Universität Karlsruhe (TH)
Research University · founded 1825

Carnegie Mellon

Cognitive Systems Lab
Prof. Dr.-Ing. Tanja Schultz
Fakultät für Informatik
Universität Karlsruhe (TH)

ACT-R Laboratory
Prof. John R. Anderson
Department of Psychology
Carnegie Mellon University

Predicting the BOLD Response: A Computational Model of Humans Solving an Arithmetic Task

Diploma Thesis

by

cand. inform.
Anne K. Porbadnigk

Advisors:
Prof. Dr.-Ing. Tanja Schultz (Universität Karlsruhe)
Prof. John R. Anderson (Carnegie Mellon University)

Referees:
Prof. Dr.-Ing. Tanja Schultz
Prof. Dr.-Ing. Rüdiger Dillmann

Date of Registration: 2009-03-02
Date of Submission: 2009-05-19

This is to certify that I developed and wrote the thesis at hand independently and that I indicated all references and resources that were used in the process.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, 2009-05-19

Abstract

In recent years, functional Magnetic Resonance Imaging (fMRI) has proven to be a particularly valuable instrument for cognitive neuroscience, due to its high spatial resolution and good signal-to-noise ratio. On the other hand, cognitive architectures have been developed as a way of specifying general theories of cognition. Then, these can be implemented as computational frameworks for investigating questions of artificial intelligence, for instance. One of its main exponents is ACT-R ('Adaptive Control of Thought - Rational'). Recently, ACT-R models have been used not only to make predictions about the behavior of humans, but also to predict their brain activity while solving given tasks, specifically their hemodynamic responses (blood oxygen level dependent responses; BOLD). This approach allows validating assumptions underlying ACT-R with neuroimaging data, as well as gaining insights into the cognitive functions of the brain.

This thesis is based on an fMRI study of 16 subjects that were scanned while solving an arithmetic task, for which the results had to be reported manually or verbally. The experiment mainly served as a feasibility study for using verbal protocols in an fMRI scanner. Verbal protocols were used only once before in ACT-R fMRI research, but on a different domain. The main contribution of this thesis is the development of a computational model of the subjects, implemented within the cognitive architecture ACT-R, that is capable of predicting the BOLD responses of the subjects. As a first step, a range of different solution strategies were tested as well as components for introducing variability in the model. Experiments showed that a probabilistic mixture model that utilizes two different strategies with a specific probability is capable of capturing the behavioral data of the subjects best. This probabilistic model was then used in a second step to generate predictions of the BOLD responses for eight pre-defined regions of interests (ROIs) in the brain. These were then validated on neuroimaging data. The results demonstrate that the model is capable of successfully predicting the BOLD responses for all of the ROIs, except for the caudate nucleus. For the four perceptual-motor brain regions, the shapes of the BOLD responses can be predicted accurately, but with an offset in timing. In contrast, the predictions for the four regions involved in cognitive control show partly different patterns, but align well with the actual imaging data. The results strongly indicate that the shift in timing is caused by anticipatory activity, which comes expected for the manual and vocal region. Surprisingly, this also includes the auditory ROI.

Zusammenfassung

In den letzten Jahren hat sich die funktionelle Magnetresonanztomographie (fMRT) auf Grund ihrer hohen räumlichen Auflösung und des guten Signal-Rausch-Verhältnisses als ein für kognitive Neurowissenschaften bemerkenswert nützliches Instrument erwiesen. Zugleich wurden kognitive Architekturen entwickelt, um allgemeine Kognitionstheorien zu spezifizieren. Diese können dann als System implementiert werden, um beispielsweise Fragestellungen der künstlichen Intelligenz zu untersuchen. Eine der Hauptvertreter kognitiver Architekturen ist ACT-R ('Adaptive Control of Thought - Rational'). In letzter Zeit wurden ACT-R Modelle nicht nur benutzt, um Vorhersagen über das Verhalten von Menschen zu treffen, sondern auch um ihre Gehirnaktivität beim Lösen von vorgegebenen Aufgaben vorherzusagen, genauer gesagt die hämodynamische Antwort ihrer Gehirne (blood oxygen level dependent response; BOLD). Dieses Verfahren ermöglicht es sowohl, die Annahmen, die ACT-R zugrunde liegen, mit Hilfe neurowissenschaftlicher Daten zu validieren, als auch neue Erkenntnise ber die kognitiven Fgkeiten des Gehirns zu gewinnen.

Diese Diplomarbeit basiert auf einer fMRT Studie, bei der 16 Probanden gescannt wurden, während sie Arithmetik-Aufgaben lösten, wobei sie je nach Aufgabenstellung manuell oder verbal ihre Antwort abgaben. Das Experiment diente vorwiegend als Machbarkeitsstudie für die Verwendung von verbalen Protokollen in fMRT Studien. Diese wurden bisher erst einmal in der ACT-R fMRT Forschung verwendet, wobei die gestellte Aufgabe jedoch andere kognitive Ressourcen in Anspruch nahm. Der hauptsächliche Beitrag dieser Arbeit besteht in der Entwicklung eines Modells der Probanden, die die Arithmetik-Aufgaben lösen, welches innerhalb der kognitiven Architektur ACT-R implementiert wurde und in der Lage ist, die BOLD-Antworten der Probanden vorherzusagen. Zuerst wurden eine Reihe verschiedener Lösungsstrategien getestet, ebenso wie verschiedene Komponenten, mit denen Variabilität in das Modell eingeführt wurde. Experimente zeigten, dass ein probabilistisches Mixtur-Modell, welches zwischen zwei verschiedenen Strategien mit einer festgelegten Wahrscheinlichkeit wählt, die Verhaltensdaten der Probanden am besten erfassen kann. Dieses probabilistische Modell wurde dann in einem zweiten Schritt verwendet, um Vorhersagen von BOLD-Antworten für acht in ihrer Lage vordefinierte Gehirnregionen zu generieren, die dann mit den neurowissenschaftlichen Bilddaten validiert wurden. Die Ergebnisse zeigen, dass das Modell in der Lage ist, die BOLD-Antworten erfolgreich für alle Regionen vorherzusagen, bis auf den Nucleus Caudatus. Für die vier Gehirnregionen, die Wahrnehmung und Motorik widerspiegeln, kann das Muster der BOLD-Antworten genau vorhergesagt werden, wenn auch mit einem zeitlichen Versatz. Im Gegensatz dazu zeigen die vier Regionen, welche kognitive Kontrollfunktionen haben, teilweise abweichende Muster, jedoch ähnliche Zeitverläufe wie die neurowissenschaftlichen Bilddaten. Die Resultate geben starken Grund zur Annahme, dass die zeitliche Verschiebung durch Erwartungspotentiale verursacht wird, was für jene Regionen zu erwarten war, die bei manueller und verbaler Aktivität angeregt werden. Erstaunlicherweise umfasst dies jedoch auch die auditorische Region.

Acknowledgements

First of all, I want to thank my advisor at Carnegie Mellon University, Prof. John R. Anderson, for his guidance, ideas and time. It was both a pleasure and a great honor to work with him. Moreover, I am very grateful for having been granted the chance of working in John's laboratory at Carnegie Mellon University as a visiting researcher. Thanks also to my advisor at the University of Karlsruhe, Prof. Tanja Schultz, for her patience and time.

The project was a combined effort of a group of people, each of whom contributed a substantial part. Jennifer Ferris and Yulin Qin organized the experiments, ran the subjects and transcribed the verbal protocols. The software both for recording the data and for analyzing the voice recordings was written by Shawn Betts. Jon Fincham handled the fMRI data as well as all the preprocessing steps. Special thanks to all of them for their dedication and excellent work as well as for their help and support. I also want to express my gratitude to Dan Bothell who was of invaluable help for any and all questions related to ACT-R and Lisp. Finally, I want to thank all the other members of both John Anderson's and Tom Mitchell's fMRI groups for the valuable discussions and their feedback. Many thanks to all who contributed by comments and proof-reading, in particular to Jennifer Ferris. Finally, I want to thank Sophia Heller for her constant help and encouragement as well as for being the most supportive friend one could imagine.

This thesis was financially supported by a scholarship of the Baden-Württemberg Program via the interACT exchange program between the University of Karlsruhe (Germany) and Carnegie Mellon University.

Contents

List of Figures

List of Tables

1. Introduction

The brain may be one of the most challenging, but potentially also most rewarding frontiers that scientists are facing today. Questions about certain cognitive functions of the brain and their neural representation can be approached from a wide array of different perspectives, using a variety of techniques.

On the one hand, functional neuroimaging methods have revolutionized our ability to investigate cognitive brain functions during the last two decades. Functional Magnetic Resonance Imaging (fMRI) has proven to be a particularly important research technique in cognitive neuroscience and a multitude of other disciplines, due to its high spatial resolution and good signal-to-noise ratio. On the other hand, a wide variety of computational models of cognition have been developed since Alan Newell's call for cognitive architectures as a way of specifying general theories of cognition [91]. Research on cognitive architectures is concerned with the creation and understanding of models that show capabilities similar to those of humans, which is a central goal in both artificial intelligence and cognitive science. One of the main proponents is ACT-R ('Adaptive Control of Thought - Rational') in which cognition emerges through the interaction of a number of components. These so-called modules have been associated with certain brain regions. Based on the standard models of the hemodynamic response, the activity patterns of the modules can be converted into predictions for the hemodynamic response (blood oxygen level dependent response; BOLD). This mapping function can be used to convert the behavior of a complex cognitive model into predictions for the BOLD responses of subjects as they solve arithmetic problems, for instance.

1.1 Motivation

What is the benefit of linking ACT-R to brain imaging data? Many of the assumptions on which ACT-R has been built have been justified by behavioral data. However, even though a given ACT-R model may produce similar behavioral data as is shown by subjects for a specific task, it is hard to justify the detailed assumptions made by the modeler about which modules are active at which points of time during the course of events. This assumptions-to-data ratio improves a lot by using brain

imaging data, which allows us to track changes of certain brain regions over time, that are associated with modules in ACT-R. Furthermore, this analysis can reveal potential shortcomings of the theory which can then be addressed.

On the other hand, brain imaging data provides a huge amount of information to be analyzed. Thus, it is important to be aware of which questions to ask during such an analysis. Even a thorough analysis might not necessarily help in explaining the time course of brain activity in certain regions. These can be explained to a certain degree based on the assumptions underlying the architecture of ACT-R. Also, the predictions of different models can be compared with the BOLD responses shown by the subjects, thus allowing inferences about the cognitive processes and subprocesses taking place during an experimental task.

1.2 Purpose of this Work

Overall Aim of the Project

The intention of this project is to achieve a better understanding of human cognition in a complex task. The step taken here is to develop the methodology of linking cognitive architectures to fMRI data. Furthermore, it is a feasibility study on whether it is possible and sensible to use verbal protocols in an fMRI scanner. While we are focused on mathematical problem solving and ACT-R, the methods should generalize to other aspects of cognition and other theoretical frameworks.

This thesis is based on an fMRI study of 16 subjects that were scanned while solving an arithmetic task. The task required adding two fractions (100 trials per subject) and reporting the results either verbally or manually. In order to track the timing of the mental steps the subjects had to take to solve the task, they were asked to give four responses, two intermediate results and the final fraction (numerator and denominator). The manual condition was added as a control condition to the verbal protocols. As a second variable, the fractions either contained only integers or letter to number mappings that were learned before, thus manipulating the level of difficulty. The experiment aimed at investigating the following hypotheses:

Hypothesis 1. It is possible to use verbal protocols in the given fMRI study.
Hypothesis 2. The output modality (verbal or manual) does not affect the BOLD responses of the predefined prefrontal, parietal, anterior cingulate or caudate ROI (region of interest).
Hypothesis 3. The BOLD responses for the verbal and the manual output modality are affected in a similar way by the level of difficulty (manipulated by including substitutions or not).
Hypothesis 4. The BOLD responses shown by the subjects during the arithmetic task can be predicted by an ACT-R model for eight predefined regions of interest.

Fractional arithmetic is a domain where procedure and conceptual factors come together in young children. The given task was chosen for this experiment as it seemed to contain just enough complexity to explore the use of verbal protocols, while being restricted enough for a detailed analysis.

Aim of the Thesis

The aim of this thesis was to build a model that is capable of predicting the BOLD response of subjects who are solving a given arithmetic task. Therefore, the focus of attention was on investigating Hypothesis 4. While doing so, we wanted to detect possible shortcomings of ACT-R as a model of human cognition that could then be addressed. Thus, the major work was conceptual and required model development, statistical analysis, and computational experiments. There is already an extensive strand of research on predicting brain imaging data with ACT-R. However, this is only the second time that verbal protocols are used and the first time on this domain. The given experiment allows to compare the behavior of the verbal and auditory module of ACT-R with activity in the associated brain regions. It is important to note that the work presented here is a confirmatory study that uses predefined regions, rather than taking an exploratory approach that would involve finding regions whose activity reflects a certain cognitive function.

1.3 Structure of the Thesis

The structure of the thesis is outlined in the following.

- Chapter 2 provides background knowledge relevant for understanding the research presented in this thesis. In the first section, functional neuroimaging techniques are explained and compared, with focus on fMRI. Second, cognitive architectures are introduced, in particular the computational framework of ACT-R on which the model is based that was developed for the arithmetic task. Besides the general theory of ACT-R, a detailed explanation is provided of how the activity patterns of the modules in ACT-R are converted into predictions of the BOLD responses. Finally, an overview is given of related work.
- Chapter 3 outlines the experimental setup as well as the arithmetic task that subjects were asked to solve under different conditions. Also, definitions are provided that are used throughout the rest of the thesis.
- In chapter 4, the computational model is described that was developed for the arithmetic task. Besides data structures and information processing, different answering strategies are discussed with which the model was provided.
- Chapter 5 provides details about the database, a brief interpretation of the behavioral data and a discussion of potential sources of variability.
- Chapter 6 is one of the main chapters of this thesis, as it explains the steps taken in order to refine and evaluate the model based on the subject data. This includes the adjustment of timing and variability of the model based on the behavioral data as well as the methodology for the evaluation on the fMRI data.
- The results of this evaluation on the fMRI data are analyzed in chapter 7. Furthermore, the remaining hypotheses are discussed.
- Finally, a summary is provided in chapter 8 along with ideas for future work.

2. Background

The following chapter provides background information relevant to the study and the model on which this thesis is based. First, section 2.1 introduces neuroimaging techniques with particular focus on fMRI, the technique used for this thesis. Section 2.2 gives a brief introduction to cognitive architectures. Subsequently, a description of the cognitive architecture ACT-R is provided (section 2.3), followed by an explanation of the methodology used to calculate a prediction of fMRI data with ACT-R (section 2.4). Finally, related work is described in section 2.5.

2.1 Neuroimaging Techniques

A wide range of imaging techniques exist for studying the structure and the function of the brain. The corresponding techniques are referred to as structural and functional neuroimaging, respectively. These two types of neuroimaging have been combined to varying degrees of integration, in order to study questions such as brain connectivity (for an overview of current research trends along these lines, refer to [110]).

In the following, the principles underlying a number of functional neuroimaging techniques are discussed briefly, with focus on BOLD fMRI, the neuroimaging technique employed in this study. First, an explanation of the physical principles and technical aspects of magnetic resonance imaging (MRI) is provided (section 2.1.1). Even though standard MRI actually provides structural information about the brain, it is the basis of functional MRI (fMRI). fMRI obtains information about the function of the brain and is described in the following section (refer to section 2.1.2). Subsequently, an overview of other functional neuroimaging techniques is provided (section 2.1.3), as well as an assessment of fMRI for the task at hand (section 2.1.4).

There is a vast field of literature on functional neuroimaging. [48] provides an overview over neuroimaging methods and their use in neuropsychology, while a more technical overview can be found in [50] and [88], among many others. Particularly good and detailed explanations of fMRI and its underlying principles can be found in [67].

2.1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that uses strong magnetic fields and electromagnetic high frequency pulses perpendicular to this field to obtain images of biological tissues. To be more precise, the response of protons to these external factors is monitored, as they are the most common type of nuclei in the human body due to their prevalence in water molecules. Thus, the signal obtained from a voxel during MRI is the sum of the water signals from the different materials in the voxel, e.g. tissue, blood, or bulk cerebrospinal fluid (CSF), leading to different contrasts in the resulting image.

The following sections provide an explanation of the physical principles on which MRI is based, as well as some technical aspects (mostly based on [67]). A detailed description of the components of MRI scanners and image reconstruction techniques is beyond the scope of this thesis and can be found in [50], for instance.

Spins

Elementary particles, i.e. particles that are basic building blocks which cannot be divided any further, have the quantum-mechanical property of possessing a spin. Each type of particle has a characteristic spin quantum number s with

$$s = \frac{n}{2} \tag{2.1}$$

with $n \in N$.

Based on this spin quantum number, the spin angular momentum S can be determined according to the following formula:

$$S = \hbar\sqrt{s(s+1)} \tag{2.2}$$

with S : spin angular momentum
$\hbar$: reduced Planck's constant
s : spin quantum number

If the particles are charged, a non-zero magnetic dipole moment μ is linked to their spin, according to the following formula. For these particles, the g-factor assumes a value $g \neq 1$ ($g = 1$ for exclusively orbital rotations).

$$\mu = g\frac{q}{2m}S \tag{2.3}$$

with μ: magnetic dipole moment
g : g-factor
q : charge of the particle
m : mass
S : spin

Composite particles, such as atomic nuclei or protons are also ascribed a spin, which is the sum of the spins of the elementary particles that form the composite. Thus, composite particles can also have a non-zero magnetic dipole moment μ.

This is the case for protons (hydrogen nuclei, ^1H), which possess a spin of $s = 1/2$. Hydrogen nuclei are the most common nuclei in the human body, as each water

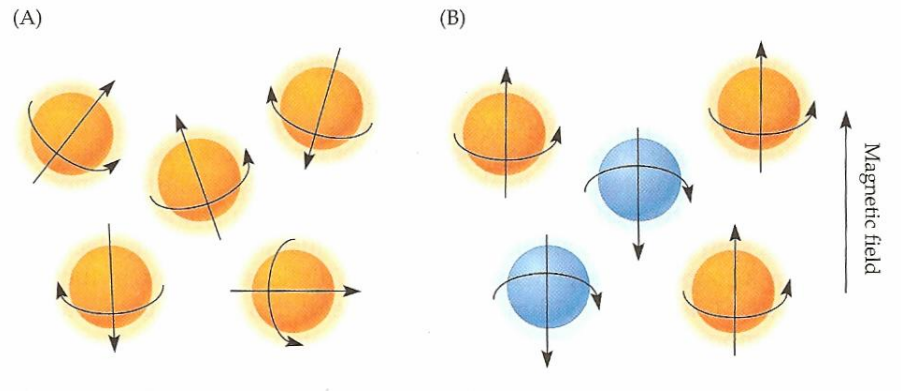

Figure 2.1: Effect of a magnetic field on the orientation of protons (taken from [67]). Without external influences, protons assume random orientations (A). If a magnetic field is applied, they align parallel or anti-parallel to it (orange and blue states in (B)).

molecule contains two of them. As mentioned before, MRI therefore monitors the responses of protons to certain external influences. The following sections explain the effects that a strong external magnetic field and electromagnetic high frequency pulses have on protons.

Nuclear Spin Magnetization

Without external influences, protons assume random orientations in space (see Figure 2.1 (A)), such that no overall net magnetization exists. However, if an external magnetic field B_0 is applied, protons align either parallel or antiparallel to it, as can be seen in Figure 2.1 (B) (orange and blue states, respectively). The state of protons that are aligned in parallel to the field is also referred to as low-energy state, while the anti-parallel state is also called high-energy state. Protons are slightly more likely to assume a low-energy than a high-energy state, resulting in an overall longitudinal magnetization parallel to the external magnetic field. The direction of the net polarization is given by the so-called nuclear spin magnetization vector M.

Besides the longitudinal magnetization, the application of this external magnetic field has a second effect: the protons also start precessing around an axis parallel to the external magnetic field with their specific Larmor frequency, also called eigen frequency. These physical effects can be observed for any nucleus containing an odd number of protons or neutrons (i.e. any nucleus possessing a non-zero spin) that is exposed to an external magnetic field; the respective Larmor frequencies depend on the specific type of nuclei.

Nuclear Magnetic Resonance and Relaxation

The two potential states that protons assume upon the application of an external magnetic field are separated by a specific splitting energy, corresponding to a radio frequency photon. If a radio frequency pulse (RF pulse, excitation pulse) is now applied which has the same frequency with which the nuclei of interest are precessing, i.e. their Larmor frequency, some of the spins in the low-energy state absorb this

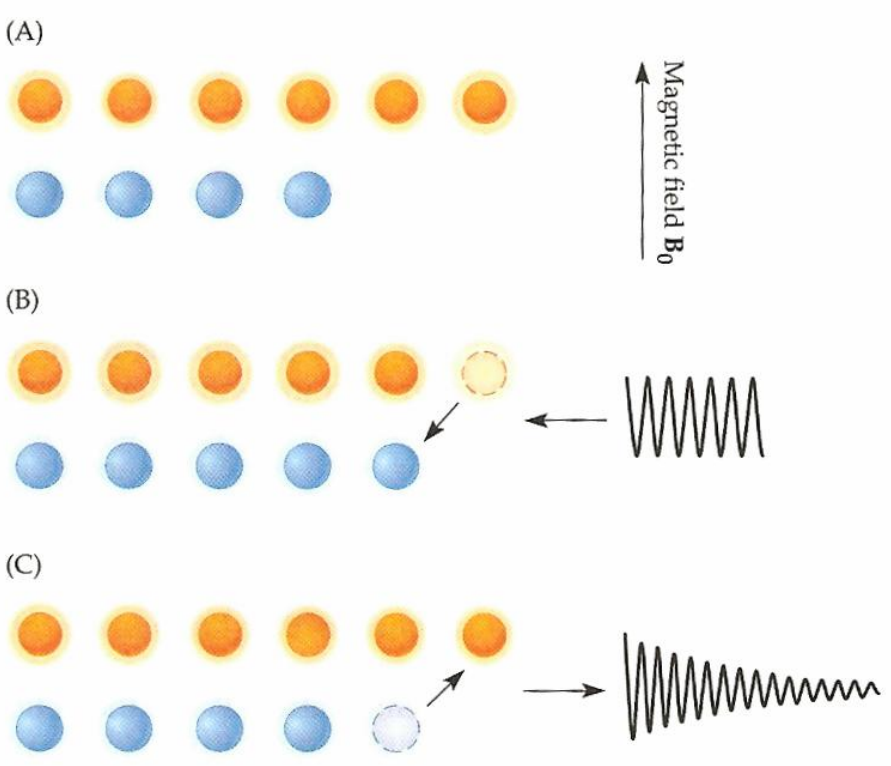

Figure 2.2: Changes of energy states of spins due to energy absorption or transmission (taken from [67]). Spins assume a high-energy state (blue) or, more frequently, a low-energy state (orange), when an external magnetic field is applied (A). The application of an excitation pulse with a certain frequency causes some spins to switch from their low-energy state to a high-energy state, as they absorb energy (B). The opposite effect can be observed when the pulse is turned off as some spins release energy and return to a low-energy state (C).

energy. As a result, they switch from their low-energy state to the high-energy state (see Figure 2.2 (B)). When the pulse is turned off, the opposite effect can be observed, as some spins in the high-energy state release energy and return to a low-energy state (Figure 2.2 (C)). This effect is known as Nuclear Magnetic Resonance (NMR) and is one of the main principles behind MRI.

The excitation pulse is usually applied in a right angle to the direction of the magnetic field in order to obtain the strongest signal possible. As a result of this, the spin magnetization vector that had so far been parallel to the external magnetic field (due to the small excess of protons in the low-energy state), will flip to the side by 90 degrees, leading to a transverse magnetization component.

When the RF pulse is switched off, protons return to their equilibrium state, an effect that is referred to as relaxation. It is crucial for the technique of MRI that the relaxation of protons occurs at different speeds, depending on which tissue they are part of. It is the different rates of spin relaxation that are being measured in MRI. Depending on which contrast the experimenter is interested in, different parameter settings are used for scanning, thus manipulating the contrast between different kinds of tissue in the resulting MR image.

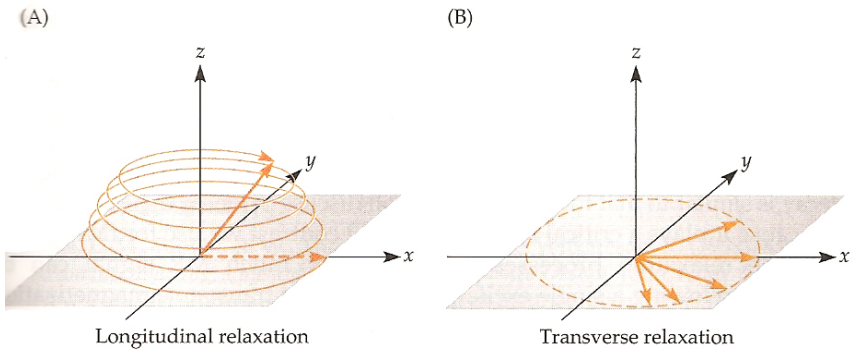

Figure 2.3: Longitudinal and transverse relaxation of spins when the RF pulse is switched off that was applied perpendicular to B_0 (taken from [67]). The z-axis points in the direction of the external magnetic field B_0.

T_1 and T_2 Relaxation Processes

Relaxation also refers to the recovery of the spin magnetization vector M back to its original orientation, after it was flipped to the side by 90 degrees (if the RF pulse was applied perpendicular to the magnetic field B_0). The two main relaxation processes are called T_1 and T_2 relaxation, referring to the relaxation of the longitudinal and the transverse components of M. Accordingly, these relaxation processes are also called longitudinal and transverse relaxation, which will be used here as synonyms for T_1 and T_2 relaxation. In the literature, these processes are also referred to as spin-lattice relaxation (T_1 relaxation) and spin-spin relaxation (T_2 relaxation).

These two types of relaxation can be seen in Figure 2.3 (A) and (B) for the case that the RF pulse is applied perpendicular to the external magnetic field B_0. In this Figure, the z-axis denotes the direction of B_0 and thus also the direction of the spin magnetization vector M before the application of the RF pulse. As a result of the excitation pulse, the spin magnetization vector M is tilted into the xy-plane.

The longitudinal relaxation time T_1 characterizes the recovery of the z-component of the spin magnetization vector M (denoted with M_z) back to its equilibrium state. More precisely, T_1 is the decay constant in the following equation, describing the discovery from $M_z(0) = 0$ to the equilibrium value $M_{z,equ}$:

$$M_z(t) = M_{z,equ}(1 - e^{-\frac{t}{T_1}}) \tag{2.4}$$

While T_1 is the decay constant of the z-component of the spin magnetization vector M, T_2 is the corresponding decay constant for the xy-component of M. This xy-component, denoted by M_{xy}, is perpendicular to the external magnetic field and decays from its initial value $M_{xy}(0) \neq 0$ to zero, according to the following equation:

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T_2}} \tag{2.5}$$

So while the longitudinal magnetization component M_z drops to zero after the application of the RF pulse ($M_z(0) = 0$) and then recovers to its non-zero value, the opposite is the case for the transverse component that has a non-zero value $M_{xy}(0)$ when the pulse is switched off and then recovers back to zero. This transverse magnetization component $M_{xy}(0) \neq 0$ results in an oscillating magnetic field that can

be recorded. In theory, it decays exponentially with decay factor T_2 according to equation (2.5) and is referred to as free induction decay (FID). This decay is due to the loss of phase coherence in the xy-plane. Initially, that is right after the application of the RF pulse, all the spins are in phase such that there is a non-zero net magnetization in the xy-plane. However, this coherence and thus the signal are lost after a short amount of time.

As mentioned before, equation (2.5) only gives the theoretical decay in an idealized environment in which all protons precess with the same frequency. In reality, inhomogeneities in the magnetic field B_0 lead to interferences and thus an actual relaxation time that is shorter than T_2. The decay constant for this process is referred to as T_2^* and stands in the following relationship to T_2:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B_0 \tag{2.6}$$

with γ gyromagnetic ratio
ΔB_0 local differences in the strength of the magnetic field B_0

The values of the three decay constants T_1, T_2 and T_2^* depend to a large extent on the type of tissue. However, the strength or irregularities of the magnetic field B_0 generally affect T_2 to a much smaller degree than the other constants. According to [82], the following relationship holds between the three decay constants:

$$2T_1 \geq T_2 \geq T_2^* \tag{2.7}$$

MRI Scanners

MRI scanners are composed of the following parts, out of which the magnet is the most expensive and most space-consuming component (for a picture of an MRI scanner, refer to Figure 3.1):

- a strong magnet for providing the external magnetic field B_0 (clinical magnets range from 0.1 to 3 Tesla),
- shim coils for correcting inhomogeneities in B_0,
- a gradient system for spatial encoding of the MR signal,
- a radio frequency system, both for providing the excitation pulse and recording the MR signal.

The need for a strong magnet as well as for a radio frequency system should be obvious after the explanations of the last sections. Homogeneity of the magnetic field B_0 depends on the precision of the magnet and is crucial for acquiring good images. However, the field is distorted when a human is placed in the scanner, in particular at boundaries between air and tissue. In order to correct the resulting inhomogeneities of the magnetic field, shim coils are needed.

In order to spatially encode the imaging volume, three orthogonal magnetic gradient coils are employed (in x-, y- and z-direction). As a result, the magnetic field varies linearly across the volume as does the Larmor frequency that can now be described as a function of the position in space (x-, y-, z-axis). As a consequence, stronger gradients result in a better resolution.

Image Acquisition and Contrasts

In most cases, the 2D Fourier Transform (2DFT) with slice selection or the 3D Fourier Transform Technique (3DFT) are employed for constructing MR images. Brain imaging requires a very rapid acquisition of images, which is why a different technique is used here, called echo-planar imaging (EPI). According to this scheme, a number of gradient echos with different spatial encodings is employed after each excitation pulse.
For an overview of preprocessing methods applied to the raw data prior to the actual image analysis, refer to [54], for instance.

One of the distinctive features of MRI that sets it apart from other imaging techniques is its remarkable contrast resolution. Essentially, the signal and thus the contrast depends on the strength of the MR signal that is recorded by the RF system at specific voxels within the subject in the scanner. For most MR techniques, the contrast is influenced by the following effects:

- relative differences in density of the nuclei being monitored,
- differences in T_1, T_2 and T_2^* relaxation times,
- parameters of the pulse sequence, such as echo and repetition time.

It is important to note that T_1, T_2 and T_2^* relaxation times can in turn be affected by other aspects, such as the blood oxygenation on which BOLD fMRI is based (see section 2.1.2).

Depending on which contrasts the experimenter is interested in, pulse sequences with certain characteristics are employed. Important scanning parameters for adjusting for the type of contrast are the echo time (TE) and the repetition time (TR). The pulse sequence is repeated after the repetition time TR which usually ranges between 1000 ms and 3000 ms, while TE is generally between 5 ms and 100 ms. An overview of the four basic types of MR images and their dependency on TE and TR is provided in Table 2.1.

Table 2.1: Basic types of MR images. Pulse sequences are abbreviated as GRE (gradient echo sequence) and SE (spin echo sequence); TE stands for echo time, TR for repetition time.

MRI Type	TE	TR	Sensibility for
T_1-weighted	short	short	gray vs white matter
T_2-weighted	long	long	water, fluids
T_2^*-weighted	long	long	venous blood
spin density weighted	short	long	amount of available spins

2.1.2 BOLD fMRI

Since the first BOLD fMRI studies were published in 1992 (refer to [23, 77, 94]), this technique has become a very popular and valuable tool in cognitive neuroscience and many other disciplines to study the function of the brain. Standard MRI scans of the brain deliver information about the structure of the brain. Functional magnetic resonance imaging (fMRI) is a variation of MRI which instead focuses on obtaining information about the function of the brain. For this reason, it measures the hemodynamic response of the brain that is assumed to be related to brain activity. Nearly all fMRI studies, as well as the study analyzed in this thesis, are based on a measure called blood oxygenation level dependent contrast (BOLD contrast) that is explained in the following.

Neuronal Activity and BOLD Effect

The central hypothesis behind fMRI is that regional neuronal activation results in a regional increase in oxy- to deoxyhemoglobin. In other words, it is assumed that brain activity and the oxygenation of blood are correlated, i.e. that active brain regions need more oxygen than non-active regions. The BOLD contrast is then used to track the changes in the balance of oxyhemoglobin to deoxyhemoglobin. Even though this hypothesis is widely assumed to be true, it has been argued that this assumption might be oversimplified (for instance, see [102]).

It is important to note that BOLD fMRI does not measure neuronal activity itself, but the metabolic demands of neurons that are active. Since the brain does not have extensive energy stores, it depends on being provided continuously with energy via the vascular system, i.e. blood supply. The primary sources of energy are glucose and oxygen, which are used to create ATP (adenosine triphosphate) in brain cells. Oxygen is attached to hemoglobin molecules and then exchanged for carbon dioxide once it is used in the brain. Oxygenated hemoglobin (Hb) and deoxygenated hemoglobin (dHB) exhibit a different behavior in magnetic fields, since the first is diamagnetic, whereas the latter is paramagnetic.

Completely deoxygenated blood is 20% more susceptible to magnetic fields than fully oxygenated blood. The paramagnetic nature of deoxyhemoglobin leads to local field distortions in and around blood vessels which affects the spins of the hydrogen nuclei. Brain activation is assumed to result in a change of the ratio of oxygenated to deoxygenated hemoglobin and thus to induce a change in the MRI signal. This has become known as the Blood Oxygenation Level Dependent (BOLD) effect [93]. Ogawa et al. showed that changes in blood oxygenation can be visualized using T_2^*-weighted MR images. These images show a slightly stronger MR signal for oxygenated blood than for deoxygenated blood which is referred to as the BOLD contrast. In order to show the T_2^*-based contrast in blood, strong magnetic fields are necessary ($\geq 1.5\ T$). Only for these, a significant difference can be found between the transverse relaxation values for oxygenated and deoxygenated blood, on which the signal is based.

