentum \mathbf{I} and magnetic moment μ are bound through

$$\mu = \gamma \mathbf{I}, \quad (2.1)$$

where γ is a so called gyromagnetic ratio characteristic for each atom nucleus.

The magnetic spin in constant magnetic field is unstable, precessing along an axis parallel with the outer magnetic field with a frequency ω_0 . This rotation can be visualized as a movement along the surface of a cone with central axis parallel to the axis of the external magnetic field (denoted z for the purpose of this thesis). The precession frequency ω_0 is a so called Larmor frequency and is given by

$$\omega_0 = -\gamma \mathbf{B}_0 \quad (2.2)$$

where \mathbf{B}_0 is the magnetic induction of the external magnetic field.

Hamiltonian H of the nuclear magnetic spin in stationary external magnetic field \mathbf{B}_0 oriented along the axis z of the laboratory coordinate system is given by

$$H = -\gamma\hbar I_z B_0 = \hbar\omega_0 I_z, \quad (2.3)$$

where \hbar is Planck constant and I_z is the z component of the nuclear magnetic spin operator, which can take on any of the $2I + 1$ values:

$$I_z = -I, -I + 1, \dots, I - 1, I. \quad (2.4)$$

This set of states in means of energy is called Zeeman's multiplet.

Transitions between energy levels are induced by using an additional circularly polarized magnetic field \mathbf{B}_1 . This magnetic field is perpendicular to the stationary field \mathbf{B}_0 and its intensity is much lower. The transition is possible only if the atom nuclei are in resonance with the oscillating magnetic field \mathbf{B}_1 . That means that the frequency of the additional oscillating magnetic field used for excitation has to be equal to the Larmor frequency of the nuclei. This ability of the nuclei to accept energy of the magnetic field and temporary take on a state with higher energy is called nuclear magnetic resonance (NMR).

2.1.2 NMR Theory

The theory of nuclear magnetic resonance can only be thoroughly described through quantum mechanics. For the purpose of this thesis a much easier, yet not as exact description through classic physics will be given.

The theory of quantum mechanics closely describes the magnetic spin system of the nucleus and the mechanics of its transitions among energy levels. However, for the external field an approach of classic physics is used [5]. In this thesis, the measured samples are large inhomogeneous groups of molecules and macromolecules—a whole set of non-isolated magnetic spins. In this case both, the magnetic spin Hamiltonian and the wave function, get very complicated and for the description of the NMR experiment a large number of approximations have to be done.

The approach of classic physics is a phenomenological description of the magnetic resonance effect which consists of a set of first order linear differential equations known as Bloch equations. The Bloch equations describe the motion of macroscopic magnetization vector \mathbf{M} given by the sum of all the magnetization vectors of the nuclei within the measured sample. The downside of the macroscopic, classic approach is the inability to describe all the spin interactions on atomic level. For example the relaxation time

constants are averaged over all nuclei of the same type. However, for the purpose of this thesis such an approximation is acceptable.

2.1.3 Magnetization Vector and Bloch Equations

Upon insertion of the measured sample to the external static magnetic field \mathbf{B}_0 in the state of thermodynamic equilibrium a non zero magnetic moment is present. It exists due to anisotropic distribution of the magnetic spin polarization of the nuclei and can be expressed as a vector of macroscopic magnetization \mathbf{M}_0 . This anisotropy is caused by interaction of the spin system with its neighboring molecule environment (grid). Magnetization vector can be expressed as

$$\mathbf{M}_0 = \chi_{jad}\mathbf{B}_0, \quad (2.5)$$

where χ_{jad} is the magnetic susceptibility. It is a macroscopic quantity that embodies microscopic contribution of all the nuclear spins within the sample and is given by

$$\chi_{jad} = \frac{\mu_0 N_V \hbar^2 \gamma^2 I(I+1)}{3k_B T}, \quad (2.6)$$

where μ_0 is the permeability of vacuum, N_V is the number of nuclear spins in unit volume, k_B is the Boltzmann constant, and T is the thermodynamic temperature of the system.

In the state of thermodynamic equilibrium the magnetization vector only has a non zero z component (parallel to \mathbf{B}_0). The components perpendicular to the external field sum up to zero due to the precession motion. The excitational oscillating \mathbf{B}_1 field can be used to deflect the magnetization vector from the equilibrium. The time evolution of the magnetization vector in laboratory coordinates during the excitation by a series of radio frequency pulses can get very complicated due to time dependence of the total external magnetic field. Therefore, the equation describing the evolution of the magnetization vector is described in a coordinate system rotating with a frequency ω_{ref} . In this system the magnetic spins rotate on a new frequency ω ,

$$\Omega = \omega_0 - \omega_{ref} \quad (2.7)$$

If the frequency of the rotating coordinate system is equal to the Larmor frequency, the effect of \mathbf{B}_0 cancels out and the total external magnetic field is time independent and equal to \mathbf{B}_1 .

As stated above, the time evolution of the magnetization vector is described by phenomenological Bloch equations:

$$\frac{dM_x}{dt} = \gamma(\mathbf{M} \times \mathbf{B})_x - \frac{M_x}{T_2}, \quad (2.8)$$

$$\frac{dM_y}{dt} = \gamma(\mathbf{M} \times \mathbf{B})_y - \frac{M_y}{T_2}, \quad (2.9)$$

$$\frac{dM_z}{dt} = \gamma(\mathbf{M} \times \mathbf{B})_z - \frac{M_z - M_0}{T_1}, \quad (2.10)$$

where \mathbf{B} is the total local magnetic field in the rotating coordinates (z is parallel to \mathbf{B}_0). Local field means that this field depends on the type and location of the nucleus and reflects the electron structure around the nucleus. T_1 and T_2 are constant quantities introduced phenomenologically to characterize relaxation mechanisms within the studied spin system. Relaxation is a process in which the spin system tries to return back to the state of equilibrium. T_1 constant is a so called longitudinal or spin-grid relaxation time and expresses the rate at which the longitudinal (parallel to z) component of the magnetization vector \mathbf{M} returns to its value at the state of equilibrium (M_0). This relaxation is caused by the interaction (energy exchange) between the spin system and the atomic grid. T_2 constant is a so called transverse or spin-spin relaxation time and expresses the rate at which the transverse (perpendicular to z) component of the magnetization vector diminishes. The transverse relaxation is a result of the spin-spin interactions within the measured sample.

As stated earlier, deflection of the magnetization vector is induced by application of radio frequency (RF) pulses. These RF pulses are often very short compared to the relaxation times so that these effects during the excitation can be ignored. The measurements itself take place at the times when the \mathbf{B}_1 RF pulses with frequency set to Larmor frequency are off. Setting the frequency of the rotating coordinates ω_f to the frequency of the RF pulse leads to a constant \mathbf{B}_1 . Solving the Bloch equations shows that the magnetization vector \mathbf{M} will be precessing with angular frequency $\omega = \gamma B_1$. The orientation of the magnetization vector once the RF pulse is turned off depends on the length and the phase of the RF pulse. In experiments used for the purpose of this thesis two types of pulses were used. Both are oriented along the x axis but differ in the flip angle

$$\beta = \int \gamma B_1 dt \quad (2.11)$$

First is a so called $\frac{\pi}{2}$ pulse ($\beta = \pi/2$) at the end of which the magnetization vector is oriented in the direction of the y axis (perpendicular to both \mathbf{B}_0 and \mathbf{B}_1). The other is a so called π pulse ($\beta = \pi$) which deflects the magnetization vector to the opposite direction to \mathbf{B}_0 (the $-z$ direction).

The signal measured after the application of the RF pulse (during the relaxation processes) is called free induction decay (FID) and is usually measured in the transverse plane (perpendicular to \mathbf{B}_0). From the Bloch equations can be derived that the decay of the transverse component of the magnetization vector after application of a $\frac{\pi}{2}$ RF pulse is

$$M_{\perp}(t) = M_0 e^{-t/T_2} \quad (2.12)$$

The time varying magnetization vector induces electric currents in a set of coils used as sensors for the measurement of the FID signal. The signal detected $s(t)$ during the synchronous quadrature detection is complex

$$s(t) = n e^{-t/T_2} e^{i(\Omega_0 t + \phi)}, \quad (2.13)$$

where n is proportionate to the number of spins within the sample, Ω_0 is the Larmor frequency in the rotating coordinate system, and ϕ is the initial phase.

However, this only applies for a sample which consists only of magnetic spins with the same Larmor frequency and which is placed in a perfectly homogenous external magnetic field. In reality the decay of the magnetization vector is faster due to magnetic field inhomogeneities and spin variety within the sample which result in variance of the Larmor frequencies. Various Larmor frequencies make the spins come out of phase. Therefore, the overall magnetization vector decays faster. In reality the signal decays with transverse relaxation time T_2^* ,

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{\text{inhomog}}} \quad (2.14)$$

To compensate for the field inhomogeneities and reduce their effect on the transverse relaxation time (and the peak width in the resulting spectra) a special combination of RF pulses is used. These sequences are called stimulated and spin echo sequences. For the purpose of this study only spin echo sequences were used and will be further explained in detail.

2.1.4 Spin Echo Sequence

The spin echo sequence consists of two RF pulses (Fig. 2.1). First a $\frac{\pi}{2}$ RF pulse is applied followed by a π RF pulse in time τ after the first pulse. After application of the first pulse the signal decays, at the time of the second pulse no signal can be detected, in time $t = 2\tau$ the signal reappears in a so called spin echo. The time $TE = 2\tau$ is called echo time. The first pulse flips the magnetization vector to the transverse plane. The spins come out of

phase due to field inhomogeneities. After the second pulse the spins start to get back to phase. Once the spins are in phase the signal is measured. The magnitude of the transverse magnetization vector in time $t = TE$

$$M_{\perp}(t = TE) = M_0 e^{-2\tau/T_2} = M_0 e^{-TE/T_2}. \quad (2.15)$$

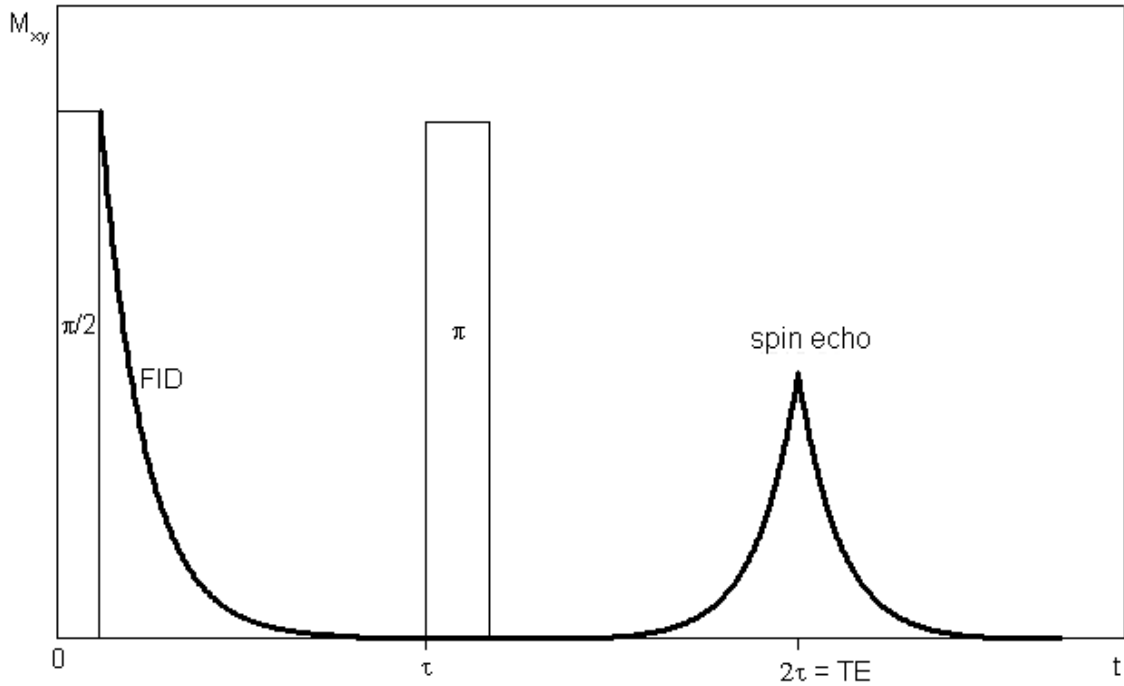


Fig. 2.1. Schematic of the spin echo sequence. Source: [5].

2.2 NMR Spectroscopy

Experiments in nuclear magnetic resonance have shown that even spins with the same value of gyromagnetic ratio γ differ slightly in their Larmor frequency. Spin-spin interactions show to locally affect the external magnetic field \mathbf{B}_0 . Therefore, the spins within the sample appear to be in a slightly different local magnetic field B_{loc} and precess with a slightly different Larmor frequency. In biological tissue the main interactions are chemical shift effect and J-coupling and are discussed further below. The dipole-dipole interaction, in this case, sums up to zero.

2.2.1 Chemical Shift Effect

Chemical shift effect in NMR spectroscopy enables the main qualitative differentiation of signals from individual atom nuclei. The electron cloud around the nucleus provides some sort of a shield against the external magnetic field and that way reduces the magnetic induction of the field to

$$B_{loc} = (1 - \sigma)B_0, \quad (2.16)$$

where σ is a so called shielding ratio, which describes the chemical neighborhood of the nucleus. Larmor frequencies of nonequivalent nuclei in the sample are then

$$\omega_k = \gamma(1 - \sigma_k)B_0. \quad (2.17)$$

Because the Larmor frequency is dependent on the external static magnetic field (property of the machine used for experiment) a new parameter (independent of the external field) for description of the chemical neighborhood of the nucleus is defined. It is called chemical shift (denoted as δ in this thesis) and is defined as

$$\delta_k = \frac{\omega_k - \omega_{ST}}{\omega_{ST}} 10^6, \quad (2.18)$$

where ω_{ST} is the standard resonance frequency. The units of the chemical shift quantity are called parts per million (ppm). In ^1H nuclear magnetic resonance spectroscopy the resonance frequency of N-acetyl aspartate (NAA) is used as a standard ($\omega_{NAA} = 2.01$). For the purpose of this study only ^1H nuclear magnetic resonance was considered. In ^1H NMR most of the signal measured comes from hydrogen nuclei. The signal from hydrogen is much stronger than from any other atom nuclei. Therefore, if there are any ^1H atoms in the sample no other signals can be detected. This is the case of human brain mostly due to high content of water. However, there are other types of NMR, such as ^{31}P spectroscopy, where the atoms of phosphor are the main source of signal. NAA has a significant peak in the NMR spectra from human brain approximately in the middle of the metabolite part of the spectrum. The function of NAA and the shape of the NMR spectra are both discussed further in more detail.