Shape of the BOLD Response

It has been shown that a sudden change in neural activity results in a change of the signal which develops and decays over several seconds [22]. The BOLD response usually follows a typical time course, as described in the following: If neuronal activity increases above baseline level, the inflow of oxygenated blood to this region increases due to the higher metabolic demand. About 2 s after the onset of neuronal activity, for example due to the presentation of a certain external stimulus, the BOLD signal increases above baseline, peaking about 5 to 6 s after the onset. This peak can turn into a plateau, if the neuronal activity continues. Typically, this is followed by a decline until second 10 after the last stimulus onset with an ensuing prolonged undershoot, i.e. a drop of the signal below baseline level. Sometimes, a small initial dip before the rise can be observed in the signal as well, which is slightly smaller than the undershoot.

This so-called Hemodynamic Response Function (HRF) is usually modeled with a standard gamma function, which has been widely used for describing the hemodynamic response (refer to [30, 41, 42, 59], for instance). A whole strand of research is dedicated to estimating the HRF (for details, refer to section 2.5.2). It should be noted that the signal can vary substantially both within and across subjects. As it was shown in an event-related reaction time task, for instance, the shape of the BOLD response of the central sulcus varies significantly across subjects and to a lesser degree within subjects [1]. Another study that included different activation tasks (visual, motor and simple cognitive tasks) also showed that session context has significant effects on the BOLD signals obtained [85].

fMRI Studies

Two main different approaches need to be distinguished in BOLD fMRI studies, exploratory and confirmatory studies. For both, fMRI data is acquired under certain experimental conditions. Exploratory studies then make use of statistical analysis in order to define in which regions of interest activity was triggered. Most commonly, Statistical Parametric Mapping is used for this approach (for an introduction, refer to [54], for instance). In contrast, confirmatory studies investigate pre-defined regions and aim at confirming their assumed role and activity in a task. The latter approach has been taken for the study on which this thesis is built on.

For conducting an fMRI study, two properties of the scanner used are of particular importance. First, since a transient phenomenon is to be captured, image acquisition needs to be quick. Typically, echo planar imaging (EPI) is employed to acquire rapid images. Second, as mentioned before, particularly strong magnetic fields are necessary (i.e. $\geq 1.5\ T$) in order to ensure that a good contrast is obtained for the T_2^*-weighted images.

2.1.3 Comparison of Functional Neuroimaging Techniques

Nowadays, a wide range of functional neuroimaging methods is being employed to study the functionality of the brain, such as the following techniques:

- magnetoencephalography (MEG)
- (extracranial) electroencephalography (EEG)
- positron emission tomography (PET)
- functional magnetic resonance imaging (fMRI)
- functional near infrared spectroscopic imaging (fNIRS)

With the exception of PET, these techniques are all non-invasive. Despite its invasiveness, PET is included here for historical reasons, as it can be seen as the predecessor of fMRI. A number of different aspects can be considered when comparing different imaging techniques, despite their invasiveness. Two characteristics of particular importance are the spatial and temporal resolution of a technique. Spatial resolution determines to which degree the separation of adjacent brain regions is feasible. Temporal resolution affects to which degree the timing of neuronal activity can be estimated from the measured data. In this context, it is important to note that neuronal events have time constants on the order of milliseconds.

The neuroimaging techniques listed above can be divided into two groups: On the one hand, PET, fMRI and fNIRS are based on changes in blood flow to the brain, that is on hemodynamic data (group A). MEG and EEG on the other hand record the electrical and magnetic effects of brain activity, i.e. neuronal data (group B).

The first group of techniques (A) is based on the assumption that an increase in cerebral blood flow to a certain brain region is correlated with neuronal activity in this region, as active regions are assumed to need more oxygen than non-active regions. This assumption is central for BOLD fMRI; a more detailed explanation can be found in section 2.1.2. Since the measurement is based on blood flow, a slow physical property, the temporal resolution of these techniques is worse than that of group (B), ranging within the order of seconds instead of an order of several hundred milliseconds. On the other hand, the spatial resolution of group (A) is superior to that of group (B); with the latter, it proves hard to localize the source of neural activation. It should also be pointed out that the methods in group (A) measure brain activity indirectly (hemodynamic data), whereas the techniques of group (B) provide direct measurements of brain activation (neuronal data). Even though these two types of data are tightly coupled, there may be discrepancies between neuronal and hemodynamic data. Thus, the techniques in group (A) and (B) can be used to obtain complementary data. Further individual advantages and disadvantages of the different methods are discussed in the following sections.

An overview of these neuroscience techniques with regard to the range of their temporal and spatial resolution can be found in Figure 2.4. The figure also contains invasive methods which are not considered further here, as they are not suitable for the strand of research discussed here. In the following, the list of non-invasive techniques mentioned above is discussed in more detail.

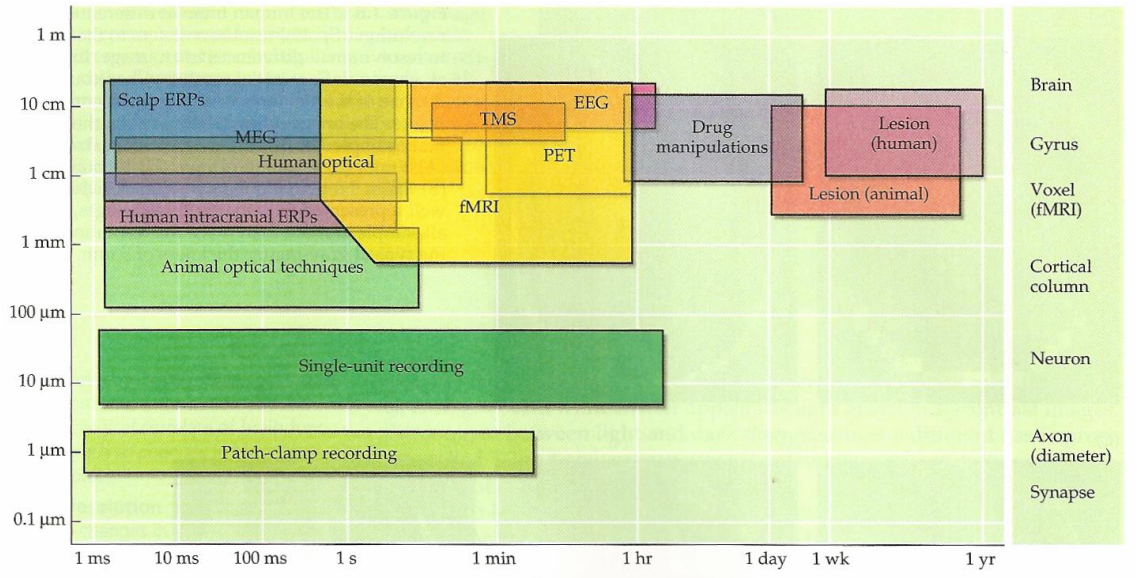

Figure 2.4: Differences between neuroscience techniques with regard to temporal and spatial resolution (picture taken from [67]). The fields named 'scalp ERPs' and 'humanoptical' correspond to the methods '(extracranial) EEG' and 'fNIRS', respectively.

Extracranial Electroencephalography (EEG)

Electroencephalography (EEG) measures electrical activity of the brain. To be more precise, the synchronous activity of cortical neurons is recorded. This signal is assumed to be generated by large dipole layers of these synchronized cortical neurons which extend over large areas of the cortical surface. For extracranial EEG, electrodes are applied on the skull, as opposed to measuring inside of the brain.

Just as MEG, EEG features a high temporal resolution on the order of milliseconds which is one of the main advantages of these two techniques. Another advantage of EEG, in particular in contrast to fMRI, is the comparatively low price for the equipment as well as its mobility. Out of the other imaging techniques, only fNIRS has a comparably portable and affordable hardware. The main drawback of EEG is the low spatial resolution (on the order of centimeters) which renders the localization of the sources difficult or even impossible. Furthermore, even though the EEG hardware has some advantages, it requires the application of conductive paste to the skin of the subject (unless dry electrodes are used), which requires some time. Also, an EEG cap exerts pressure on the subjects' head and can therefore cause headaches. It should also be noted that the signal depends on electrical and geometrical properties of the brain and skull of a subject. One of the most critical aspects might be the high level of electrical background noise that is superimposed on the signal of interest and can cause a significant amount of artifacts, for instance caused by eye blinks.

Magnetoencephalography (MEG)

Magnetoencephalography (MEG) measures magnetic fields induced by synchronized neuronal activity in the cortex. As the signal is weak while the ambient noise is high,

this technique requires recording devices with an extremely high sensitivity, so-called superconducting quantum interference devices (SQUIDs).

As mentioned before, MEG provides an extremely high temporal resolution with a precision of less than 1 ms, similar to EEG. MEG has further advantages that are not shared by EEG and are outlined in the following. First, no direct skin contact is required which facilitates the measurement. Second, the signal obtained with MEG is less dependent on the geometry of the subject's head than EEG. Third, the spatial resolution of MEG is much better than that of EEG, even though both methods do not provide a high precision in comparison to the methods of group (A). Also, EEG and MEG share the disadvantage that they can only measure cortical signals and not those deep in the brain. Thus, overall brain imaging is beyond the reach of both MEG and EEG. In contrast to EEG which can be made portable up to a certain degree, this is not feasible for MEG due to the hardware used: MEG is based on the usage of SQUIDs which are space-consuming and in need of constant cooling. Another disadvantage is the high cost of SQUIDs.

Positron Emission Tomography (PET)

Positron Emission Tomography (PET, [101]) is included here despite its invasive nature, as it used to be the main functional neuroimaging technique before the advent of fMRI. PET and fMRI are both capable of delivering 3D images of functional processes, in contrast to the other techniques mentioned here. Radioactive tracers with short half-lives (on the order of about 120 min) are injected in order to measure changes in the brain, including blood flow and glucose metabolism. As the radioactivity of the tracer decays, information can be obtained about the concentration of the tracer in tissues of interest: In a process called positron emission decay, the tracer emits a positron that is annihilated upon encountering an electron, producing a pair of gamma photons in the process. These gamma photons move in opposite directions and are recorded by the coincidence detector of the PET system.

The underlying assumption of PET is that brain activity correlates with high radioactivity in the brain, due to the injected tracers. Thus, just as fMRI, this technique is an indirect measurement of brain activity, in the sense that it is based on the flow of cerebral blood.

PET has the advantage of being a very versatile imaging technique, as it can be used to monitor particular body functions by using specific radiopharmaceuticals. Furthermore, its spatial resolution is very high, comparable to the resolution of fMRI (voxel side length: about 3 mm). However, as mentioned before, this is a highly invasive technique in contrast to the others mentioned here, as the subject is exposed to ionizing radiation. Furthermore, the production of the radiopharmaceuticals requires expensive hardware and their short half-life usually requires them to be produced on-site, such that they can be injected as soon as possible.

Functional magnetic resonance imaging (fMRI)

Functional MRI measures the changes in the ratio of oxy- to deoxyhemoglobin, based on their different magnetic properties by applying a strong external magnetic field and a high frequency pulse perpendicular to it. For a detailed explanation of fMRI, refer to section 2.1.2.

Both PET and fMRI are capable of measuring activation in central regions of the brain, i.e. 3D maps of brain activity can be obtained. This is not feasible with fNIRS or techniques based on neuronal data (EEG, MEG). As PET exposes subjects to ionizing radiation, fMRI is the only non-invasive technique that can be used to obtain 3D information of brain activity. Another advantage of fMRI, that is shared by the other techniques in group (A), is its high temporal resolution. The voxels used in fMRI are typically 3 to 5 mm on a side for full-brain studies (for the study here, a voxel is defined as being 3.125 mm long and wide and 3.2 mm high). The temporal resolution of fMRI is limited by the speed of blood flow and is determined by the repetition time TR which typically ranges between 1 and 3 s (2 s for the study discussed here). Another disadvantage of fMRI is the expensive, space-consuming and immobile scanner that is required for measurements. In addition, it should be noted that subjects are not allowed to move at all while being scanned, since this leads to artifacts. Also, it has to be made sure that the subject is not claustrophobic.

Functional near infrared spectroscopy (fNIRS)

Functional near infrared spectroscopy (fNIRS, [124]) can be used to measure the concentration of oxy- and deoxyhemoglobin in the cerebral blood. Thus, just as fMRI, it is based on the hemodynamic and the metabolic responses. Also, it has signal characteristics that are very similar to those obtained with fMRI. The technical setup involves a number of light sources omitting near infrared light (700 nm to 900 nm) that is sent through cortical tissue and recorded by a detector array. While this type of light can transcend most tissue, it is partly reflected, absorbed and scattered by hemoglobin. These effects, however, are different between oxy- to deoxyhemoglobin. Thus, if the hemoglobin level is measured continuously, this technique can be used to detect relative changes in the concentration of oxy- and deoxyhemoglobin at a given location.

Since this imaging technique is based on the same phenomenon as fMRI, it can replace this technique for certain setups. An advantage over fMRI is the portability of an fNIRS system, which is simply mounted on the head of a subject. Also, the setup does not require conductive paste, as is the case for most EEG systems. However, it shares the disadvantage of EEG that it can only obtain a signal from cortical tissue and thus cannot be used to obtain a full brain scan which is possible with fMRI. Furthermore, the data is less detailed than that of EEG.

2.1.4 Assessment of fMRI for the given Task

The poor temporal resolution of fMRI is often cited as one of its major drawbacks. However, the nature of the arithmetic task which we examine in this study is relatively complex. As pointed out in [16], tasks like this have the advantage that the problem solving time is relatively long. Thus, time series can be acquired that are long enough to reveal part of the temporal structure of the task in the BOLD response. Additionally, this type of task tends to lead to relatively strong responses and thus a good signal-to-noise ratio in the signal.

As will be explained in more detail in section 2.4.1, the activity of modules in the computational model used here is associated with activity in small, predefined regions of interest in the brain. In order to monitor this activity, a high spatial

resolution is required, which can currently only be obtained with methods based on the hemodynamic response (PET, fMRI, fNIRS). Out of these methods, only PET and fMRI can currently provide the required information about brain activity in central regions of the brain. PET in turn is usually ruled out due to its invasive nature.

Therefore, despite the potential drawbacks, fMRI is an appropriate brain imaging technique for the task at hand.

2.2 Cognitive Architectures

In his influential book 'Unified Theories of Cognition', Alan Newell called for cognitive architectures as a way of specifying general theories of cognition [91]. Different definitions have been given for cognitive architectures, such as the following:

- *A cognitive architecture specifies the underlying infrastructure for an intelligent agent. ... An architecture includes those aspects of a cognitive agent that are constant over time and across different application domains.* [79].
- *A cognitive architecture is a computational modeling platform for cognitive tasks.* [123]
- *A cognitive architecture is a specification of the structure of the brain at a level of abstraction that explains how it achieves the function of the mind.* [5]

The last definition may be the most suitable for explaining the research presented in this thesis, even though it is not the most general.

Newell pointed out that only those cognitive models should be allowed by an architecture that are cognitively plausible, i.e. which correspond to human behavior. In contrast, models which do not fulfill this precondition should be rejected by the architecture. In other words, research on cognitive architectures is concerned with the creation and understanding of models that show capabilities similar to those of humans, which is a central goal in both artificial intelligence and cognitive science. Two important aspects of cognitive architectures should be pointed out that distinguish them from expert systems which are restricted to narrowly defined contexts. First, cognitive architectures aim at covering a wide range of tasks and domains. Second, they offer a single structure with certain theoretical commitments.This structure can be iteratively refined by integrating specific findings in it, after which the structure can be tested again. Thus, cognitive architectures follow Newell's call for research at the systems level rather than at the level of specialized components [90].

A wide and diverse range of cognitive architectures has been proposed which serve different purposes (oftentimes for historical reasons), some only as an abstract set of guidelines while full implementations exists for others. Prominent examples are ACT-R [5, 9], Soar [78, 91], EPIC [76], Icarus [114], Prodigy [36, 86] and Clarion [121], to just mention a few. [79] gives a very good overview of cognitive architectures and related research challenges. Other comparisons of cognitive architectures can be found in [106, 122], for instance. ACT-R is the only cognitive architecture that has been linked to actual brain imaging data so far. As this thesis is based on the ACT-R theory, this cognitive architecture will be described in more detail in the following section.

2.3 Introduction to ACT-R

ACT-R stands for 'Adaptive Control of Thought - Rational', where 'rational' indicates the influence that rational analysis has had on the ACT theory [3]. The history of ACT-R can be traced back to the HAM theory of memory (Human Associative Memory, [10]) and early ACT theories. ACT-R has been implemented in a computational framework, the newest version of which is ACT-R 6.0 that was used for this thesis. Both the framework and the models are written in Lisp. Within this simulation environment, cognitive models can be developed that can predict the behavior of humans and give insights into the reasons for their behavior [123]. This possibility to collect quantitative measures and directly compare them to the behavioral data of human subjects sets ACT-R apart from other cognitive architectures. Since ACT-R has both symbolic and subsymbolic components, it is referred to as a hybrid cognitive architecture: In addition to the symbolic production system which forms the core of ACT-R, subsymbolic components exist which are represented in mathematical equations.

In the following, the distinction between procedural and declarative knowledge in ACT-R is explained (section 2.3.1), before the modular structure of ACT-R 6.0 and the communication between modules is outlined in section 2.3.2. Finally, mathematical details are provided on how the time is calculated that is needed for retrieving procedural knowledge or uttering a response (section 2.3.3 and 2.3.4, respectively). This section only explains those aspects of ACT-R that are relevant for the research behind this thesis, leaving out learning mechanisms, for instance. Good introductions to the theory behind ACT-R can be found in [5, 9, 123], while the reference manual provides all technical details [28].

2.3.1 Procedural and Declarative Knowledge

One of the fundamental assumptions of the ACT-R theory is the distinction between two types of knowledge, procedural and declarative knowledge, which are stored in two separate memory structures in ACT-R [2]. Procedural and declarative memory differ not only in their contents, but also in how knowledge is represented.

Declarative memory contains facts about the world, such as the knowledge that 2+3=5, which are stored in the form of so-called chunks. In order to access declarative knowledge, it has to be explicitly requested. Chunks have an activation value associated with them that determines the time needed to retrieve them. In contrast, procedural knowledge contains information about how things are done, such as how an addition with carry is performed, for instance. This is captured in productions which have an IF-THEN structure, also called left-hand side (LHS) and right-hand side (RHS). If the conditions on the LHS match a given situation, the rule becomes active (it 'fires') and the actions on the RHS are executed. Each of these productions maintains a utility value that determines which rule fires in case there is a tie.

2.3.2 Modules in ACT-R

One of the main ideas behind ACT-R is that '*cognition emerges through the interaction of independent modules*' [5]. The modular structure of ACT-R restricts the way in which the production system can access information. The most current version of ACT-R (6.0) is composed of eight modules which can be seen in Figure 2.5.

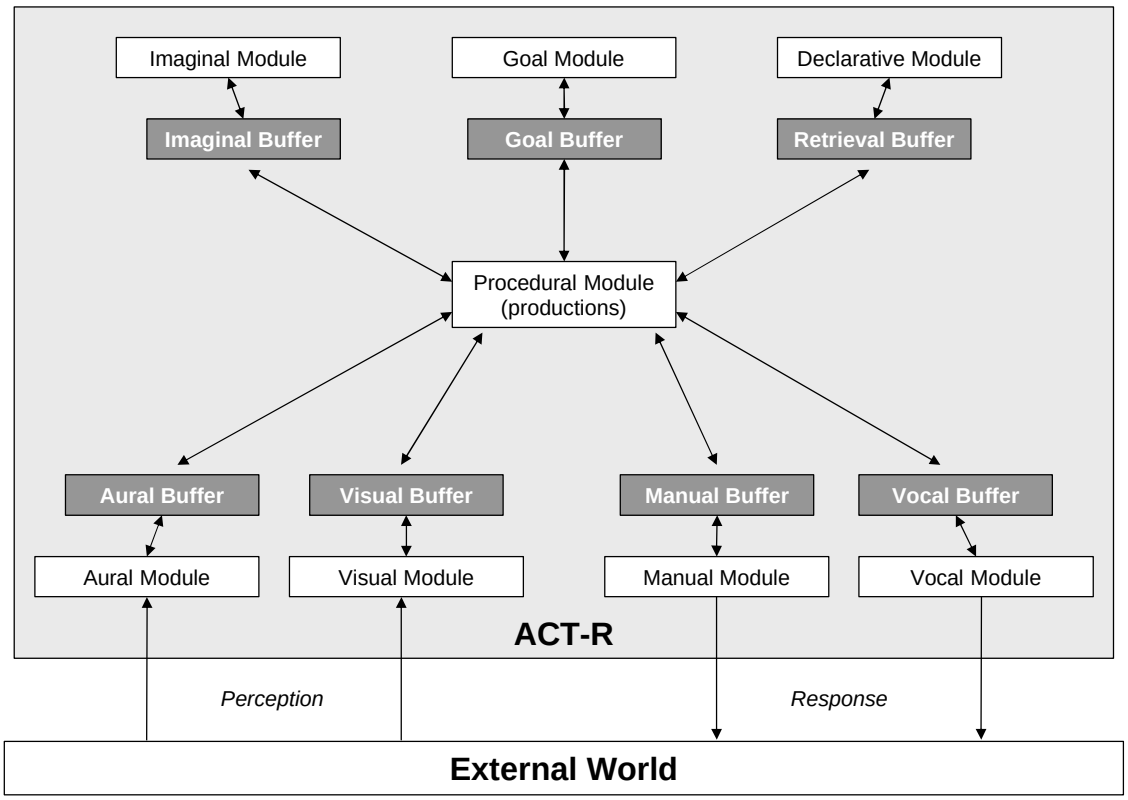

Figure 2.5: The eight modules of ACT-R 6.0 and their relationship to each other and the external world.

Modules in ACT-R 6.0

The modules fall in two different categories, modules with cognitive functions and perceptual-motor modules. Two of the cognitive modules were already mentioned in the previous section: the declarative module/memory, where declarative knowledge is stored, and the procedural module/memory, which contains procedural knowledge in the form of productions. In fact, these productions can be seen as rules for behavior which are selected when the situation is appropriate. For instance, if the goal is to solve the equation $x - 5 = 8$, the production might fire that the model wants to retrieve the result of the addition $8 + 5$. This request would be sent to the declarative module which retrieves declarative information from memory, in this case the information that $8 + 5 = 13$. Now that the problem is solved, the goal of the model might change (e.g. from 'calculate' to 'respond') which would induce the activation of the goal module. The task of the goal module is to keep track of the current intentions of the model. In contrast to this, the imaginal module is used to hold mental representations of the problem which becomes necessary when performing intermediate steps, among others. In the example mentioned before, the model might be solving a system of equations and needs to remember that $x = 13$ to solve the next equation $x + y = 20$. Another important difference between the goal and the imaginal module is that a time cost of 200 ms is associated with modifying the content of the imaginal buffer, whereas this is not the case for the goal buffer.

Besides the cognitive modules described so far, ACT-R has four perceptual-motor modules which constitute an interface to the external world or rather to a simulation thereof. These are to a substantial part based on the EPIC (Executive-Process Interactive Control) architecture [75, 76]. The visual and the aural module perceive the environment of the model, that is they process information about what the model sees and what it hears. A model can respond via the vocal or manual module which enable it to speak and to use a keyboard or mouse.

For the model described later on, it is important to note that the current implementation of the visual module does not model eye movements, as it is mainly a model of visual attention. Thus, there is no difference in time cost if the model moves its gaze to objects further away compared to objects close-by. This aspect is considered in a system called EMMA (eye movements and movements of attention, [111]), but is not yet integrated in the default vision module of ACT-R.

Communication between Modules

The modules can progress autonomously and can thus work in parallel, unless a module is waiting for information from another module. Parallelism within a module is supported as well; however, simultaneous processes can interfere with each other. Nonetheless, serial bottlenecks restrict the information processing as explained in the following.

Except for the procedural module, each module has a buffer associated with it (see Figure 2.5) that serves as an interface with that module since modules cannot be accessed directly. Modules communicate by putting information in their respective buffers. For instance, if the visual module has perceived new input on a screen, it encodes this information and makes it available for the other modules by placing it in the visual buffer. However, a buffer can only hold one chunk at a time. A chunk

is a small piece of information such as a fact ('5 + 8 = 13' in the example above) that was retrieved. This constitutes a serial bottleneck in information processing as it restrains how fast information can be communicated among modules.

At any point in time, the current state of ACT-R is represented by the contents of the buffers. The central procedural system operates in recognize-act cycles. This means that it recognizes a specific state of the system (represented by a specific pattern in the buffer contents) and acts accordingly by sending requests to modules or placing chunks in buffers. However, only a single production rule can fire at a given point in time. Moreover, this has a time cost of 50ms associated with it which seems to have become a consensus for many production-system architectures [14]. Thus, the fact that communication between modules depends on the procedural module can be seen as the overall bottleneck of the system. Unlike the other modules, the procedural module is not associated with a cortical structure but the basal ganglia.

2.3.3 Retrieval of Declarative Knowledge

As mentioned before, the declarative module can be seen as the declarative memory of the model, where knowledge is stored in chunks. The time needed for retrieving a chunk (retrieval time RT) depends on a number of factors and is calculated as follows:

$$RT = Fe^{-(f*A_i)} \tag{2.8}$$

with RT : time needed to retrieve the chunk [sec]
F : latency factor parameter (: lf)
f : latency exponent parameter (: le)
A_i : activation of the chunk

The activation A_i of a chunk is in turn determined by the following equation.

$$A_i = B_i + S_i + P_i + \epsilon_i \tag{2.9}$$

with B_i : base-level activation of the chunk
S_i: spreading activation value
P_i: partial matching value
ϵ_i: noise value, ϵ_i = transient noise + permanent noise

For the model described in this thesis, only A_i and ϵ_i are of importance. The noise value ϵ_i is calculated as the sum of two components, a transient component and a permanent component. While the latter is associated with a specific chunk, the transient component is computed each time a retrieval request is made, independent of what chunk is retrieved. For our modeling approach, the transient noise is sufficient. The transient noise is generated from a logistic distribution with $\mu = 0$; the parameter : ans determines the s value of the distribution.

A retrieval failure occurs if no chunk could be retrieved or the potential chunks to be retrieved are below the activation threshold. This failure is indicated after the following amount of time has passed:

$$RF = Fe^{-(f*\tau)} \tag{2.10}$$

with RF : time until a failure is indicated [sec]
F : latency factor parameter (: lf)
f : latency exponent parameter (: le)
τ: retrieval threshold parameter (: rt)

2.3.4 Uttering a Response

The time cost associated with uttering a response (articulation time) consists of the preparation and initiation time (the time before the output is detectable) plus the time needed to actually speak the word. Thus, it can be calculated according to the following equation.

$$A(x) = I + P(x) + S(x) \tag{2.11}$$

with $A(x)$: articulation time of a word x
I : initiation time
$P(x)$: preparation time for word x
$S(x)$: time to speak word x

The initiation cost I is always 50 ms by default. The preparation cost $P(x)$ is 150 ms if the model has not spoken before and 100 ms when the model has spoken before, but is not just repeating the word. In the latter case, there is no preparation cost. The time for actually uttering the word depends on the length of the word to be spoken. By default, ACT-R needs 50 ms per letter in the word to 'speak' it. As a matter of fact, ACT-R does not use a text-to-speech system, as it is only the timing of the response that is simulated, not the utterance itself.

2.4 Predicting a BOLD response with ACT-R

Traditionally, ACT-R models have been used to model and predict behavioral data of subjects. However, linking ACT-R to brain imaging data, specifically fMRI, is beneficial in two ways: On the one hand, fMRI findings can be used as evidence for assumptions in the architecture or point towards shortcomings of the theory which can then be addressed. On the other hand, assumptions underlying the architecture of ACT-R may explain the time course of brain activity in certain regions.

First, section 2.4.1 explains the mapping of the different modules of ACT-R to specific brain regions, as well as the interpretation and limitations of these mappings. Subsequently, section 2.4.2 provides details on how the BOLD prediction is calculated from the activity of these modules. A very good overview on this topic can be found in [13]. For more detailed information, refer to [5].

2.4.1 Mapping of Modules to Brain Regions

The basic idea behind predicting a BOLD response with ACT-R is to relate the activity of the modules to certain regions of interest (ROIs) in the brain. As described in section 2.5.1, previous research has led to the definition of specific regions of interest that are associated with the modules of ACT-R. These mappings can be found in Table 2.2 and are described in more detail in the next section. Each region is 5 voxels long, 5 voxels wide and 4 voxels high, with the exception of the Caudate (4x4x4 voxels) and the Anterior Cingulate Cortex (ACC) (5x3x4 voxels). For the research described here, a voxel is defined as being 3.125mm long and wide and 3.2mm high. Table 2.2 only give the Talairach coordinates for the left hemisphere since this is the region we are trying to predict. The location of these brain regions is illustrated in Figure 2.6 and explained in the following. These mappings are visualized in Figure 2.7 which also illustrates how these components are coordinated by the procedural module.

Table 2.2: Mapping of ACT-R modules to Regions of Interest (adapted from [5] (p.75)). The given Talairach coordinates describe the location of the regions in the left hemisphere. In order to obtain the x-coordinates of regions on the right-hand side, invert the sign of the x-coordinates provided here. (PPC = Posterior Parietal Cortex; LIPFC = Lateral Inferior Prefrontal Cortex; ACC = Anterior Cingulate Cortex).

		Talairach coordinates		
Module	ROI	X	Y	Z
(1) Visual	Fusiform	-42	-61	-9
(2) Aural	Auditory	46	-22	9
(3) Manual	Motor	-41	-20	50
(4) Vocal	Motor	-43	-14	33
(5) Imaginal	Parietal (PPC)	-23	-63	40
(6) Declarative	Prefrontal (LIPFC)	-43	23	24
(7) Goal	ACC	-7	10	39
(8) Procedural	Caudate	-14	10	7

Explanation of the Mappings

Both the visual and the aural module are associated with regions that reflect advanced processing of either visual or aural signals. For the visual module, a region in the fusiform gyrus in the temporal lobe was identified. Although large parts of the brain are concerned with processing visual signals, this region seems to show best when information in the focus of attention is processed. Other research has confirmed the role of this region for perceptual recognition [62, 84]. The aural module is mapped to a region in the secondary auditory cortex that is involved in advanced auditory processing.

The manual and the vocal module are mapped to different parts of the motor and sensory cortex. While the motor module is associated with a region along the central sulcus which is known to represent the hand (including parts of Brodmann Areas 3 and 4 at the central sulcus), the vocal module is mapped to a region further down the motor strip devoted to the face and tongue.

The imaginal module is associated with a posterior region of the parietal cortex which includes parts of Brodmann Areas 39 and 40 at the border of the intraparietal sulcus. In previous studies by John Anderson's laboratory, this region was shown to reflect representational changes in different tasks, i.e. it showed activity when mental representations had to be transformed [4, 6]. Other groups have also shown that this region is involved in visual-spatial and verbal representations, reflecting mental imagery [40, 45, 46, 105, 128].

For the declarative module, a region of the prefrontal cortex has been identified that includes parts of Brodmann Areas 45 and 46 around the inferior frontal sulcus. Both retrieval and storage operations seem to be reflected in this region, which is confirmed by a large number of studies on memory [33, 34, 52, 81, 125, 126].

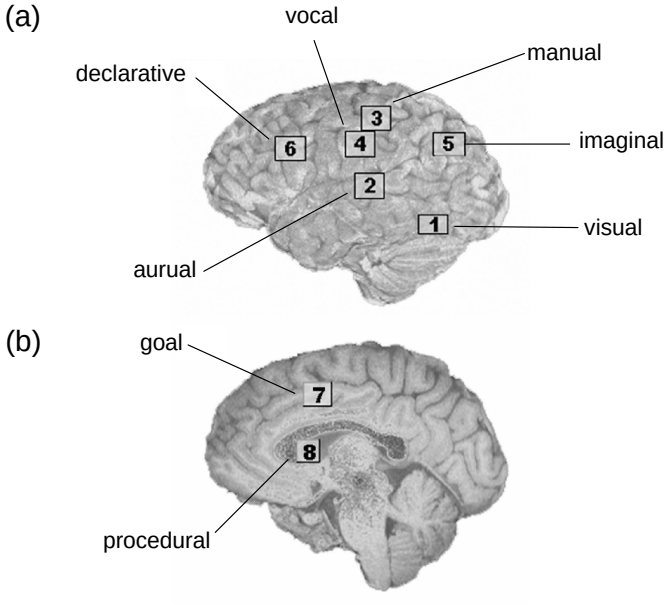

Figure 2.6: Locations of the eight ROIs associated with ACT-R modules, divided into regions close to the surface of the cortex (a) and those located deeper in the brain (b) (adapted from [5] (p.75)). The numbers refer to the corresponding mappings, given in Table 2.2.

The goal module is associated with a region of the anterior cingulate cortex (ACC) that includes parts of Brodmann Areas 24 and 32. Other research has confirmed the role of this region for internal control [29, 49, 98], which is consistent with the function of the goal module. For diverging theories on the role of the ACC, refer to section 2.5.1.

Due to its central and special role among the modules in ACT-R, the procedural module is not associated with a cortical structure, but the basal ganglia, in particular the head of the caudate. The procedural module is responsible for selecting cognitive actions depending on the current cognitive state, a role that is assumed to be similar to that of the basal ganglia (refer to [53, 65, 104], for instance).

Despite the fact that other research has shown similar associations, it should be noted that the exact functions of the regions in the parietal cortex, prefrontal cortex and ACC (associated with the declarative, imaginal and goal modules) are the subject of ongoing research. For more details on previous research associating ACT-R modules with certain regions of interest in the brain, refer to section 2.5.1.