2.2.2 J-coupling

J-coupling or indirect spin-spin interaction is mediated through chemical bonds (electrons). Chemical bonds affect the distribution of the electrons around the bound nuclei

and therefore affect the local magnetic field and the Larmor frequency. It is characteristic for splitting the signal only from nonequivalent nuclei within one molecule to several signals. The number of multiplets X into which the signal is split is independent of the external static magnetic field and is equal to

$$X = 2nI + 1, \quad (2.19)$$

where n is the number of nonequivalent nuclei with the quantum number I interacting with the given nucleus.

2.2.3 Spectra Acquisition

As stated above, real sample due to mutual interactions contains N types of nuclei (protons in case of ^1H NMR) with corresponding Larmor frequencies $\omega_1, \dots, \omega_N$ and transverse relaxation times T_{21}, \dots, T_{2N} . The signal of free induction decay is therefore equal to

$$s(t) = \sum_{k=1}^N n_k e^{-t/T_{2k}} e^{i(\Omega_k t + \phi_k)}, \quad (2.20)$$

where n_k is proportionate to the number of spins of the given type, Ω_k is the Larmor frequency corresponding to ω_k in the rotating coordinate system and ϕ_k is the corresponding initial phase.

For better interpretation, the Fourier transform of the signal is often used. The Fourier transform (FT) transforms the signal from time to frequency domain. The resulting function $s(\omega)$ is called a spectrum. Due to the linear behavior of the Fourier transform the FT of a sum of signals is a sum of FT of the individual signals.

$$S(\omega) = \sum_{k=1}^N S_k(\omega) = \sum_{k=1}^N FT[n_k e^{-t/T_{2k}} e^{i(\Omega_k t + \phi_k)}](\omega), \quad (2.21)$$

where

$$S_k(\omega) = FT(n_k e^{-t/T_{2k}}) = n_k e^{i\phi_k} \cdot T_{2k} / (1 - i(\omega_k - \omega) \cdot T_{2k}). \quad (2.22)$$

The real part of the spectrum is the absorption Lorentz curve while the imaginary part is a dispersion Lorentz curve. Usually the real part (absorption curve) is used for the purposes of NMR spectroscopy, its peaks are narrower and diminish faster on the side lobes. This helps to distinguish various peaks in the spectra especially if their resonance frequencies are similar and the peaks overlap. To get the absorption part of the spectrum the phase term needs to be negated by multiplication by a term with an opposite phase. This multiplication is called phase correction. Phase correction is usually applied manually by

tuning the phase of the correction term until the peaks are narrow. It is also possible to use absolute spectra. Absolute spectra have wider peaks but do not require phase correction.

Some interesting properties of the absorption curve are:

1. individual peaks reach their maxima for $\omega = \omega_k$ and the maxima values are $n_k T_{2k}$
2. line width of the individual peaks are equal to $2/T_{2k}$
3. the surface under the individual peaks (integral of s_k over ω) is equal to $n_k \pi$ and can be used to calculate concentrations of the spins of the given type within the sample.

2.2.4 NMR Spectra Localization

In order to determinate metabolic concentration correctly, it is necessary to measure spectra from a predefined region. Specifically it is important to measure a signal from very small volumes (voxels) which can be selected and localized according to previously measured MR images. Localized in vivo ^1H NMR spectroscopy includes single voxel spectroscopy (SVS) and spectroscopic imaging (CSI).

A spectrum from a single voxel (SVS) can be acquired by measuring the signal from the point of intersection of three perpendicular planes excited by the use of gradient magnetic field and RF pulses. This can be done with a PRESS (Point Resolved Spectroscopy) or STEAM (Stimulated Echo Acquisition Mode) measuring sequence in which only a preselected part of the sample is excited by selective RF pulses.

In spectroscopic imaging (CSI) signals from multiple regions within one plane are measured by sequential changing of the gradient values in two perpendicular directions. For all of the sequences mentioned RF pulses with a rectangular (or as close to as possible) profile in the frequency domain are used (usually a sinc function). These pulses enable the excitation of a cubical volume in the previously selected exact location.

Only PRESS sequence will be further described as it was used in the experimental measurements for this thesis.

2.2.5 PRESS Sequence

For volume selection frequency selective RF pulses combined with application of additional gradient magnetic field are used. Signal measured in the PRESS sequence (Fig. 2.2) is a spin echo signal discussed earlier. This sequence consists of three selective pulses,

one $\pi/2$ pulse and two π pulses. Simultaneously with the first RF pulse a gradient magnetic field G_x in the direction of the x axis is applied. This way only spins in a plane perpendicular to the x axis are excited. After time $TE_1/2$ gradient magnetic field G_y in the direction of the y axis is applied simultaneously with the second RF pulse. This way spins in the plane perpendicular to the y axis are excited. In the region of intersection of the two excited planes spin echo signal will emerge in time TE_1 . However, this signal is not measured. In time $TE_1+TE_2/2$ a gradient field G_z in the direction of the z axis is applied simultaneously with the last RF pulse. In time TE_1+TE_2 the spins from the point of intersection of the three planes will come back to phase and generate a spin echo signal only from the volume defined by all three planes.

In practice along with the gradients and RF pulses mentioned above additional gradients are applied to reduce possible effect of undesirable signals. Gradients G_i can be applied along any x , y , or z axis.

2.2.6 2D PRESS CSI

The plane selection in 2D PRESS CSI (2D Point Resolved Spectroscopy Chemical Shift Imaging) sequence (Fig. 2.3) is analogical to the selection in the SVS PRESS sequence itself. Unlike the PRESS sequence, here in addition N equidistant additional magnetic field gradients G in the direction of the phase coding are switched on for time τ . In case of 1D CSI with gradient G_x , the precession frequency ω of all the spins varies along the x axis,

$$\omega(x) = \gamma x G. \quad (2.23)$$

After time τ , all spins will again precess with the same frequency. However, their phase will be different according to the difference in frequency at which they were precessing for the time τ . This will result in a phase shift $\Theta(x)$ at the end of the phase coding,

$$\Theta(x) = -\gamma x G \tau. \quad (2.24)$$

The phase shift of the preceding spins is distributed along the x axis. By introducing the quantity k_l for each value of the gradient G_l of the l -th step of the phase coding

$$k_l = \frac{\gamma}{2\pi} G_l \tau = \frac{\gamma}{2\pi} l \Delta G \tau = -N/2 \dots N/2 - 1 \quad (2.25)$$

where N is the number of steps in the phase coding, the spatially dependent phase shift $\Theta_l(x)$ corresponding to the l -th step of the phase coding expressed as

$$\Theta_l(x) = -2\pi k_l x \quad (2.26)$$

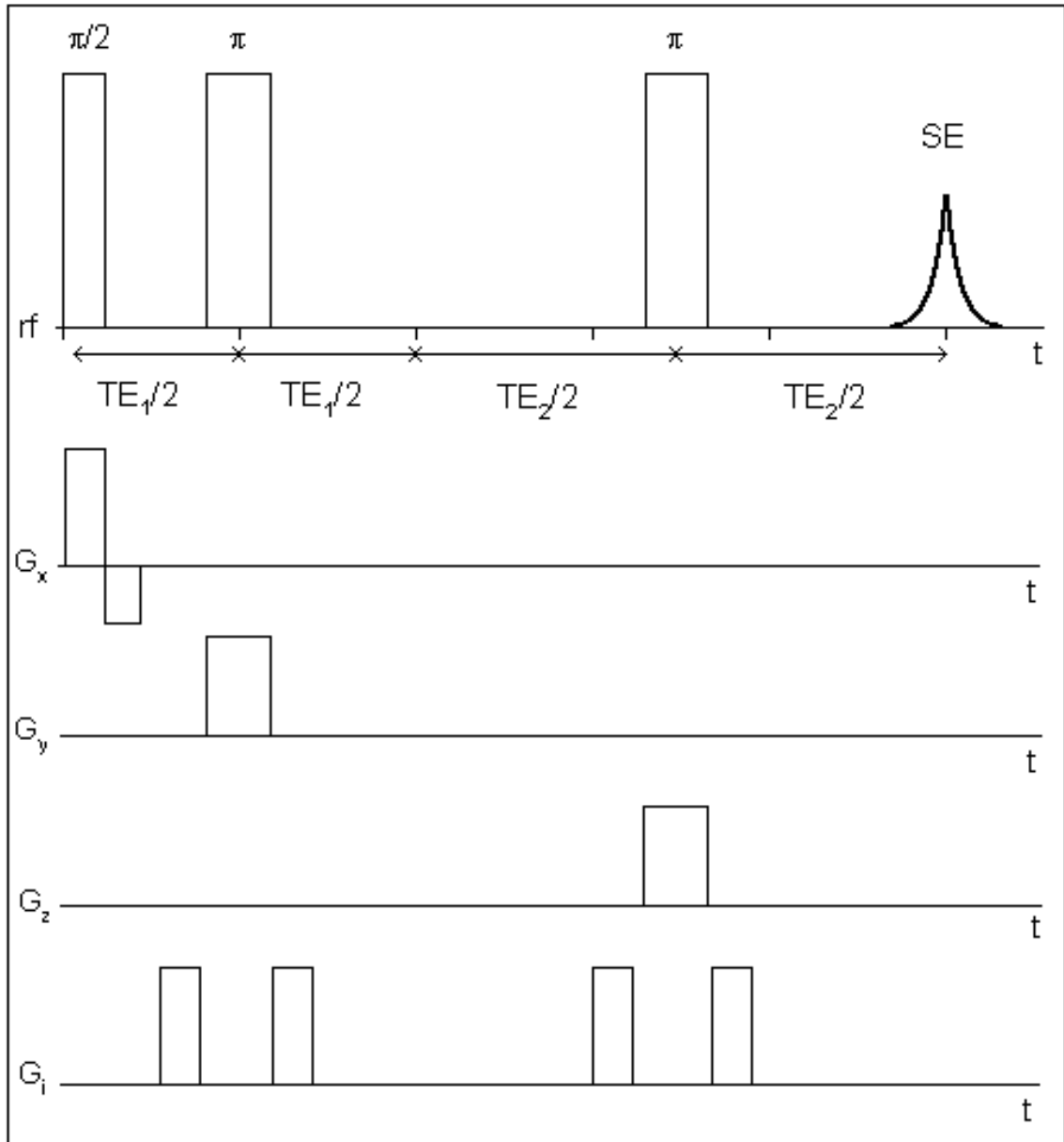


Fig. 2.2. Basic scheme of the PRESS sequence. Source: [5].

The signal measured is $S(t, k_l)$ is therefore a function of k_l

$$S(t, k_l) = \int s(t, x) e^{-i2\pi k_l x} dx, \quad (2.27)$$

where $s(t, x)$ are the individual signals distributed along the x axis.

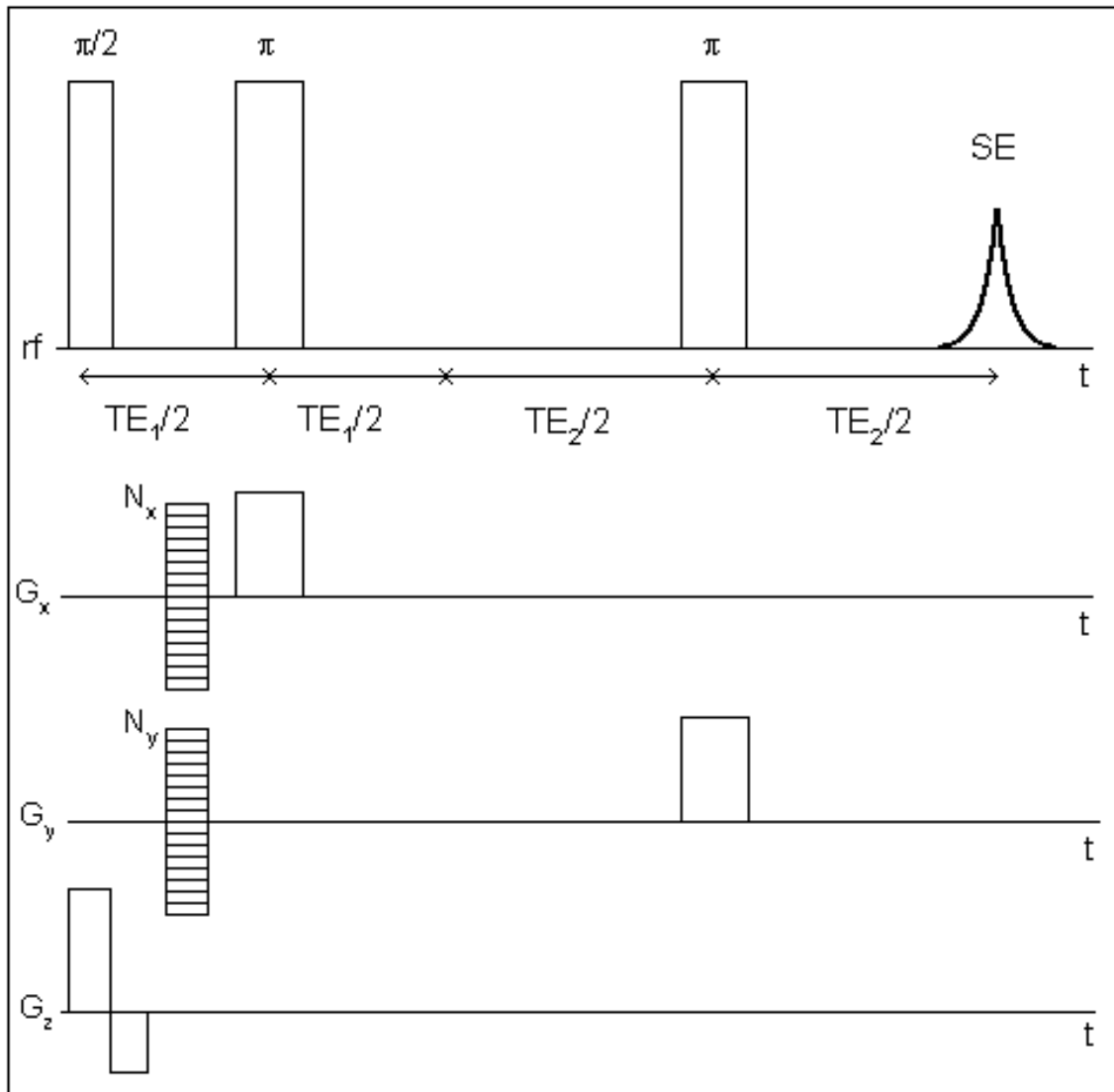


Fig. 2.3. Basic scheme of the 2D PRESS CSI sequence. Source: [5].