Interpretation and Limitations of Mappings

Although several studies by John Anderson's laboratory have provided evidence for these associations between modules and the same predefined cortical regions (see section 2.5.1), they are still the result of work in progress and thus potentially subject to corrections. However, it is important to point out that we are taking a confirmatory approach here in contrast to the more typical exploratory approach. For the latter, the aim of the analysis is to distinguish which regions show significant changes in activation during a given task which is then taken as the basis of the interpretation. In contrast, we use regions which were defined and associated with

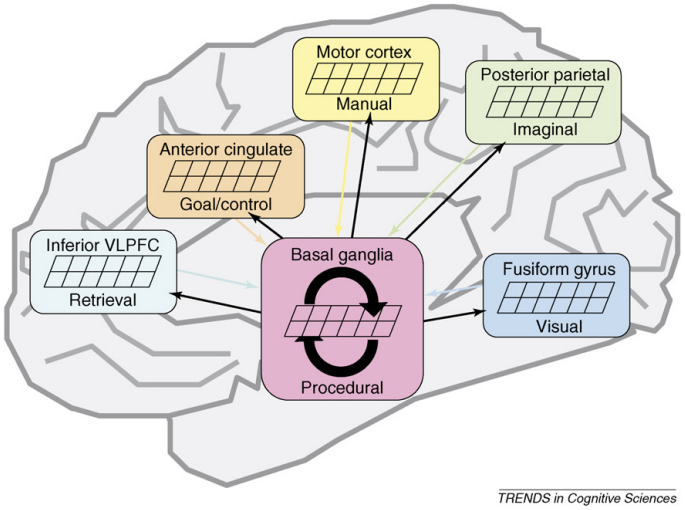

Figure 2.7: Illustration of the association of six ACT-R modules with brain regions and how these are coordinated by the procedural module (taken from [13], VLPFC = ventrolateral prefrontal cortex). The vocal and aural module are missing in this illustration.

ACTR modules before performing this experiment. Thus we do not run the risk of correcting for false alarms.

It is important to restrain from jumping to simple conclusions based on these one-to-one mappings of modules / functions to specific locations in the brain. Two arguments should be taken into account when working with these associations. First, there is no claim that a given function only drives activation in the one ROI that has been identified in ACT-R fMRI research. Quite in contrast, similar processes tend to be distributed in the brain, resulting in more than 30 regions involved in visual processing, for instance. Second, there is no assumption that the brain regions mentioned here are restricted to performing only one architectural function. For instance, other research has provided evidence that regions which are mapped to different modules in ACT-R fMRI research (specifically prefrontal/retrieval, parietal/imaginal, ACC/goal, caudate/procedural) have the tendency to show joint activation (for instance, refer to [35, 45, 113]). Nonetheless, these regions are engaged by factors that are systematically different and a range of studies conducted in John Anderson's laboratory have shown that these differences can be predicted based on the activity of the associated modules.

Finally, as mentioned before, these regions are restricted to the left hemisphere. This obviously makes sense for the mapping of the manual module to the motor region, since subjects in our experiments use their right hand to enter their results manually (refer to section 3.3.4 for details on how responses are given). Furthermore, findings by John Anderson's laboratory indicated that both the parietal and prefrontal regions (corresponding to the imaginal and the declarative module) show stronger responses in the left hemisphere for symbolic tasks. The left caudate and left ACC (procedural and goal module) tend to show a stronger response than their counterparts on the right-hand side as well [4]. Even though this is not the case for the remaining three ROIs, they are restricted to the left-hand side for reasons of consistency, such that only ROIs in the left hemisphere are taken into consideration.

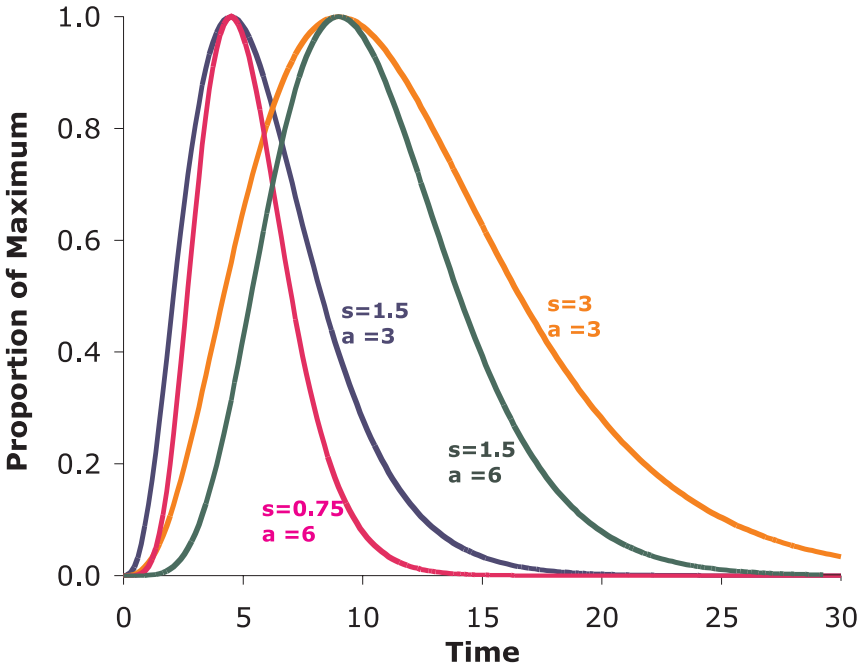

Figure 2.8: Effect of different choices of the exponent a and the time scale s on the shape of the hemodynamic curve (taken from [4]). The magnitude parameter m was adjusted such that each function reaches a maximum of 1.0.

2.4.2 Calculation of a BOLD Prediction

Given these mappings, the idea is that the engagement of a module corresponds to an activity in the respective brain region at the same time, driving the metabolic demand and thus producing a hemodynamic function. This function is modeled with a standard gamma function, which has been widely used for describing the hemodynamic response (see [30, 41, 42, 59], for instance). This curve is defined as follows:

$$H(t) = m \left(\frac{t}{s}\right)^a e^{-(t/s)} \tag{2.12}$$

with $H(t)$: hemodynamic function
m : magnitude parameter, determining the magnitude of the response
s : scaling factor for the time
t : time units that have passed since the engagement of the module
a : factor that determines the shape of the curve (the higher a, the steeper the slope of the function)

The impact of different choices for the exponent a and the time scale s can be seen in Figure 2.8. The parameter m that determines the magnitude of the response is adjusted such that the peak of all curves reaches 1.0 to facilitate comparison. As can be seen, the larger the value of s, the longer the duration of the BOLD response, while a larger value of a results in a quicker rise and fall of the function.

How does this relate to modules in ACT-R? The engagement of a module results in a BOLD response as described in equation (2.12) with a delay of t time units after the actual activity of the module. The function shows a slow response to a given

activity, peaking at $a * s$. These parameters are set to a=6 and s=0.75 s by default for all modules such that the time to peak is 4.5 s (pink curve in Figure 2.8.

Just as in the corresponding brain region, the BOLD response accumulates over the course of time whenever the module is active. Thus, we arrive at the prediction of the cumulative BOLD response by convolving the hemodynamic function with a demand function $D(t)$, that gives the probability that the region/module is active at time point t. This results in the following equation:

$$B(t) = \int_0^t D(x)H(t-x)dx \tag{2.13}$$

with $B(t)$: cumulative BOLD response
$D(t)$: demand function giving the probability that the region is engaged at time t
$H(t)$: hemodynamic function

The area under the curve is proportional to the time that the region/module is engaged in total. So if a module shows activity for T seconds, the area under the curve of the BOLD response equals $T * m * s * \Gamma(a)$. Hence, the BOLD response allows us to directly verify assumptions about the amount of time that a module is active during a task.

In summary, an ACT-R model of a given task results in predictions about the time course of activation, i.e. which module is active when and for how long (refer to Figure 4.5 for an example). This time course corresponds to demand functions $D_i(t)$ for each of the modules i . Based on this, the BOLD responses can be predicted by convolving each of these demand functions with the hemodynamic function. An example of a demand function $D(t)$ for the auditory cortex is given in plot (a) of Figure 2.9. The BOLD prediction that results from the convolution of $D(t)$ and $H(t)$ can be found in plot (c) of the same figure.

The approach described here has the disadvantage that it is not possible to predict a negative BOLD response: Since the demand function $D_i(t)$ reflects the engagement of a given module, it can never become negative. This is in contrast to what is shown in Figure 2.9 where $D(t)$ takes negative values as well.

2.5 Related Work

Several groups have attempted to predict BOLD responses of human subjects and a substantial amount of research has been completed on using ACT-R for this purpose. In the following, an overview is given over both of these (refer to section 2.5.2 and 2.5.1, respectively), as well as on the challenge of using verbal protocols in an fMRI study (section 2.5.3).

2.5.1 Previous Work on Linking ACT-R to fMRI

In the following, an overview is given about previous work on associating ACT-R modules with brain regions. Much of the research on linking ACT-R to brain imaging data investigated tasks in the math domain, with particular focus on algebraic problem solving ([4, 72, 99, 100, 120], for instance) besides arithmetic tasks [15, 108].

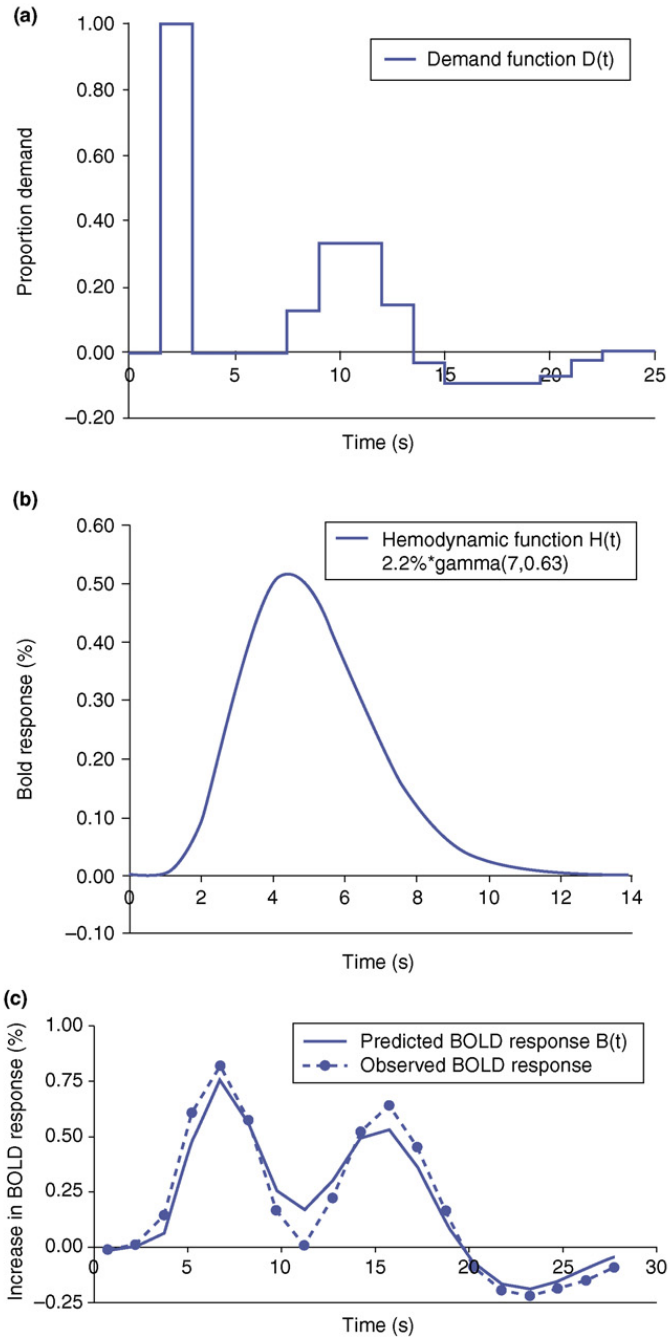

Figure 2.9: Illustration of the prediction of the BOLD response for the auditory cortex, based on the experiment in [16] (illustration taken from [13]). Plot (a) illustrates the demand function $D(t)$ for the aural module in ACT-R, i.e. the proportion of time that it will be active, according to a given ACT-R model. Plot (b) shows the estimated hemodynamic function $H(t)$. In plot (c), the observed BOLD response is plotted as well as the prediction which is calculated by convolving the demand function $D(t)$ in (a) with the hemodynamic function $H(t)$ in (b) (according to equation (2.13)).

The following is mostly based on reviews provided in [4, 5, 13]. The first section describes the first studies that linked ACT-R with fMRI data and which were focused on the parietal, the prefrontal and the motor cortex (associated with the imaginal, declarative and motor module, respectively). The following section provides some of the insights gained in previous studies concerning the mappings of the ACC to the goal module and the caudate nucleus to the procedural module. Finally, a study is described in more depth that is particularly relevant to this study as it was also concerned with the modality-specificity of cognition.

First Studies: Parietal, Prefrontal and Motor Cortex

The original publication behind the series on linking ACT-R to brain imaging data [17] was based on imaging studies for two different tasks: One task involved algebra equation solving in adults [19], while the other was concerned with the manipulation of abstract symbols [27]. An exploratory analysis led to the definition of three mappings of modules to regions of interests: the parietal (imaginal module), prefrontal (declarative module), and motor region (manual module). For subsequent confirmatory studies, these regions were kept fix such that they could be verified. The following studies [18, 20, 100] clarified the relations between the activation in these regions and led to a better separation of the activities in the parietal and prefrontal cortices. The association of the prefrontal cortex with retrieval as well as that of the parietal cortex with representational changes were confirmed. Furthermore, it could be shown that the factor that correlates to activation in these regions is the number of times the buffers are accessed (for retrieving information or changing a mental representation), rather than the time these pieces of information or representations are kept in mind.

The distinction between the prefrontal and parietal region was confirmed by Sohn et al. [118, 119]. Their research showed that, although both the parietal and prefrontal ROI show activity during retrieval, it is the latter which responds when the difficulty of the retrieval process is increased (based on the fan effect). This research was continued and extended by [11] where both the difficulty of the representation and that of the retrieval operation were varied (using retention delay), during a task where subjects studied and generated paired associates. It was shown that it was the parietal region that responded stronger when the pairs had to be constructed, but not the prefrontal. These significant differences between prefrontal and parietal ROI confirmed the assumed roles of these two regions. This was also the case when repetition was manipulated [43]. In a Tower of Hanoi task that required handling mental representations of future problem states, the parietal region also showed a stronger response than the prefrontal ROI [6], which supported the assumptions made about its role. All of the above studies influenced the time needed to retrieve knowledge from declarative memory.

As mentioned in section 2.4.1, the role ascribed to the parietal ROI is consistent with investigations by many other groups that found it to be related to retrieval factors [33, 34, 52, 81, 125, 126]. Similarly, there is strong evidence by a number of studies that the parietal ROI is involved in visual-spatial and verbal representations, reflecting mental imagery [40, 45, 46, 105, 128].

Even though the first ACT-R fMRI studies only described results for the prefrontal, parietal and motor ROIs, data from other regions was collected as well and was interpreted later. Most notably, the relation between a region in the anterior cingulate

cortex (ACC) and the goal module was investigated, as well as the association of the procedural module with a region in the head of the caudate nucleus. These two relations will be discussed in the following.

Anterior Cingulate Cortex

The ACC is assumed to play a role in motor control, arousal and cognition [96]. The exact role of the ACC in cognition is a subject of debate in the scientific community. A range of partly contradicting theories have associated the ACC with different aspects of cognition. Among others, activity in the ACC has been suggested to reflect the following:

- error detection [51], as error-related negativity seems to originate from the ACC [47],
- task difficulty [97], thus encompassing raised activity during hard trials which nonetheless do not lead to errors,
- conflict among potential responses, but not conflict detection [37, 83],
- learned error probability for a given trial [31],
- transitions between mental states, thus controlling cognition [29, 49, 98].

The theory that the ACC reflects state transitions is similar to the role that has been suggested by ACT-R fMRI research. Research in John Anderson's lab found that task difficulty influences activity in the ACC, but only where raised task difficulty involved a larger number of mental steps to be taken. For instance, a study where children learned equation solving over six days showed that activity in the ACC is related to task difficulty, if the latter is defined as above [4]. These findings contradict theories on error detection and response conflict for two reasons: Not only was this result also found for trials that do not involve errors, but the effect also shows before subjects enter their response. Another study provided strong evidence that activity in the ACC is related to controlling attention instead of conflict monitoring [116].

Even though the findings are by and large consistent with the error-likelihood theory, some ACT-R fMRI studies such as [117] found that the ACC activation fluctuates within trials. This effect cannot be explained by the error probability for a given trial, but suggests that changes in cognitive control occurred during the trial, thus supporting the theory that the ACC controls cognition.

Caudate Nucleus

As explained in section 2.3.2, the role of the procedural module differs from those of the others: it selects the rules to be fired, depending on the contents of the buffers of the other modules. For this reason, the procedural module is not associated with a cortical structure, but with the head of the caudate nucleus, which is part of the basal ganglia. This is consistent with other publications that assume that procedural memory can be associated with the basal ganglia ([21, 64], for instance). Also, it has been suggested that this region is part of a neural selection system [53, 65, 104],

which resembles the role of the procedural module in ACT-R. It should be noted that other cognitive functions have been attributed to the basal ganglia as well, such as procedural learning and reinforcement learning. However, these more specific theories can be united by the role which is assumed for the basal ganglia in ACT-R fMRI research [13].

For relatively simple experiments, the activity of the caudate can be predicted well based on the activity of the procedural module, as reported in [99]. However, the pattern of activity showed a high similarity with that of the parietal, prefrontal and the ACC in this study. This was not the case for more complex experiments, such as the study described in [15] which involved a complex series of mental calculations. Here, the caudate deviated from the other ROIs by showing a raised level of activity at the beginning and ending of the task. For this experiment, the activity in the caudate ROI could not be predicted with an ACT-R model. [120] investigated algebraic problem solving and the interaction between task- and control-related information in particular. For this task, the activity of the caudate could be predicted successfully: the activity of the caudate increased, as more information processing steps had to be taken, independent of the type of steps to be taken. The increasing demand of control (with rising difficulty) showed a comparable impact on the caudate and the ACC.

The prediction for the caudate ROI is based on the activity of the procedural module, i.e. on the number of productions that fire over the course of time. The activation of a production oftentimes leads to changes in the problem representation, i.e. activity of the imaginal module. For this reason, a strong correlation can oftentimes be found between the activity of these two ROI (as in [4]), unless input is presented in different modalities [118]. However, it might be more successful to use the amount of information conveyed by the productions that fire as basis for this prediction, as suggested by [120].

Modality-specificity of Cognition

[16] describes an ACT-R fMRI study that is concerned with the issue of modality-specificity of cognition, among others, and is therefore of special interest for the work presented here. Subjects were asked to perform an information-processing task, based on the task used in [18]. A major difference to [18] consisted in the manipulation of the input and output modality: the input was provided either visually or as an audio signal, subjects were instructed to respond either verbally or by keying the result in (only the output modality was varied in the study of this thesis). Furthermore, the subjects were either delayed by having to retrieve previously learned substitutions or not, similarly to the study presented here. However, there was also a condition in which they were purposefully delayed in order to trigger a rehearsal of the response, unlike the given study. As one of the results of the study, it was concluded that the perceptual regions (auditory cortex, fusiform gyrus) seemed to only reflect perception. In contrast to this, the two motor regions seemed to additionally engage in rehearsal, besides their expected motor control function. The four cortical regions of interest (parietal, prefrontal, anterior cingulate and caudate) did not seem to be affected by the input or output modality. However, a study on complex mental calculations found that the caudate ROI reflected the onsets of visual stimuli, while the parietal region was involved both during processing visual input and while manual output was generated [15].

2.5.2 Other Computational Models for fMRI Data

Since the advent of fMRI as a technology for studying brain activity, a multitude of studies have been run, investigating a broad range of questions. In the following, an overview is given over research that investigates the Hemodynamic Response Function. Subsequently, two approaches are described in more detail that propose computational models for predicting previously unseen fMRI data. Finally, these approaches are compared to the ACT-R fMRI approach.

Estimating the Hemodynamic Response Function

One strand of research is focused on estimating the so-called Hemodynamic Response Function (HRF) that describes the hemodynamic response of a given brain region to a short stimulus. Among others, a robust estimate of the HRF is crucial for research on neurovascular coupling, that is concerned with the question how metabolic and electrical brain signals relate to each other (measured by fMRI and EEG, for instance). The resulting models of the HRF differ in various aspects which will be discussed in the following:

- linear versus nonlinear models,
- parametric versus non-parametric approaches,
- aspects of processes that need to be known (onset times, number of processes etc.).

It is usually assumed that the brain can be considered to be a linear and time-invariant system (LTI), which has been shown to be a useful assumption [30, 32, 59]. Modeling the brain as an LTI implies that the superposition principle holds. This means that the BOLD response resulting from a given task in an fMRI experiment can be estimated for each brain region as a linear addition of the HRFs in that brain region. Each of these HRFs results from the activation of the brain region due to a mental process that occurs while the task is solved by a subject.

However, it has been argued that hemodynamic responses might add in a nonlinear fashion ([109], for instance). Even though nonlinear aspects have been observed, these are usually attributed to neuronal habituation effects, an idea that was brought up by [30]. Habituation refers to the phenomena that the magnitude of the response of neural cells to a stimuli diminishes when the stimuli is repeated. The findings reported in [42] support the linear systems approach, as even for the rapid succession of isolated trials the HRFs seemed to add linearly. Moreover, the paper provides a model and proposes methods for handling hemodynamic responses that overlap in time. However, the model is restricted in the sense that it is based on the assumption that both the onset time and the identity of the mental processes are known.

Both parametric and non-parametric methods have been proposed for estimating the HRF. Parametric approaches based on the LTI assumption estimate the HRF with a particular function, such as a gamma or Gaussian density. In order to account for variations in shape across different regions of the brain, this function is then sometimes biased to a certain degree. Early examples in the literature include [41, 56, 103]. A prominent method is called Statistical Parametric Mapping (SPM)

which refers to the combined use of the general linear model and Gaussian random fields. In essence, SPM is a mass univariate approach in that a statistical process is assigned to each voxel. Thus, spatially extended statistical processes are modeled and an unlikely diversion of the overall brain activity is then interpreted as being caused by certain cognitive processes, for instance [54–56].

As for non-parametric approaches, it has been proposed, among others, to estimate the HRF in a Bayesian framework, using temporal prior information on the cerebral hemodynamic response [60, 92]. Two of the main restrictions of these approaches, their restriction to synchronous event-related paradigms as well as their restriction to only estimate one HRF resulting from one condition, have been addressed in [39]. In synchronous paradigms, stimuli are aligned with the rate at which images are acquired, in contrast to asynchronous paradigms where this is not the case.

These Bayesian approaches are based on the assumption that all processes underlying the HRFs are known, including their onset times. This is not required for so-called Hidden Process Models (HPMs), as described in [68, 69]. Hidden Process Models provide a probabilistic framework which can be used to decompose the fMRI signal, recorded while a subject solves a cognitive task, into the hemodynamic responses of single processes. HPMs are capable of learning the timing of each of these overlapping processes as well as their contribution to the overall signal. This approach can be seen as a generalization of the General Linear Model and is a special case of Dynamic Bayes Networks.

Predicting Brain Activation Patterns

Besides approaches that focus on finding the neural representation of certain processes through statistical analysis, computational models have been proposed for estimating / predicting fMRI patterns. Models that are capable of estimating cerebral activation may not only provide important insights into the neural representations of certain processes, but also a better understanding of brain activations. In the following, two recent computational models are described that stand out due to their ability to predict fMRI data even for previously unseen stimuli.

Mitchell et al. presented a computational model that is capable of predicting brain-wide activation patterns that are assumed to encode the semantics of a given concrete noun (rather than the visual features of the word) [87]. The model is trained on the fMRI data of a few dozen words as well as a huge word corpus that provides it with the context in which these words occur. Based on these, it can predict the activation pattern even for previously unseen words with an accuracy high above chance level. The main assumption behind this work is that the neural representation of concrete nouns is correlated with the frequency with which the noun occurs with certain 'feature words' in the huge text corpus. The prediction is then calculated as a weighted sum of fMRI images, one per feature word, depending on how likely the noun is to co-occur with these feature words.

The work by Gallant et al. presented in [73] has shown that it is possible to identify which candidate image a subject is currently presented. The model for this task estimates a quantitative receptive-field model for each voxel, based on a Gabor wavelet pyramid. The model is built upon fMRI data that was recorded in the visual cortex (early visual areas V1, V2, V3) while a set of 120 natural images was

presented to the subjects. Subsequently, the model can be used to predict activation patterns for a range of novel images. These predictions are then compared with the actual brain activation patterns that were recorded and the best matching pattern is selected as hypothesis. It should be noted, that even though the accuracy of the prediction is high above chance level even for novel images, the model is subject-specific as it is trained and tested on data from the same person.

The ACT-R fMRI Approach in Context of other Models

It should be noted that each of the models described here takes a statistical approach to analyzing and predicting fMRI data. They are based on theoretical assumptions about the functioning of the brain, but were mostly developed for the purpose of interpreting and modeling fMRI data. In contrast to this, ACT-R stands out as a complex theory of cognition. This theory and its implementation as a computational framework were developed and evaluated on behavioral data over decades and is only now used for modeling and predicting brain imaging data. As a consequence, ACT-R research is concerned with the types of processes or subprocesses taking place during as task as opposed to estimating the shape of the hemodynamic response function. Also, as this approach is based on a general theory of cognition, it lends itself to investigations of a much broader scope of questions than the approaches described in [73, 87], for instance, which focus on the semantics of words and the perception of natural images. In contrast to the subject-dependent approach presented in [73], the ACT-R model presented here is subject-independent up to a certain degree, as it attempts to predict data averaged over 16 subjects.

Both here and in previous research linking ACT-R to fMRI data, a LTI approach has been taken, i.e. it is assumed that the HRFs add in a linear fashion. In a way, one could call the approach taken here parametric in the sense that gamma functions are used for estimating the HRF (see section 2.4.2). However, the focus is usually not on finding the most appropriate parameters of the gamma function (experiments with these parameters might still bring new insights; see section 8.2).

In contrast, modeling fMRI data with ACT-R is focused on investigating which types of subprocesses trigger the HRFs during a cognitive task. This is being done within the boundaries of the ACT theory of cognition, i.e. within the computational framework of ACT-R. This framework allows alternative models of these processes to be formulated and evaluated on the data. In a way, this is a similar approach as HPMs that allow testing different hypotheses about the number and types of mental processes during a given cognitive task. Both ACT-R fMRI research and HPMs are concerned with mental chronometry, i.e. they are concerned with decomposing a cognitive task into separate processes or steps, albeit at different levels of granularity. Hidden Process Models focus on high level cognitive processes such as 'Comprehend picture', whereas an ACT-R model would subdivide these processes down to a level of if-then rules, resulting in a much higher number of processes taking place. Ideas on how these two approaches could be combined are given in section 8.2.

2.5.3 Using Verbal Protocols in an fMRI Study

The ability of employing verbal protocols successfully in an fMRI study is very desirable for advancing our understanding of language production. Also, this methodology is of great importance for studying the question of modality-specificity of cognition, i.e. the question whether cognitive steps for solving a task are influenced by the modality in which the response has to be given. Neither of these questions can be studied in animal models as speech is unique to humans.

Unfortunately, using overt speech in an fMRI scanner can lead to artifacts. The studies described in [24] and [26] were among the first to investigate systematically the use of verbal protocols in fMRI studies and the resulting artifacts. Two types of artifacts have been identified.

First, overt speech can result in a repositioning of the head and thus lead to spatial misalignment between images as well as to distortions of the signal. However, studies have shown that this factor is usually negligible [95]. Second and more importantly, the BOLD signal can be disturbed by the so-called susceptibility effect. This effect is caused both by continuous changes in volume of air in the lungs of the subjects during articulation, as well as the tongue and jaw movement related to uttering a word. These factors lead to inhomogeneities in the magnetic field which in turn causes warping or blurring of the image. These artifacts are difficult to handle and cannot be corrected after the acquisition of the signal, as the effects are not distributed uniformly over the image. Also, the artifacts are strongly correlated with the task. It should be noted, however, that techniques have been proposed to deal with the susceptibility effect (among others in [25]), such as using an event-related design with a specific stimulus and interstimulus duration. For instance, an event-related fMRI study in which subjects had to speak briefly showed that brain imaging data can be obtained with fMRI that is relatively artifact-free, if the experimental setup is chosen with care [95].

Historically, covert speech was frequently used as a substitute for overt speech, as it is oftentimes assumed that these two actions share the same neural correlates, except for those related to motor execution. However, this assumption has been criticized by several groups ([24, 61, 66] among others), who claim that brain activation resulting from overt verbal responses is different from that of covert verbal responses. Evidence for this assumption has been reported by [38, 89], for instance. Moreover, it has been argued in [66] that a continuous feedback loop exists between auditory feedback and speaking, which is missing in the case of covert speech and may thus lead to a different brain activation pattern. In addition, it generally seems to be the case that the magnitude of the BOLD signal in the covert speech condition is much smaller than that in the overt speech condition (supported by [38], for instance). However, [66] also found the opposite for Broca's region, depending on which words the subjects were asked to speak.

Thus, even though the use of covert speech circumvents the issue of artifacts related to overt speech, it cannot fully substitute overt speech in experiments. Since the experiment discussed in this thesis is concerned with the interplay of the BOLD responses in two ROI related to speech and audition, among others, the use of covert speech was not an option here. However, it should be noted that an experiment involving overt speech needs to be designed with particular care to minimize artifacts.

3. Experimental Setup

The motivation behind the experiment is provided in section 3.1, followed by details about the recording equipment and software (section 3.2). Section 3.3 explains the task and the automatic problem generation as well as the four experimental conditions and definitions used later on in this thesis. The actual recording procedure is described in section 3.4 which provides details about the subjects, outlines the training and scanning session as well as the timeline for one recording.

3.1 Motivation and Hypotheses

A number of different questions were targeted in this experiment, for some of which an ACT-R model was needed, while this was not necessarily the case for others, as they could be answered based on the subject data alone. Nonetheless, these questions show the rationale behind the 2xw2 design of the experiment and are thus provided here. For the sake of clarity, the main questions are formulated here as hypotheses and will be discussed in more detail at the end of this section. It should be pointed out that Hypothesis 4 can be considered the central hypothesis of this thesis.

Hypothesis 1. It is possible to use verbal protocols in the given fMRI study.
Hypothesis 2. The output modality (verbal or manual) does not affect the BOLD responses of the predefined prefrontal, parietal, anterior cingulate or caudate ROI, i.e. the brain regions that are associated with the ACT-R modules of cognitive control (declarative, imaginal, goal and procedural module).
Hypothesis 3. The BOLD responses for the verbal and the manual output modality are affected in a similar way by the level of difficulty (manipulated by including substitutions or not).
Hypothesis 4. The BOLD responses shown by the subjects during the arithmetic task can be predicted by an ACT-R model for the eight predefined regions of interest.

It should be noted that Hypotheses 1 to 3 are qualitative predictions, whereas Hypothesis 4 formulates the more general assumption, central to this thesis, that the patterns of activation (i.e. the BOLD response) and the effects shown by subjects

can be predicted by an ACT-R model. In order to investigate these questions, two variables were varied (for more details, refer to section 3.3.4):

- output modality: verbal or manual (Hypothesis 1, 2, 3, 4)
- difficulty level: low or high (Hypothesis 3, 4)

As problem domain, fractional arithmetic was chosen, since extensive knowledge exists in John Anderson's laboratory on examining mathematical problem solving with ACT-R fMRI studies. In fractional arithmetic, procedure and conceptual factors come together in young children. Additionally, there was an interest in whether the algorithmic character associated with algebra would show up in this domain.

In the following, the hypotheses are explained as well as the expected results.

Hypothesis 1

To a large degree, this was a feasibility study for verbal protocols (Hypothesis 1), as we wanted to examine their usefulness in the scanner and investigate whether they would lead to artifacts in the signal. The manual condition was added as a control condition. This was a logical choice since most previous studies in ACT-R fMRI research used manual output. Manual output has the advantage of providing a clear and well-defined time stamp for each response that can easily be recorded by the program that runs the interface of the experiment.

Hypothesis 2

A more generalized way of phrasing Hypothesis 2 would be to say that we expected that the output modality would not have an impact on the cognitive steps taken by the subjects to solve the given arithmetic problems. Thus, we assumed that the response times in the manual versus the verbal condition would only differ to a small degree, reflecting the difference in time caused by reporting the result in different ways (results can usually be spoken out loud faster than entered manually). More importantly, we expected to obtain

- similar BOLD responses for both output modalities (verbal, manual) for the brain regions associated with the cognitive modules (imaginal/parietal, declarative/prefrontal, goal/ACC, procedural/caudate),
- BOLD responses that differ between the two output modalities (verbal, manual) for the brain regions associated with the perceptual-motor modules (visual/fusiform, aural/auditory, manual/motor, vocal/motor).

We expected that the latter would differ since the answers were reported in a different way (affecting the manual and vocal module) which would then lead to different perceptions of the answer (visual, aural). In contrast to this, we did not expect major differences for regions associated with the cognitive modules, because we assumed that the output modality would not affect the cognitive steps taken (Hypothesis 2). This assumption is consistent with previous findings of an ACT-R fMRI study where the input and output modalities were varied as well ([16], see also section 2.5.1).

Hypothesis 3

For investigating Hypothesis 3, we introduced a difficulty measure which was manipulated in order to test whether this would have the same impact on task performance for both output modalities. The level of difficulty was raised by asking subjects to solve problems where the numerators were replaced by number-to-letter substitutions. The subjects were trained on on these substitutions before the scanning session (see section 3.4.2). We expected the subjects to need more time for solving problems that included number-to-letter substitutions, but assumed that the extent of this effect would be the same for each of the output modalities (Hypothesis 3).

It should be noted that the investigation of Hypotheses 1 to 3 does not necessarily require a model. However, a model and its prediction are capable of providing helpful hints which behavioral and brain imaging data can be expected based on the ACT theory and its cognitive architecture. In particular, the investigation of Hypotheses 2 and 3 benefit from having a model for the data: A model is a way to formalize the cognitive steps that we assume to be taken by the subjects. Thus, if the model provides a good fit to the subject data, it can provide us with ideas for why the hypotheses are true or not, i.e. why the output modality might or might not have have an impact on the cognitive steps taken or why the level of difficulty effects both output conditions alike.

Hypothesis 4

Hypothesis 4 can be considered the central hypothesis of this thesis. As mentioned before, it is crucial to note that we are taking a confirmatory rather than an exploratory approach, since we use predefined regions (see section 2.4.1) whose activity we want to predict with an ACT-R model. If this prediction is successful, this is not only a confirmation of the predefined mappings, but also of the ACT theory and architecture as such. However, it is important to note that a failure to do so could be the result of different factors: the model might not be refined enough, an aspect might be missing in the ACT-R architecture, or this might point to an error in the mappings of modules to brain regions, among others. It should be noted beforehand that most previous ACT-R fMRI studies with complex experimental tasks found discrepancies between the BOLD prediction of the procedural module and the actual activity in the caudate ROI. Thus, we expect this to be the case for this study as well.

In this context, it should also be pointed out, that extensive efforts have been made before to use the ACT-R modules for predicting brain activity in certain regions. However, previous experiments did not include verbal protocols. [16] was the only study where this was the case, but a different domain was investigated, focusing on language rather than on arithmetic. Thus, this study brings a novel aspect to the strand of fMRI ACT-R research, as the experimental data allows to test whether the activity of the verbal module predicts related brain activity correctly in a mathematical context. Directly linked to this, the activity of the aural module in reaction to the verbal module in ACT-R can be investigated, as well as the question whether this activity corresponds to actual brain imaging data.

3.2 Recording Equipment and Software

3.2.1 fMRI Scanner and Preprocessing Steps

The following description is based on [8], since the same scanner and the same preprocessing steps were used. The recordings were done at the Brain Imaging Research Center (BIRC) in Pittsburgh (Pennsylvania, USA) which is a joint venture between Carnegie Mellon University and the University of Pittsburgh.

Images were acquired using gradient echo-planar image (EPI) acquisition on a Siemens 3T Allegra Scanner (see Figure 3.1) using a standard RF head coil (quadrature birdcage), with 2 s repetition time (TR), 30 ms echo time (TE), 70 degree flip angle, and 200 mm field of view (FOV). The experiment acquired 34 axial slices on each TR using a 3.2mm-thick, 64x64 matrix. The anterior commissure - posterior commissure (AC-PC) line was on the 11th slice from the bottom scan slice. Acquired images were analyzed using the NIS system [1]. Functional images were motion-corrected using 6-parameter 3D registration (AIR, [127]). All images were then co-registered to a common reference structural MRI [2] by means of a 12-parameter 3D registration [127] and smoothed with a 6 mm full-width-half-maximum 3D Gaussian filter to accommodate individual differences in anatomy.

In other words, each trial resulted in a time series with a temporal resolution of 2 s. The length of the time series was not fixed, but depended on the length of the given trial, as subjects were given as much time as they needed to solve a problem. This was followed by a cool down period of a minimum of six scans, plus whatever time there was remaining (< 2 s) such that the next stimulus onset would occur at the next scanner pulse. Thus, this usually resulted in 7 cool down scans.
Each slice that is recorded is divided into a 64x64 grid. The x- and y-sizes of the recorded voxels measure 3.125 mm (=200 mm / 64), since we use a field of view (FOV) of 200 mm. 34 of these matrices (one per slice) are acquired every 2 s. Thus, every 2 s, the MR signal for each of 64x64x34 voxels is recorded, where dimensions of each voxel are 3.125 mm x 3.125 mm x 3.2 mm.

3.2.2 Audio Recording and Analysis

Audio Recording

Audio recordings for the fMRI sessions were conducted with a microphone that was installed at the inner top of the head coil, hooked into a set of headphones as can be seen in Figure 3.2. This way, interferences from respiratory motions could be avoided. The recording was done in two forms, both of which used the same microphone. First, the microphone recorded to a program called Sound Forge (SONY) on a console in the control room. Second, a method was created to simultaneously record to a laptop. The latter was used as a backup and contained a continuous recording of the entire session, whereas the main recording (to the console) only contained the audio signal during the blocks in which problems where solved.

[1] NeuroImaging Software package, developed at the Laboratory for Clinical Cognitive Neuroscience, University of Pittsburgh, and the Neuroscience of Cognitive Control Laboratory, Princeton University (http://kraepelin.wpic.pitt.edu/nis/index.html)

[2] The Talairached reference image that was used is available from the website http://act-r.psy.cmu.edu. This is the reference brain used for all studies from John Anderson's laboratory.

Figure 3.1: Preparation of one of the subjects for the fMRI scan.

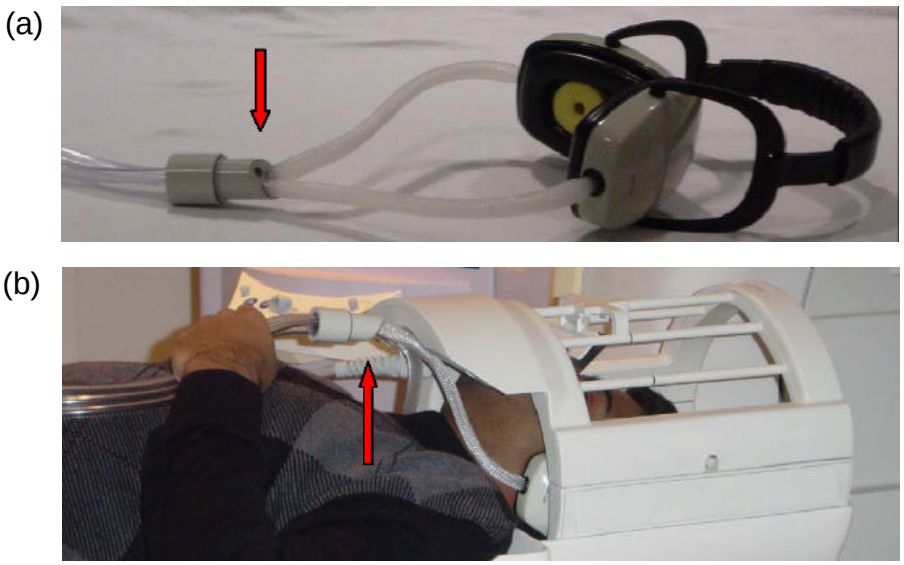

Figure 3.2: Setup of the microphone for recording the audio signal in the scanner (taken from [70]). In both images, the red arrow points to the microphone. In Figure (a), the set of headphones can be seen into which the microphone is integrated. Figure (b) shows where the microphone was located during the recording.

A program named Voice Elicitor was used to extract the voice from the scanner noise in this signal (refer to [71] for more details). Essentially, Voice Elicitor employs a segmented active noise cancellation technique. The reference noise used for this method is dynamically updated to account for the changes in the acoustic noise in the scanner due to slight movements of the subject. First, a version of Voice Elicitor was used that was created by the BIRC. Later, this was replaced by a corresponding program written by Shawn Betts.

Analysis of Audio Recordings

The files created by Voice Elicitor were analyzed with Annotate Wave, a program that was created by Shawn Betts as well. Annotate Wave selects audio wave forms that resemble a vocal response. Thus, the process of annotating the file is made half-automatic and the process of transcribing the verbal protocols is facilitated. Subsequently, an experimenter has to go through the data in order to check the segmentations and transcribe the content of what was said by the subject. This was done by Jennifer Ferris for the most part. These protocols were then synchronized with the behavioral log file created by the program that controlled the experiment. This log file contains the times and content of the manual responses of the subjects, i.e. when the subjects clicked on which buttons. Thus, the synchronized log file shows the time course of both the verbal and the manual responses.

3.3 Corpus Production

In the following, the arithmetic task is explained as well as the motivation behind the instructions and the automatic problem generation (section 3.3.1 to 3.3.3). The 2xw2 design of the experiment is outlined in section 3.3.4, where the different experimental conditions are explained. Finally, definitions regarding response times, blocks and recordings are provided that are used throughout this thesis (section 3.3.5 and 3.3.6).

3.3.1 Description of the Task

A pilot study showed that the original task, that included multiplication and then the addition of fractions, was too open. On this basis, the fractional arithmetic task was developed that is described in the following.

The subjects are asked to solve arithmetic problems. For each problem, they have to add two given, single-digit fractions. Thus, each problem has the following form:

$$\frac{n1}{d1} + \frac{n2}{d2} = ?$$

with $n1$, $d1$, $n2$, $d2 \in \mathbb{N}, < 10$

While doing so, they are asked to follow the four steps outlined in Table 3.1 and report the result of each of these steps. The first two steps result in two intermediate results, the last two steps in the final fraction. Before giving a rationale behind these instructions and their order (section 3.3.2), these steps will be explained, using the following example:

$$\frac{2}{9} + \frac{3}{8} = \frac{16}{72} + \frac{27}{72} = \frac{43}{72}$$

Table 3.1: Steps to be taken by the subject for the problem $\frac{n1}{d1} + \frac{n2}{d2} = ?$.

Description of the step	Mathematical operation
1. calculate the first numerator	$num1 = n1 * d2$
2. calculate the second numerator	$num2 = n2 * d1$
3. sum these two new numerators up	$newNum = num1 + num2$
4. calculate the new denominator	$newDenom = d1 * d2$

Due to the way in which problems are generated (see section 3.3.3), the fractions first need to be extended such that they share the same denominator. The subjects are instructed to report the numerators of these two new fractions, first the one of the left fraction (here, $num1 = 2 * 8 = 16$) and then the one of the right fraction (here, $num2 = 3 * 9 = 27$). Since the fractions are chosen such that the denominators do not have a common factor, this always requires a multiplication for each of the intermediate results. Then, the subjects have to enter the numerator and denominator of the resulting fraction ($newNum = 16 + 27 = 43$ and $newDenom = 9 * 8 = 72$ in the example), which requires an addition for the new numerator and a multiplication for the new denominator. It should be noted, that subjects were not explicitly told beforehand that the denominators do not have a common factor.

3.3.2 Motivation for the Instructions

The timing of the responses of the subjects allows us to make assumptions about the time needed to perform certain cognitive steps. Therefore, we specifically asked for the two intermediate steps to be taken and reported, in addition to asking the subjects to report the final fraction.

Furthermore, it was necessary to guide the subjects as of when to perform which mathematical operation such that the resulting brain imaging data of different subjects and trials were comparable to each other and could be aggregated (using a method called event-locked averaging, as described in section 6.4.1). However, the pilot study had shown that subjects tend to employ a wide variety of strategies to solve the arithmetic task when they are given the freedom to do so. Thus, the interface was designed in a way that limited this freedom and asked subjects specifically to follow four steps in a certain order (see Table 3.1). Of course, multiple orders exist in which these four steps could be potentially taken.

There were multiple reasons for choosing this order of steps for the experiment. First, it seemed to be the most popular strategy among the subjects of the pilot study. Then, the order was motivated by the fact that a result has to be reported for each step. We asked the subjects to first calculate the new numerator (steps 1 to 3) and then to perform the simpler calculation of the new denominator (step 4) as we wanted to keep the user interface as simple as possible, not showing too much information at any point in time. In this way, the subjects entered the numerator and denominator of the final fraction right after each other and had a visual feedback for the final fraction. An alternative and potentially more intuitive order of steps would have been to first calculate the new denominator (step 4) and then the new numerator (step 1 to 3). However, this would have meant that the denominator

would have needed to be shown the whole time to provide visual feedback of the final fraction which we wanted to avoid.

We assume that we can treat the times when the results are reported as the time needed to perform specific cognitive steps. It may be argued that subjects could take steps in a different order while still reporting the results in the order we asked for. Also, they might perform multiple calculations before reporting the results. The model partly accounts for the latter, as different answering strategies were implemented (refer to section 4.5 for details). Concerning the first criticism, it should be taken into account that performing the steps in a different order requires to keep the results in mind. It can be assumed that subjects would prefer to follow the steps suggested to them instead of taking their own approach in order to avoid unnecessary memory retrievals. Also, this approach would lead to a very long answering time for the first result compared to the others (as multiple calculations would need to be completed before the first answer was given), which is not indicated by the subject data (see section 5).

3.3.3 Automatic Problem Generation

The problems are generated and chosen in the following way: First, a denominator is selected for the first fraction that is greater than 2 and smaller than 10. Then, based on that number, a denominator is picked for the second fraction such that they have no common factors. Finally, a numerator is picked for each fraction that is smaller than and that has no common factors with the denominators. All denominators and numerators are single digit numbers. Those fraction pairs whose sum has a numerator less than 100 and whose intermediate numerators are both double digits are collected for use as problems. Out of these 220 potential problems, problems are picked at random to fill the sessions, the only constraint being that a subject encounters a problem only once. This also excludes the problem $m+n$ if the subject has already solved $n+m$, where n and m are fractions. This results in a list of 110 problems. Out of these, 48 require an addition with carry when calculating *newNum*, while the addition does not involve a carry for the other 62 problems.

3.3.4 Conditions

For the reasons outlined in section 3.1, two variables are varied during the recordings, resulting in a 2xw2 design of the experiment:

- output o: the way in which the four results should be reported ($o \in \{verbal, manual\}$),
- mapping m: whether the problems contain number-to-letter substitutions or not ($m \in \{sub, nosub\}$).

Each of these two variables can take two different values, as specified in Table 3.2 and as will be explained in the following two sections. Thus, a problem has to be solved in either of these four conditions: *verbal-nosub*, *verbal-sub*, *manual-nosub*, *manual-sub*.

Table 3.2: Variables in the experiment: output o and mapping m.

Variable	Value	Meaning
o	*manual*	The subjects need to type in all four results.
	verbal	The subjects speak the results of the first two steps out loud (*num1*, *num2*), while they enter the results of the last two steps manually (*newNum*, *newDenom*).
m	*nosub*	The given problems only consist of numbers.
	sub	The numerators of the fractions are replaced by number-to-letter substitutions that have to be re-substituted to solve the problem (mappings are given in Table 3.3).

Output

As pointed out before, the manual condition was included as a control condition to the verbal protocols (for more details, refer to section 3.1). The two methods for output only differ in how the subject is supposed to report the first two results; the last two (*newNum*, *newDenom*) are always entered manually. Therefore, the answering time of the last two steps of a *verbal* trial are directly comparable to the corresponding steps of a *manual* trial. The main reason for always having the subjects enter the last two results manually was to obtain a relative gold standard of how long it took a subject to solve a given problem. Since we wanted to have good timing data for the final result, manual responses were more appropriate than verbal responses: The time stamp of a click on a button can be determined very precisely (and thus delivers a very clear endpoint of the task), whereas a verbal response has a certain duration and less well-defined starting and ending points.

In the *verbal* condition, the subjects give the first two responses by speaking them out loud, whereas they are entered manually in the *manual* condition. For entering a result manually, the subjects use a mouse with their right hand in order to click on buttons shown on the screen. These buttons are arranged in the same way as the keys of the num block of a standard keyboard (see Figure 3.3) in order to minimize eye movement and the time needed to move the cursor of the mouse to a button. Once a number is entered by a mouse click, it cannot be deleted anymore. Including a function for deleting numbers would not have made sense for this experiment, since we are only concerned with problems that were solved correctly from the start. Only in this case can we make plausible assumptions about the cognitive steps taken by the subject to solve a given problem. The model is based on these assumptions and cannot capture the random thoughts and emotions that the subject might have upon entering a wrong number.

The output variable (o) is varied between recording blocks, which means that the subjects are instructed at the beginning of each block to use either the *verbal* or the *manual* method. Depending on the subject ID, the subjects are asked to use the *manual* method for the first block (if the subject ID is odd) or the *verbal* method (otherwise). For the subsequent blocks, the output condition is alternated (e.g. *manual-verbal-manual-verbal...*).

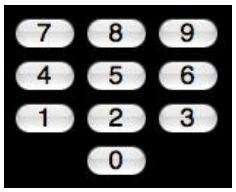

Figure 3.3: Layout of the number pad for keying in results.

Mapping

This variable was introduced for various reasons. First and most importantly, we wanted to have a measure of difficulty that we could manipulate to investigate the effect of varying difficulty on the two different output modalities, as explained in section 3.1.
Second, we wanted to further test the theory that the prefrontal cortex is associated with the retrieval activity of the declarative module. The substitution condition (*sub*) requires two additional retrievals from memory, compared to the no-substitution condition (*nosub*), as subjects need to remember two letter-to-number substitutions that they learned previously in order to solve the given problem. This should lead to a higher activation of the prefrontal cortex in the substitution condition, but not the no-substitution condition (*nosub*).
Third, even though the complexity of the task results in a fairly long problem solving time already, the additional retrievals due to the introduction of substitutions slow the subjects down and thus help spreading the cognitive steps further apart time-wise. In this way, the substitutions facilitate teasing apart the cognitive steps to perceive the temporal structure of the BOLD response. Hence, this condition contributes to further alleviating the drawback of fMRI as being a brain imaging technique with low temporal resolution.

Table 3.3 shows the number-to-letter substitutions that were used. The subjects are trained on these mappings before the recording starts, as explained in section 3.4.2. Mappings only affect the numerators of the two given fractions. In the substitution condition, these numerators are replaced by letters that have to be mentally substituted to solve the problem. The rest of the problem stays the same, i.e. the results are to be entered as integers. For instance, a problem like this might be shown:

$$\frac{p}{9} + \frac{t}{8} = ?$$

The subjects then recall the mapping they learned previously, i.e. that $p = 3$ and $t = 2$, and solve the problem accordingly. When no substitution is used (*nosub* condition), the problem is presented with normal integers.

The mapping variable (m) is varied within each block of ten problems, such that each block contains five problems with substitutions and and five without. The problems are then presented in random order.

3.3.5 Response Times

The behavioral data can be characterized in two different ways, either by using absolute or the relative response times. These are defined in the following and are visualized in Figure 3.4. Response times refer to either a verbal or a manual response.

Table 3.3: Letter to number mappings used for the substitution condition.

Number	Corresponding mapping
1	n
2	p
3	t
4	k
5	a
6	h
7	u
8	b
9	r

Definition. The absolute response time $AbsT(x)$ of an answer x is defined as the time that has elapsed from the beginning of the task (stimulus onset at time S) until the answer x is given ($x \in \{num1, num2, newNum, newDenom\}$).
For a verbal response, this is calculated as the difference between the times of stimulus onset and the time when the utterance ends (V_{end}):

$$AbsT_{verbal}(x) = V_{end} - S$$

If the answer is given manually, two digits need to be entered by clicking on them, since every result is a double digit number. The time of the second click (M_2) is used for calculating the absolute response time:

$$AbsT_{manual}(x) = M_2 - S$$

Definition. The relative response time $RelT(x)$ of an answer x is defined as the time that has elapsed since the previous answer has been given
($x \in \{num1, num2, newNum, newDenom\}$).
Thus, the relative response times are calculated as follows:

$$\begin{aligned} RelT(num1) &= AbsT(num1) \\ RelT(num2) &= AbsT(num2) - AbsT(num1) \\ RelT(newNum) &= AbsT(newNum) - AbsT(num2) \\ RelT(newDenom) &= AbsT(newDenom) - AbsT(newNum) \end{aligned}$$

Definition. The problem solving time T is defined as the time that is needed to complete a problem, that is the absolute time needed until the fourth answer is given. Thus, T corresponds to $absT(newDenom)$.

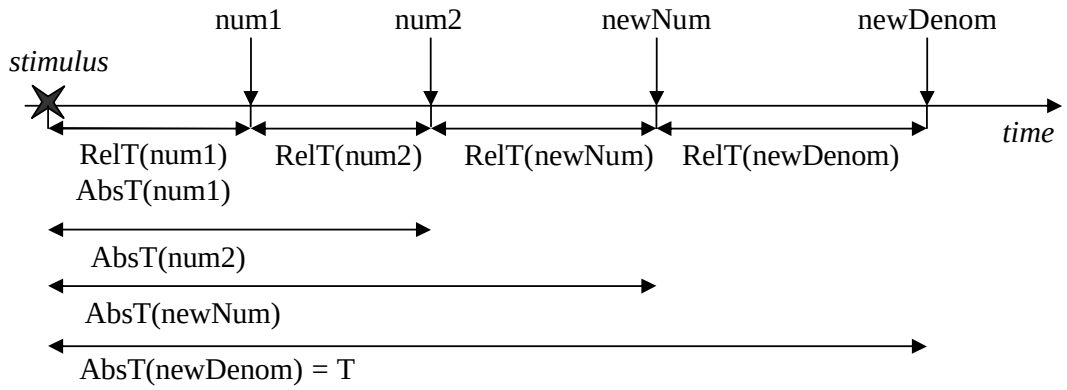

Figure 3.4: Time line for solving one problem, with the definitions of relative and absolute response times (RelT and AbsT respectively) as well as the problem solving time T.

3.3.6 Blocks and Recordings

A recording for a given subject is composed of various blocks which will be explained in the following.

Definition. A block $L(o)$ of length n is defined as an ordered set of triples

$$((o, m_i, p_i))_{i \in \{1,...,n\}}$$

where $m_i \in \{sub, nosub\}$ defines whether the numerators were replaced by a mapping, $o \in \{manual, verbal\}$ defines the output modality for that block and p_i is the arithmetic problem to be solved. Thus, each triple (o, m_i, p_i) corresponds to one arithmetic problem and the conditions in which it has to be solved.

The value of n is always set to 10, that is each block consists of 10 problems. As mentioned before, the output variable was varied between blocks.

Definition. A recording $R(s)$ of a subject s is defined as a set of size r

$$R(s) = (L_1(o_1), L_2(o_2), ..., L_r(o_r))$$

where $L_i(o_i)$ is a recording block ($i \in \{1, ..., r\}$) and

$$r = \begin{cases} 4 & \text{in the training session (recording outside the scanner)} \\ 10 & \text{in the scanning session (recording in the scanner)} \end{cases}$$

Thus, if the subjects were recorded in the scanner, they had to solve 100 problems (10 blocks with 10 problems each). In this case, the recording can be described in the following way (the example below is for a subject with an odd ID, thus starting with the *manual* condition):

$(L_1(manual), L_2(verbal), L_3(manual), L_4(verbal), L_5(manual), L_6(verbal),$
$L_7(manual), L_8(verbal), L_9(manual), L_{10}(verbal))$

In detail, the first block might take the following form. As explained in section 3.3.4, each block needs to contain five substitution problems ($m = sub$) and five problems where normal integers are used ($m = nosub$).

$$\begin{aligned} L_1(manual) = ((&manual, sub, p_1), (manual, nosub, p_2), (manual, nosub, p_3), \\ &(manual, nosub, p_4), (manual, sub, p_5), (manual, sub, p_6), \\ &(manual, nosub, p_7), (manual, sub, p_8), (manual, sub, p_9), \\ &(manual, nosub, p_{10})) \end{aligned}$$

3.4 Recording Procedure

In the following, details about the subjects are provided and the procedure of learning, training and scanning sessions is outlined as well as the timeline for one recording.

3.4.1 Subjects

After piloting efforts, data was obtained from 21 subjects. However, during the recording of five of these, technical issues occurred. Thus, the database contains data from the remaining 16 subjects out of which 7 are female and 9 are male. The subjects ranged between 19 and 28 years of age, with an average age of 22.07 years. All subjects had normal or corrected-to-normal vision (using contact lenses) and were right-handed. Subjects were paid for their participation. Participants were paid 75 US$ at the completion of the study (65 US$ for the scanning session and 10 US$ for the behavioral training session outside of the scanner). All of them gave informed consent before the beginning of the recordings. Also, they were told that they could quit the experiment anytime, if they so wished.

The subject data was split into a development set and a test set (see section 5.1.1). Out of the 16 subjects considered, the first nine were used in the development set, while the last seven subjects constituted the test set. A brief overview over the two data sets can be found in Table 3.4. More details about the subjects are provided in Table A.1 in the Appendix. Even though it would have been very interesting to interview the subjects on which strategy they used, among others, this was not checked with them. Since a relatively large amount of time had passed between the fMRI recordings and the analysis, asking the subjects in retrospect did not make sense anymore.

Table 3.4: Overview of subjects in the database.

			Gender	
Data Set	Number of Subjects	Mean Age	Female	Male
development set	9	22.89	4	5
test set	7	20.83	3	4
complete database	16	22.07	7	9

3.4.2 Learning, Training and Scanning Sessions

First, each subject has a learning and training session in which s/he memorizes the mappings first and then practices the task. Each participant completed his/her training within 24 hours before the scanning session, in the ACT-R lab at Carnegie Mellon University (outside of the scanner). For our modeling efforts, we are only concerned with the data collected in the scanning session, both with the behavioral and the fMRI data.

As a first step, the subject is asked to learn the nine letter-to-number mappings shown in Table 3.3 that have to be memorized for the task. After they are shown these mappings in a randomized order, they continue with a learning session during which they are prompted to type in the correct numbers when given a letter. The learning session consists of at least three blocks, depending on the performance of the subject. During each block, each mapping is tested until the subject has recalled it correctly once. Whenever the subject cannot recall a mapping correctly, the correct answer is shown. The prompts are given in a randomized order. As mentioned before, each subject has to complete at least three learning blocks. However, during the last block, the subject has to get every mapping right the first time. Thus, additional blocks will be added to the learning session until this is the case.

After the learning session, a training session is started where subjects get to practice responding in the four different conditions. They solve four blocks with 10 problems each, alternating between verbal and manual output:

$(L_1(manual), L_2(verbal), L_3(manual), L_4(verbal))$

After the training session is completed successfully, the actual scanning session is recorded, usually on the next day. Before the actual scanning session is started, structural scans are taken for registration with the reference brain and for checking for anomalies. The session itself consists of 10 blocks, each of which is composed of 10 problems. After the structural scans and at the end of each block, the experimenter talks to the participant to make sure that s/he is fine and to see whether s/he has any questions. Then, s/he is asked if s/he is ready to continue.

3.4.3 Recording of one Problem

The program used to present the problems and record the answers of the subjects was written by Shawn Betts. The problem itself is presented for as long as the subject needs to solve it, thus resulting in a variable numbers of scans. The layout of the user interface can be seen in Figure 3.5. These screenshots only show the central parts of the actual screens. The two fractions to be added are always shown in white font on the top of the screen. Depending on the step, boxes are shown underneath where responses have to be entered manually (see (a) or (b) in Figure 3.5). The number pad below these boxes is used for entering manual input.

Before a problem is presented, a crosshair is shown for two seconds. For the *manual* condition, this is followed by showing the problem and two boxes where the two first responses are supposed to be entered manually ($num1$, $num2$). These are located next to each other and underneath the corresponding fraction (see screen (a) in Figure 3.5). After having typed in the second digit, there is a delay of 250 ms where the subject sees his/her final input before it disappears. Once $num2$ is entered, two

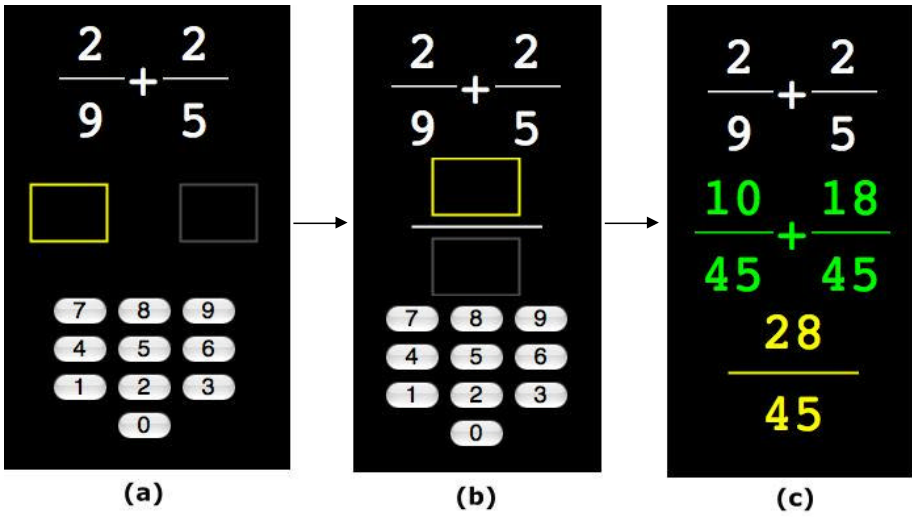

Figure 3.5: Sequence of screens from the user interface utilized for the experiment. Screen (a) is only shown in the manual condition, since *num1* and *num2* are supposed to be entered manually. Screen (b) and (c) are used in both the manual and the verbal condition. On screen (b), the resulting fraction has to be entered. Screen (c) shows the result in yellow font and the steps that were required to calculate it (in green or red, depending on whether the response of the subject was correct or not). Whenever a result has to be entered manually, the current box is highlighted with a yellow frame.

new boxes are shown where the resulting fraction should be entered (see screen (b) in Figure 3.5) which also disappear 250 ms after the last result is entered. Since the first two responses are spoken out loud in the verbal condition, the subjects are directed to this screen (b) directly after the crosshair. No feedback is provided for the two intermediate results ($num1$, $num2$), but once the final fraction is entered, the correct fraction is shown in yellow font. In addition, the steps are presented that had to be taken to calculate the resulting fraction, either in green or red font, depending on whether the problem was solved correctly or not (see screen (c) in Figure 3.5). This feedback is displayed for five seconds.

4. Computational Model

This chapter provides an explanation of how the task at hand was modeled within the framework of ACT-R 6.0. The purpose of the ACT-R model presented here is to simulate how a subject solves a given problem correctly . Therefore, this model is later only compared to behavioral and fMRI data that was recorded when subjects solved a problem correctly.

ACT-R 6.0 and thus also the model presented here are written in Common Lisp (for details on ACT-R 6.0, refer to the reference manual [28]). The model was implemented and tested using Allegro Common Lisp by Franz Inc. [1]. The code of the model will be provided on the website of John Anderson's laboratory [2].

The first part of this chapter outlines the information processing of the model: Section 4.1 explains the virtual interface used by the model (visual and manual), followed by details on how declarative knowledge is encoded in the model (section 4.2). The centerpiece of the chapter is section 4.3, where the information flow of the model is explained when it solves a given arithmetic problem. Limitations and restrictions of the model are pointed out in section 4.4.
The second part of the chapter is dedicated to the strategies of the model. First, an explanation of the different strategies is given that were implemented and compared with each other (see section 4.5; for a comparison with the subject data, refer to section 6.2). Second, the implications of these strategies for the activity of the modules and buffers is explained (section 4.6). The different courses of activity have far-reaching consequences, as this forms the basis of the BOLD predictions, as outlined in section 2.4.

4.1 Interface for the Model

The interface for the ACT-R model approximates the actual interface used for running the subjects (see Figure 3.5). A screenshot of the interface for the model can be seen in Figure 4.1. The problem is shown in the middle of the screen, as is the

[1] http://www.franz.com/
[2] http://act-r.psy.cmu.edu

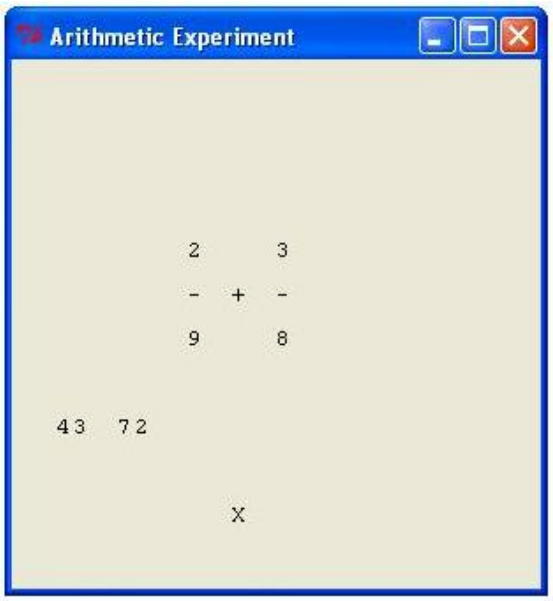

Figure 4.1: User interface for the ACT-R model.

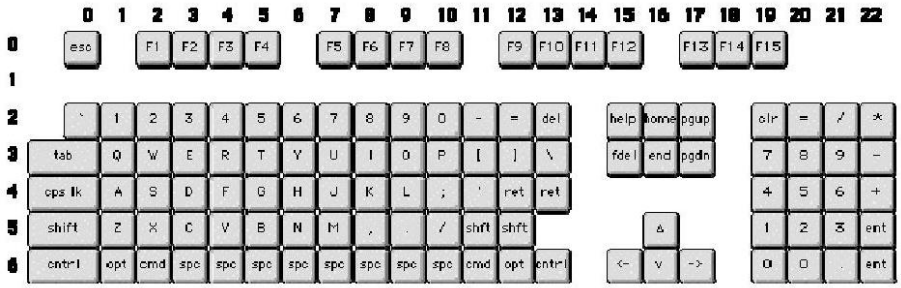

Figure 4.2: Virtual keyboard used by ACT-R (taken from [28]).

case for the subjects. As the results are entered manually, they appear in one row underneath the problem, from left to right. In the given example, the model was run in the verbal condition such that only 47 (*newNum*) and 72 (*newDenom*) had been entered manually.

In order to enter their manual response, the subjects need to use a mouse to click on the buttons of the NUM block on the screen. Thus, they need to turn their attention to this part of the screen. We did not focus on modeling how a result was keyed in exactly and had our model use a virtual keyboard for reporting results manually (see Figure 4.2). However, we still wanted to reflect that visual activity. Therefore, an X is shown on the screen of the interface instead of a NUM block to which the model turns its attention whenever a subject would look at the NUM block.

It should be noted that eye movement is not yet part of the default implementation of ACT-R (see section 2.3.2). This means that moving one's gaze to nearer objects takes the same amount of time as moving it to further objects. Thus, neither the order in which the model looks at information on the screen nor the location where the information is presented influences the time which the model spends gazing at the screen.

4.2 Declarative Knowledge for the Task

As explained in section 2.3.1, we distinguish between two types of knowledge in ACT-R, declarative and procedural knowledge which are encoded in chunks and production rules, respectively. Here, a description of declarative knowledge is given, whereas the production rules and the information flow are described in section 4.3.