In case of 2D CSI phase coding is applied along two perpendicular directions. For example while measuring signals from a transverse plane (presuming z is parallel to \mathbf{B}_0) phase coding gradients in directions x and y are applied. By incremental change of these gradients signals from individual voxels within the preselected plane are measured. This way N_x by N_y signals are acquired.

In practice 2D PRESS CSI sequence usually follows a sequence for outer volume

suppression (OVS) which suppresses the signal from the region surrounding the selected CSI region which might otherwise reach into the measured area.

The data measured is reconstructed through 2D Fourier Transform from the frequency domain (so called K-space) to the spatial domain separately for each timepoint.

2.3 NMR Spectra Processing

The task of quantitative analysis of the NMR signal is mostly to measure the integral intensity of the signal, the frequency components of the signals and the frequencies of the maxima and the transverse relaxation time T_2 . In medical practice NMR spectroscopy is used to distinguish and identify signals of various metabolites in the spectra.

2.3.1 Metabolites of the Human Brain

Human brain contains wide variety of metabolites which can be detected through NMR spectroscopy. The information about metabolite content in the brain tissue is helpful in diagnosing and classifying human brain diseases and disorders such as brain tumors, epilepsy, multiple sclerosis and many other. The main metabolites of the human brain which can be detected in NMR spectra are N-acetyl aspartate (NAA), choline (Cho), and creatine along with phosphocreatine (Cr+PCr).

N-Acetyl Aspartate

N-acetyl aspartate has a dominant narrow peak in NMR spectra from human brain. Its resonance frequency is 2.01 ppm which comes from CH_3 group a much weaker component of the NAA signal is caused by CH_2 groups. The role of NAA in human brain is unclear. It occurs almost exclusively in neuronal tissue, specifically in cell sap of neurons. Therefore, the presence of NAA is often associated with the presence of living neural cells.

Choline and Phosphocholine

Choline resonates on frequency of 3.22 ppm. The strongest sharpest peak comes from N- CH_3 groups. Choline derivatives are mostly located in cell membrane or in the myelin sheathing of the axon. From the strength of the choline signal mostly the metabolic activity of the neuronal membrane can be estimated. In case of rapid cell death and membrane decay, choline washes out of the membranes into the intercellular space while the intensity

of its signal increases. Therefore, rapid increase of choline signal is often associated with cell membrane decay. The signal of choline and phosphocholine have a very close resonance frequency and are hard to distinguish in NMR spectra. Because of this the signals are often considered to be one and processed together.

Creatine and Phosphocreatine

These two metabolites have two strong peaks, one from a CH₃ group with the resonance frequency at 3.01 ppm and the other from a CH₂ resonating at 3.90 ppm. Creatine and phosphocreatine play a very important role in energetic metabolism of the human body. They participate in the synthesis of ATP,



This reaction is extremely important because it allows the rephosphorylation of ADP to form ATP and runs even anaerobically. Increased signal of creatine and phosphocreatine is often associated with increased metabolic activity in the measured region. Same as the above case of choline and phosphocholine, creatine and phosphocreatine have very similar resonance frequencies and are again usually processed as one.

2.3.2 Determination of the Signals of Metabolites

The accurate assessment of the signals of the individual metabolites present in the measured signal is a key part of the quantitative analysis. In principle, the measured signal can be processed in both time or frequency domain using parameterized or non-parameterized methods and imposing a different degree of prior knowledge about the evaluated signals [6]. Each approach mentioned has its advantages and disadvantages, and the selection of the processing method has to be considered for the given study (long or short TE , ^1H or ^{31}P spectroscopy) and requirements (absolute quantification, processing speed). Several programs enabling estimation of signal strengths or the concentrations of metabolites directly have been presented. To quantify ^1H SI spectra, the SI data interface CULICH, using the LCModel program for spectra fitting, has been developed [7], [8]. The K-space 2D-FFT zero filling interpolation, K-space filtering and grid shift have been performed according to the CULICH algorithm. The target concentrations used as targets for the training of the artificial neural network were acquired using the LCModel algorithm. Therefore, a brief description of the CULICH preprocessing methods and signal analysis by the LCModel program is given below.

2.3.3 Description of the CULICH Preprocessing

CULICH is a program developed for chemical shift spectroscopic image processing and result visualization. Currently, CULICH is used for data analysis of spectroscopic data in IKEM; therefore the same spatial interpolation algorithm and grid shift were performed on the input data as a part of the preprocessing process so that the results of the artificial neural network can be compared to those of CULICH.

Because the point spread function is not rectangular and extends over many voxels, the signal measured at the given position (the given voxel, in reality) is always contaminated by the signals from other locations in the sample. This signal ‘bleeding’ leads to the loss of localization precision, and it is the determining factor for the resulting resolution of the SI experiment. The elimination of this effect by the deconvolution of the measured concentrations into the known point spread function and the true metabolite signal is generally not possible, because all spectra in the SI grid are never available with the sufficient quality and deconvolution methods are very prone to noise. Even if the relative contributions of distinct voxels are small, the contamination by lipid signal can be severe due to big differences in metabolite and fat concentrations (1:103). Moreover, because the frequency of the lipid resonances in distant areas are often shifted (owing to different magnetic field strengths), signal bleeding may result in the spoiling of the whole spectral range.

For a fixed size of field of view, the profile of the point spread function improves if the number of phase-encoding steps is increased. However, due to time constraints and sensitivity reasons, a compromise in the number of encoding steps is necessary. Even if the number of phase-encoding steps is limited, other possibilities of improving the point spread function shape are available. Because the FT of the product of two functions is the convolution of their FT, multiplying measured k-space data with a filter function will influence the resulting point spread function. This post acquisition k-space filtering, also called apodization, is realized by the multiplication of the measured k-space data with symmetrical filters having the maximal value in the center of the k-space and smoothly decreasing toward its edges. The operation is performed for all time points. By using a proper filter function $w(l, m, n)$, the point spread function profile can be improved. In this case the Hamming filter

$$w(l) = 0.54 + 0.46 \cos\left(\frac{\pi l}{2l_{max}}\right), \quad (2.29)$$

where l ranges over N_x sampled values, and l_{max} stands for the maximal sampled value of l . Apart from the k-space filtering two additional preprocessing steps are usually performed: Zero filling and grid shifting.

After spatial reconstruction of SI data has been performed, spectra in all voxels of the spectroscopic grid are available. It is possible to increase the number of voxels artificially after the measurement. This operation, called zero filling, consists in appending zeros symmetrically to the signal in K-space prior to the FT. Zero filling represents an interpolation method and does not affect real spatial resolution. On the other hand it represents a FT-tailored interpolation which helps to improve the readability of spectroscopic images.

Another unique feature of the FT reconstruction is the possibility of adjusting the exact position of the grid after the measurement. This operation is based on the shift theorem of FT. A multiplication of all k-space values by a proper phase factor before the DFT results in shifting all positions of the voxels (the whole spectroscopic grid). This is very useful since the partial volume effects play an important role due to the large voxel size. By means of grid shifting, the area of interest can be centered to a specific voxel.

2.3.4 Description of the LCModel Program

The LCModel program is a commercial program, dedicated originally to single voxel spectra processing. The program decomposes the measured in vivo spectrum into spectra of individual metabolites measured in vitro, which form a basis set (see LCModel User's Manual [7] for more details).

To ensure the correct decomposition of the spectrum, the basis set must contain all metabolites detectable in the in vivo spectrum. Since the concentrations of all in vitro spectra in the basis set are known (Fig. 2.4), the basis set also serves as a concentration reference, enabling the calculation of concentrations of in vivo metabolites in the spectrum. The concentrations are determined as follows: First both the in vivo data and the in vitro basis set are zero filled to double the number of points in the time domain and Fourier transformed. The discrete in vivo spectrum $Y(\mathbf{v}_k)$ is then modeled using the basis set metabolites M_l as

$$\hat{Y}(\mathbf{v}_k) = e^{-(\phi_0 + \mathbf{v}\phi_1)} \left[\sum_{j=1}^{N_B} \beta_j B_j(\mathbf{v}_k) + \sum_{l=0}^{N_M} C_l \sum_{n=-N_S}^{N_S} S_n M_l(\mathbf{v}_k - n, \gamma_l, \epsilon_l) \right] \quad (2.30)$$

with the constraints

$$C_l \geq 0 \quad (2.31)$$

$$\gamma_l \geq 0 \quad (2.32)$$

$$(2.33)$$

$$\sum_{n=-N_S}^{N_S} = 1, \quad (2.34)$$

where C_l are concentrations of the N_M metabolites measured in vitro.

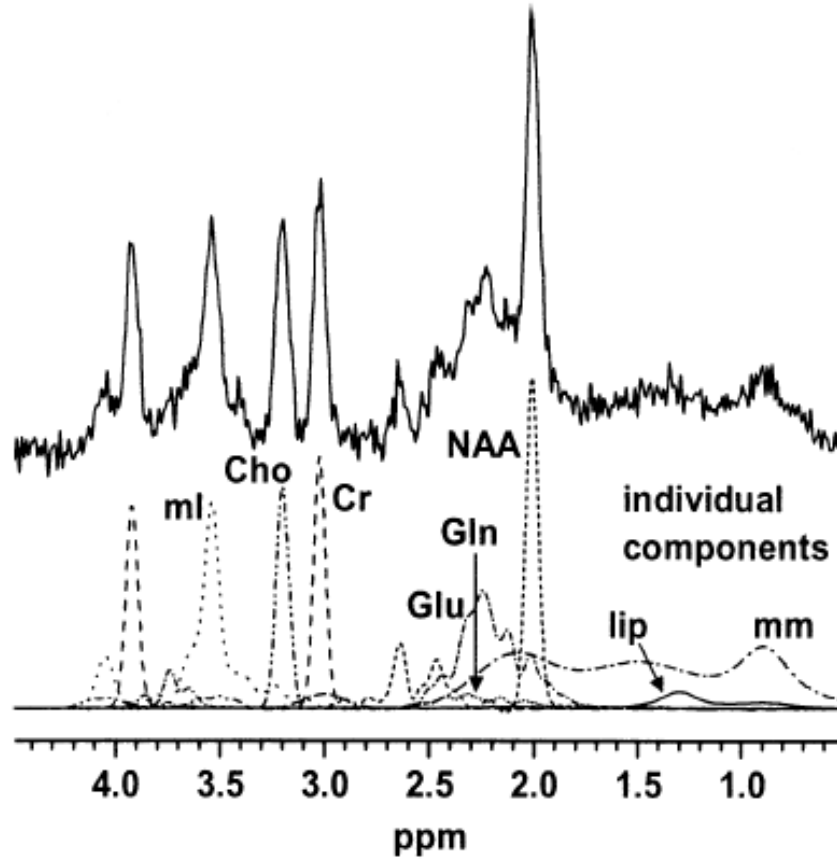


Fig. 2.4. The decomposition of the spectrum (upper picture) into individual metabolites included in the basis set (lower picture). Source: [4].

The N_M metabolite spectra in the basis set, $N_l(\nu, 0, 0)$, are broadened to account for shorter T_2 times in vivo (the parameter γ_l) and shifted to account for small errors in referencing the spectra (parameter ϵ_l).

The basis set metabolites $M_l(\nu_l, \gamma_l, \epsilon_l)$ with the parameters γ_l and ϵ_l are calculated from

$$M(\nu_l, \gamma_l, \epsilon_l) = FT(m_l(t)e^{-t(\gamma_l + i\epsilon_l)}), \quad (2.35)$$

where FT denotes discrete Fourier transform and $m_l(t)$ is the inverse DFT of the model spectrum $M_l(\nu_l, 0, 0)$. The sum

$$\sum_{n=-N_S}^{N_S} S_n M_l(\nu_{k-l}, \gamma_l, \epsilon_l) \quad (2.36)$$

is a discrete convolution of the basis set metabolites M_l with the lineshape coefficients S_n to account for field inhomogeneities and eddy currents. The baseline is modeled by N_B cubic B-splines, $B_j(\mathbf{v})$ with equally spaced knots. Φ_0 and Φ are zero and first order phase corrections.

The model parameters are estimated using a Marquard modification of a constrained Gauss-Newton least squares method

$$\frac{1}{\sigma^2(Y)} \sum_{k=1}^N \{Re[\hat{Y}(\mathbf{v}_k) - Y(\mathbf{v}_k)]\}^2 + T = \text{minimum}, \quad (2.37)$$

Besides the usual squared difference between the model function and the measured spectrum resolved in N data points, the term T containing additional contributions imposing the smoothness and zero boundary conditions on lineshape coefficients S_n and the smoothness of spline baseline in the minimized function is considered. The calculated concentration values are provided along with estimates of their Cramer-Rao bounds.

The scaling of basis set metabolites and of the measured in vivo spectrum is performed automatically by LCMoel according to the volumes of the voxels and corresponding transmitter reference amplitudes. The remaining corrections have to be performed individually by the user. The standard output of LCMoel analysis is shown in a figure below. Original phased spectrum, fitted spectrum, residuum, baseline and estimated concentrations with their Cramer-Rao bounds can be seen in the standard output (Fig. 2.5). In addition, all estimated parameters as well as details of the spectra fitting are available in the special text files.