Chunks

Declarative knowledge about basic arithmetic (numbers, additions and multiplications of single digit numbers) as well as the mappings is encoded in corresponding chunks of the following form:

```
(chunk-type mapping value string)
(chunk-type addition-fact arg1 arg2 sum)
(chunk-type mult-fact arg1 arg2 product)
(chunk-type number tens ones value string text)
```

Mapping chunks contain the relation of a letter (slot *string*) and its corresponding number (slot *value*), as it was learned by the subjects. These chunks need to be retrieved in the substitution condition where numerators are replaced by letters. Basic arithmetic facts, that is the addition and multiplication of single digit numbers, are encoded as chunks as well. Thus, the outcomes of these operations can simply be retrieved from memory and do not need to be calculated. Each of these chunks has two slots for arguments and one for the resulting outcome.

Finally, the concept of numbers is encoded in chunks which may be the least intuitive. For instance, the following chunk corresponds to number 16:

```
(sixteen isa number tens one ones six value 16 string "16"
text "sixteen")
```

When the model sees a visual stimulus such as '16' (*string*), it needs to retrieve which value corresponds to this visual representation (*value*). Furthermore, the model has to know which *tens* and *ones* this number has, both for mathematical operations and for manual output where the tens have to be entered first. Finally, the word 'sixteen' (*text*) is needed in the verbal condition as the model needs to know which letters to speak.

Knowledge about Additions

Knowledge about additions is modeled both in declarative and procedural memory. It is a widely reported empirical phenomena that two different strategies are mainly used for solving simple arithmetic problems: either computing the answer or directly retrieving the result. Generally, children choose their strategy based on the size of the given problem: for small problems, they will rather retrieve the result rather than compute it; for larger problems, the situation is vice versa [115]. In contrast to this, it was shown in [57] that younger and certainly older adults almost exclusively use the retrieval strategy (88% and 98%, respectively). [63] reported that adults used a multitude of procedures other than retrieval in order to solve addition and multiplication tasks. However, adults with a higher proficiency in mathematics, which is the case for our subjects as well, were more likely to use retrieval.

Thus, I assume for this model that the results are simply retrieved for operations involving single digit numbers, as mentioned above. Knowledge about double digit numbers, however, is encoded in form of production rules.

4.3 Information Flow in the Model

The way how the model solves the arithmetic task can be divided into several steps that occur repeatedly. These steps differ depending on the condition (*verbal-nosub*, *verbal-sub*, *manual-nosub*, *manual-sub*), while the order in which these steps are executed is determined by the answering strategy that is employed. In this section, the steps are explained that need to be taken. Section 4.5 contains details about the strategies employed.

The organization of the model is visualized in Figure 4.3 (a) for the basic strategy and in Figure 4.3 (b) for the convolution strategy. In the following, I will explain the flow of information along the lines of the basic strategy since it is more intuitive and straight-forward. The differences between the strategies will be explained in section 4.5.

Before explaining the organization of the model in more detail, I will give a brief description of Figure 4.3. The figure gives an overview over the basic strategy (left-hand side) and the convolution strategy (right-hand side). The column in the middle, separating both strategies and indicated by '*', shows whether the operations are the same for both strategies ('same', light gray) or not ('different', dark gray). For each strategy, the goal state is given in the leftmost columns of the figure while the operand that is currently in the center of attention is provided in the column next to it. 'Manual' and 'Verbal', the titles of the next two columns, indicate which output method is used. Each line in these two columns corresponds to a production rule that fires. The color coding works as follows: Yellow indicates visual perception and encoding of operators on the screen, green stands for calculations and orange for activity related to output.

It should be mentioned that the names of the production rules as they can be found in the code of the model differ slightly from those in Figure 4.3. Some of the rules in Figure 4.3 have been given the same name to illustrate that they serve the same purpose. For instance, the production rule *find* occurs four times, for each numerator and denominator to be retrieved from the screen. However, in reality, it corresponds to four slightly different production rules since the model looks for the numbers in four different spots. As explained in section 2.3.2, the current implementation of ACT-R does not model eye movements, so the model needs the same amount of time for finding the numbers on the screen, no matter where it has to look for information.

4.3.1 Use of Goal and Imaginal Buffer

The task is already subdivided by the four responses that need to be given (*num1*, *num2*, *newNum*, *newDenom*). Each of these need to be calculated and then reported, resulting in eight different subgoals, which are reflected in the use of the goal buffer. The chunks put in the *state* slot of the buffer are shown in the leftmost column of Figure 4.3. For instance, the chunk *num1-calculate* means that the aim of this step

(a) Basic Strategy

Goal state	Operand	Verbal	Manual	*
num1-calculate	n1	find attend perceive (retrieve-mapping) encode		same
	d2	find attend perceive encode		
	num1	calculate		different
num1-out	num1	respond-verbal	encode	
		attend-sound	attend-num-block respond-tens attend-output respond-ones attend-output	
num2-calculate	d1	find attend perceive encode		
	n2	find attend perceive (retrieve-mapping) encode		
	num2	calculate		
num2-out	num2	respond-verbal	encode	
		attend-sound	attend-num-block respond-tens attend-output respond-ones attend-output	
sum-calculate	newNum	start-addition add-ones (process-carry) add-tens add-tens-done		same
sum-out	newNum	process-sum attend-num-block respond-tens attend-output respond-ones attend-output		
denom-calculate	newDenom	calculate		
denom-out	newDenom	process attend-num-block respond-tens attend-output respond-ones attend-output		

(b) Convolution Strategy

Goal state	Operand	Verbal	Manual
num1-calculate	n1	find attend perceive (retrieve-mapping) encode	
	d2	find attend perceive encode	
	num1	calculate	
		remember	
num2-calculate	d1	find attend perceive encode	
	n2	find attend perceive (retrieve-mapping) encode	
	num2	calculate	
		remember verbal	remember manual
num1-out	num1	respond-verbal	attend-num-block
		attend-sound	respond-tens attend-output respond-ones attend-output
num2-out	num2	respond-verbal	respond-tens
		attend-sound	attend-output respond-ones attend-output
sum-calculate	newNum	start-addition add-ones (process-carry) add-tens add-tens-done	
sum-out	newNum	process-sum attend-num-block respond-tens attend-output respond-ones attend-output	
denom-calculate	newDenom	calculate	
denom-out	newDenom	process attend-num-block respond-tens attend-output respond-ones attend-output	

Figure 4.3: Organization of the ACT-R model for both the basic and the convolution strategy (left-hand side and right-hand side, respectively). For more details on the strategies, see section 4.5. The column in the middle (indicated by '*') shows whether the operations are the same for the two strategies (light gray) or different (dark gray). For each strategy, the goal state and the current operand are given in the two columns on the left. Each line in the next two columns corresponds to a production rule that fires; if a rule is written in white font, the goal state is changed. 'Manual' and 'Verbal' indicate which output method is used.

is to calculate the first numerator, while the chunk *num1-out* reflects the goal to report it. Since the notation in the code differs slightly from the definitions given here, the name of the operand according to the notation used in this thesis is given in the column called 'Operand'. If a production rule results in a change of the goal state, it is printed in white font in Figure 4.3.

Although not shown in the figure, it is important to note that the imaginal buffer is effectively used as working memory for this task, that is for remembering intermediate results as well as information found on the screen. For this purpose, a chunk is placed in the imaginal buffer that is updated as new information needs to be remembered. This process involves a time cost of 200 ms by default, depending on the imaginal delay parameter (*:imaginal-delay*). For experiments with the value of the imaginal delay parameter, refer to section 6.3. Only when the model decides that a piece of information is not relevant anymore, that information is deleted without a time cost (issuing an *=imaginal* request).

4.3.2 Visual Perception and Encoding of Operators

The first step of the model and the subjects is to find the information on the screen that is relevant for calculating $num1$ and encode it (yellow block in Figure 4.3). This step is the same for both the manual and the verbal condition. As can be assumed after some practice, the model knows already where to look at the screen to find the information it needs. Thus, it sends a request to the visual-location buffer with the approximate coordinates of where it expects to find the information (*find*). Then, it moves its attention to the item found in this spot (*attend*) and sends a request to declarative memory to find out what this item is (*perceive*). If it is is a number, it stores this information in the imaginal buffer (*encode*). If it is a mapping, one more step is required as the corresponding chunk with information about this mapping needs to be retrieved from declarative memory (*retrieve-mapping*) before it can be encoded and remembered. This can only occur for numerators, as this is the only place where mappings can occur.
Given a fraction $\frac{n1}{d1} + \frac{n2}{d2}$, the model looks first at $n1$ and then at $d2$ for the calculation of $num1 = n1 * d2$. As mentioned before, the order of where to look first ($n1$ or $d2$) does not influence the behavior of the model as ACT-R does not incorporate eye movement in its current version.

4.3.3 Calculation of Intermediate Results and Responses

Once the model has encoded both operands from the screen, it retrieves a chunk from working memory that contains the result of the multiplication of these two numbers, a so-called multiplication-fact. Once the information is available after the necessary retrieval time (according to equation (2.8)), the corresponding response is given. For this, it is necessary to retrieve a *number* chunk from declarative memory that contains information such as which word corresponds to the symbol (for *verbal* responses) and which tens and ones the number has (for *manual* responses).

In the verbal condition, only one production rule is necessary to model the response (*respond-verbal*) as the result only has to be prepared and spoken. However, one more production rules is needed to capture that the model hears itself speaking (*attend-sound*).

The manual condition involves more steps and production rules. First, the *number* chunk that was retrieved needs to be remembered (*encode*). The model only includes a crude approximation of the fact that subjects have to look at the NUM block for finding the correct buttons, essentially by turning its attention to an X on the screen (see Figure 4.1). This is ensured by the production rule *attend-num-block*. Since all results are double-digit numbers, two buttons need to be pressed (*respond-tens*, *respond-ones*). Instead of clicking on buttons, the model presses keys on a keyboard (see section 4.1). We expect subjects to get visual feedback from the screen after entering a response manually. Thus, the model turns its attention to the screen as soon as something new appears there which is only the case when a result is keyed in (*attend-output*).

The rules governing perception of the verbal and manual output (*attend-sound*, *attend-output*) have a special role, since they only fire when the model perceives the output, i.e. when it hears itself speaking or sees a response appear on the screen that was keyed in. Thus, although Figure 4.3 suggests that they fire right after the rule for the corresponding response, this does not have to be the case. The model continues with the task if a response takes a while to be perceived. For this purpose, the utility of the perception rules is set to a high value, such that these rules fire as soon as an output is detected, independent of whether another rule (for continuing with the task) applies as well.

The steps described so far model the visual perception and encoding of operators with an ensuing calculation and response. According to the basic strategy, these are the same for *num1* and *num2*. After these four blocks are finished (*num1-calculate*, *num1-out*, *num2-calculate*, *num2-out*), the model moves on to the next subtask, the calculation of the new numerator.

4.3.4 Calculation of the Final Result

The calculation of the new numerator (*newNum*) requires the addition of two double digit numbers, *num1* and *num2*. Since the result cannot simply be retrieved as is the case for the single digit additions, the model has to go through a couple of steps for performing the calculation. At this point, it has both *num1* and *num2* in its imaginal buffer, since it remembers them from the previous calculations. First, the ones have to be added and then the tens which means that the model needs to retrieve the corresponding addition-facts (*start-addition*, *add-ones*, *add-tens*). As explained in section 3.3.3, some problems involve a carry in the addition while others do not. The difference between these two cases involves only one more production rule (*process-carry*) and the retrieval of one more addition-fact if the carry occurs. The result of the addition is again stored in the imaginal buffer (*process-sum*) before it is reported manually just as described for the first two results.

Then, the new denominator *newDenom* is calculated by retrieving the multiplication-fact with the result of the calculation $d1 * d2$, both of which are still memorized by the model and can thus be found in the imaginal buffer. One might argue that it seems more plausible to look at the screen for this information. This was implemented in an alternative model. However, it did not fit the data such that we kept the current strategy. As usual, a number chunk needs to be retrieved which gives meaning to the retrieved result (*calculate-newDenom*). The result is then stored in the imaginal buffer (*process-newDenom*) which corresponds to the *encode* production rules for

num1 and *num2*). Finally, this result is keyed in as well with the standard output procedure described in section 4.3.3.

4.4 Limitations of the Model

The model described here comes with a number of limitations that should be pointed out. First, as mentioned before, we did not try to model those cases when subjects fail to solve the task correctly. As can be seen in Table 4.1, most problems were solved correctly by the subjects (93.75% correct, averaged over conditions and subjects). Data on the performance of individual subjects is provided in Table A.2. Trials were subjects failed to solve the given problem correctly were not considered for comparison with the model. As the interface did not allow for correcting answers that were entered wrong unintentionally, these trials were excluded as well as they also fell into the category of problems solved incorrectly. Thus, the model does not reflect either of these possibilities.

Table 4.1: Overview of the percentage of trials solved correctly by the subjects ('-s' stands for 'substitution', '-ns' for 'no substitution').

	Percent correct				
Data Set	Total	Verbal-ns	Verbal-s	Manual-ns	Manual-s
development set	94.33	95.56	90.22	93.78	97.78
test set	93.00	91.43	90.86	96.00	93.71
complete database	93.75	93.75	90.50	94.75	96.00

Second, the model does not incorporate retrieval failures or the fact that subjects might forget intermediate results needed to proceed (*num1*, *num2*). Subjects might encounter these errors but still complete the task correctly. Even though this probably occurred in the experiments, we capture this partially by introducing variability in retrieval times (see section 6.3) and using relatively low base-level activation for the chunks.

Furthermore, it can be argued that the retrieval of mappings is likely to become increasingly easier over time which would lead to a speedup in retrieval time for mappings. In contrast to this, it could be expected that subjects become tired or bored of the task over time, leading to longer retrieval times as well as to a general deceleration in performing mathematical operations. Neither of these effects are directly incorporated in the model. However, the focus of this work is on predicting the BOLD responses. For this purpose, we only compare data from the model to experimental data that is averaged over subjects and trials. Thus, any changes of retrieval times or problem solving time are averaged out anyways. The model is then fitted to the averaged data.

Moreover, it should also be noted that the type of movement performed by the subjects and the model are different (mouse movement versus key strokes), as well as the movement in space since the layout of the keys differed (NUM block versus linearly aligned number keys). This was still acceptable as this aspect was not in the focus of attention.

Finally, the production rules of the model only account for reporting double-digit numbers. This could be easily extended to reporting numbers which have fewer or more digits. However, this was not necessary since the problems considered for this experiment were chosen such that only double-digit numbers occurred (see section 3.3.3).

It should be pointed out that the purpose of the model is to give an idea about the approximate times at which the different modules are active and the duration of their activity. Essentially, it is used as a tool for predicting the BOLD response rather than for predicting or modeling the behavioral data exactly. Thus, the shortcomings and restrictions mentioned here can be tolerated, as they do not limit the purpose of the model.

4.5 Answering Strategies

In order to solve the question at hand, several strategies seem plausible. Also, the high variability in response times, both within and between subjects indicates that different strategies are employed by the subjects (see section 5.2). There are a number of potential strategic choices which are listed in the following (see Table 4.2) and explained in more detail in the following sections.

Table 4.2: Potential strategies considered for the model.

Question	Strategy	Meaning
Point of time when $num1$ is reported	*basic*	Report $num1$ as soon as it is calculated.
	convolution	Calculate $num1$ and $num2$, then report both.
	mixture	Utilize *basic* or *convolution* strategy with a certain probability.
Visual activity before calculating $newDenom$	*re-encode*	Look at the screen to get the information for calculating $newDenom$.
	encode-once	Remember the information needed for calculating $newDenom$.

It is important to note that the interface used in the experiment triggered certain answering strategies and made others rather unlikely, since they required more cognitive steps. This was done on purpose with the aim of acquiring data that was comparable (for more details, refer to section 3.3.2). Subjects were not trained on a certain strategy in the training session, as the interface already restricted their answering strategies. As argued in section 3.3.2, the interface guided subjects through the four steps, leaving only a few answering strategies open, mainly the question of when to report the results.

Other strategies within the restrictions of the interface could be thought of. However, we assume that we have captured the main strategies in the model (described in the following). This is supported by the fact that the model manages to capture the behavioral data well, not only of the development set but also of the previously unseen test set (see section 6.2). It can generally be argued that alternative models

could be produced within the architecture of ACT-R. However, the ACT theory and the computational framework of ACT-R allow only for a small number of plausible models. One could use the BOLD responses to select the most appropriate model or strategy in retrospect, based on the predictions of the BOLD responses by the different models. This is certainly a possibility, but left for future work since this thesis was concerned with the prediction of the BOLD responses without prior knowledge of this data. Furthermore, it should be taken into account that the subject data shows artifacts in the *verbal* conditions such that it would only make sense to adjust the model to the data collected in the *manual* conditions, if at all.

4.5.1 Point of Time when *num1* is Reported

Basic Strategy

The most straight-forward approach to solving a given problem is to adhere to the following scheme, that is to iterate over the list of answers to be given and report each answer as soon as it is calculated (see Algorithm 1 and Figure 4.3 (a)).

Algorithm 1 basic strategy

for all x $(x \in (num1, num2, newNum, newDenom))$ **do**
  calculate(x)
  report(x)
end for

However, the results *num1* and *num2* are needed for calculating *newNum*. For the manual condition, each of them will only stay on the screen for a short amount of time after the answer has been entered. Thus, when using the basic strategy, the subject/model has to keep $num1$ in mind while calculating $num2$. After having reported $num2$, both $num1$ and $num2$ have to be retrieved in order to calculate $newNum$. This additional challenge can be altered by using what is called convolution strategy in the remainder of this thesis.

Convolution Strategy

Instead of reporting *num1* right when it is calculated, this strategy suggests to keep the result of *num1* in mind, calculate *num2* and then report both of them after each other. This corresponds to the scheme outlined in Algorithm 2 and is visualized in Figure 4.3 (b).

Algorithm 2 convolution strategy

calculate($num1$)
calculate($num2$)
report($num1$)
report($num2$)
for all x $(x \in (newNum, newDenom))$ **do**
  calculate(x)
  report(x)
end for

Note that the difference between basic and convolution strategy concerns only the first two answers (*num1*, *num2*); the last two answers are given in the same way (*newNum*, *newDenom*).

Mixture Model

It seems reasonable to assume that strategies vary between and within subjects, such that both strategies are employed. Therefore, we emulated a mixture model that would choose either strategy with a certain probability.

Definition. A mixture model is defined as $M(x, y)$, where x is the probability that the basic strategy is chosen and y is the probability that the convolution strategy is employed to solve the problem (x, y being floating-point numbers with $x, y \in [0, ..., 1]$ and $x + y = 1$).

For instance, the model $M(\frac{2}{3}, \frac{1}{3})$ would use the basic strategy with the probability $\frac{2}{3}$ and utilize the convolution strategy with probability $\frac{1}{3}$.

4.5.2 Visual Activity before Calculating *newDenom*

Encode-once Strategy

The calculation of the denominator of the resulting fraction (*newDenom*) requires the multiplication of the denominators of the two given fractions ($newDenom = d1 * d2$). Both are available on the screen during the whole time needed to solve the problem. Since the subject has processed them previously for calculating *num1* and *num2*, it seems plausible that the subject keeps them in mind when s/he processes the visual information the first time s/he encounters it. We refer to this strategy as *encode-once*, since the information is encoded only once and then retrieved later from the imaginal module for the calculation of *newDenom*.

Re-encode Strategy

However, the *encode-once* strategy requires to keep two additional pieces of information available in working memory, leading to a higher workload. So we assume a different strategy could be used, according to which the subject does not try to remember the information, but looks at the screen another time to retrieve these pieces of information when they are needed. This is referred to as *re-encode* strategy.

For simplicity, only the encode-once strategy is shown in the visualizations of the information flow in the model (Figure 4.3).

4.6 Activity of Modules and Buffers

The activity of the modules and thus the buffers depends both on the strategy of the model and the condition (i.e. *verbal-nosub*, *verbal-sub*, *manual-nosub*, *manual-sub*). The conditions determine the amount of activation whereas the strategies influence its timing. This is visualized in Figure 4.4. For each of the eight major steps identified in section 4.3, the figure shows how often the buffer of a given module is engaged. For the retrieval buffer, it also provides information about how often a

Step	retrieval				goal	imaginal	manual	visual	vocal	aural	manual-nosub	manual-sub	verbal-nosub	verbal-sub
	number	mapping	add-fact	mult-fact										
calculate num1	3	1		1	1	3		2						
calculate num2	3	1		1	1	3		2						
output num1 (manual)	2				1		2	3						
output num2 (manual)	2				1		2	3						
output num1 (verbal)					1				1	1				
output num2 (verbal)					1				1	1				
calculate newNum	1		2 (carry) or 3 (no carry)		1	4								
output newNum (manual)	2				1		2	3						
calculate newDenom	2			1	1	1								
output newDenom (manual)	1				1		2	3						

Figure 4.4: Buffer use and retrievals for each of the eight major steps. The steps involving manual output are colored yellow, those for verbal output in orange. Blue font and boxes indicates that the step is only required in the substitution condition. Additions-facts are abbreviated with 'add-facts', multiplication-facts with 'mult-facts'.

request for certain chunk types is issued. The effect of these retrievals on response times is discussed in more detail in section 6.2.1.

An example of the activity of the modules over time can be seen in the trace of the model (Figure 4.5). It visualizes the engagement of the modules for the *verbal-nosub* condition, when the basic strategy is employed for solving the problem $\frac{2}{9} + \frac{3}{8}$. Each column corresponds to the activity of a module and is color coded: declarative/retrieval in red, imaginal in blue, manual in green, goal in black, vocal in magenta, procedural in brown and visual in light green. The longer the duration of the activity of a given module, the longer the block, as the timeline is given on the y-axis. If the module is only active for an instant, a horizontal line is shown instead of a block.

For instance, the leftmost column shows retrieval activity (red blocks), that is the retrieval of mappings, numbers, addition-facts and multiplication-facts. In this example, the longest blocks correspond to the retrieval of multiplication-facts, the shorter ones to that of addition-facts and the very short ones that only appear as a horizontal line are related to the retrieval of numbers (no mappings needed to be retrieved for this problem). The difference in retrieval times is explained in section 6.2.1.

Based on the trace, the activity and information flow of the model (see section 4.3) can be monitored, as each block visualizes activity of the corresponding module: first, a new goal state is set (*num1-calculate*, black block). This evokes visual activity, as two numbers $(n1, d2)$ have to be found on the screen (two green blocks). Each

of these is remembered temporarily, leading twice to activity in the imaginal buffer (two blue blocks). The long activity of the retrieval buffer is due to the retrieval of a multiplication-fact ($n1 * d2$, long red block). Finally, the vocal module becomes active when the response is given (yellow block), which leads to the engagement of the aural module shortly afterwards (block in magenta). The same pattern is then repeated for calculating the second numerator. Since the last two answers ($newNum$, $newDenom$) are given manually, the manual module becomes active for each of these responses instead of the vocal and aural module. The manual module becomes active twice since two digits need to be entered for each response (two green blocks). The procedural module (indicated by 'production' in the figure) shows activity whenever a new rule fires.

The trace also gives a visual representation of the demand function $D(t)$ that shows the activity of each buffer, as described in section 2.4.2 and shown in Figure 2.9. This demand function is then convolved with the hemodynamic function in order to calculate the prediction of the BOLD response for each module, according to equation (2.13).

For the given problem, the model needed about 16 seconds, based on the parameter settings described in section 6.2.1. The time course of activation is mainly determined by the structure of the model, though, that is the number and nature of the rules which fire when the model solves a problem. As pointed out in section 2.3.2, the modules work in parallel, unless a module has to wait for information from another module.

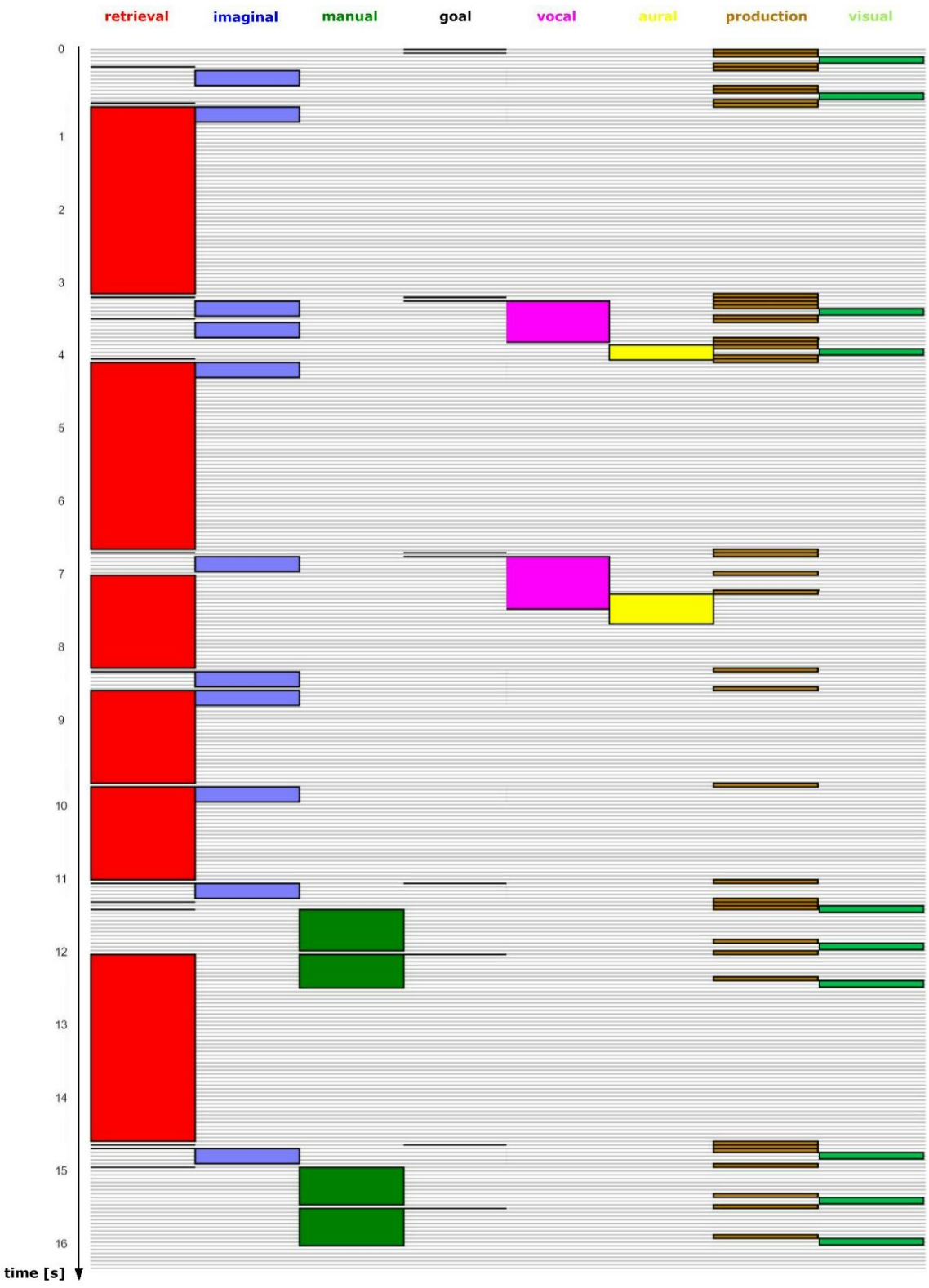

Figure 4.5: Trace for the *verbal-nosub* condition (basic strategy, problem $\frac{2}{9}+\frac{3}{8}$). The activity of the different buffers is shown over time (vertical time line, time is given in seconds). Each column corresponds to the activity of a module and is color-coded. The longer the duration of the activity of a given module, the longer the block. If the module is only active for an instant, a horizontal line is shown instead of a block.

5. Subject Data

In this chapter, the subject data will be discussed briefly. First, some details are provided about the database, which was split into a development set and a test set (section 5.1). Then, a brief interpretation of the behavioral data is given in section 5.2. Finally, potential sources of variability are discussed (section 5.3).

5.1 Database

5.1.1 Development Set and Test Set

Data was obtained from 40 subjects in total. However, the first half was partly a pilot study and was not recorded consistently such that this data could not be compared. Out of the 21 subjects considered for the study (IDs 20 to 40), the data of five had to be excluded since the data was incomplete due to technical issues: For subjects 20 and 31, the verbal protocols were not recorded correctly, while the fMRI data was incomplete for subjects 22, 33 and 38.

Thus, the database that was used here includes the data of 16 subjects. The subject data was split into a development set and a test set. The development set was used to fit the model to the behavioral data, specifically to the timing of the responses given and the variability in problem solving time. Eventually, the model was validated on the complete database (development set and test set), as the model's prediction of the BOLD response was compared with the BOLD prediction shown by the subjects. A detailed discussion of the approach taken can be found in chapter 6.

The development set consists of the data of the first nine subjects in the database, while the last seven subjects formed the test set. The data was split in this way for practical reasons. Since the modeling efforts ran in parallel to the scanning sessions, the data of the first subjects available was used for data fitting. The test set was intended to be bigger, but more subjects than expected had to be excluded due to technical issues during the recording (as mentioned above, the fMRI data was incomplete).

5.1.2 Datapoints Excluded

Out of the 100 trials collected from each subject, some had to be excluded. On the one hand, only those problems were considered that were solved correctly. On the other hand, some problems were identified as outliers and therefore not considered. An overview of the resulting number of data points for each condition and each subject can be found in Table 5.1.

Table 5.1: Overview of the subject database. For each subject, the number of correctly solved problems (out of 100) are given as well as the number of outliers. Only those trials which were solved correctly and not classified as outliers were included in the database The conditions are abbreviated as follows: 'V'= verbal, 'M'= manual, '-s'= substitution, '-ns'= 'no substitution'. 'Dev' stands for development set, 'test' for test set.

				Database [# trials]				
Set	Subject ID	Correct	Outliers	Total	V-ns	V-s	M-ns	M-s
dev	21	96	22	74	21	17	21	15
dev	23	88	1	87	23	17	22	25
dev	24	98	1	97	24	23	25	25
dev	25	95	0	95	24	22	24	25
dev	26	89	9	80	18	18	21	23
dev	27	96	4	92	23	24	23	22
dev	28	99	0	99	24	25	25	25
dev	29	92	0	92	24	23	21	24
dev	30	96	3	93	23	25	21	24
test	32	93	0	93	24	24	24	21
test	34	94	1	93	21	24	24	24
test	35	92	3	89	24	20	20	25
test	36	93	1	92	22	23	25	22
test	37	88	13	75	19	19	16	21
test	39	98	2	96	25	23	24	24
test	40	93	2	91	21	23	23	24
development set		849	40	809	204	194	203	208
test set		651	22	629	156	156	156	161
complete database		1500	62	1438	360	350	359	369

Incorrectly solved Problems

As mentioned before, we only modeled how the arithmetic task is solved correctly. Thus, only those trials were used which were completed correctly by the subjects. The percentages for each subject can be found in Table A.1. On average, the subjects solved 93.75% of the problems correctly in the scanning session and 86.56% in the training session. The training session was not considered further. Development set and test set differed only slightly in the amount of problems that were solved correctly in the scanning session (94.33% versus 93.00%). 51 data points were excluded from the development set and 49 from the test set as the problem was not

solved correctly, leaving 849 data points in the development set and 651 in the test set (see Table 5.1).

Outliers

Some of the data points had extraordinarily long problem solving times that clearly showed that something had gone wrong during these trials, even though the result was correct. For instance, a subject might have forgotten the first result ($num1$) which is needed to calculate the third result ($newNum$) and was thus required to recalculate this intermediate result to proceed. Since the model presented here does not attempt to capture these mistakes, these outliers are not considered. For the remainder of this thesis, an outlier is defined as follows.

Definition. A data point i in the subject database is defined as being an outlier by having a problem solving time T_i which fulfills the following inequality:

$$T_i \geq \text{cut-off}(c)$$

where c is the condition in which the data was recorded
($c \in$ (*verbal-nosub*, *verbal-sub*, *manual-nosub*, *manual-sub*))

Table 5.2: Cut-off values for the problem solving time. $\bar{T}$(c) refers to the mean problem solving time in a given condition c for the complete database. A detailed breakdown of the means and variances of the different data sets can be found in Table A.4.

Condition c	cut-off(c) [s]	$\bar{T}$(c) [s]
verbal-nosub	23	12.25
verbal-sub	27	14.63
manual-nosub	25	14.01
manual-sub	32	16.50

The cut-off values for each condition are defined in Table 5.2. They correspond roughly to $2 * variance(c) + mean(c)$ for each condition c, based on the complete dataset. Table A.3 shows how many outliers were found per subject. Table 5.1 shows the effect on the database. As can be seen, 62 data points in the complete dataset were identified as outliers, leaving 1438 data points to be considered for the comparison with the model. The term 'data point' here refers to a trial, i.e. a recording of behavioral and fMRI data while a subject solved one problem.

5.2 Interpretation of the Behavioral Data

In the following, the behavioral data is interpreted briefly. Although the data is split into a development set and a test set for comparison with the model, the complete database is considered here.

Figure 5.2 shows the mean of the absolute and the relative response times (plot (a) and (b), respectively). Absolute time is defined as the time that has elapsed since stimulus onset, whereas relative time is the time since the last response was given (for the exact definition, refer to section 3.3.5).