2.4 Artificial Neural Networks

An artificial neural network (ANN), often just called a neural network (NN), is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons (perceptrons) and processes information using a connectionist approach to computation. In many cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

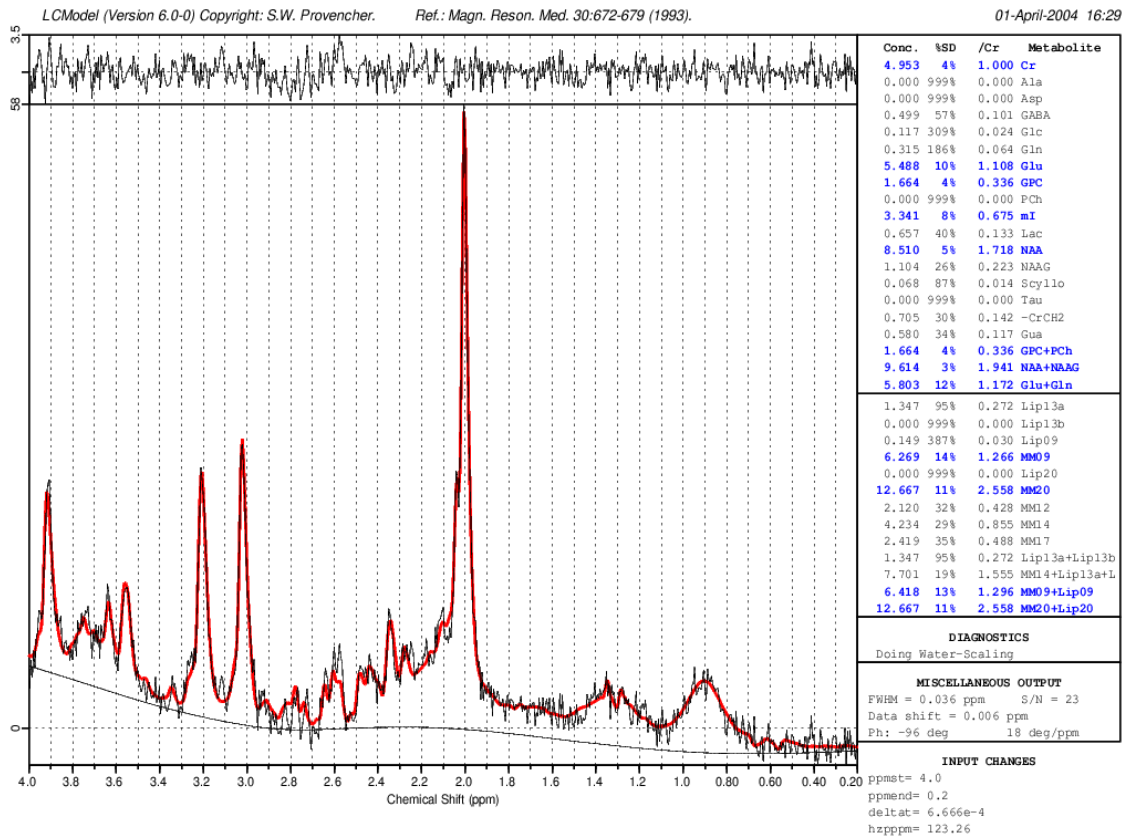


Fig. 2.5. The standard ‘one page output’ showing the results of the LCModel analysis. The fit of the spectrum is shown by the thick line and the baseline by the solid black line under the fitted spectrum. The residuum is shown in the upper part of the figure. On the right the estimated concentrations along with Cramer-Rao bounds and the ratio to creatine are shown for all metabolites in the basis set. Source: [4].

There is no precise definition as to what a neural network is, but most definitions agree that it involves a network of simple processing elements (neurons), which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. The original inspiration for the technique was from examination of the central nervous system and the neurons (and their axons, dendrites and synapses) which constitute one of its most significant information processing elements. In a neural network model, simple nodes (called neurons or perceptrons) are connected together to form a network of nodes—a neural network. While a neural network does not have to be adaptive per se, its practical use comes with algorithms designed to alter the strength (weights) of the connections in the network to produce a desired signal flow.

These networks are also similar to the biological neural networks in the sense that functions are performed collectively and in parallel by the units, rather than there being a clear delineation of subtasks to which various units are assigned.

In modern software implementations of artificial neural networks the approach inspired by biology has more or less been abandoned for a more practical approach based on statistics and signal processing. In some of these systems neural networks, or parts of neural networks (such as artificial neurons) are used as components in larger systems that combine both adaptive and non-adaptive elements. While the more general approach of such adaptive systems is more suitable for real-world problem solving, it has far less to do with the traditional artificial intelligence connectionist models. What they do, however, have in common is the principle of non-linear, distributed, parallel, and local processing and adaptation.

Neural network models in artificial intelligence are usually referred to as artificial neural networks (ANNs); these are essentially simple mathematical models defining a mathematical function mapping a set of inputs on a set of outputs. Each type of ANN model corresponds to a class of such functions.

There are various types of ANN, for the purpose of this thesis simple feedforward artificial neural networks were trained by a supervised learning algorithm called scaled conjugate gradient backpropagation.

2.4.1 Feedforward Artificial Neural Network

The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are

no cycles or loops in the network.

2.4.2 Supervised Learning

In supervised learning, we are given a set of example pairs (a set of targets \mathbf{x} and a set of outputs \mathbf{y}), and the aim is to find a function $f(\mathbf{x})$ in the allowed class of functions that matches the examples (find $f(\mathbf{x})$ so that $\mathbf{y} = f(\mathbf{x})$). In other words, we wish to infer the mapping implied by the data. The learning process is usually iterative. In order to quantify the progression of the learning process a cost function needs to be defined. The cost function is related to the mismatch between our mapping and the data and it implicitly contains prior knowledge about the problem domain.

A commonly used cost is the sum of squared error which tries to minimize the overall error between the network's output, $f(\mathbf{x})$, and the target value \mathbf{y} over all the example pairs. A common training algorithm for training neural networks which employs gradient descent for the class of neural networks called Multi-Layer Perceptrons to minimize this cost is called backpropagation.

Tasks that fall within the paradigm of supervised learning are pattern recognition, classification, regression, and function approximation.

2.4.3 Linear Perceptron

The Perceptron (Fig. 2.6) is a classifier that maps its input \mathbf{x} (a vector) to an output value $f(\mathbf{x})$ (a single value). It is a simple decision machine. The perceptron can be viewed as a classifier choosing its decision based on a set of advice from its advisors and adding its own bias. The input vector can be viewed as a set of votes of the advisors. The classifier has a certain trust in each of the advisors which affects how important the opinion of the advisor is for the final decision. It weighs the opinions of its advisors based on the trust it has in them and adds its own opinion (bias). The decision is made according to the sum of the weighted votes. In practice the perceptron consist of a set of weights (a weight vector \mathbf{w}) used for the calculation of a weighted sum of the inputs (input vector \mathbf{x}). To the total of the sum a bias is added. Sometimes the result is run through an activation function which maps the result to an interval within set bounds or a discrete set of output values (\mathbf{y}).

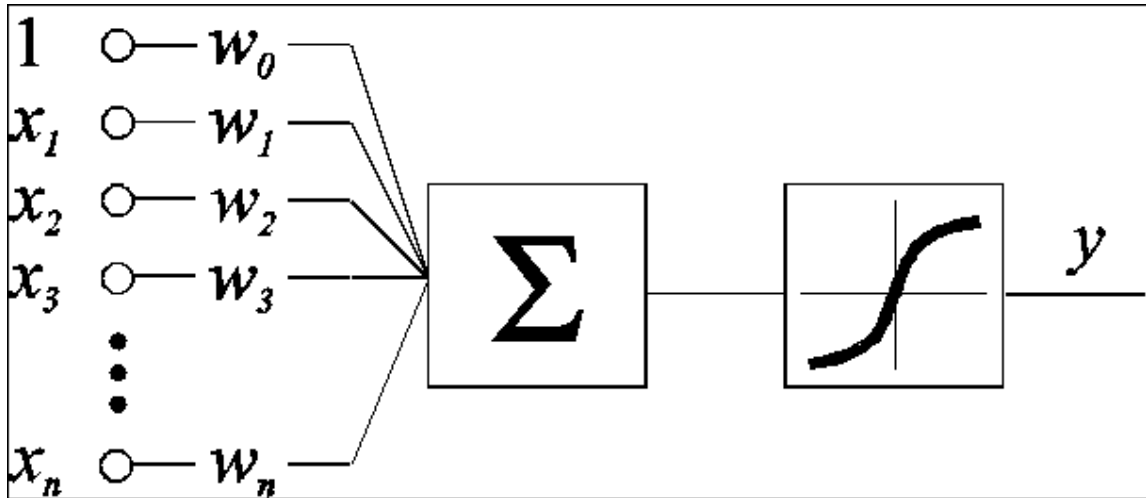


Fig. 2.6. Basic scheme of a perceptron. Image taken from the web.

Simple Perceptron Classifier

A single perceptron can work as a classifier used for classification into 2 classes 1, -1 with inputs being real vectors from \mathbf{R}^n vector space defining a decision hyperplane. Vector \mathbf{x} (x_1, \dots, x_n) is the input vector defining a single point in the \mathbf{R}^n vector space that needs classification. Output y is the decision of the classifier (either 1 for class one or -1 for class two). Vector \mathbf{w} (w_0, w_1, \dots, w_n) defines the weights of the perceptron corresponding to the values in the input vector, w_0 is the bias of the perceptron. For the purpose of easier notation a vector \mathbf{x}' is defined, where $\mathbf{x}' = (1, x_1, \dots, x_n)$. The activation function of the classifier is a simple step function $f(t) = -1$ for $t < 0$ and $f(t) = 1$ otherwise. This function maps the outputs of the weighted sum onto the set of classes 1, -1.

The perceptron classification function $y(\mathbf{x})$ can be expressed as

$$y(\mathbf{x}) = f(w_0 + \sum_{j=1}^n w_j \cdot x_j) = f(\mathbf{w} \cdot \mathbf{x}'), \quad (2.38)$$

$$f(t) = \text{sign}(t) \quad (2.39)$$

where the “.” Operator in $\mathbf{w} \cdot \mathbf{x}'$ denotes a dot product of the two vectors.

Learning of a Perceptron

To get the classifier to work properly it needs to be trained on a set of examples, e.g. a set of input vectors with a known affiliation to one of the classes. In the training (learning)

process the weight vector \mathbf{w} of the perceptron is adjusted so that all the examples are classified correctly.

The perceptron can be viewed as a hyperplane dividing the input vector space into two subspaces which define the two classes. Therefore, only linearly separable classes can be correctly classified by a perceptron classifier.

The learning process is iterative, in each step the hyperplane is deflected so that the new example lies on the correct side of the hyperplane. At first the weight vector is picked randomly, then in each step the weights are adjusted following this pattern:

$$\mathbf{w}_{new} = \mathbf{w}_{old} + \alpha(y_t - y) \mathbf{x}', \quad (2.40)$$

where \mathbf{w}_{new} is the new vector of perceptron weights, \mathbf{x}' is the extended input vector \mathbf{x} , y_t is the correct target class for the input vector \mathbf{x} , y is the output class assigned by the perceptron to vector \mathbf{x} according to the old weight vector \mathbf{w}_{old} , and α is an optional constant called learning rate which affects the importance of each new example. From this pattern it is obvious that the weight vector is adjusted only if y_t is not equal to y , only if the classification is incorrect. This procedure is performed for each example from the training set. If the two classes are linearly separable then the perceptron learning algorithm will always converge to a solution.

Since the inputs are fed directly to the output unit via the weighted connections, the perceptron can be considered the simplest kind of feedforward neural network.

2.4.4 Multi-Layer Perceptron

A multilayer perceptron is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output. It is a modification of the standard linear perceptron in that it uses three or more layers of neurons (nodes) with nonlinear activation functions, and is more powerful than the perceptron in that it can distinguish data that is not linearly separable, or separable by a hyperplane.

If a multilayer perceptron consists of a linear activation function in all neurons, that is, a simple on-off mechanism to determine whether or not a neuron fires, then it is easily proved with linear algebra that any number of layers can be reduced to the standard two-layer input-output model. What makes a multilayer perceptron different is that each neuron uses a nonlinear activation function which was developed to model the frequency of action potentials, or firing, of biological neurons in the brain. This function is modeled in several ways, but must always be normalizable and differentiable.

The main activation functions used in current applications are sigmoids, a differentiable variation of a step function such as

$$f(t) = \frac{1}{1 + e^{-at}} \quad (2.41)$$

or a hyperbolic tangent.

$$f(t) = \tanh(t). \quad (2.42)$$

The multilayer perceptron (Fig. 2.7) consists of an input and an output layer with one or more hidden layers of nonlinearly-activating nodes. Each node in one layer connects with a certain weight w_{ij} to every other node in the following layer.

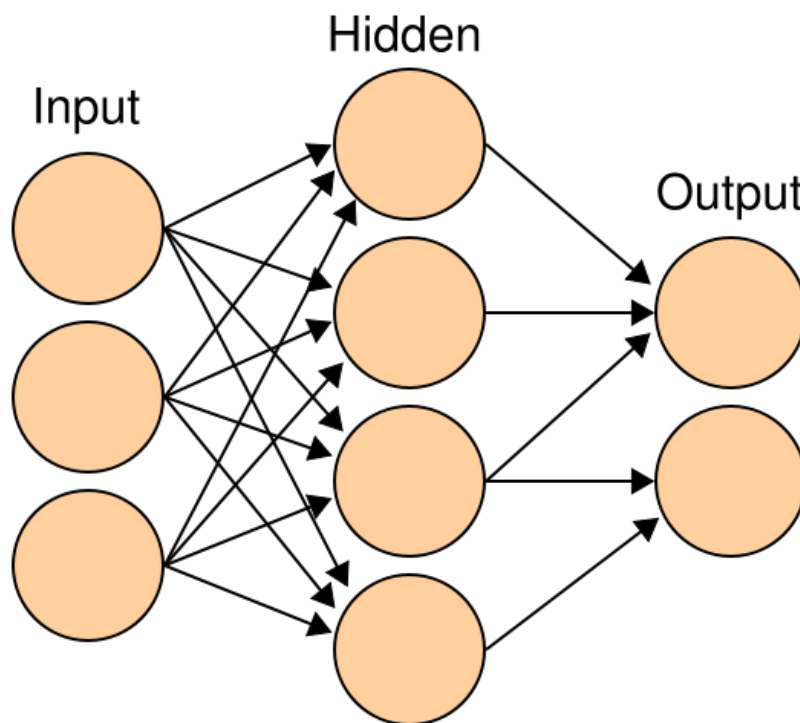


Fig. 2.7. Illustration of a 3-layer neural network. Image taken from the web.

2.4.5 Learning through Backpropagation

Learning occurs in the perceptron by changing connection weights (or synaptic weights) after each piece of data is processed, based on the amount of error in the output compared to the expected result. This is an example of supervised learning, and is carried out through backpropagation, a generalization of the least mean squares algorithm in the linear perceptron.

To be able to adjust the weights a measure of error needs to be defined. In this case of a multi-layer perceptron the output error is a function of all the weights in all the perceptrons in all the layers. A backpropagation algorithm adjusts the weights in the output layer first and then propagates backwards through the hidden layers adjusting the weights of all the other perceptrons. The correct weights are such that minimize the output error. Therefore, the task of the backpropagation algorithm is to find the minimum of the error function by adjusting the weights. It is based on gradient minimization method, in each step a gradient of the error function is calculated and the new weights are modified in the direction of the greatest descent of the error function.

A common measure of the output error is the sum of squared error E over all N inputs used for training

$$E = \sum_{i=1}^N (y_i - t_i)^2. \quad (2.43)$$

Then the new vector of weights $\mathbf{w}(k+1)$ can be calculated from the old vector $\mathbf{w}(k)$

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \alpha \frac{\partial E}{\partial \mathbf{w}}, \quad (2.44)$$

where α is a constant called the learning rate which defines by how much the new weights will change in the direction of the greatest descent of the output error. So that the gradient can be calculated, all activation functions of the perceptrons have to be differentiable. This explains why sigmoids have replaced the step function used in single perceptron models.

First the error of the neurons in output layer is calculated. Then the error term for hidden neurons is calculated by summing the error terms for each output neuron influenced by the hidden neuron, weighting each of the error terms by the weight connecting the hidden neuron to the output neuron. This weight characterizes the degree to which hidden neuron is responsible for the error in the output neuron.

The training process is repeated for each of the input examples from the training set and then for the whole set again until the convergence criterion is met. Some tasks allow for the error to descent to zero for most a certain threshold is set. If the error descent below the threshold or if it does not noticeably decrease for a number of iterations the training is terminated.

Here is the basic scheme of the backpropagation algorithm:

1. Choose initial weight vector \mathbf{w}_k and set $k = 1$.
2. Determine a search direction \mathbf{p}_k and a step size α_k so that

$$E(\mathbf{w}_k + \alpha_k \mathbf{p}_k) < E(\mathbf{w}_k).$$

3. Update vector: $\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha_k \mathbf{p}_k$.
4. If $E(\mathbf{w}_k) > T$ threshold then set $k = k + 1$ and go to 2. Else return \mathbf{w}_{k+1} as the desired minimum.

2.4.6 Scaled Conjugate Gradient Backpropagation

Scaled conjugate gradient backpropagation is a variation of a backpropagation algorithm that optimizes memory usage and step size selection using principals of conjugate gradient minimization developed in numerical mathematics for multivariate function minimization [9]. It is useful especially for large scale neural network training where classic backpropagation algorithms might run out of system memory. It also speeds up the convergence of the training process considerably. The conjugate gradient method chooses the search direction and the step size more carefully by using information from the second order approximation of the output error as well.

2.5 Hankel-Lanczos Singular Value Decomposition

HLSVD is an accelerated approximation of singular value decomposition (SVD) through Lanczos algorithm applied on a Hankel matrix. SVD is a method of identifying principal components of a non-square matrix. Let's assume a signal of length N which can be expressed as a sum of M various damped sinusoids. It was shown that if such a signal is rewritten in a matrix \mathbf{X} so that the i -th row of the new matrix is equal to the signal rotated to the right by $(i-1)$ elements. This way an N by N matrix is formed. This new matrix has M non-zero principal components. The first row of each principal component corresponds to one of the damped sinusoids which formed the base signal.

SVD is not an elementary operation and for large matrices take in order of minutes to run. If the number of damped sinusoids in the signal is known (M) a much smaller matrix with the same principal components can be formed. This matrix is called Hankel matrix. The Hankel matrix \mathbf{H} has $N - M + 1$ rows and M columns. In the first row are the first M elements of the signal. In the second row is the second to $(M+1)$ st element and in the

$(N - M + 1)$ st row are the last M elements of the vector:

$$\mathbf{H} = \begin{pmatrix} x_1 & \dots & x_M \\ x_2 & \dots & x_{M+1} \\ \vdots & \ddots & \vdots \\ x_{N-M} & \dots & x_{N-1} \\ x_{N-M+1} & \dots & x_N \end{pmatrix}$$

If M is a lot smaller than N then the size of the Hankel matrix \mathbf{H} is a lot smaller than the full matrix \mathbf{X} and SVD runs a lot faster. To restore the vector corresponding to the principal component the first column and the last row are read (elements x_1 through x_N).

For very complex signals even the Hankel matrix is too large to be fully decomposed. Lanczos algorithm was developed to approximate the first largest singular values and the principal components associated with them.

Chapter 3

Experimental Part

In the experimental part the protocols and equipment which were used to measure the data that were further processed for the neural network training and performance testing are described.

3.1 Simens Magnetom Trio-3T

For the purpose of magnetic resonance spectroscopy measurements in studied subjects Magnetom Trio-3T designed by Siemens was used. Through superconductive coils cooled by liquid helium it generates static magnetic field of magnetic induction 2.89 T. Full body coil was used for excitation. The spectroscopic data was measured by an 8-channel Transmitter/Receiver head spectroscopic coil with proton resonance frequency of 123.26 MHz.

3.2 Institute for Clinical and Experimental Medicine

The data used for the purpose of this thesis were measured at the department of magnetic resonance at the Institute for Clinical and Experimental Medicine (IKEM) in Prague. Data were measured on both healthy volunteers and patients. All the data from patients were taken as a part of standard examination protocol used in IKEM; no additional burden was put on the patients in order to gather data for the purpose of this thesis. Data for ANN modeling and testing were taken from clinical examinations of patients and healthy volunteers with written consent. The measurement procedures followed a protocol approved by the Ethic Committee.

3.3 MR Examination

Before the examination all the patients and volunteers were informed of the measurement procedure. Most of all it was confirmed that the subject does not have any ferromagnetic implants, cardio stimulator or a pacemaker.

During the measurement the examined subject is required to lie on his back back in the tunnel of the MR scanner. His head is placed in the head spectroscopic coil. For the purpose of best magnetic field homogeneity the center of the table is moved into the tunnel so that the center of the spectroscopic coil is in the center of the body coil. For safety measures each subject holds a signal balloon which can be pressed to signal any distress or danger.

To localize the subject within the scanner a localizer sequence was performed. Then proper MR images were taken. Lastly 2D CSI spectroscopic image was taken.

3.4 MR Image Acquisition

MR images were taken in 3 perpendicular directions: sagittal, frontal, and transversal. In each direction multiple slices were measured. The images were measured using a turbo spin echo sequence (TSE). The slice thickness was 4 mm and the distance between two equidistant slices was 10 mm. The MR images were used for proper volume of interest (VOI) selection for the spectroscopic image acquisition. Once all the spectroscopic data was measured and processed the results were projected onto the nearest MR image for the purpose of visualization.

3.5 Spectroscopic Image Acquisition

Based on the MR images a slice for excitation is selected. In all the experiments used in this thesis the slice thickness is 15 mm. Once the slice is selected the region of interest within this slice is chosen. The volume of interest is divided based on the phase coding step sizes into separate voxels. In this study the voxel length and width was set to 10 mm. With slice thickness of 15 mm it means each voxel was 1.5 ml in large. The volume of interest standard size is a grid of 8 by 8 voxels but for some of the experiments it was adjusted slightly to fill the region needed for proper examination of the subject.

Once the volume is selected the magnet needs to be tuned to local inhomogeneities of the magnetic field in the region of interest. The tuning (shim) is done by a set of 16

coils. First the magnetic induction of the magnetic field is mapped. Then in an automatic process the inhomogeneity function is being minimized as a function of electric currents in the shimming coils. This is a key part of the measurement. If the field is not homogeneous enough the peaks in the NMR spectra are too broad to be distinguished and properly identified. If the automatic shimming procedure fails the currents in the shimming coils were adjusted manually. During the measurement additional sequence for outer volume suppression (OVS) was applied.

A standard 2D PRESS CSI sequence was used for all the spectroscopic images. Repetition time (TR) was set to 1510 ms while the echo time (TE) was 30 ms.

3.5.1 Water Suppression Sequence

The signal of hydrogen bound in molecules of water is the strongest component of the FID signal from human brain tissue. With no water suppression the FID consists almost only of the water component. For the purpose of this thesis water suppression was done by application of an additional water saturation pulse during the 2D PRESS CSI sequence. In water saturation an RF pulse is used to keep the hydrogen protons in the water molecules fully excited so that they can not react to the next excitation pulse and that way contribute to the FID signal.

For each subject two measurements were taken, one with water suppression using a water saturation RF pulse, one with no water suppression. The signal of water is used as a scale and allows quantifying the concentration of metabolites from the signal with suppressed water by comparing it to the strength of the signal of unsuppressed water.

3.5.2 Spectroscopic Data Averaging

Because the MR signals measured in vivo have a low signal to noise ratio, signals acquired through more measurements are averaged for better results. The signal of unsuppressed water is so strong that only one acquisition needed to be taken compared to four averages without the suppression. The number of averages is limited by the total time of the measurement. A 2D PRESS CSI sequence with 4 averages used in this study takes 7 minutes and 15 seconds to run. The maximum time spent in the magnet at once is 1 hour. Within this time all the examinations prescribed by the doctor have to be done including possible manual shimming which might be quite time demanding.

3.5.3 Delta Frequency

The volume of interest is excited by a sequence of radio frequency pulses. The position of the excited volume is coded through frequency band of the excitation RF pulses. Since the resonance frequency varies based on the type of the proton (or based on the metabolite in which the proton is bound) the position of the excited volume varies slightly for each proton type (metabolite). To minimize this effect the carrier frequency of the RF pulse, is set to the central frequency of the metabolite spectrum. The difference between the resonance frequency of water in vivo and the carrier frequency of the RF pulse is called the delta frequency. For the purpose of this study the delta frequency was set to -2.2 ppm (close to the resonance frequency of NAA) for measurements with water suppression and to 0 ppm with no water suppression (so that the resonance frequency of water is in the center).

3.6 Data Export

Once the data is measured it is exported from the scanner to sPACS, a database server, in dicom format used for the purpose of medical imaging. Further processing of the magnetic resonance signals was done on a PC in mathematical software called MATLAB. In order to access the data in MATLAB a script widely used in IKEM designed to read dicom files and decode the stored data was used.

3.7 Metabolite Concentration Quantification Toolbox

An application was designed in MATLAB which uses several dicom files from the scanner as inputs: a set of T_2 weighted MR images measured in slices parallel to the slice of the chemical shift image, a 2D chemical shift image with water suppression, and a 2D chemical shift image with no water suppression. The application processes the 2D CSI measured with water suppression (the metabolite spectrum), creates a voxel grid and assigns concentrations of NAA, choline, and creatine metabolites to each voxel. The 2D CSI image with no water suppression (the spectrum of water) is used to estimate the water content in each voxel. The signal from each voxel is divided by the water content in that voxel, this means that final signal expresses the ratio between the metabolite content and water content in the particular voxel—the metabolite concentration. Once the concentration of

the metabolites are calculated for each voxel the result is plotted as a 2D map over a MR image measured in the closest plane to the center of the CSI slab.

The application also allows optional uploading of corresponding results acquired from the commercial LCMModel software for comparison of the two methods. Once the LCM results (targets) are uploaded an error map is calculated and the differences between the two sets of results are processed statistically. The calculated error as well as the target concentrations from LCM can be also plotted over the same MR image for comparison.

3.8 Spectroscopic Data Processing

Several steps of signal processing of the spectroscopic data are done before the data is ready to be fed into the artificial neural network (illustrated in Fig. 3.1). First the 2D CSI signal is interpolated to double number of voxels in each direction in within the measured slice. Then the residual signal of water is removed from the signal. Then the signal is filtered to reduce noise level. Then the signal is referenced to the signal of water. Now the spectra are calculated and in the next step the baseline correction is applied to the spectra. Now the data is ready to be used for ANN training, or if the network is already trained, it can be fed into the network to obtain the desired concentration maps. Finally the maps are visualized over a MR image. The procedure is summarized in a diagram in Fig. 3.2. Below each step is described in detail.

3.8.1 2D Data Interpolation

The measured signals were stored by the 3T Siemens Trio Tim system in spatial domain. To virtually increase spatial resolution (for the purpose of better transparency of the result, the resolution of the MR experiment is not affected) the data were interpolated. In order to interpolate a two dimensional inverse Fourier transform was preformed, then the data were symmetrically padded by zeros to double resolution, then again transformed back to spatial domain by a 2D-FFT. That way the size of the matrix from one subject was 32 by 32 voxels with a signal of 2048 time points each. The principals of the 2D FFT-based interpolation are described in the theoretical part of this thesis.

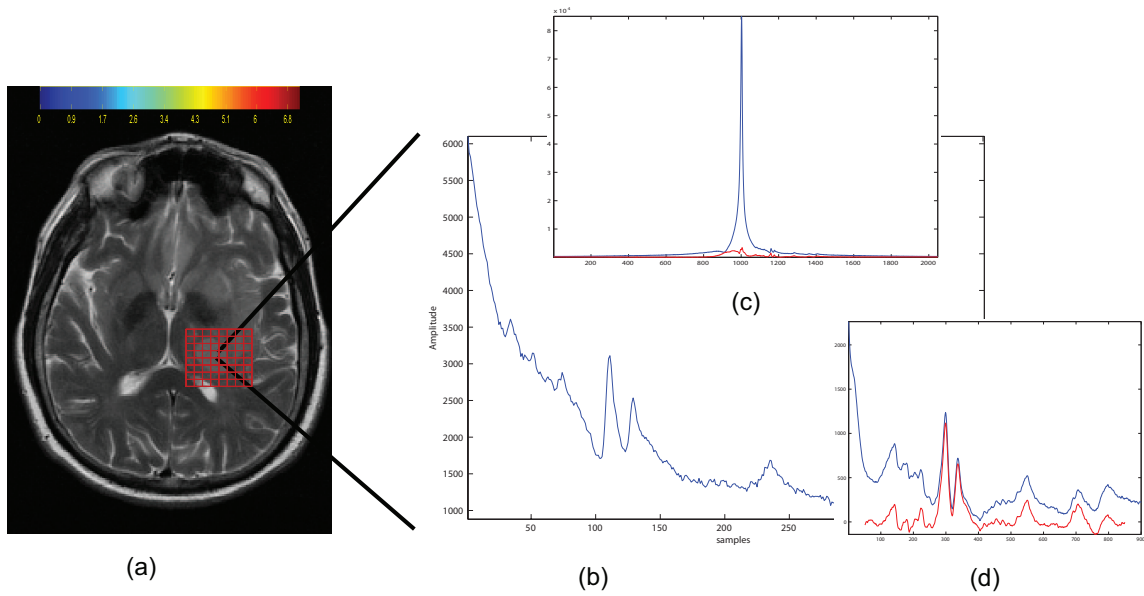


Fig. 3.1. Absolute NMR spectrum illustration. (a) Spatial configuration (16x16 voxel) of NMR recording. (b) example of NMR raw spectra. (c) after HLSVD water suppression apodisation and zero filling. (d) Same spectrum after baseline correction.