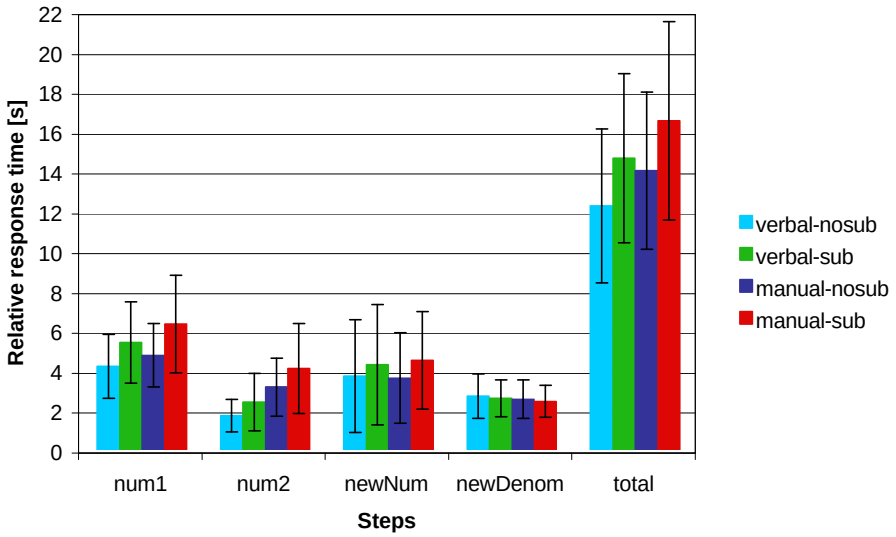

Figure 5.1: Mean relative response time and variance of the subjects in the database.

Absolute Times

With regard to the absolute times, two effects are obvious in the data (Figure 5.2, plot (a)). First, having to retrieve mappings from memory (substitution conditions) slows subjects down as could be expected. The effect seems independent of the output modality, although the difference is slightly bigger for the manual condition (2.2 s difference on average for manual and 2 s for verbal). Second, it takes subjects longer to respond in the manual condition than in the verbal condition and slightly more so when a substitution is involved. As can be seen, subjects are fastest in the *verbal-nosub* and slowest in the *manual-sub* condition, while their absolute response times in the conditions *manual-nosub* and *verbal-sub* are almost indistinguishable.

Relative Times

The relative response times give more insights into the differences between the four steps to be taken in this task. With regard to the the first two responses, reporting *num1* obviously takes more time than reporting *num2* in all conditions, although this is more pronounced for the verbal conditions. This results in a dip in the plot of the relative times (plot (b) in Figure 5.1). It is important to note that both calculations require the exact same steps, such that the faster response for *num2* cannot be attributed to a faster calculation. The drop in response time can probably partly be related to the shorter preparation time needed when an answer has just been given in the same output modality. However, the difference between the response times seems too big to just be explained by this (ranging from 1.6 s in the *manual-nosub* conditon to 3 s in the *verbal-sub* condition).

Rather, it can be hypothesized that the subjects use a different answering strategy for these two responses. It seems reasonable to assume that subjects may withhold the first answer (*num1*) until the second answer is calculated (*num2*). Then, they are reported one after each other, resulting in a long response time for *num1* (during which both results are calculated) and a short response time for *num2* as the result is already known. The reasoning behind this is explained in section 4.5.1 where a more formal description of this strategy can be found (Algorithm 2). In our model, we

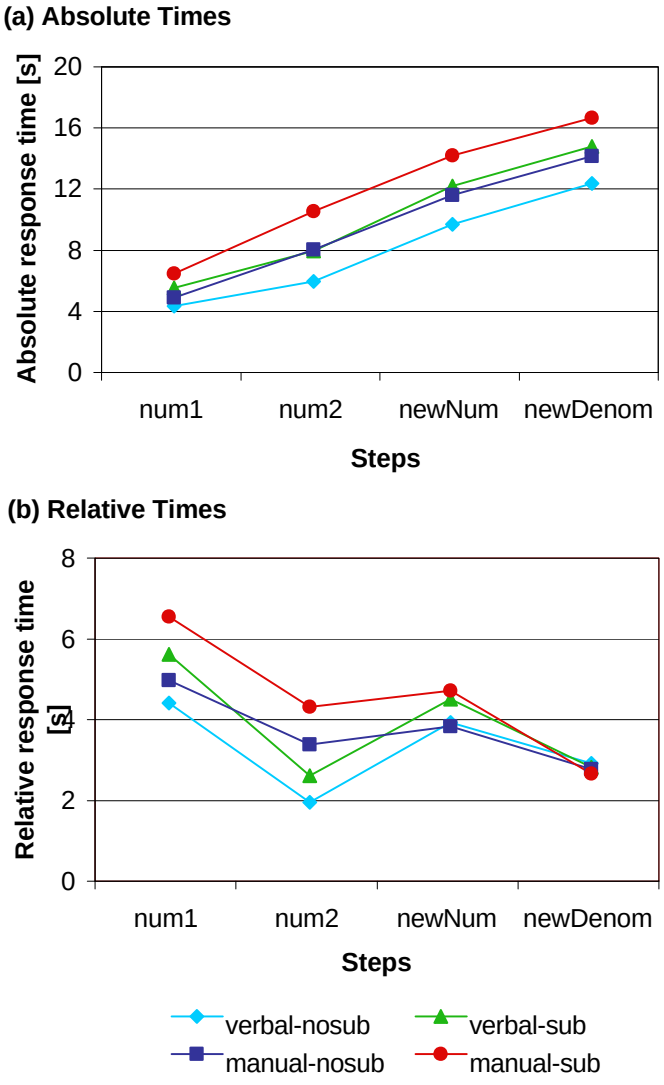

Figure 5.2: Mean response times for the subjects in the database, depending on the condition. Both the absolute response times (a) and the relative response times are given (b).

refer to this strategy as *convolution* strategy whereas the straight-forward approach of reporting results right away is called *basic* strategy.

The last two responses (*newNum*, *newDenom*) require the same steps for either condition, since they have to be entered manually and also do not involve any mapping. However, the relative time needed for reporting *newNum* differs between conditions. As can be seen in Figure 5.1, the response time for this step also shows a high variability in the subject data. In order to calculate *newNum*, subjects need to add the two previous results (*num1* and *num2*) which need to be kept in mind for this purpose. The high variability in response time seems to indicate that subjects sometimes forget the previous results and have to re-calculate them. This is supported by the fact that subjects need more time to respond in the substitution condition, although no substitution is involved in this step (it is, however, when a previous result needs to be re-calculated).

In contrast to this, the subjects show almost no difference in the relative response time for *newDenom*. Also, the variability in the subject data is the lowest for this step. This can be attributed to the fact that this is the easiest step to calculate in the task. However, this also supports the hypothesis that subjects do not always report *num1* right away. If that was the case, the response times for the *manual-nosub* condition would be comparable since the same steps are required (multiplication of two numbers on the screen and manual response). However, the mean of the relative response for $num1$ is 4.78 s for *manual-nosub*, whereas it is 2.56 s for *newDenom* (averaged over all conditions).

5.3 Sources of Variability

We assume a number of different sources for variability in the subject data as outlined in the following, specifically variability due to

- the range of different problems to be solved,
- a variety of solution strategies,
- non-deterministic retrieval times for task-related knowledge.

A certain variability in response times is introduced by having a subject solve 100 different problems (the problem generation is described in section 3.3.3). These different problems will lead to slightly different response times and BOLD responses for two reasons. First, different responses require different mouse movements and/or speaking slightly longer or shorter words. Second, the calculation of the new numerator as the sum of the two first responses ($newNum = num1 + num2$) requires a carry for some problems, which is not the case for others. In the latter case, the response times as well as the problem solving time T will be slightly shorter.

A more potent source of variability might be the strategy that subjects choose for solving the problem. The data indicates that subjects do not always seem to utilize the same strategy, resulting in a shift of the response times for the intermediate results while the problem solving time remains similar. Figuratively speaking, the plots of the absolute times show a different pattern when a different strategy is

applied. For our model, we assume two different strategies and a mixture strategy, as described in section 4.5.

Finally, the subjects need to retrieve different pieces of knowledge from their memory which fall in the following three categories:

- letter to number mappings that were previously learned (e.g. '$t = 2$'),
- facts about multiplication (e.g. '$2 * 8 = 16$'),
- facts about addition (e.g. '$6 + 7 = 13$').

Although it can be assumed that retrieving an item from each category is linked to a certain standard retrieval time, this time will probably vary even for the same item both within and between subjects. For instance, a subject might have trouble remembering which number was linked to the letter t, while remembering the mapping $t = 2$ instantly the next time s/he encounters it.

Besides the factors discussed here, it should be noted that even within the exact same task, the hemodynamic response can show substantial variability both between and across subjects (refer to section 2.1.2 for more details).

6. Model Refinement and Evaluation

In this chapter, the steps are explained that were taken in order to refine and evaluate the model, based on the subject data. These steps are summarized in Table 6.1 and are explained in detail in the following sections. However, to give an overview of the methodology used, the steps are outlined first in section 6.1. Thus, this section also serves as an overview over this chapter.

Table 6.1: Methodology used for adjusting and evaluating the model (absT = absolute response time, relT = relative response time, T = problem solving time, dev = development set).

			Data sets used for	
Factor	data type	comparison of	data fitting	evaluation
timing	behavioral (mean)	absT, relT	dev	test
variability	behavioral (distribution)	T	dev	test
brain activity	fMRI (mean)	BOLD response	-	all

6.1 Overview of the Methodology

Even though I distinguish between two data types in the following (behavioral and fMRI), it should be clear that these are two parts of the same recording: While a subject solves a task in the scanner, both his/her behavioral data (i.e. the response times) and his/her brain activity are recorded at the same time.

As a first step not mentioned in the table, the ACT-R model was developed based on what seemed reasonable assumptions and without consulting the subject data (for a description of the model, refer to chapter 4). This first version of the model was deterministic in its temporal behavior and a certain variability only occurred

since it was run on 110 different problems (for the problem generation, refer to section 3.3.3).

Then, the timing of the responses and the variability of the model were adjusted based on the behavioral data. For this purpose, the subject data was split into a development set for data fitting and a test set for evaluation (9 and 7 subjects, respectively).

First, the timing of the model was adjusted (see section 6.2). For this purpose, I compared the absolute and relative response times that the model needed for the four different responses (*num1*, *num2*, *newNum*, *newEnum*) with the corresponding mean response times of the subjects in the development set. Based on this comparison, I adjusted the base-level activation of the chunks in the declarative memory of the model and determined which strategies were most appropriate for further use.

As a next step, I introduced variability in the model (refer to section 6.3). To be more precise, the aim was to match the variability that subjects showed in their problem solving time T. While mean response times were used for the first step, this step required matching the distribution of the individual times in the development set. For this purpose, I experimented with parameters that determine the time needed for the retrieval of knowledge and a parameter that determines the timewise behavior of the imaginal module.

Once these two steps were completed, the resulting model was evaluated on the data of the subjects in the test set. Finally, the model was used as a basis to predict the BOLD response for different regions of interest. This prediction was then compared with the actual brain imaging data of the subjects, this time using the whole dataset (development and test set combined). Again, the data was averaged per condition over subjects and model runs, respectively, using a method called event-locked averaging (see section 6.4). It should be pointed out that the first two factors, the response times and the variability, mainly serve as a tool for predicting the third factor, the brain activity that subjects show while solving the arithmetic task. The comparison of the prediction of the model and the BOLD responses shown by the subjects is provided section 7.1 of the next chapter.

6.2 Timing of the four Responses

As a first step, the mean absolute and relative response times that the model needed for the four different responses (*num1*, *num2*, *newNum*, *newDenom*) were compared with the corresponding mean response times of the subjects in the development set. For this purpose, the model was run on all 110 problems and the mean of the response times was taken. The model and subject data were compared based on the correlation coefficient and the mean deviation. In order to fit the response times of the subjects, two steps were taken. On the one hand, I adjusted the base-level activation of the chunks containing facts about multiplication, addition or the mappings (for more details on these chunks, refer to section 4.2). On the other hand, I determined which strategies fit best (for more details on strategies, refer to section 4.5).

6.2.1 Base-Level Activation and Retrieval Times

The types of chunks that encode declarative knowledge in the model developed for this thesis are explained in section 4.2 (numbers, mappings, addition-facts, multiplication-facts). As explained in section 2.3.3, the higher the base-level activation of a chunk, the shorter its retrieval time which is calculated according to equation (2.8).

Moreover, it is important to note which response times are influenced by the retrieval times of which chunks. As can be seen in Figure 4.4, each response time depends on the retrieval of number chunks. In contrast, addition-facts only need to be retrieved for the calculation of *newNum* (with a stronger impact on problems were a carry is required since this requires the retrieval of an additional addition fact). The retrieval time for multiplication-facts affects the other three responses (*num1*, *num2*, *newDenom*). Mappings are only retrieved in the substitution conditions and their retrieval influences only the response times of the first two responses (*num1*, *num2*).

The values for the base-level activation were determined empirically, trying to match the response times of the subjects in the development set. At the same time, I paid attention to stay within reasonable bounds, depending on how easy I assumed that certain facts to be retrieved by subjects. The values that worked best for fitting the current data can be found in Table 6.3. As a starting point, the base-level activation values of a previous simple model of the addition of fractions were used that I had developed for a different experiment and fitted to a different dataset (see Table 6.2). Also, the same latency factor was used as in this model (: $lf = 3.05$). Given these two parameters, the retrieval times are calculated based on equation (2.8).

Table 6.2: Values of the base-level activation for the conversion task (previous experiment, latency factor $lf = 3.05$).

Type of chunk	base-level activation	retrieval time [sec]
number	9	0
addition-fact	0.5	≈ 2
multiplication-fact	0.05	≈ 3

Table 6.3: Values of the base-level activation in the arithmetic task that worked best for matching the response times of the subjects in the development set (latency factor $lf = 3.05$).

Type of chunk	base-level activation	retrieval time [sec]
number	10	0
mapping	1	1.122
addition-fact	0.9	1.24
multiplication-fact	0.2	2.497

The previous experiment (conversion task) was concerned with the addition of fractions were either the left, the right or both fractions needed to be converted. However, no mappings were involved. Since neither the visual perception of the problem

on a screen nor the output of the response were modeled for this experiment, the values for the base-level activation needed to be higher than for the current experiment to account for the extra time needed for perception and output.

The base-level activation of the number chunks are set high to reflect that it takes almost no time to retrieve the meaning of a number. For instance, when a subject wants to report a result verbally, it can be assumed that s/he finds the corresponding word for a given integer (e.g. "sixteen" for 16) almost instantly. In contrast to this, the base-level activation of the multiplication- and addition-facts as well as the mappings are set to considerably lower values, thus resulting in longer retrieval times. Using the base-level values specified in Table 6.3, the retrieval of mappings and addition-facts takes almost the same amount of time (1.122 s and 1.24 s, respectively), whereas it takes much longer to retrieve a multiplication-fact (2.497 s). These times seem fairly high. However, they will change in the next step, when variability is introduced; thus, they will be discussed and compared with experimental data in section 6.3.2.

It has been shown empirically that the outcome of specific operations can be retrieved faster than others due to the so-called problem-size effect. This means, that it takes subjects longer to solve simple arithmetic problems that involve larger numbers (a short description of this empirical phenomena can be found in [80]). However, as argued earlier in section 4.4, we are only concerned with data averaged over all 110 different problems such that these differences do not need to be taken into account.

6.2.2 Comparison of Different Strategies

Experiments were run with all strategies described in section 4.5, resulting in four different models that could be compared to the subject data in the development set. Additionally, mixture models were emulated, that chose either of the strategies with a certain probability. For each condition, both the absolute and relative response times of the models are compared with the subject data. Figure 6.1 shows the comparison of absolute times, Figure 6.2 refers to relative times.

Re-encode versus Encode-once Strategy

This strategic choice is concerned with the last response given (*newDenom*). The model can either look at the screen again in order to re-encode the necessary information $d1$ and $d2$ (re-encode), or it can keep this information in mind (in the imaginal buffer) after having looked at it earlier during the task (encode-once).

The strategy only influences the time of the last response *newDenom*. The time needed for this response mostly depends on the retrieval time of the outcome of the multiplication $d1 * d2$. This sounds like an easy parameter to adjust. However, the calculation of *num1* and *num2* also depend on this retrieval time such that the retrieval time for multiplication-facts cannot be changed arbitrarily (see Figure 4.4). It turns out that even though the re-encode strategy may sound more intuitive, the encode-once strategy leads to a much better fit. Thus, only this strategy was used in the following models.

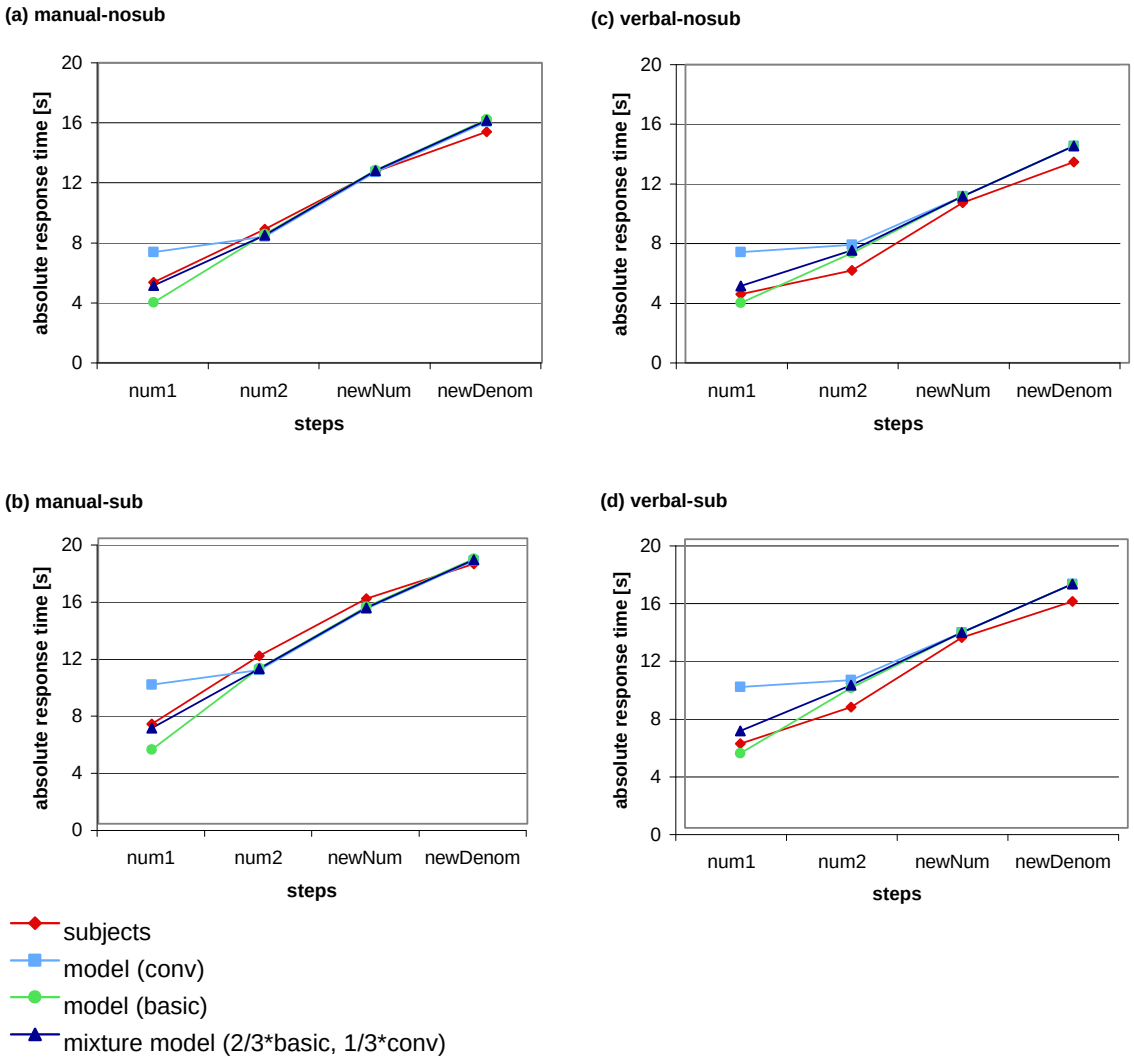

Figure 6.1: Absolute response times for the subjects in the development set and different models for the four different conditions. The times for the basic and convolution strategy are shown as well as the mixture model $M(\frac{2}{3}, \frac{1}{3})$.

Basic versus Convolution Strategy

Basic and Convolution strategy differ in when the first response is given (*num1*). Either it is reported right after its calculation (basic) or after the second response *num2* has been calculated as well, such that the two results can be reported one after the other (convolution).

This strategic choice affects only the time of the first response (*num1*) which is influenced by the retrieval time for multiplication-facts and, in case of a substitution, the retrieval of a mapping. Although there does not seem to be a huge difference between the times of the two models and that of the subjects concerning the absolute times, the difference becomes quite obvious when analyzing the relative times. As defined in section 3.3.5, absolute time refers to the time from stimulus onset, relative times to the time that has elapsed since the last response.

As can be seen in Figure 6.1 and Table 6.4, neither model seems to be much off for the absolute times. Both show a high correlation with the subject data (0.983 for basic and 0.957 for convolution), even though the mean deviation is at 0.933 and 1.706 s, respectively. However, the relative times show that neither the assumption of the basic strategy nor the convolution strategy produce a good fit to the first two response times (see Figure 6.2):

- When employing the basic strategy, the model needs almost the exact same amount of time for reporting *num1* and *num2*. This does not correspond to the subject data where the time for reporting $num2$ is smaller than that for $num1$.
- Using the convolution strategy results in too short of a response time for $num2$, while the time for $num1$ is too long. Figuratively speaking, the convolution strategy captures the 'dip' that can be seen in the response times of the subjects (for $num2$), but is too extreme.

Although the first two response times are not modeled particularly well with these two strategies, both strategies produce a fairly good fit to the last two responses. It should be pointed out that these two strategies do not differ for the calculation and output of the last two responses (for reasons explained above, the encode-once strategy is used for the calculation of the last response, *newDenom*).

Mixture Model

Given that neither the basic nor the convolution strategy were able to capture the first two response times well, the use of a mixture model seems justified. As explained in section 4.5, the mixture model simulates that the subjects chose either strategy with a certain probability. This means that the model decides for each problem anew and then sticks to the chosen strategy until this particular problem is solved. The probabilities were determined experimentally, such that they would produce a good fit to the mean response times of the development set.

Interestingly, the data indicates that the subjects use each strategy in a similar manner, independent of the condition. In general, the more complex convolution strategy seems to be less likely to be employed than the more straight-forward basic

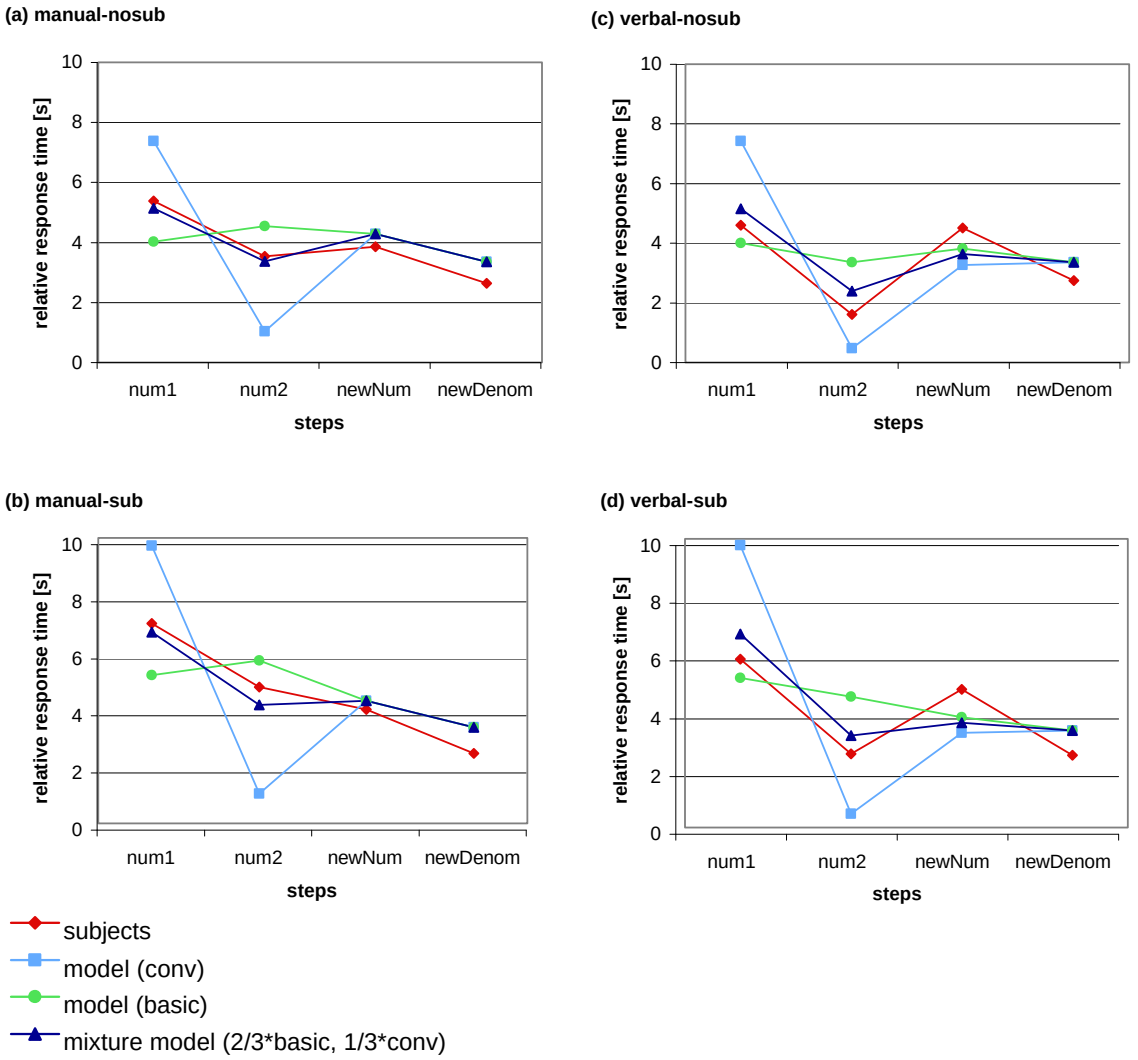

Figure 6.2: Absolute response times for the subjects in the development set and different models for the four different conditions. The times for the basic and convolution strategy are shown as well as the mixture model $M(\frac{2}{3}, \frac{1}{3})$.

strategy. The experiments showed that the mixture model $M(\frac{2}{3}, \frac{1}{3})$ provides a good fit to the subject data, i.e. the model which chooses the basic strategy $\frac{2}{3}$ of the time and utilizes the convolution strategy $\frac{1}{3}$ of the time. For simplicity, $M(\frac{2}{3}, \frac{1}{3})$ is referred to as 'mixture model' in the following. As can be seen in Figure 6.2, the response times produced by the mixture model are a bit too high for the verbal conditions whereas they are a bit too low for the manual conditions.

Table 6.4: Overall correlation and mean deviation between the response times by the models and the data of subjects in the development set. Absolute times refer to the time from stimulus onset, relative times to the time that has elapsed since the last response.

	Absolute Times		Relative Times	
Model	Correlation	Mean Dev.	Correlation	Mean Dev.
basic strategy	0.983	0.933	0.681	1.087
convolution strategy	0.957	1.706	0.759	2.051
mixture model $M(\frac{2}{3}, \frac{1}{3})$	0.986	0.801	0.889	0.685

However, the mixture model captures the shape of the curve much better than the basic or the convolution strategy. As can be seen in Table 6.4, the correlation between the relative times of the subjects and the mixture model is 0.889 and thus higher than that for the two single strategies (0.681 for basic and 0.759 for convolution). Also, the mean deviation is much lower (0.685 versus 1.087 and 2.051, respectively). As mentioned before, the absolute times are less informative than the relative times and both the basic and the convolution strategy achieved a high correlation already, when only absolute times are taken into consideration. The mixture model still shows a slightly better correlation (0.986) and also a lower mean deviation (0.801) than the two other strategies.

6.3 Variability in Problem Solving Time

As described in section 5, the subjects showed a high degree of variability in the problem solving time T, that is the amount of time needed to complete a problem (as defined in section 3.3.5). In order to match the distribution of problem solving times shown by the subjects, I introduced variability in the previously deterministic model. As for the first step, the fit was assessed based on correlation coefficient and the mean deviation. Again, I used the development set for adjusting the model and evaluated on the test set. However, in contrast to the first step, I did not use the mean times, but the distribution of the 900 individual times in the development set (9 subjects solving 100 problems each).

6.3.1 Purpose and Parameters

I experimented with parameters that determine the time needed for the retrieval of knowledge and a parameter that determines the timewise behavior of the imaginal module (see Table 6.5). It needs to be pointed out that this simulates just one of the potential sources in variability which are described in section 5.3. In a nutshell, variability is caused by external and subject-related factors, where the latter

are more interesting for us. Externally, it is induced by using 110 different problems. More importantly, it is also caused by different solution strategies used by the subjects and by individual differences in the time needed to retrieve task-related knowledge. These are two factors that need to be included in the model. The first factor was discussed in the previous section where a mixture model was proposed that choses between two strategies with certain probabilities. However, providing the model with two different strategies for solving a problem captures the variability only partly. Furthermore, it does not affect the problem solving time T much, but rather the distribution of the four response times within that time span. Thus, we are concerned with the second factor here, the variability of retrieval times that influences the variability of T.

For this purpose, I experimented with the parameters listed in Table 6.5. These can be divided into two groups, one that changes the behavior of the imaginal module and one that affects the declarative module. As explained before, the imaginal module is used to hold mental representations of the problem and is effectively used as short term memory in this model (see section 4.3.1). The task of the declarative module is to retrieve declarative information from memory (mappings, addition- and multiplication-facts in this task).

Table 6.5: Parameters adjusted for introducing variability in problem solving time.

		Value		
Parameter	Name	Default	Range	Module affected
:imaginal-delay	imaginal delay	0.2	≥ 0	imaginal
:lf	latency factor	1.0	≥ 0	declarative
:ans	activation noise	NIL	≥ 0	declarative
:rt	retrieval threshold	0.0	any value	declarative

The imaginal delay influences the behavior of the imaginal buffer. It specifies the time needed to complete a request or modification request to the imaginal buffer (in seconds). Thus, it can only take positive values. As mentioned before, it takes 200 ms to change a chunk in the imaginal buffer by default.

The other three parameters (latency factor, activation noise, retrieval threshold) affect the declarative module. The latency factor and activation noise influence the time needed to retrieve a chunk from declarative memory as specified in equation (2.8). The latency factor corresponds to F in the equation, whereas the activation noise has an impact on A_i. More specifically, the activation noise corresponds to the s value used to generate the transient noise added to the activation equation (2.9). The activation noise can be seen as a way to mirror variability within a subject, whereas the latency factor represents variability between subjects. Finally, the retrieval threshold specifies the minimum activation that a chunk needs to have to be retrieved. It corresponds to τ in the equation for retrieval failure (2.10). Except for the latency factor which had been set to 3.05 (see section 6.2.1), none of the other parameters had been changed compared to the default settings of ACT-R.

6.3.2 Experiments with Variability

The first step for matching the variability of T was to bin the problem solving times: For both the subject data and the model data, 1 second bins were used for the problem solving time, such that T was rounded down to the next integer. In other words, T was replaced by $\lfloor T \rfloor$. Since the aim was to match the variability of the 9 subjects in the development set, each solving 100 problems (25 per condition), the model was run 250 times for each of the four conditions ($= 10 * 25$). Among the list of 110 potential problems as defined in section 3.3.3, problems were chosen at random; each problem had the same probability of 1/110 to be chosen. The distribution of problem solving times shown by the subjects in the development set can be seen in Figure 6.3 (plot (a)).

Variability due to Problem Selection

As a baseline for comparison, I considered the deterministic model first, where variability is only caused by the differences in the arithmetic problems that have to be solved. Two factors influence the response time: the fact that some problems require a carry in the addition and the nature of the perceptual-motor modules.

As described in section 3.3.3, 48 out of the 110 problems involve a carry in the addition of *num1* and *num2*, while the others do not require this step. The difference between these two cases is that an addition with carry requires the retrieval of an additional production rule and that of one more addition-fact. Having one more production rule fire involves a time cost of 0.05 s. For reasons described in section 6.2.1, the retrieval of an addition-fact takes 1.24 s for the given model. Thus, it takes the model 1.29 s more for performing an addition with carry (1.24 s+0.05 s).

In addition, the problem solving time is influenced by more subtle differences due to the characteristics of the output the model wants to give. The time needed for uttering a response depends on the length of the word, while the time for entering results manually depends on the position of the buttons to be clicked. The latter will not be analyzed here, since we only used an approximation of the manual response (for details, refer to section 4.1). As mentioned before, each response was a double-digit number such that always two digits needed to be entered (by clicking on two buttons). Thus, the time needed for responding manually only depends on which two digits need to be entered.

Due to the implementation of the vocal module in ACT-R, the time needed to utter a response depends on initiation and preparation cost as well as the time to actually speak the word (see equation (2.11)). The initiation cost is always 0.05 s. For the first answer, the module needs 0.15 s to prepare. Since the module is already prepared after having given the first answer ($num1$), it needs 0.05 s less to prepare for the second answer ($num2$). These are all default values in ACT-R that were not changed. The time for actually uttering the word depends on the length of the word to be spoken. Due to the problem selection, the words for $num1$ and $num2$ range between 3 and 11 letters in length. Thus, the time for actually uttering a result varies between 0.15 s (3*0.05 s) and 0.55 s (11*0.05 s). The activity of the module therefore ranges from 0.30 s up to 0.75 s, as can be seen in Table 6.6.