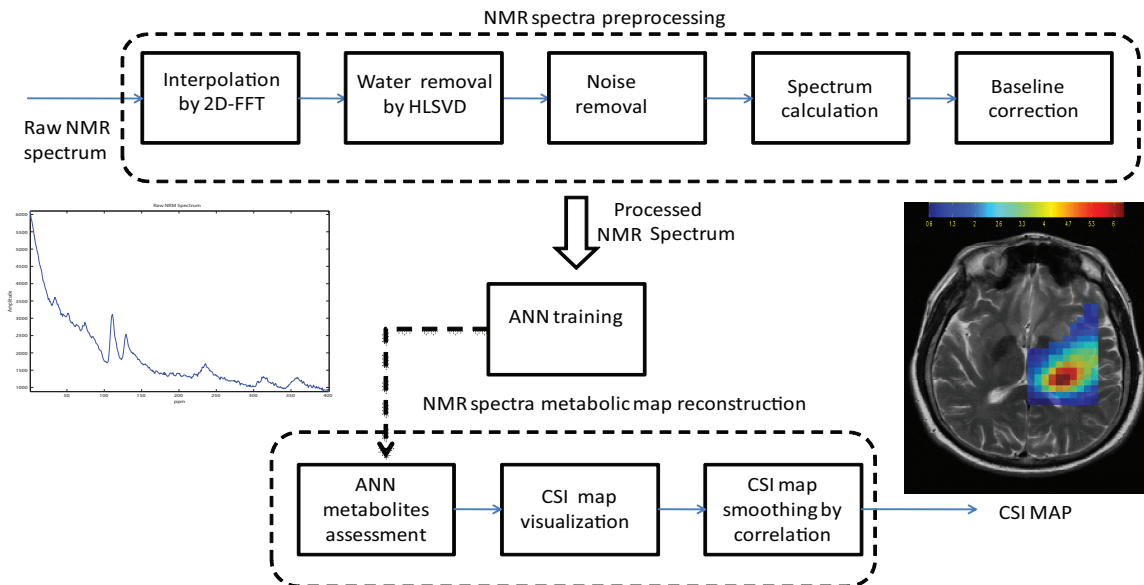


Fig. 3.2. Algorithm flow-chart diagram.

3.8.2 HLSVD Residual Water Signal Removal

Since the signal of free induction decay is a sum of damped sinusoids (one from each type of protons from the measured sample) the HLSVD algorithm can be used to identify each component and quantify the concentrations of the metabolites associated with it. An open source application which implemented this algorithm is called MRUI [10]. The performance of MRUI on in vivo noisy signals measured with low external field homogeneity is poor. However, this algorithm is very strong in residual water component identification.

Even after application of water suppression a significant remainder of the signal of water persists in the signal. To remove the remaining water component HLSVD was used. The most significant principal component of the signal after SVD shows to be a good approximation of the remaining water signal [11]. In order to quickly calculate the first singular value and the corresponding vectors, the Lanczos algorithm was applied on a Hankel matrix composed for the signal from each voxel. Once the component corresponding to the first singular value is calculated it is subtracted from the signal (Fig. 3.3).

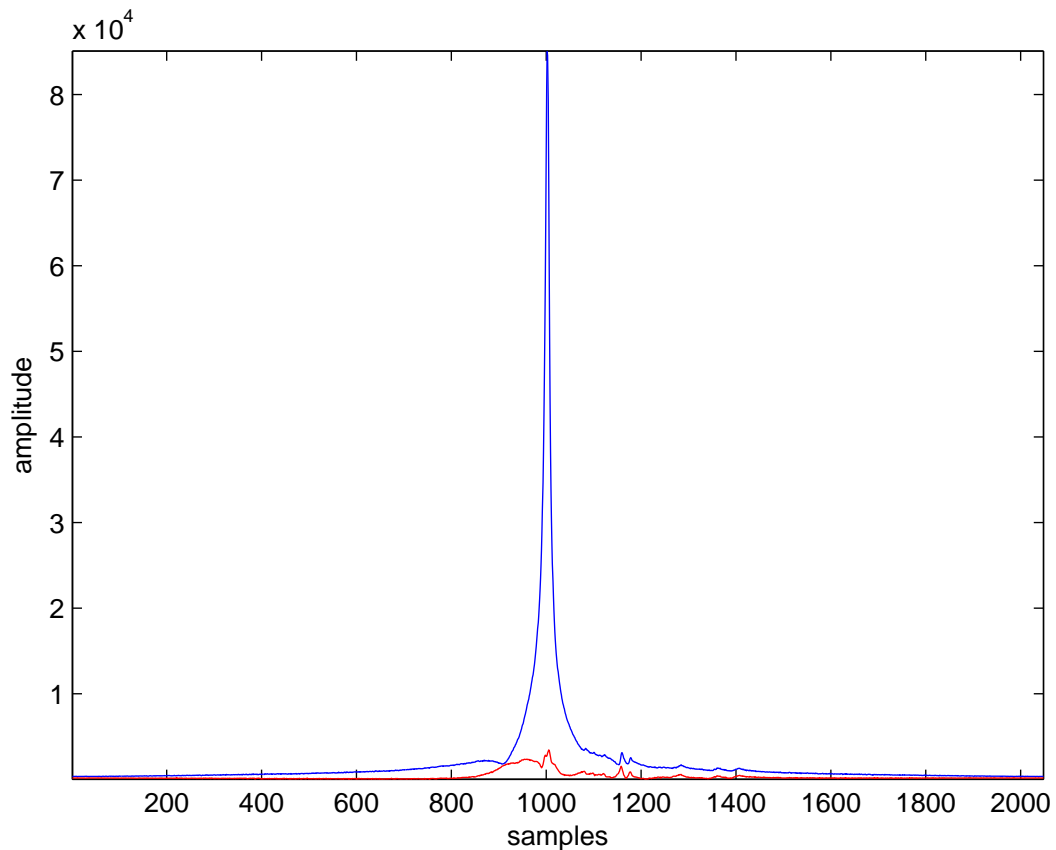


Fig. 3.3. An example of raw NMR spectrum (blue) and NMR spectrum after HLSVD residual water signal removal (red), the frequency axis is in samples.

3.8.3 Noise Removal

To reduce noise the signal was multiplied by a time damped exponential $e^{-t/500\text{ms}}$.

The significant signal components are time damped as well. After multiplication by a damped exponential the significance of noise does not increase as rapidly with time. Basically, each time point of the signal is weighted based on the SNR in that point. The noise level is constant throughout the whole experiment while the signal diminishes with time. Therefore the weighting function follows the absolute amplitude evolution of the signal. The spectra calculated after this noise reducing procedure are significantly smoother. However, the signal is damped as well and the metabolite peaks in the spectrum are widened. Therefore, the time constant has to be selected carefully. For the purpose of this thesis several values were tested, the value of 500 ms showed a good compromise between peak broadening and noise level reduction.

3.8.4 Water Referencing

The same processing is done for the signals measured with no water suppression so that the amplitudes of both signals are processed equally. Discrete fast Fourier transform (DFFT) of the reference 2D-CSI signal of water, measured with no water suppression, is then calculated and integrated across all frequencies for each voxel separately. The integral corresponds to the water content in the particular voxel. For each voxel the amplitudes of signal with water suppression are divided by the integral of the water spectrum of that voxel.

3.8.5 NMR Spectrum Calculation

After removing the remaining water components, noise reduction and water referencing the signal is padded with zeros to double resolution in time domain. Then discrete fast Fourier transform is performed and an absolute value is taken.

3.8.6 Baseline Correction

The spectrum now consists of a few metabolite peaks and an underlying baseline. The points of the spectrum that are part of the peaks are usually a part of relatively a steep slope. To identify the location of the peaks for each point of the spectrum a standard deviation of the points in its near neighborhood is calculated. If the standard deviation is above

a certain threshold (set to 20% of the maximum amplitude in the metabolite spectrum) the point is classified as a part of a metabolite peak. The remaining points are classified as a part of the baseline. Through these a polynomial of the order of 6 is fitted and subtracted from the spectrum resulting in baseline correction (see Fig. 3.4). This is important since the spectrum also contains signal components from larger macromolecules which are characteristic with low amplitude very wide peaks. If these peaks are not removed from the signal they can significantly affect the metabolite concentration estimation. The baseline correction is applied only to a part of the spectrum, a wider neighborhood of the part of the spectrum which contains all of the metabolite peaks. The polynomial fit error is largest on the borders of the fitted area. Therefore, the area of interest is increased slightly for the purpose of polynomial fitting (by 50 points on each side).

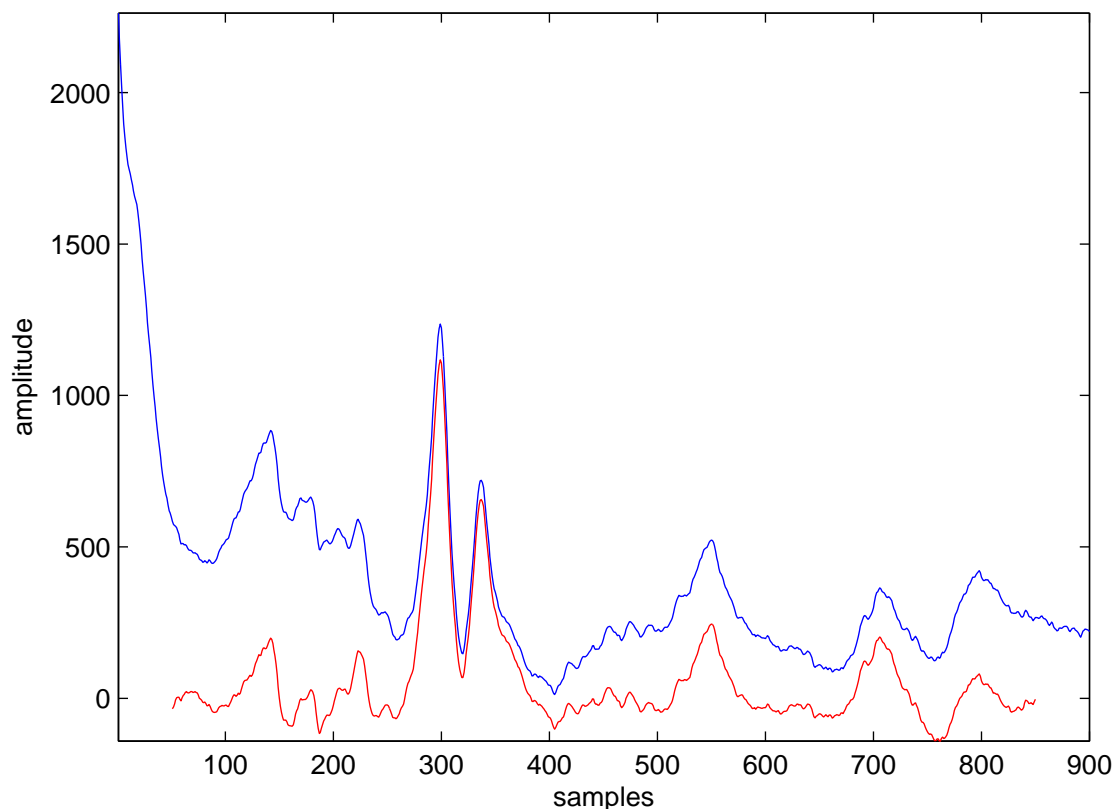


Fig. 3.4. An example of NMR spectrum after HLSVD residual water signal removal (blue), and the same spectrum after baseline correction (red), the frequency axis is in samples, frequency band selection corresponds to the band used for ANN training.

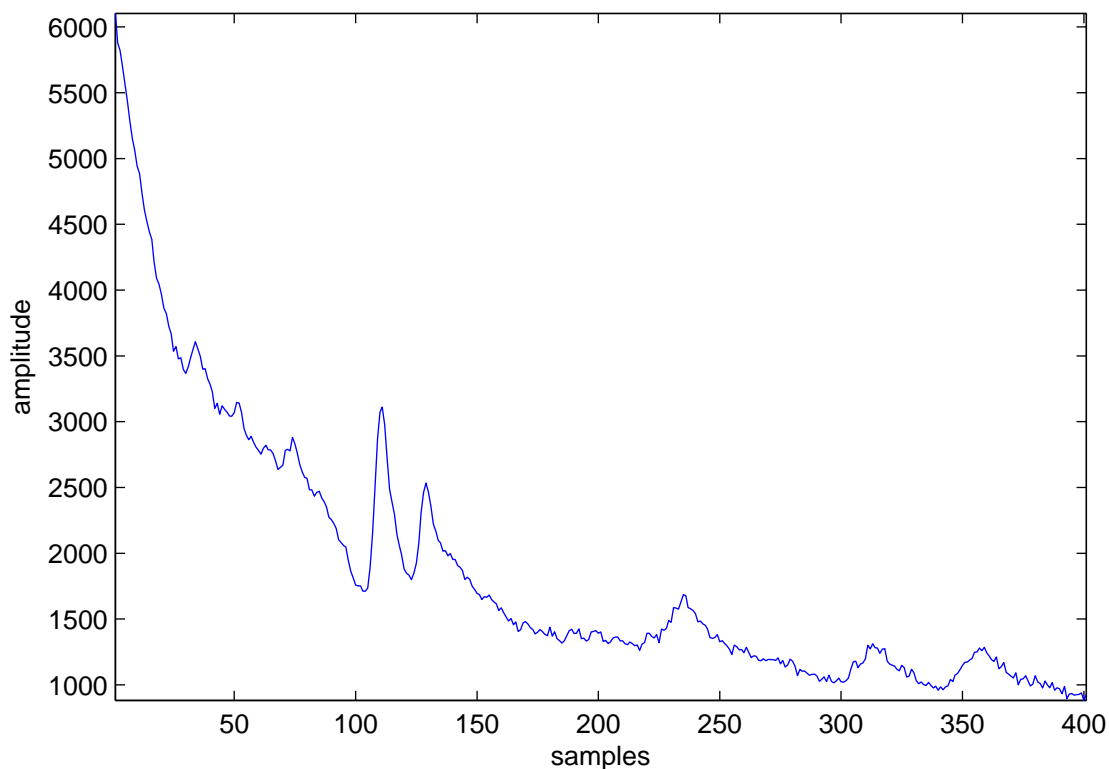


Fig. 3.5. An example of unprocessed NMR spectrum, the frequency axis is in samples, frequency band selection corresponds to the band used for ANN training.

3.8.7 Spectral Band Selection

Finally only 800 points of the spectrum in frequency band from approximately 0 ppm to 4 ppm (all the metabolite peaks are within this band) is cut. Therefore, the final vector representing the metabolite spectrum used as an input for the neural network is 800 points long.

A raw NMR spectrum within this frequency band and a fully-processed spectrum within the same frequency band are shown in Fig. 3.5 and Fig. 3.6, respectively.

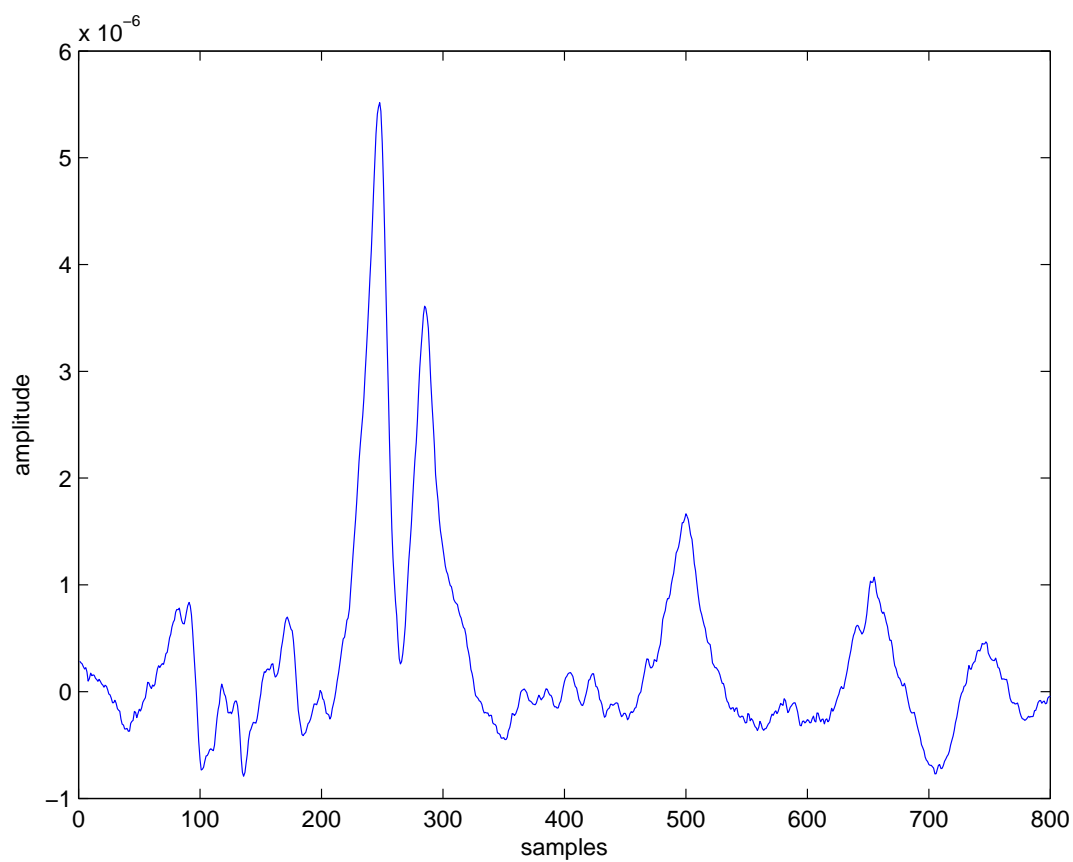


Fig. 3.6. An example of fully processed NMR spectrum, the frequency axis is in samples, frequency band selection corresponds to the band used for ANN training.

3.9 Training Data

Now that the preprocessing of a NMR chemical shift image is done enough data needs to be collected and processed to allow training of an artificial neural network. For best results the training set needs to be as large as possible with maximum possible variety to maximize the generalization ability of the network. For best results 11 2D chemical shift images from 10 subjects (over 2000 spectra altogether) were used for training. 7 subjects were healthy volunteers, one was a patient diagnosed with a brain tumor, and two were patients diagnosed with epilepsy. The quality range of the spectra was as wide as possible matching the quality range of the measurements that are found sufficient by the MR department in IKEM for the purpose of medical diagnosis.

3.10 ANN Set-up

The set-up of the artificial neural network was determined experimentally. Networks with two to four layers were tested. A network with only two layers (a perceptron) was tested mostly to verify the complexity of the problem and to justify application of a multi-layer perceptron neural network. Because of the size of the dataset larger four layer networks were very time and memory consuming to test and in any of the set-ups tested did not deliver acceptable results. In the three layer models a variety of hidden layers were tested ranging from 30–300 neurons. The smallest number of neurons in the hidden layer that allowed for the training algorithm to converge was 85 neurons. The initial idea was to train one network with three output neurons, one for each metabolite. But separate networks for each metabolite were tested later on and showed much better results and convergence ability of the training process.

Finally the following set-up was chosen: The neural network has three layers. 800 points in the input layer are necessary to meet the length of the input vector. The number of neurons in the hidden layer was set to 100. The output layer consists of only one neuron. The output is a real number and is directly equal to the approximation of the concentration of the metabolite for which the network was trained. Concentrations are in laboratory units (no correction for T_1 , T_2 , white and gray matter, and CSF or field inhomogeneity were done). The output function for the hidden layer is `tansig`. There is no output function in the output layer (it remains purely linear). Three networks were trained to calculate concentrations of three main metabolites in the NMR spectra of human brain (NAA, choline, and creatine). As a convergence criterion the sum of squared error was used. Scaled conju-

gate gradient backpropagation algorithm was used to train the network. Several variations of the backpropagation training algorithm were tested. The scaled conjugate gradient backpropagation proved to cope with the large scale of the data the best (memory-wise as well as time-wise). The convergence criterion was set to total sum of error squared of 0.5 in all 700 inputs used for training. The time needed for the training algorithm to converge was approximately 10 minutes and 3000 to 5000 iterations were needed.

3.10.1 Training Dataset Reduction

Because of bad homogeneity of the magnetic field, some of the signals were not good enough to be processed correctly. While such signals were present in the training set the training algorithm could not converge. To enable the convergence of the training algorithm the number of vectors in the training set was reduced to 700 spectra randomly selected from the spectra of the training subjects. Moreover, the spectra too different from an average spectrum from the training set was restricted from the training. Even though, large variety of the training data is good for generalization ability of the network, some spectra were totally out of phase or just of such bad quality that including them in the training set prevents convergence of the training procedure of the neural network. The similarity was determined by the correlation coefficient of the spectrum and an average spectrum calculated from a large set of manually picked spectra used for training (a standard good quality spectrum). The right threshold needed to detect useless spectra was found experimentally and set to 0.8.

3.10.2 Application of the Artificial Neural Network

Once the networks were trained they were ready for testing. For a CSI signal from a MR tomograph to be analyzed by these networks it needs to be preprocessed by the same procedures as were all the training signals. Once the set of spectra (one for each voxel from the volume of interest) is calculated it is fed one by one to the three neural networks to calculate the concentrations of NAA, choline, and creatine respectively for each voxel separately.

3.11 Data Visualization

When the concentrations are calculated a metabolite map can be plotted over a corresponding MR image to demonstrate the spatial distribution of the metabolite. In the dicom file downloaded from the scanner a precise location of the excited volume is noted. This information is used to calculate the distance of each MR image slice measured from the central plane of the chemical shift image. The MR images are measured from slices 4 mm thick while the CSI slice is 15 mm thick. Moreover, the center of the MR image does not have to match the center of the CS image. This has to be taken in consideration while evaluating the images; the MR image serves mostly better orientation and special localization within the CS image.

Once the MR image closest to the CSI slice is selected the metabolite map is plotted over it. Some of the spectra from the measured volume of interest might be of very low quality—usually the spectra from border areas or some areas of local field inhomogeneities which might be caused by bones, ferromagnetic metal deposit, tooth fillings, brain cavities, etc. Such spectra were not included in the training data set and therefore the result of the artificial network analysis can not be predicted. These results have very low information. Moreover, these results can be numerically incomparable to the rest of the data and affect the colormap scale of the plotted metabolite map in such a way that no other data is readable. To avoid this effect the correlation coefficient of each spectrum with the previously introduced standard spectrum (an average spectrum from a large number of spectra in the training dataset) is calculated. The user can then manually pick a threshold. Spectra from each voxel with the correlation coefficient lower than the selected threshold are not shown in the plot.

Chapter 4

Results

4.1 Performance Test Data

Some data from healthy volunteers was used for optimization of all the parameters defining the procedure such as network topology, noise reduction filter time constant determination, convergence criterion value determination and other. To avoid dependence between the data used for optimization and testing a completely new dataset was measured for the purpose of testing of the developed application.

The performance of the system was tested on three healthy volunteers, one patient with a progressive brain tumor, one patient with a suspicion for a brain tumor and one patient with a suspicion for a metabolic disease.

4.2 Test Criteria

The main aim of the artificial neural network analysis was to duplicate the results of the LCMModel algorithm. Therefore the test criteria were set to describe the similarity between the two algorithms. The main criterion is the visual similarity between the metabolite maps calculated by each method. This similarity does not have to match in colors since those are affected by the color scale which is defined by the maximal and minimal values within the dataset. It is important that the areas with lower and higher metabolite concentrations match for both algorithms.

Along with the visual criterion a few statistical values are calculated to describe each method. To show the difference between the two methods an error map is calculated, it is simply the map of absolute difference between the results delivered by the neural network

and the LCMoel algorithm. Then the average values of both methods are compared to determine whether the neural network approximation is biased. To show how good an estimate of the LCMoel results is the neural network able to deliver the standard deviations of the methods are compared. Along with it an average relative error is calculated. All the averages are calculated over all spectra from the CSI voxel grid.

4.3 Test Results

The spectra of healthy volunteers are usually very similar in each voxel of the excited volume of interest. Therefore, this data is good for general performance test but it can not be used to demonstrate the ability of the network to distinguish the difference between spectra. Based on the results the data used for performance testing can be divided in three groups.

First group of high quality spectra from two healthy volunteers measured from a large area at the top of the frontal lobe of the brain did not show almost any variety. This data was analyzed correctly with maximum relative error of 10%, the estimate did not show to be biased (the average concentrations varied by $\pm 5\%$ and the standard deviations were similar as well). This is a very good result. However, this was expected for this type of data, such data can be processed successfully by the means of much simpler algorithms (such as simple maximum amplitude detection) in shorter computational time.

Next group of the test subjects is represented by one patient with a suspicion for a brain tumor. The measurement quality was very low and even the results from the LCMoel were questionable. The comparison of the ANN approach and LCMoel was possible but the meaning of the result is questionable.

The last group of test subjects was of most significance. The key test subject was a patient diagnosed with a brain tumor. The LCMoel identified the tumor as progressive. Another good test subject was a 7-month old patient with a suspicion for a metabolic disease. LCMoel showed no significant asymmetry in the metabolite map. Data from one of the volunteers was measured from cerebellum and was of poor quality. LCMoel was able to successfully identify the concentrations of metabolites from most of the volume of interest. These results for these three test subjects are further discussed in more detail. Table 4.1 summarizes the statistical comparison of the LCMoel and ANN approaches for this data.

Table 4.1. Comparison of LCM and ANN algorithm performance. Subject 1 is a patient diagnosed with a brain tumor, subject 2 is a 7-month old patient with a suspicion for a metabolic disease, and subject 3 is a healthy volunteer. High error and discrepancies in mean or variance are usually caused by a small number of voxels near the border of a measured volume with the poorest signal quality.

| Subject | Method | Choline | | | Creatine | | | NAA | | |
|---------|--------|---------|------|-------|----------|------|-------|------|------|-------|
| | | avg | std | er(%) | avg | std | er(%) | avg | std | er(%) |
| 1 | LCM | 1.40 | 3 | | 5.16 | 1.83 | | 6.86 | 1.89 | |
| | ANN | 1.42 | 2.56 | 17.7 | 5.42 | 3.15 | 34 | 6.75 | 2.20 | 13.50 |
| 2 | LCM | 1.63 | 0.46 | | 5.37 | 1.89 | | 5.16 | 1.5 | |
| | ANN | 1.58 | 0.45 | 8.90 | 5.15 | 2.10 | 12.90 | 5.23 | 1.29 | 12.5 |
| 3 | LCM | 1.48 | 0.61 | | 7.45 | 2.43 | | 6.36 | 2.07 | |
| | ANN | 1.51 | 0.46 | 19.1 | 7.63 | 2.32 | 15.9 | 6.13 | 1.67 | 16.2 |

4.3.1 Patient Diagnosed with a Brain Tumor

An MRI scan showed pathological tissue in the brain of the patient, possibly a brain tumor. High concentration of choline and low NAA points out low content of healthy neurons, rapid cell membrane decay along with high amount of creatine showing intense metabolic processes. The pathological tissue is probably a progressive brain tumor. Both LCM and ANN models show the location and point out the nature of the pathology (Fig. 4.1).

4.3.2 7-Month Old Patient with a Suspicion for a Metabolic Disease

Analysis of 7-month old patient with a suspicion for a metabolic disorder is reported in Fig. 4.2. In this examination it is vital to detect possible asymmetry in metabolite distribution in the two hemispheres. Both LCM and ANN models showed good symmetry and almost exactly the same space distribution. The diagnose would not differ no matter which model is used. However, ANN approach takes 5 minutes compared to several hours in case of LCM.

4.3.3 Cerebellum 2D CSI of a Healthy Adult Volunteer

The data of adult healthy volunteer was acquired from the cerebellum as can be seen in Fig. 4.3. The surrounding brain cavities and bone tissue causes magnetic field inhomogeneities and therefore the signal measured from this region is characteristic for its low

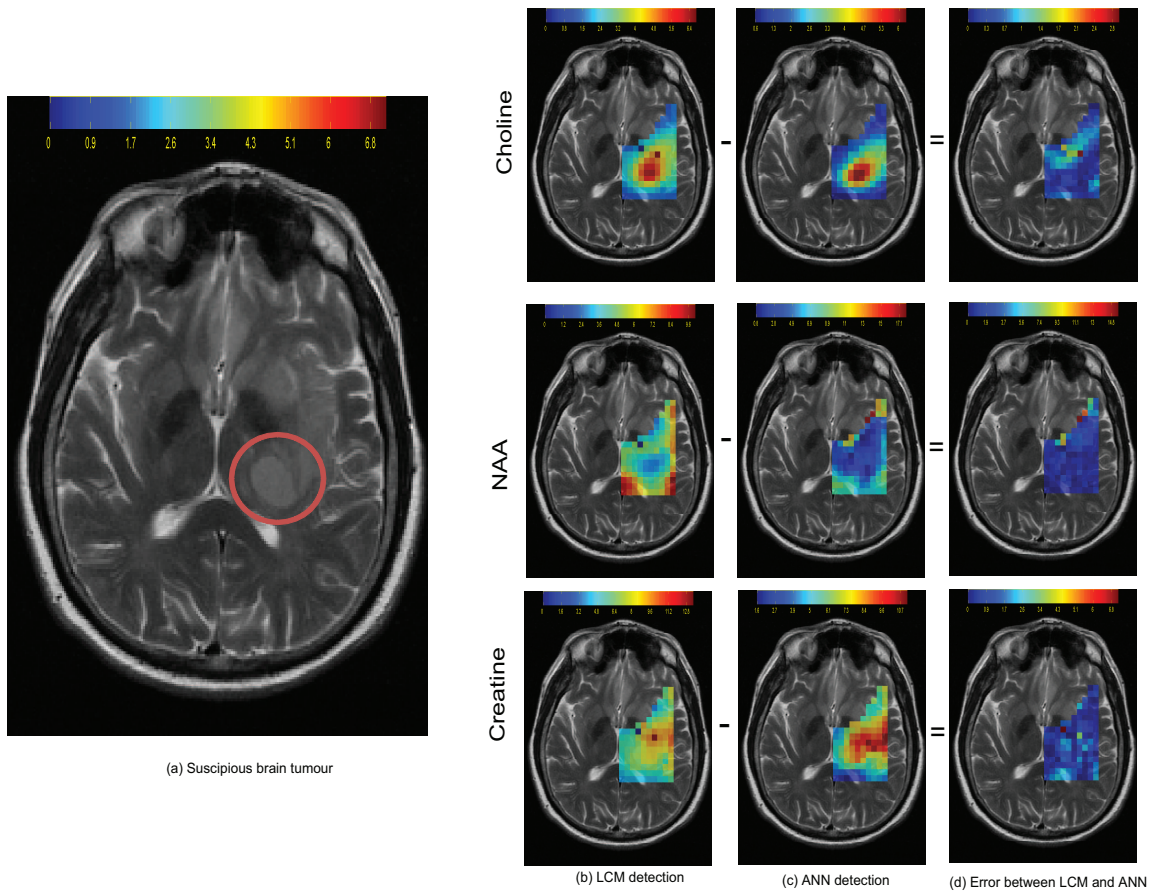


Fig. 4.1. Comparison of choline, creatine, and NAA metabolite map calculated by LCM. (a) MR image of suspicious brain tumor. (b) and ANN model (c) in case of patient with brain tumor. Tumor can be seen in the left image. The tumor was detected very precisely in both cases.