The deterministic model, run on 250 randomly selected problems per condition, leads to a distribution of times as shown in plot (b) of Figure 6.3. As expected,

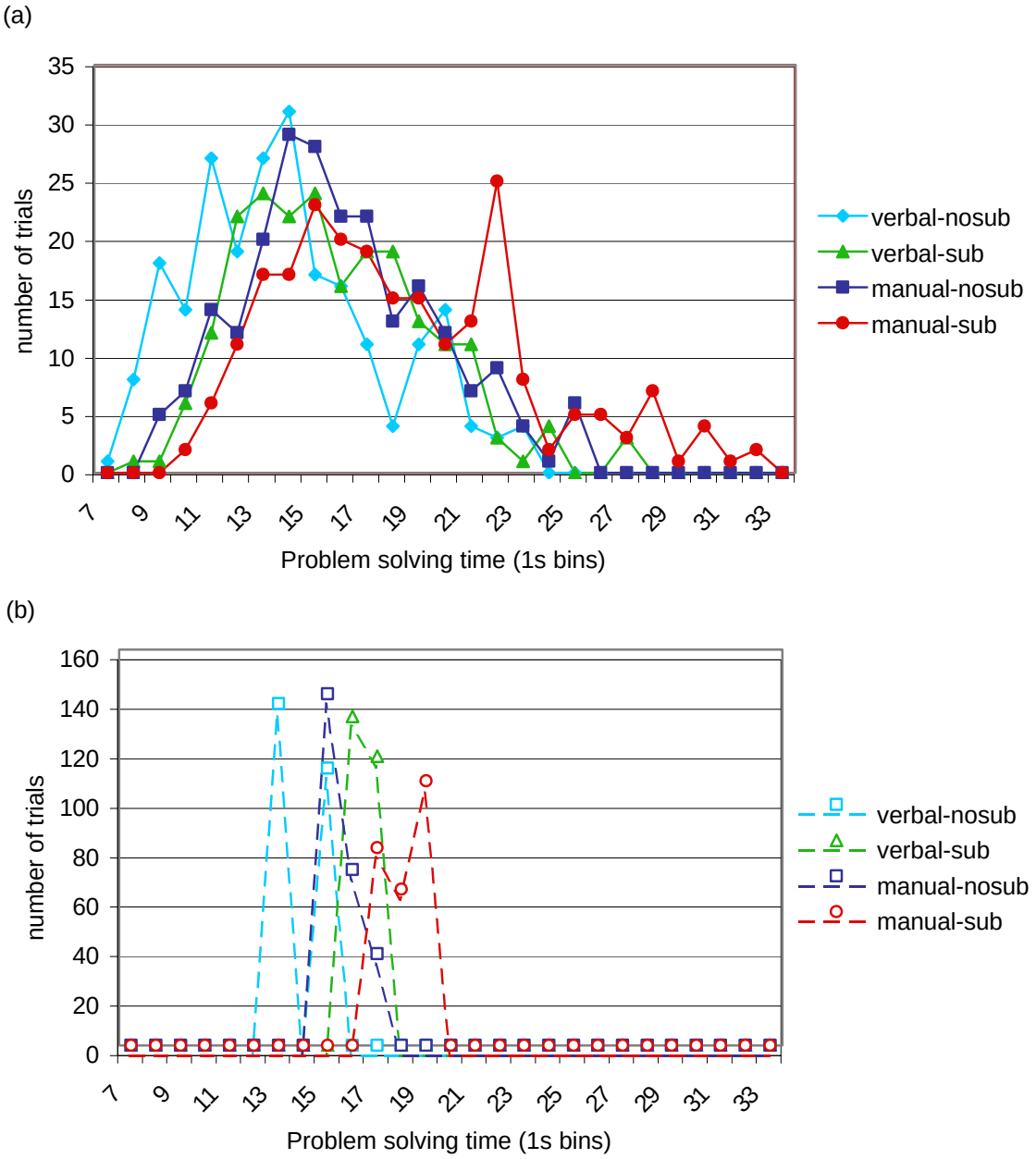

Figure 6.3: Variability in problem solving time shown by the subjects in the development set (a) and by the model (b). Even though the model was deterministic, plot (b) still shows a distribution of times due to the fact that 110 different problems were used.

Table 6.6: Time needed by ACT-R to respond verbally for the arithmetic task, where the duration of the actual utterance ranges from 0.15 to 0.55 s.

	Required Time [s]	
Factor	Minimum	Maximum
Initiation cost	0.05	0.05
Preparation cost	0.10	0.15
Duration of actual utterance	0.15	0.55
Overall time cost	0.30	0.75

the variability shown by the model is too low in comparison with the subject data (plot (a) of the same figure). However, the model already accomplishes to capture some of the characteristics that can be seen in the subject data. Interestingly, these characteristics seem to be modeled better in the manual conditions than in the verbal conditions, although the opposite could have been assumed as the process of entering a result manually is approximated by the model. In both the subject and the model data, the *manual-nosub* condition shows one peak whereas the *manual-sub* condition has two, with the latter being more pronounced. The opposite is the case for the *verbal-nosub* condition, where the first of the two peaks has a higher amplitude. Although the model shows this pattern as well, the ratio of the amplitudes of these two peaks is different. Finally, the two peak structure of the *verbal-sub* condition is only approximately reflected in the model data.

These characteristics are not attributable to the variability shown by subjects, but simply to the variability introduced externally by the problem selection.

Variability in the Information-Flow of the Model

As it turned out in the experiments, not all of the parameters described in section 6.3.1 were needed to achieve a good fit. Essentially, it seemed not to be necessary to use activation noise which influences the declarative module. Since that would have been the only reason for a retrieval failure to occur, it was not necessary to change the retrieval threshold either (as the retrieval threshold influences the time after which a retrieval failure occurs). Out of the experiments that were run, only the results of the two most promising settings are presented here (called setting A and B), the parameters of which can be found in Table 6.7. R(1.0) denotes a random number generator that produces a uniform distribution of values x, $x \in [0.0, 1.0)$.

For setting A, both the imaginal delay and the latency parameter were set randomly according to a uniform distribution; however, the random number generators used for each were independent. Furthermore, a constant parameter (0.6) was added to the latency factor. It seems reasonable that the latency factor cannot drop below a certain minimum, as it influences the retrieval time (according to equation (2.8)) which should be associated with a minimum time cost, at least for certain chunks such as facts about addition. The curves resulting from setting A can be found in plot (a) of Figure 6.4. As can be seen, the distribution of the model is wider than that of the subjects, which resembles a normal distribution.

Table 6.7: Two settings of parameters for experiments with variability of problem solving time T.

	Setting	
Parameter	A	B
imaginal delay	$1.08 * R_1(1.0)$	$0.54 * (R_{shared}(1.0) + R_1(1.0))$
latency factor	$0.6 + 3.8 * R_2(1.0)$	$0.6 + (1.9 * (R_{shared}(1.0) + R_2(1.0)))$
activation noise	default	default
retrieval threshold	default	default

Therefore, I introduced a variability component R_{shared} that is shared between the latency factor and the imaginal delay. In addition, each of them has its own variability component R_i ($i \in \{1, 2\}$). An approach similar to setting B presented here has been taken in the ACT-R models used in [12, 15]. The parameters here were chosen such that the mean values for the imaginal delay and the latency factor were the same as for setting A.

Plot (b) of Figure 6.4 shows the distribution of problem solving time T for all conditions, both for model B and the subjects in the development set. As can be seen, the model fits the subject data fairly well. This becomes more obvious in Figure 6.5 which shows the problem solving times for each condition, again for the subjects and model B. As can be seen, the distributions show a similar pattern for each of the four cases.

Remarks on the Resulting Model

Both setting A and setting B change not only the problem solving time, but also the response behavior of the model that we tried to fit to the subject data as described in section 6.2. The reason for this is that the mean value of the imaginal delay is about 0.54 s ($0.54 * (R_{shared}(1.0) + R_1(1.0))$) in contrast to the previous default value of 0.2 s. This may seem like a major deviation from the standard ACT-R settings, but the default value proves only to work well for simple tasks as opposed to comparatively complex tasks as the one modeled here. Similarly, the latency factor was set to 3.05 before, based on a previous model. However, the mean of the latency factor for the variable model is 1.9 ($0.6 + (1.9 * (R_{shared}(1.0) + R_2(1.0)))$), thus leading to faster retrievals according to equation (2.8). As mentioned before (see section 6.2.1), the deterministic model needed fairly long times for the retrieval of arithmetic facts compared with the time needed by subjects. These retrievals are more realistic in the variable model as can be seen in Table 6.8 that provides a comparison between the two models and data reported for humans [57, 63, 112].

The subjects tested in [63] needed a mean retrieval time of 1.19 s, whereas it takes the model 1.56 s to retrieve the outcome of a multiplication. Concerning additions, [57] reported a mean response time of 0.83 s for the retrieval of simple addition-facts for young adults and [63] reported a mean time of 1.0 s (the model needs 0.77 s). Furthermore, it was reported in [112] that subjects need 0.90 s more on average for an addition that involves a carry as opposed to one that does not require a carry. In the model, the difference between these two cases is that an addition with carry

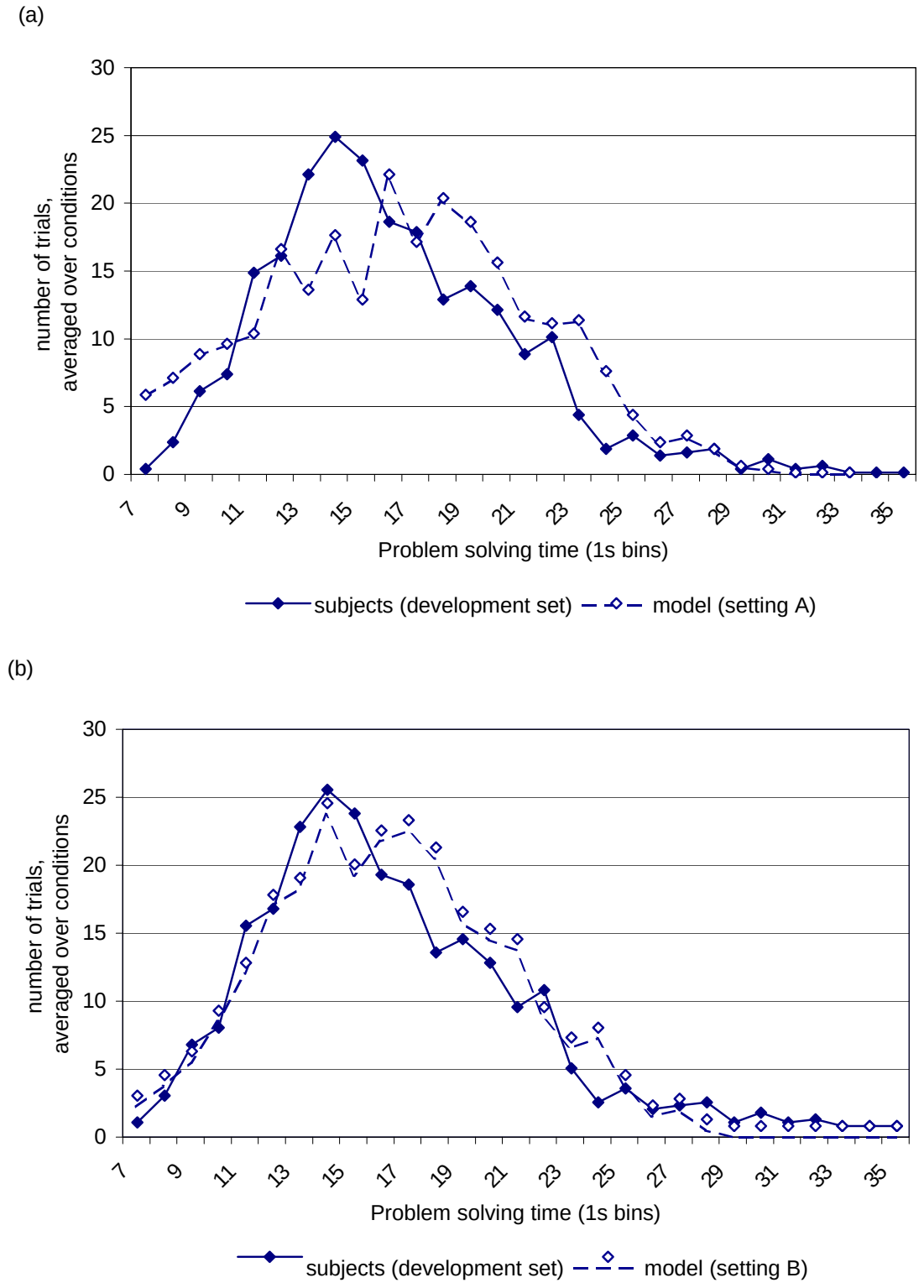

Figure 6.4: Problem solving time for model and subjects (development set), averaged over all conditions. Plots (a) and (b) show the times of the model for setting A and B (see Table 6.7), respectively, where setting B included a shared variability component between the imaginal delay and the latency factor.

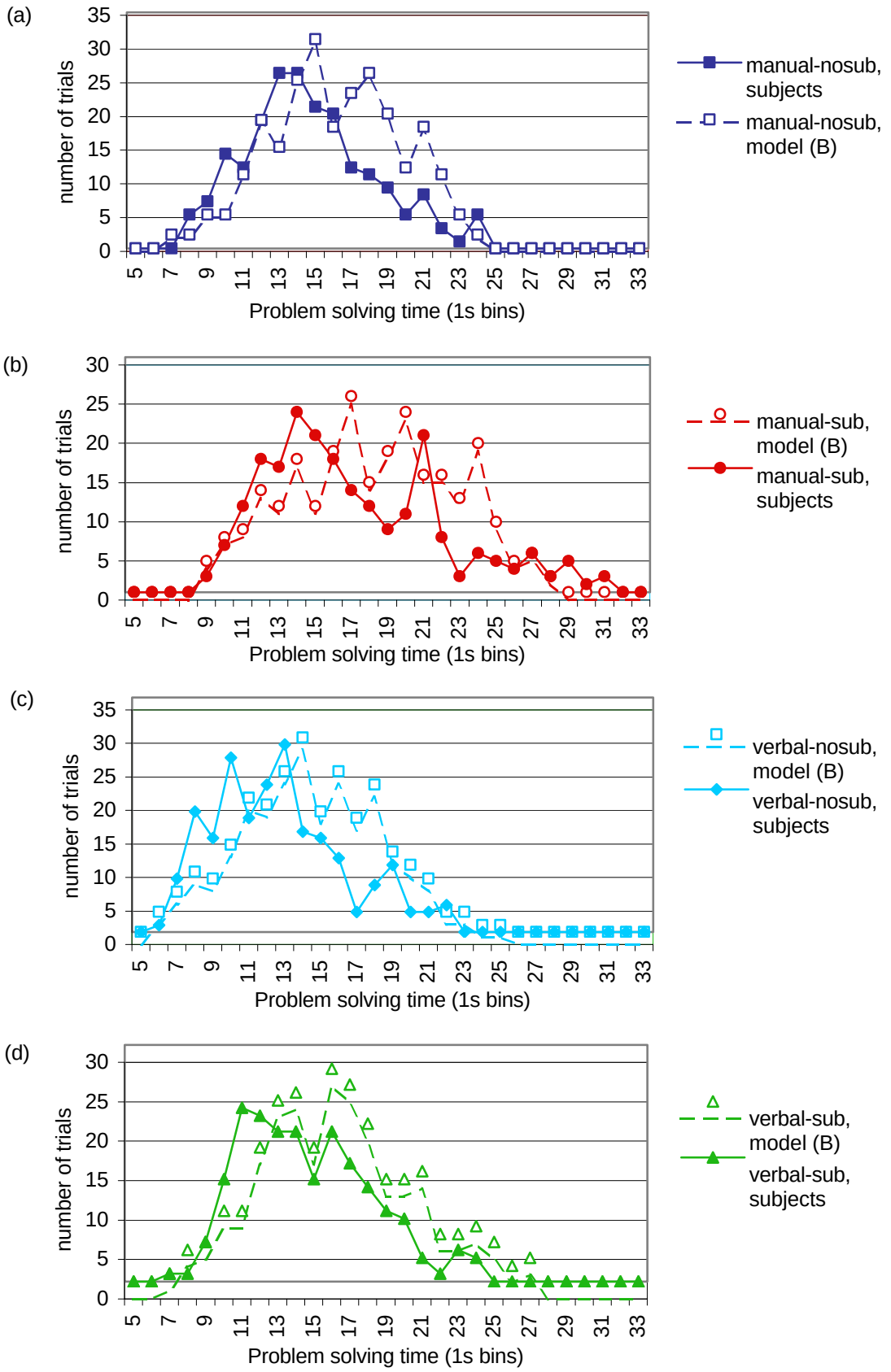

Figure 6.5: Problem solving time for each condition for the subjects in the development set and the model (setting B as defined in Table 6.7).

Table 6.8: Comparison of times required for arithmetic calculations (deterministic and probabilistic model as well as humans, as reported in [57, 63, 112]).

	Model		
Action	Deterministic [s]	Probabilistic [s]	Humans [s]
retrieval of mapping	1.122	0.70	-
retrieval of addition-fact	1.24	0.77	0.83; 1.0
retrieval of multiplication-fact	2.50	1.56	1.19
performing a carry	1.29	0.82	0.90

requires the retrieval of an additional production rule and that of one more addition-fact. Thus, the model requires 0.82 s more for performing an addition with carry which seems like a fairly good fit (0.77 s for retrieving the addition-fact and 0.05 s for having one more production rule fire). It can be concluded that the model needs slightly less time for the retrieval of additions, but more time for the retrieval of multiplications in comparison to humans, while staying within reasonable bounds.

The mean response times of the deterministic and the probabilistic models are provided in Table A.5, for the basic strategy, the convolution strategy and the mixture model $M(\frac{2}{3}, \frac{1}{3})$. The overall correlation and mean deviation between these different models and the subject data can be found in Table 6.9. The mixture model $M(\frac{2}{3}, \frac{1}{3})$ still turns out to yield the best correlation and lowest mean deviation. Although the variable mixture model leads to a higher mean deviation for absolute times (the correlation stay almost the same), it produces a better fit than the deterministic mixture model for the relative times which are more indicative for a good fit.

Table 6.9: Comparison of the overall correlation and mean deviation between the response times by the probabilistic models and the data of subjects in the development set.

		Absolute Times		Relative Times	
Model type	Strategy	Correl.	Mean Dev.	Correl.	Mean Dev.
deterministic	basic	0.983	0.933	0.681	1.087
	convolution	0.957	1.706	0.759	2.051
	$M(\frac{2}{3}, \frac{1}{3})$	0.986	0.801	0.889	0.685
probabilistic	basic	0.983	1.567	0.706	1.112
	convolution	0.969	1.59	0.785	1.621
	$M(\frac{2}{3}, \frac{1}{3})$	0.988	1.344	0.947	0.664

As mentioned before, I use the latency data for arriving at an estimate of the basic parameters in the model that mirror the variability in the subject data. Although this is an estimate that aims at fitting the data, it is important that the model is able to reproduce the distribution of problem solving times in plot (a) of Figure 6.3 as a basis of the next step, the prediction of the BOLD responses.

6.4 Prediction of the BOLD Response

After the model had been adjusted according to the behavioral data, the BOLD response was predicted for the different regions of interest (ROIs) based on the activity of the associated modules. For this purpose, the probabilistic model was used that was derived in the experiments with variability (see section 6.3.2). The aim was to predict the BOLD responses averaged over the correct trials in the database, of which about 375 exist per condition as can be seen in Table A.2. Thus, the model was run 375 times per condition on problems chosen randomly from the list of 110 problems (for the problem generation, see section 3.3.3).

Both for the model runs and the subjects, the data was averaged per condition and over trials. However, it should be noted that the data we are concerned with here are time series (in contrast to the behavioral data where only four answering times are known to us). In order to average over the time series without losing too much information, a method called event-locked averaging was used which can be seen as a combination of two standard ways of aligning fMRI data, stimulus locking and response locking (see section 6.4.1).

The prediction by the model was then compared with the actual brain imaging data of the subjects, this time using the whole dataset (development and test set combined). In contrast to the previous steps, the model was not adjusted anymore after this step as the aim was to predict previously unseen BOLD responses. However, ideas for potential changes of the model are given in section 8.2, based on the outcomes of the comparison with the subject data.

6.4.1 Event-locked Averaging

As the overall aim of this work is to compare patterns in the actual imaging data with the predicted BOLD responses, it is necessary to aggregate the BOLD responses over multiple trials. However, this is challenging due to the high variability in problem solving time, both for the individual subject data and the ACT-R data which was adapted to show a similar variability (see section 6.3). In order to still be able to aggregate both datasets and directly compare them to each other, the data had to be brought into alignment. For this purpose, a method called event-locked averaging was used [7], which had been successfully used for aligning data in previous imaging studies [8, 12, 15].

Method

The idea behind event-locked averaging is to anchor the data around observable events. Although the given experiment had multiple of such events (onset, *num1*, *num2*, *newNum*, *newDenom*, last scan), we decided to use only the trial onset, the problem solving time (the time needed to report *newDenom*) and the 15th scan as the last scan, due to the short duration of each trial. However, there is no restriction to apply this method to experiments that have more event-defined intervals (as described in [15, 107]). Since we chose three events as landmarks, two sequences resulted: the trial proper (8 scans) and an ensuing cool down period (7 scans). The lengths of these template sequences were determined by the average number of scans needed by the subjects to solve the problem (14.35 s on average, which corresponds to 8 scans as can be seen in Table 6.10). Since we usually recorded 7 cool down

scans after the problem was solved (a minimum of 6 cool down scans; for details, refer to section 3.2.1), the total number of scans was 15 on average.

Table 6.10: Relation between scan number and time (TR = 2 s).

scan number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
starting time [s]	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28
end time [s]	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30

Event-locked averaging was applied both to the actual BOLD responses shown by the subjects and the BOLD responses predicted by ACT-R, such that these time series had the above-mentioned structure (8 scans trial, 7 scans cool-down). This was done using the following definition of event-locked averaging.

Definition. The procedure of event-locked averaging (ELA) is defined as follows. Given a scan sequence S_{in} of n scans with the aim of deriving a scan sequence S_{out} of length m, the following steps are taken:

- Case 1: $n \geq m$
  - create beginning sequence S_{in}^{begin} of length $\lceil m/2 \rceil$:
    take $\lceil m/2 \rceil$ scans from the beginning of S_{in}
  - create ending sequence S_{in}^{end} of length $\lfloor m/2 \rfloor$:
    take $\lfloor m/2 \rfloor$ scans from the end of S_{in}
- Case 2: $n < m$
  - create beginning sequence S_{in}^{begin} of length $\lceil m/2 \rceil$:
    take $\lceil n/2 \rceil$ scans from the beginning of S_{in},
    duplicate the last scan in this sequence until a length of $\lceil m/2 \rceil$ is reached
  - create ending sequence $S_{in}^{end}$of length $\lfloor m/2 \rfloor$:
    take $\lfloor n/2 \rfloor$ scans from the end of S_{in},
    duplicate the first scan in this sequence until a length of $\lfloor m/2 \rfloor$ is reached
- append the two sequences to form sequence $S_{out} = (S_{in}^{begin}, S_{in}^{end})$ of length m.

In other words, if the given sequence is too long compared to the template sequence (case 1: $n \geq m$), the $(n - m)$ scans in the middle are deleted such that the desired length is reached. If m is odd, one more scan is taken from the beginning sequence. If the given sequence is too short (case 2: $n < m$), sequence S_{in} is split in the middle and the missing scans are taken from either end of the halves, thus padding the middle. If either n or m is odd, the extra scan is taken from the beginning, as in the first case.

For instance, let us assume that the given sequence S_{in} is of length $n = 4$ and has the following structure (where x_i is the ith scan):

$$S_{in} = (x_1, x_2, x_3, x_4)$$

Table 6.11: Scan sequences of length m resulting from event-locked averaging, given the input scan sequence $S_{in} = (x_1, x_2, x_3, x_4)$.

desired length m	S_{out}
2	(x_1, x_4)
3	(x_1, x_2, x_4)
6	$(x_1, x_2, x_2, x_3, x_3, x_4)$
7	$(x_1, x_2, x_2, x_2, x_3, x_3, x_4)$

The results of event-locked averaging for both cases can be found in Table 6.11 (each case either with an odd or an even desired length).

Essentially, the motivation behind event-locked averaging is to preserve the temporal structure of the beginning and the end of the sequences, while just using an approximation of the average activity in their middle. This method can be seen as a combination of two standard ways of aligning fMRI data, stimulus locking and response locking.

Effects of Event-locked Averaging on the Data

For the interpretation of the following BOLD responses, it is important to understand how the procedure of event-locked averaging affects the given data. As explained before, the data was warped around scan 8 with 7 ensuing cool down scans, resulting in 15 scans in total. This has two implications. First, the data around scan 1, 8 and 15 reflect the actual brain activity best (scans in between may have been deleted or padded). Second, only the first 8 scans were recorded during the actual trial. As mentioned in section 3.2.1, the repetition time TR was set to 2 seconds; thus, 8 scans correspond to a time span of 16 seconds (see Table 6.10). However, due to the delayed nature of the BOLD response, the brain activity resulting from an event will be seen about 4-5 seconds after the event. Thus, the effects of the last steps of the task will be seen until scan 10.

As discussed earlier (see section 6.3), the strategies used for the models lead to different response times, but do not affect the actual problem solving time T which is almost the same for each strategy. As can be seen in Table 6.12, the model needs the same number of scans for a given condition, independent of the strategy. Depending on the condition, the model solves a given problem within 7 (*verbal-nosub*), 9 (*verbal-sub*/*manual-nosub*) or 10 scans (*manual-sub*). In comparison, subjects are one scan faster on average, except for the *verbal-nosub* condition where they also need 7 scans. The effect of event-locked averaging on the scan structures can be seen in Figure 6.6. For scan sequences of length 7, the procedure will lead to the duplication of scan 5. Given a sequence of length 9, scan 6 will be deleted, whereas both scan 5 and 6 will be omitted if the scan sequence is 10 scans long. It should be pointed out that this is likely the case for the different conditions shown in Figure 6.6, but will differ from problem to problem due to the probabilistic nature of the model.

The estimated effect of event-locked averaging on the model and subject data can be seen in Figure 6.7. For both, the scanning sequence or its prediction is shown before and after event-locked averaging (ELA). The two grids on top show the effect on the

Table 6.12: Mean problem solving times T of the models and subjects, including the corresponding scan number.

	Model							
	Basic		Convolution		Mixture		Subjects	
Condition	T	Scan	T	Scan	T	Scan	T	Scan
verbal-nosub	12.76	7	12.83	7	12.79	7	12.25	7
verbal-sub	14.26	9	14.33	9	14.28	9	14.63	8
manual-nosub	14.47	9	14.30	9	14.42	9	14.01	8
manual-sub	15.97	10	15.80	10	15.91	10	16.50	9

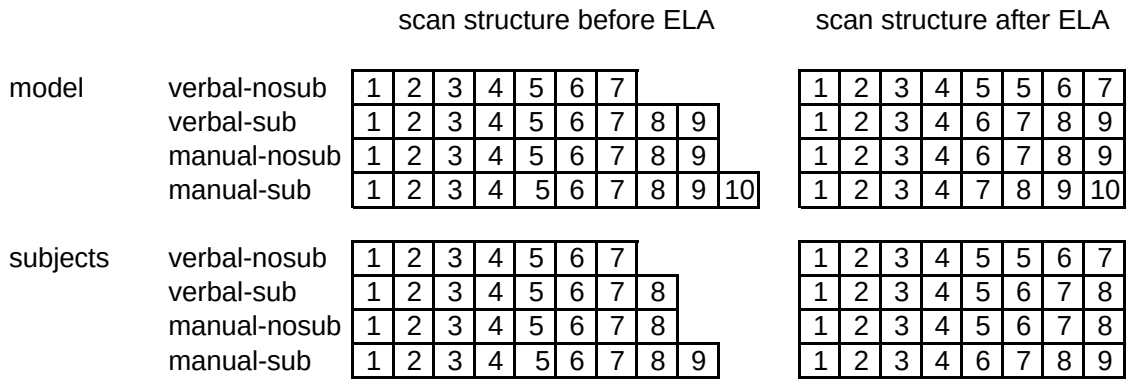

Figure 6.6: Effect of event-locked averaging (ELA) on the scan structure for each condition, both for the subjects and the mixture model. Each of the scan structures is brought to the desired length of 8 scans.

actual response times, whereas the two grids below estimate the effect on the peak times, assuming a delay of 4 s, which corresponds to 2 scans. The scans associated with certain responses are color-coded (*num1*/yellow, *num2*/red, *newNum*/ blue, *newDenom*/green). When *num2* and *num2* occur in the same scan, the block is gray.

As mentioned before, event-locked averaging only leaves the subject data for *verbal-sub* and *manual-nosub* unchanged. Both for the subjects and the model, the scans in the *verbal-nosub* condition are padded in the middle. In all other conditions, the procedure shrinks the sequences such that information might be lost. However, this affects only the first two responses given (*num1*/yellow, *num2*/red). Thus, any conclusions about these responses should be made carefully, since the corresponding scans might be distorted. However, most of the scans in which a response is given or in which we expect to see a peak are preserved. Also, the figure shows that responses *num1* and *num2* are likely to occur between scan 3 and 4, whereas it can be expected that the response *newNum* is given during the 8th scan. The model is likely to respond *newNum* during scan 5, with the exception of the *verbal-nosub* condition where the response is shifted more towards the end (scan 7). For the subject data, the response is more dispersed (between scan 5 and 7).

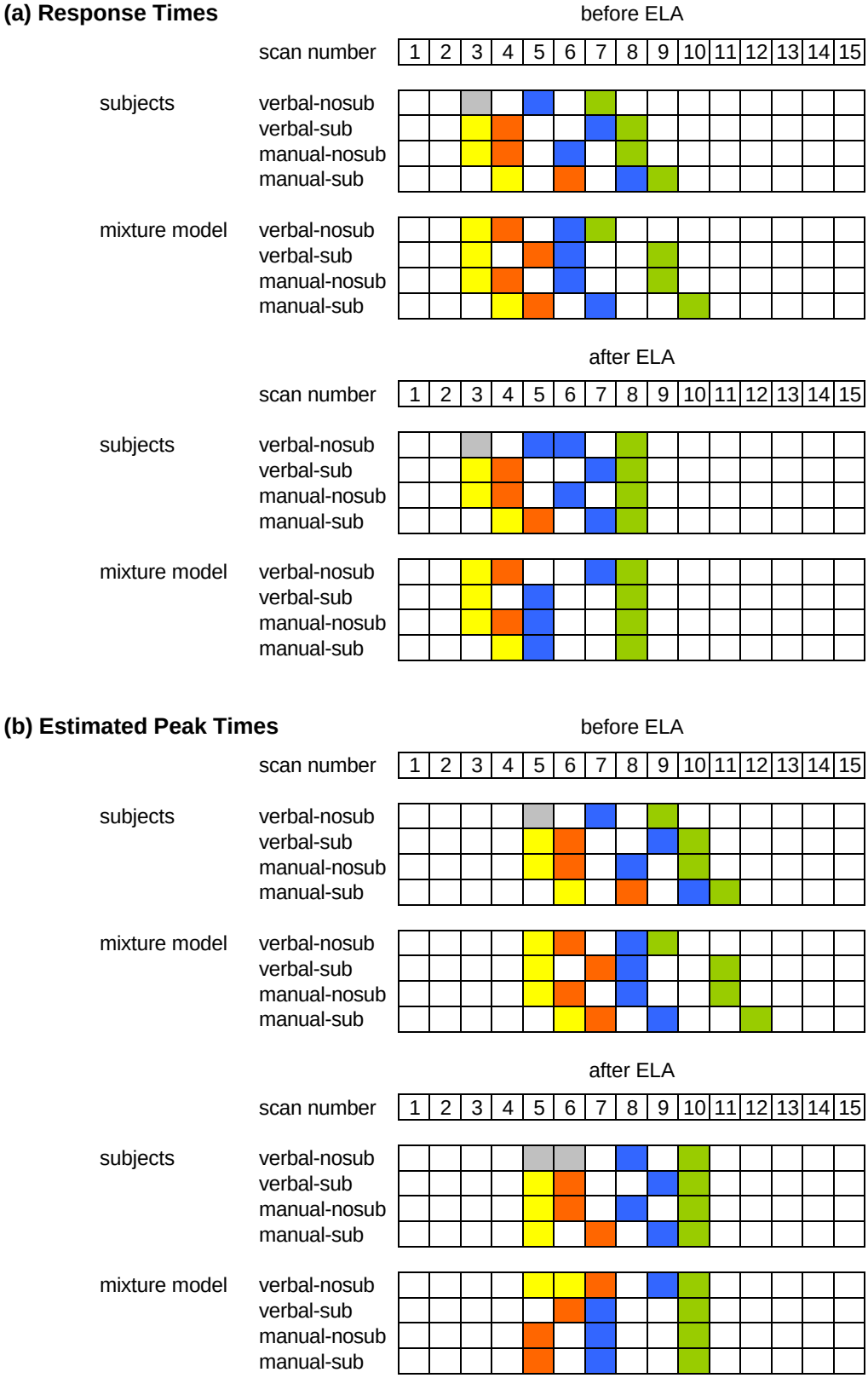

Figure 6.7: Effect of event-locked averaging on the subject data and the mixture model, both for the actual response times (a) and the estimated peak times of the BOLD responses (b). The simulated scanning sequence is shown before and after event-locked averaging (ELA). On top, the effect on the actual response times is shown. Below, the effect on the delayed peak times (4 s delay) are approximated. The scans associated with certain responses are color-coded: *num1*/yellow, *num2*/red, *newNum*/blue, *newDenom*/green. When *num2* and *num2* occur in the same scan, the block is gray.

6.4.2 Estimation of the Magnitude Parameter

The BOLD responses were calculated based on the procedure explained in section 2.4.2, that is by convolving the hemodynamic function $H(t)$ with the demand function $D(t)$ for each module. The demand function gives the probability that the region or corresponding module is active at time t. For each module, this function is based on the time course of its activity during a trial. An example of such a time course can be seen in Figure 4.5. The equations for the hemodynamic function and the BOLD response, (2.12) and (2.13), are reprinted here for clarity:

$$H(t) = m \left(\frac{t}{s}\right)^a e^{-(t/s)}$$

$$B_{model}(t) = \int_0^t D(x)H(t-x)dx$$

As explained before, the parameters s and a are set to default values in ACT-R ($a = 6$ and $s = 0.75$ s) which were left unchanged for the predictions reported here. However, the magnitude parameter m is typically adjusted to the subject data for display purposes. The main purpose for this adjustment is to enable us to directly compare the actual BOLD responses with their prediction in one plot. This is also justified by the high degree of variability in the magnitude of the BOLD signal that is typically found both between and within subjects (for instance, see [1, 85]). It should be noted that the value of the magnitude parameter does not affect the measure of correlation.