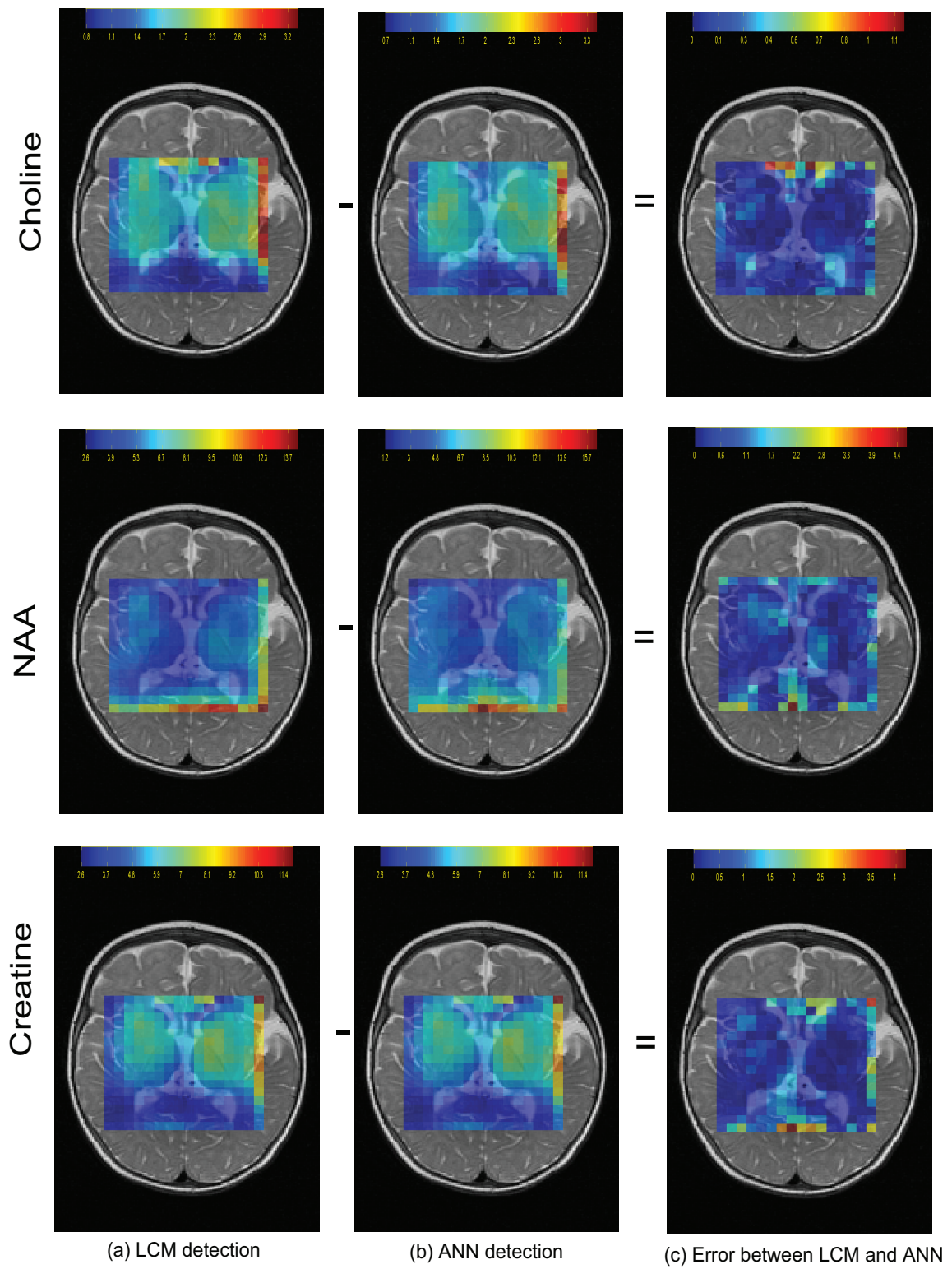


Fig. 4.2. Comparison of choline, creatine and NAA metabolite map calculated by LCM (a) and ANN model (b) in case of a 7-month old patient with a suspicion for a metabolic disorder.

quality. The results manifest the ability of the ANN model to follow the pattern of the curve fitting algorithm even in case of these low quality inputs. Although the error is significantly higher than in case of signals measured from the upper parts of the brain.

4.4 Discussion

In total over 1000 FID spectra were registered. The spatial distribution represented by the metabolite map according to LCModel curve fitting algorithm was visually very close to the approximation calculated by the neural network. For each subject the mean relative error and sum of square error was calculated. Also mean and variance of the targets and the calculated values were compared. The method did not show to be biased and the concentrations estimated by the neural networks showed to have a similar variance.

The performance of the developed application was tested on a PC system powered by an Intel Pentium 4 CPU with 512MB RAM while the LCModel application is run on a local server in IKEM. Therefore, the computational complexity is hard to compare. However, the ANN framework has reduced the computation to a couple of matrix multiplications and most of the computational time is spent on the signal preprocessing.

The LCModel license allows installation of the software to one hard drive exclusively and was bought by IKEM for a considerable amount of money. The developed ANN framework can be run on any PC which is able to run MATLAB 2006 or higher.

The LCModel performance can be rapidly increased by the means of advanced hardware server upgrade which can be expected based on the pace of development of information technology. This might seem to limit the usefulness of the ANN approximation, which might turn obsolete, an already fast algorithm would not need further speed improvements at the cost of additional error introduction. However, the LCModel algorithm is not the most advanced system used for spectral analysis. More advanced systems such as non-linear curve fitting algorithms are already being tested. Such algorithms are currently not used precisely due to their computational demands. With the advance of powerful information technology hardware development of neural network approximation of these much more complex algorithms might be considered. Moreover, with the advancement of computational hardware the new artificial neural networks might be trained on much wider set sets of training data and therefore, better generalization ability with lower output error can be expected.

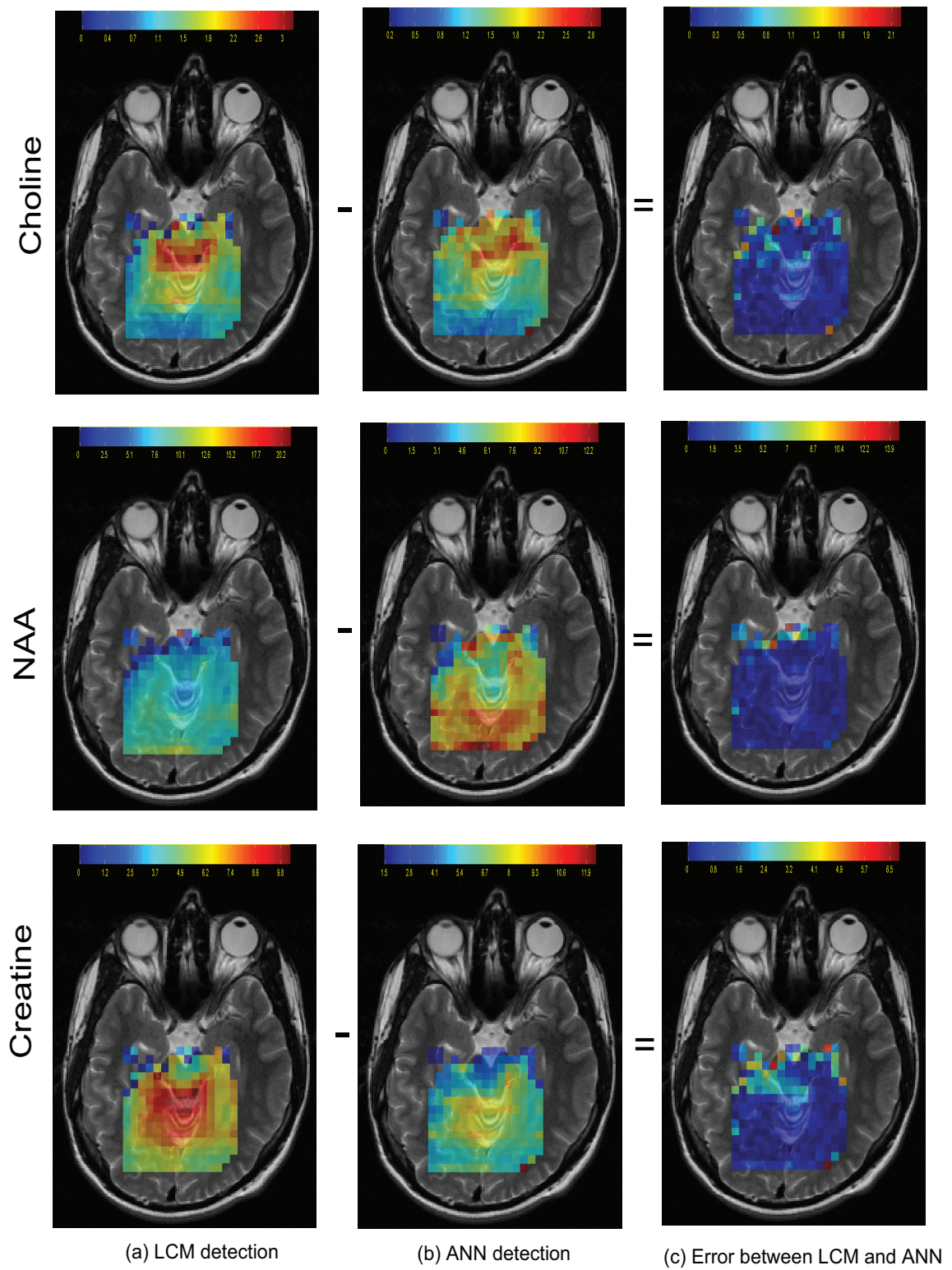


Fig. 4.3. Comparison of choline, creatine and NAA metabolite map calculated by LCM (a) and ANN model (b) in case of adult healthy volunteer.

Chapter 5

Conclusions

A preprocessing procedure was developed and an artificial neural network was trained successfully. The ability of the designed application to analyze in-vivo spectroscopic data from human brain was demonstrated.

The approximation using neural networks can give results with the mean error of under 15% within 5 minutes. Since the whole measurement is affected by a significant error due to high amount of noise and bad homogeneity of the external magnetic field the error of 15% is acceptable.

The method is limited by the error of the LCModel algorithm. Since the LCModel results are the only data used for training it is certain that the systematic error of the LCModel algorithm is included in the output of the neural network framework. To counter this effect including results acquired from multiple methods in the training dataset might be considered. However, as of now the LCModel has a dominant position in the field of in-vivo NMR spectroscopic data analysis and finding of a comparable system might be difficult.

The designed framework is a black box algorithm and has limited generalization ability. It is impossible to ensure proper function for inputs significantly different from the training set. For best results the training set needs to be as large as possible with maximum possible variety. The training dataset might be broadened by including a set of artificially generated spectra along with estimated results [2]. That way the variety of the dataset could be increased to include samples of spectra specific for various known pathologies that were not yet measured on the current scanner and therefore could not have been yet included in the dataset.

The optimal topology of the network can not be derived. Due to the size of the inputs the training process for a large training set is time and memory demanding. Testing

the performance for more possible topologies as well as reducing the length of the input vector could improve the convergence of the training algorithm and that way enable larger training sets to be used.

It was shown that an artificial neural network is able to cope with the metabolite concentration assessment problem. However, other methods such as the support vector machine might be considered as well.

The preprocessing does not count for possible phase shift. Some spectra are improperly classified because of being out of phase. Phase correction as a part of the preprocessing might be considered.

Even though the processing through the neural network is almost immediate the preprocessing of the signals takes up to a couple of seconds for each signal. Commercial software LCModel takes in order of minutes to process a signal from a single voxel. The processing of a full CSI signal takes several hours to run.

Bibliography

- [1] M. Ala-Korpela *et al.*, “Quantification of metabolites from single-voxel in vivo ^1H NMR data of normal human brain by means of time-domain data analysis,” *Magma*, vol. 3, 1995.
- [2] H. Bhat *et al.*, “Fast quantification of proton magnetic resonance spectroscopic imaging with artificial neural networks,” *Journal of Magnetic Resonance*, vol. 183, no. 1, pp. 110–122, 2006.
- [3] J. Kaartinen *et al.*, “Automated quantification of human brain metabolites by artificial neural network analysis from in vivo single-voxel ^1h nmr spectra,” *Journal of Magnetic Resonance*, vol. 134, pp. 176–179, 1998.
- [4] S. Provencher, “Estimation of metabolite concentrations from localized in vivo proton NMR spectra,” *Magn. Reson. Med.*, vol. 30, 1993.
- [5] D. Wagnerová, “Metabolic profile of human brain in vivo in mr images and spectra,” Master’s thesis, Charles University in Prague, Faculty of Mathematics and Physics, 2007.
- [6] J. B. Pouillet *et al.*, “MRS signal quantitation: a review of time- and frequency-domain methods,” *Springer*, 2008.
- [7] S. W. Provencher, *LCModel & LCMgui User’s Manual*, 2007.
- [8] F. Jírů, “Methods of the spectroscopic imaging in clinical practice and experiments,” Ph.D. dissertation, Charles University in Prague, First Faculty of Medicine, 2006.
- [9] M. F. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

- [10] A. Naressia *et al.*, “Java-based graphical user interface for MRUI, a software package for quantitation of in vivo medical magnetic resonance spectroscopy signals,” *Computers in Biology and Medicine*, vol. 31, pp. 269–286, 2001.
- [11] E. Cabanes *et al.*, “Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain,” *Journal of Magnetic Resonance*, vol. 150, pp. 116–125, 2001.