The magnitude parameter was determined experimentally for each region of interest. For this purpose, two approaches were considered. On the one hand, the parameter can be derived by determining the ratio of the maxima of the BOLD functions of the subjects and the model, averaged over conditions:

$$\widetilde{m} = \frac{1}{4} \sum_{i=1}^{4} \frac{max(B_{subj,c_i}(t_j))}{max(B_{model,c_i}(t_j))} \tag{6.1}$$

with $\widetilde{m}$: estimate of the magnitude parameter m from equation (2.12)
c_i : condition,
$c_i \in \{$*verbal-nosub*, *verbal-sub*, *manual-nosub*, *manual-sub*$\}$
$B_{subj,c_i}(t)$: BOLD response of subjects, averaged using ELA
$B_{model,c_i}(t)$: BOLD response of mixture model, averaged using ELA
t_j : scan number j, $j \in [1, 15]$

On the other hand, the magnitude parameter can be estimated such that the sum of squared deviations (SSD) is minimized (using the variables as defined above):

$$SSD = \sum_{i=1}^{4} \sum_{j=1}^{15} \left(B'_{subj,c_i}(t_j) - \widetilde{m} * B_{model,c_i}(t_j)\right)^2 \tag{6.2}$$

$$B'_{subj,c_i}(t_j) = \begin{cases} 0 & \text{if } B_{subj,c_i}(t_j) < 0 \\ B_{subj,c_i}(t_j) & \text{else} \end{cases}$$

For the latter approach, the BOLD response of the subjects was modified ($B'_{subj,c_i}(t_j)$), setting it to 0 whenever it takes a negative value. As explained in section 2.4.2, the approach taken to predict the BOLD response has the disadvantage of not being able to predict negative BOLD responses, as it is based on the activity of a given module. Therefore, the scans in which the BOLD response of the subjects takes negative values are treated as being 0 as this is the lowest value the BOLD prediction could possibly take. This issue is avoided when taking the first approach since it is based on the maxima.

The second approach, the minimization of the SSD, is usually taken in ACT-R fMRI research and has a long history of being used for the analysis of variance. This approach is suitable if the time course of the predicted BOLD response is fairly similar to that of the actual BOLD response, i.e. if onset and offset are similar and the peaks are observed at similar points in time. This procedure is capable of providing a reasonable estimate of the magnitude parameter even for rather complex BOLD patterns. However, minimizing SSD has two drawbacks: First, as described above, issues occur when comparing a prediction to a BOLD response that takes negative values; the BOLD response of the subjects needs to be treated as being 0 in these cases (see $B'_{subj,c_i}(t_j)$ above). Second, this procedure yields an unreasonably low estimate $\widetilde{m}$ when the prediction of the BOLD signal is shifted in time compared to the signal shown by the subjects, even if the two signals have a similar shape.

In the latter case, the procedure based on the ratio of the maxima (the first approach) provides a simple, but reasonable method to estimate the magnitude parameter. This works well if the prediction and the actual BOLD response have a similar shape that is not too complex (for instance, a clear one or two peak structure). Then, it is only a matter of adjusting the magnitude of the prediction to the data of the subjects such that the peaks have a similar amplitude. However, for ROIs which show a BOLD response with a more complex pattern, this is not sufficient and the minimization of the SSD yields better results.

Table 6.13 shows the magnitude parameters that were chosen for the given data. For all modules except for the aural, vocal and manual module, the minimization of the SSD was used to derive $\widetilde{m}$ for the above reasons (complex patterns and/or similar time course of the actual signal and the prediction). In contrast, the aural and vocal module showed only one peak which was reflected well in the prediction of the model. However, the prediction of this peak was shifted in time. For the manual module, the BOLD prediction was clearly shifted in time as well. Thus, the ratio of the maxima was used to estimate the magnitude parameters of these modules.

Table 6.13: Estimates $\widetilde{m}$ of the magnitude parameter for the predicted BOLD responses, derived by minimizing the sum of squared deviations (SSD) according to equation (6.2) or by calculating the ratio of the maxima according to equation (6.1).

		Based on	
Module	Estimate $\widetilde{m}$	Ratio of Maxima	SSD
Visual	12.5		x
Aural	5.1	x	
Manual	2.7	x	
Vocal	4.0	x	
Imaginal	2.3		x
Declarative	0.6		x
Goal	0.7		x
Procedural	0.9		x

7. Results and Analysis

The rationale behind devising the arithmetic experiment was to test the following hypotheses, as explained in section 3.1:

Hypothesis 1. It is possible to use verbal protocols in the given fMRI study.
Hypothesis 2. The output modality (verbal or manual) does not affect the BOLD responses of the predefined prefrontal, parietal, anterior cingulate or caudate ROI, i.e. the brain regions that are associated with the ACT-R modules of cognitive control (declarative, imaginal, goal and procedural module).
Hypothesis 3. The BOLD responses for the verbal and the manual output modality are affected in a similar way by the level of difficulty (manipulated by including substitutions or not).
Hypothesis 4. The BOLD responses shown by the subjects during the arithmetic task can be predicted by an ACT-R model for the eight predefined regions of interest.

As noted before, the focus of this thesis was on Hypothesis 4, which is therefore evaluated in most detail in section 7.1; a summary of the results is provided in section 7.1.9. The remaining hypotheses are discussed in section 7.2.

7.1 Predictions and BOLD Responses

This section is essentially an evaluation of Hypothesis 4, combined with an interpretation of the BOLD response of the subjects. Hypothesis 4 stated that the BOLD responses of the subjects can be predicted by the ACT-R model for the eight predefined ROIs.

In the following, the predictions for the eight modules will be compared to the actual brain imaging data, starting with the four perceptual-motor modules, followed by the four cognitive modules. In contrast to the first two steps taken, i.e. the adjustment to the timing of the four responses and to the variability in problem solving time, the full data set was considered (development set and test set combined). It should be noted that testing on the full data set results in the risk that the fit of the model is overestimated. However, the BOLD responses of subjects generally show a huge

variance due to inter-personal differences in brain activity, such that averaging over the full data set of 16 subjects provided us with a more generalized response. Also, the model does not attempt to reflect the brain activity of single subjects, but of a subject population in total. A look at the BOLD responses of the two subject groups, i.e. development and test set, also shows that even though they do show different characteristics, the overall picture is similar, as the averaged BOLD responses mainly differ in magnitude. The ROI corresponding to the visual module (fusiform) shows even a remarkable similarity which makes the two BOLD responses (averaged over the test set and averaged over the development set) almost indistinguishable.

For predicting the BOLD responses, I used the mixture model as well as those with just one strategy (basic and convolution), each with the probabilistic settings that were derived in section 6.3. In the detailed analysis provided in the following, the mixture model was used (and is referred to as 'model'). If references are made to the results obtained with the other two models, this is indicated. The averaged BOLD responses from the combined development and test set was used, which is referred to as 'subjects' in the following. Both the actual BOLD responses and the predictions produced by the model were averaged using event-locked averaging (ELA; see section 6.4.1).

The plots provided in the following show the observed BOLD responses (solid lines) and the predictions (dashed lines) as functions of the scan number. In each plot, the data and predictions are color-coded depending on the condition. It is helpful to use Figure 6.7 for the interpretation of the data as it visualizes into which scans the responses fall.

The correlations between the predicted BOLD responses and the actual brain imaging data are provided in Table 7.1. Since the magnitude of the BOLD curves was adjusted (see section 6.4.2), deviations are not informative and are therefore not provided here.

Table 7.1: Correlations between the predictions and the BOLD responses (for different models, all with the probabilistic setting as derived in section 6.3).

	Models		
Module	Mixture Model	Basic Strategy	Convolution Strategy
Visual	0.88	0.88	0.86
Aural	0.26	0.36	0.08
Manual	0.37	0.36	0.31
Vocal	0.56	0.64	0.31
Imaginal	0.86	0.87	0.80
Declarative	0.83	0.83	0.75
Goal	0.74	0.75	0.52
Procedural	0.40	0.41	0.23

7.1.1 Motor ROI / Manual Module

Motor activity occurs in all four conditions, since the last two responses need to be reported manually, independent of the condition. However, only the manual conditions require the first two responses to be keyed in. Motor behavior was only approximated in the model, by having it use a keyboard instead of operating a mouse and clicking on buttons.

As can be seen in Figure 7.1, the prediction for the manual module shows a relatively good fit for the *manual* condition, but is about 2 scans off. When the BOLD functions are shifted forward in time by 2 scans (that is 4 s), rise and fall of the predictions and the actual data overlap almost perfectly well up to scan 11, as can be seen in Figure 7.2. The BOLD responses drop to negative values for the subjects starting at scan 12 which cannot be predicted with the methodology used here. The subjects show a difference between *manual-sub* and *manual-nosub* which is not reflected in the model: The former condition has two peaks with a dip in activity in between, whereas the latter rises to a higher level and almost plateaus there (but still rising a bit more to a peak at scan 9). The activity predicted by the model rises just as fast as the activity in the brain imaging data for the first three scans of activity, but then continues to rise more gradually to the peak at scan 9. Interestingly, the model using the basic strategy shows more of a two peak structure, whereas the model using the convolution strategy clearly shows a plateau (plots not shown here). This pattern makes sense, since the responses for *num1* and *num2* are closer to each other and to the last two responses in time for the convolution strategy, thus resulting in sustained activity in the corresponding part of the motor cortex. For the basic strategy, this series of events is teased further apart, therefore resulting in two peaks. Thus, it could be conjectured that subjects rather use the convolution strategy in the *manual-nosub* condition and the more straight-forward basic strategy in the harder *manual-sub* condition.

Others interesting insights can be obtained by consulting the behavioral data. Since event-locked averaging was applied, the first two responses both by the subjects and the model fall in the time span during scan 3 and 4. Due to the sluggish nature of the BOLD response, it could be expected to see a peak or at least a sharp rise of activity 2 scans later. This is the case for the model, but not the BOLD response of the subjects which reaches a first peak at scan 4 already and thus precedes the model by 2 scans. It seems reasonable to attribute this to anticipatory effects as has been the case in other studies. For instance, [72] found unexpected early activity in the motor region during a study of geometry proofs. The same effect was found in [16] for a task which also included the manipulation of the output modality. Alternatively, the subjects might already move the mouse to the target before having finished their calculations. This is not be captured by the model as is uses a keyboard for manual responses. The latter explanation is made less likely as a similar pattern can be found for the second peak at scan 9 and 10 (for the subjects and the model, respectively). At this point during the task, the subjects have the mouse placed at the right spot already, such that the earlier rise in activity is likely to be caused by an anticipatory effect. The situation is comparable for the peak at scan 9 and 10 (subjects versus model) that shows in all the conditions. Given that the model keys in the last two responses at scan 5 and 8, this makes perfect sense. The response of the subjects, however, falls into scan 7 and 8 while their peak of activity occurs

already during scan 9, thus showing an early response. This is comparable to what was observed for the first peak.

Finally, two aspects can be observed in the verbal conditions. First, the subject data for *verbal-sub* and *verbal-nosub* shows differences in magnitude, albeit not in starting and ending time. In contrast to this, the model predicts almost the exact same rise for these two conditions, since its motor behavior is not influenced by the additional cognitive load of the retrieval of mappings. Second, the BOLD functions of the subjects rise much earlier than the functions predicted by the model, almost at the same time as the manual conditions, although no manual task needed be completed in these conditions. This might just reflect general activation. More specifically, the early rise could be explained by the fact that the representation of the facial muscles in the motor homunculus are located close the representation of the hands; thus, speech may drive activation in the motor ROI, that is actually associated with the use of the hands. This lack of separation of the two regions was already suggested in [16]. This effect cannot be reflected by the model, of course.

It can be concluded, that even though motor activity was only roughly modeled, it still predicts the data fairly well, apart from the anticipatory effect in all conditions and the early activation in the verbal conditions that are probably related to speech.

7.1.2 Fusiform Gyrus / Visual Module

In the computational model proposed for the arithmetic task, manual and visual activity are tightly coupled: the model looks at the screen not only to find the correct buttons to click on, but also to get visual feedback for the numbers that are keyed in. Unrelated to motor activity, the operators presented on the screen need to be visually processed for calculating the responses.

As can be seen in Figure 7.3, activity starts and ends at very similar times for the model and the subjects in all conditions. However, the curves of the model are about one scan wider as the curves flatten out towards the last scans. The model predicts the time course of the BOLD response fairly well for the two manual conditions, the only difference being that the model data shows a small dip in the plateau that is reached by the subjects.

The brain imaging data for the verbal conditions shows a clear two peak structure, that can also be observed in the model data. This seems reasonable as the verbal conditions involve less visual processing: No buttons need to be looked at for responding and no visual feedback is obtained from the response, such that the activity drops between the two bursts related to responses. However, the two maxima in the subject data have a similar amplitude, while the model predicts that the first peak is only about half the height of the second. Nonetheless, both the actual BOLD responses and the prediction peak at the same time (scan 4 and 10). Thus, the visual module shows a similar timing as the manual module, whose two peaks have the same distance from each other and occur just one scan before those of the visual module (scan 3 and 9). This seems reasonable in the sense that the visual module needs to perceive on the screen which buttons to press and then to perceive the output. The first is an activity that precedes motor activity, whereas the latter occurs afterwards. However, since we most likely see anticipatory effects in the motor ROI, the fusiform lags behind the peaks of the motor ROI.

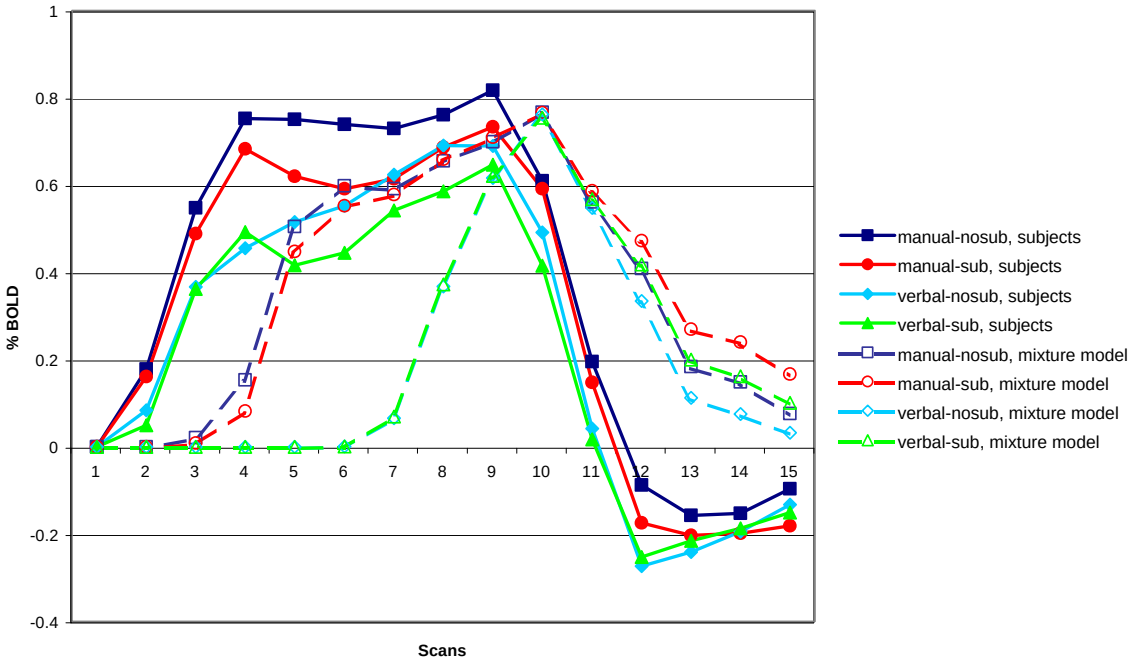

Figure 7.1: BOLD prediction for the manual module (Motor ROI).

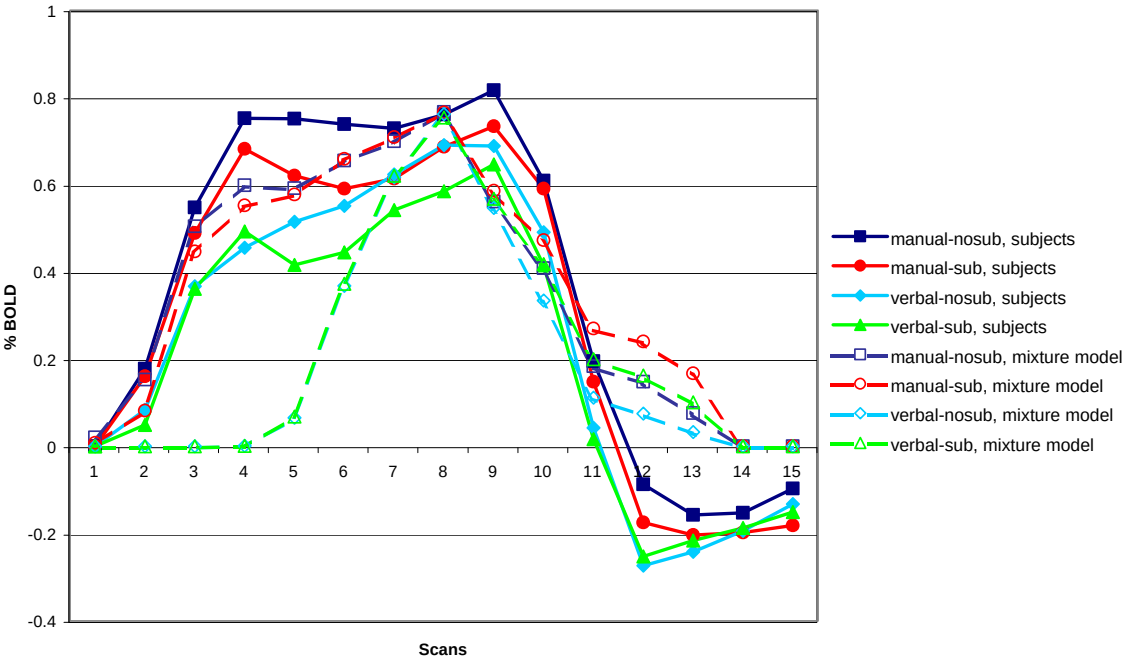

Figure 7.2: BOLD prediction for the manual module (Motor ROI), shifted forward in time by 2 scans (4 s).

In order to understand the difference in amplitude for the verbal conditions, it needs to be taken into account that the region in the fusiform gyrus considered here is involved in processing information in the focus of attention. For the model, the first peak is merely a result of the visual encoding from the screen whereas the second involves more visual processing (e.g. buttons on the screen, results typed in), thus resulting in a higher activation. However, checking whether a number was keyed in correctly is likely to engage this region of interest less than perceiving a number or a mapping with which a calculation has to be performed. It is probably due to the fact that the visual module of ACT-R does not distinguish between these different types of information processing that the predicted amplitudes differ from those shown by the subjects (the higher second peak is probably due to the fact that activity in the visual module has not decayed yet when the module is engaged a second time).

Finally, the fact that the prediction of the BOLD response at the end of a trial is overestimated seems to indicate that our assumption about the calculation of the last response (*newDenom*) is correct. We had hypothesized in section 4.5.2, that the denominators needed to calculate *newDenom* might either be re-encoded by looking at the screen again or simply retrieved from working memory (the imaginal buffer). The given model assumes the latter. Using the first (re-encode) strategy would lead to even more engagement of the visual module.

7.1.3 Vocal ROI / Vocal Module

Just as the manual and visual module influence each other, the vocal and aural modules are related. Both only become active in the model at the beginning of a trial (*num1*, *num2*) and only for the verbal conditions. Since the vocal module is only active twice per trial, one would assume that the prediction should be easy, given that we used the response times of the verbal and manual responses for fitting the model (based on the development set).

As can be seen in Figure 7.4, the two verbal responses result in a clear peak, as is the case both for the subject data and the data produced by the model. However, the magnitude of the BOLD responses differ as the activation for *verbal-sub* is higher than for *verbal-nosub* which is reversed in the model. Furthermore, the BOLD responses predicted by the model rises above base level for 8 scans, whereas the activity in the imaging data drops sharply after 6 scans. As can be seen in Table 7.2, the reason for this is that the model puts the two verbal responses further apart from each other than they are for the subjects.

However, the comparison of the prediction with the brain imaging data reveals two unexpected aspects that are not captured at all in the model: a very early onset of the BOLD response and activity in the manual conditions, although no speech is involved there.

First and most striking, the BOLD response of the subjects shows a very early rise, starting at scan 2 already, apparently before anything has been said by the subjects. For the interpretation of this finding, it is important to note that the response time for verbal responses is defined as the end of the utterance (see section 3.3.5). As mentioned before, the model predicts utterance times between 0.30 s and 0.75 s for the given problems (see Table 6.6). Let us assume that it takes both the model and the subjects about 0.5 s to utter a response. The mean verbal response times for the

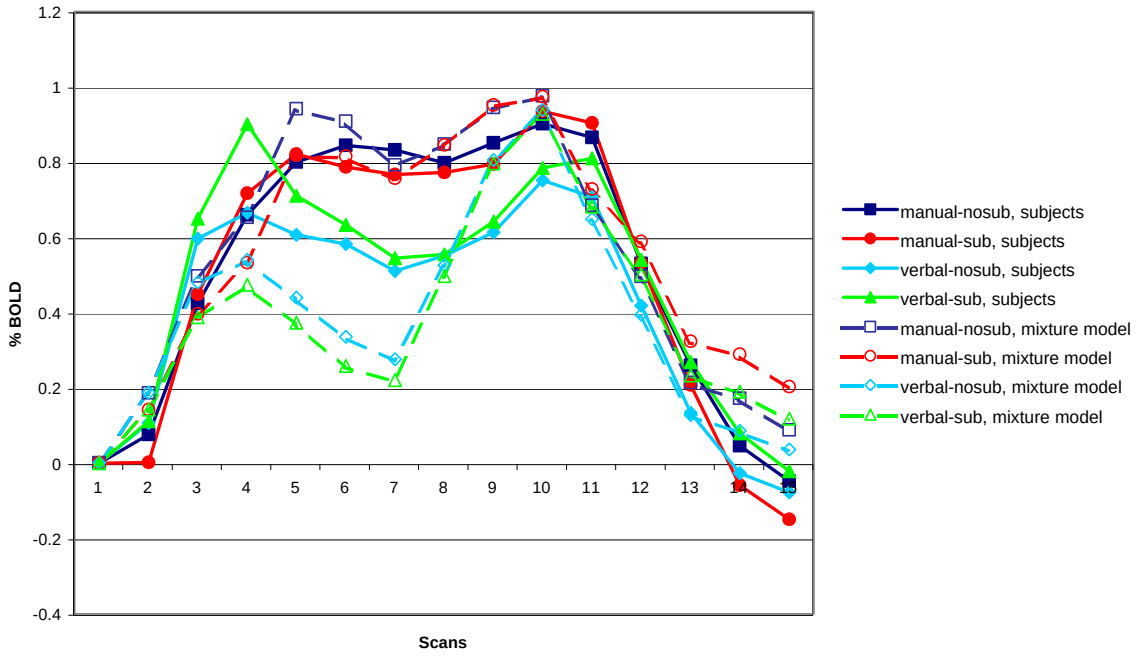

Figure 7.3: BOLD prediction for the visual module (Fusiform ROI).

subjects and the mixture model are provided in Table 7.2, as well as the estimated starting times. These times are not affected by event-locked averaging since they occur within the first four scans. As can be seen, subjects probably start speaking at the end of scan 2 (3.68 s) and in the middle of scan 3 (4.88 s) for *verbal-nosub* and *verbal-sub*, respectively. Thus, we can conclude that the BOLD response rises right when the subjects start to speak instead of being delayed as usual. Furthermore, the peak of the response is reached just a second after the completion of the verbal responses during scan 4 and thus 2 scans before the model predicts it.

This could be attributed to a synchronization issue; yet, a thorough analysis did not reveal any issue in synchronization of behavioral data with the data collected in the scanner. Furthermore, many other modules predict an onset of activity that aligns fairly well the actual brain imaging data (e.g. visual, declarative, imaginal, goal module), which we would not expect if the synchronization was off. A similar effect in this ROI, i.e. an early rise, was found in [16]. However, the effect in that study was triggered by a delay of the subjects before they were allowed to answer, such that one can assume that the region was involved in the rehearsal of the response which was not the case here.

The onset of the BOLD responses predicted by the model seems to make more sense: The model starts speaking 4 s or 5 s after stimulus onset (for *verbal-nosub* and *verbal-sub*, respectively) which is also when the predicted BOLD responses start to rise. However, a strong response should be seen at about second 8 (scan 4) which is not yet the case. As can be seen in Figure 7.4, the function reaches a peak at about second 11 (scan 6) and thus 2 scans later than expected.

Nonetheless, the shape of the BOLD response is predicted fairly accurately, as can be seen in Figure 7.5, where the prediction of the model is shifted forward in time by 4 s (2 scans).

Table 7.2: Mean verbal response times of the subjects in the complete database (absolute times), as well as the estimated onset times of the utterances (assuming an utterance length of 0.5 s). The same information is given for the probabilistic mixture model.

		Mean Response Time [s]		Estimated Onset [s]	
	condition	*num1*	*num2*	*num1*	*num2*
Subjects	verbal-nosub	4.18	5.83	3.68	5.33
	verbal-sub	5.38	7.81	4.88	7.31
Mixture model	verbal-nosub	4.41	6.59	3.91	6.09
	verbal-sub	5.41	8.09	4.91	7.59

The second effect that we found in the imaging data and which could not be predicted by the model is concerned with the manual conditions. The vocal ROI shows activity even for the manual conditions were nothing is spoken overtly, resulting in two small humps which reach their highest values during scan 4 and 11. This might be caused by the overlap of the manual and vocal ROI in the homunculus as mentioned before, leading to activation in the vocal ROI that is actually associated with the movement of the right hand (peaks in the motor ROI are at scan 3 and 9). However, the second hump is effectively too late to really be linked to motor activity during the trial, as it occurs two scans after the peak of motor activity. Alternatively, one could assume that subjects subvocalize their results, even when they are only required to enter them manually. This might partly account for the first rise in activity around scan 4. However, the second hump can again hardly be explained by that since the peak of this activation is reached at scan 11, that is 3 scans after the end of the trial. Thus, this last rise is hard to explain. Given that it occurs for all four conditions after the end of the trial proper, it does not seem to be related to any aspect specific to a condition or to the process of solving the given problem, even though the activation is stronger in the manual conditions. The most sensible explanation might be that subjects process and subvocalize the feedback given after the problem is solved (see Figure 3.5 (c)). This is supported by the sustained activation shown in the fusiform (visual activity), which only starts decaying at scan 11 (see Figure 7.3). However, this was not checked with the subjects, since a relatively large amount of time had passed between the fMRI recordings and the analysis.

7.1.4 Auditory ROI / Aural Module

The activity in the auditory ROI depends completely on the verbal responses given by the subjects, as no external auditory cues were given. The process of hearing was only captured by a single production rule in the model, that had it attend whatever unrequested sound chunk it found in its aural buffer.

First, it should be noted that although no external auditory cues were given, the fMRI scanner creates ambient noise which affects the auditory BOLD response and was not filtered out. For instance, [44] found strong responses to scanner noise in primary and secondary auditory cortex (1.5% signal change) which is relevant for this work since the auditory module is mapped to a region in secondary auditory

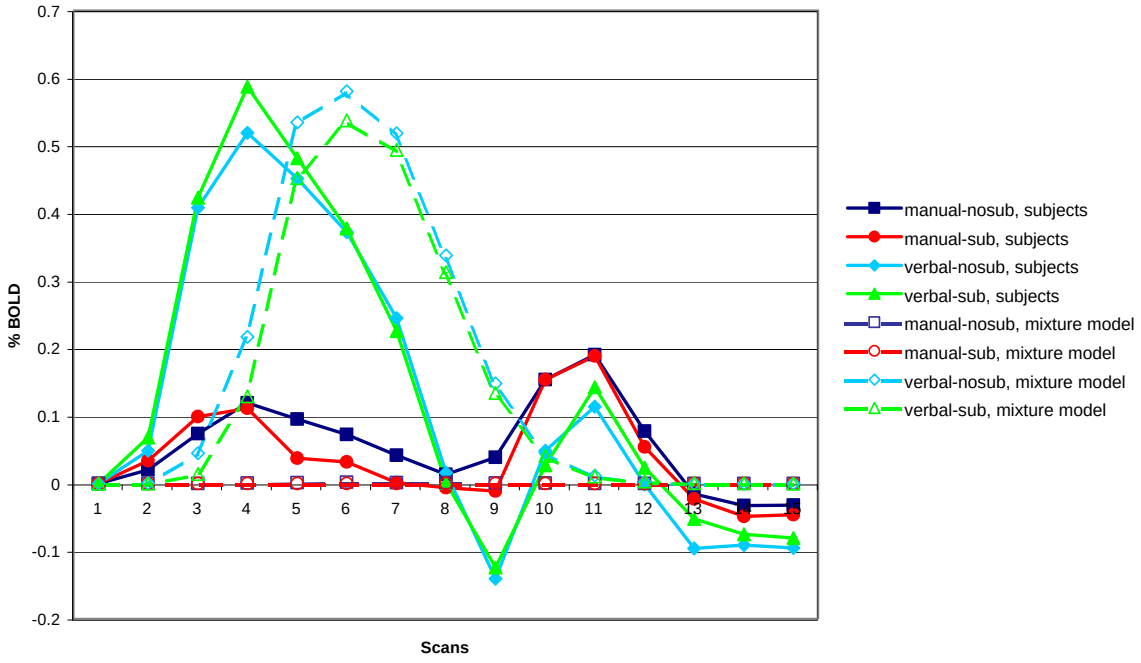

Figure 7.4: BOLD prediction for the vocal module (Vocal ROI).

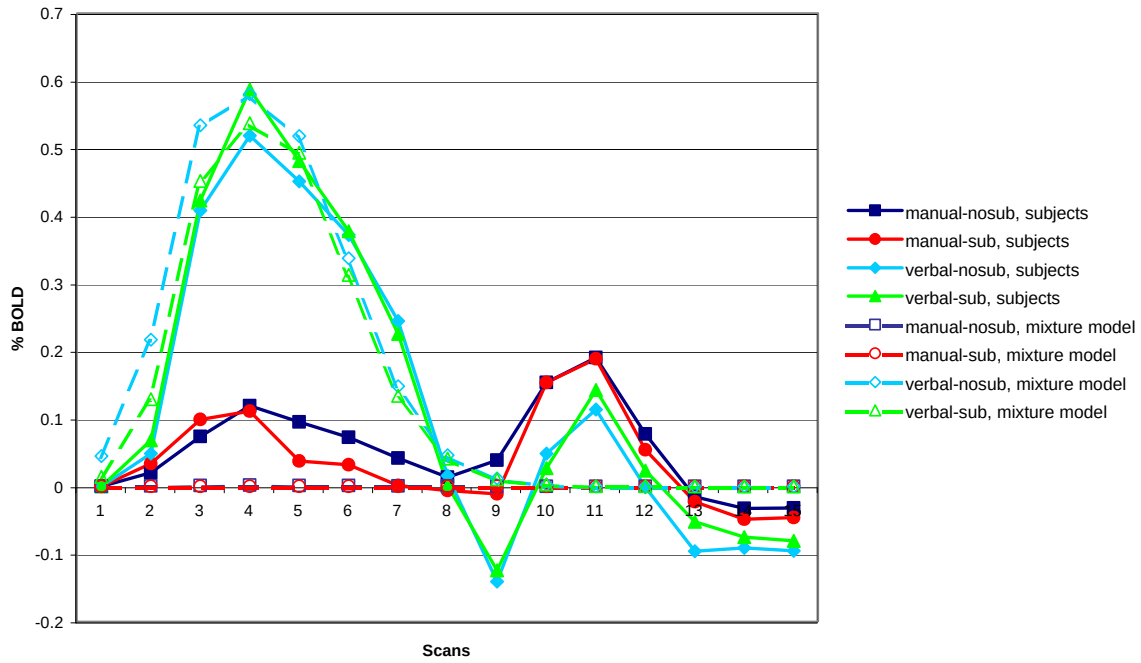

Figure 7.5: BOLD prediction for the vocal module (Vocal ROI), shifted forward in time by 2 scans (4 s).

cortex. However, the studies run in the John Anderson's laboratory typically find that the signal of the auditory cortex drops to negative values during the trial proper in all non-verbal studies. We surmise that doing the task distracts from the noise. Thus, the elevated baseline due to scanner noise returns to more normal levels during trials, thus producing a negativity (negative values in relation to the activation at the starting time of the trial). However, if an auditory signal is perceived, as is the case in the study discussed here for the verbal conditions, this adds to the negative values, thus resulting in a positive BOLD response for these conditions.

These effects can be be seen in Figure 7.6. As expected, the BOLD responses for the manual conditions drop to mostly negative values, while signal takes rises above baseline for the verbal conditions. Interestingly, the structure with two humps in the manual condition which was already found for the vocal ROI can be seen here again. This was not predicted by the model, as the model does not process any sound in the manual condition. As explained before, the responses for the verbal conditions are exclusively driven by the vocal activity of the subjects or the model and thus show a high correlation with the vocal BOLD responses (0.99 for the model and 0.85 for the subject data when only the verbal conditions are taken into account). With respect to the subject data, the BOLD responses for the aural and vocal ROI start and peak at the same time (rise at scan 2, peak at scan 4). The peak of activity coincides with the points of time when the second response is finished (5.83 s and 7.81 s for *verbal-nosub* and *verbal-sub*, see Table 7.2). However, the auditory activity drops to baseline level one scan earlier than for the vocal activity which is not reflected in the model. Corresponding similarities between the aural and vocal module can be found in the prediction of the model, as both functions rise at scan 2, peak at scan 6 and return to zero at scan 11. Thus, the model is again off by 2 scans. If shifted forward in time by 4 s (2 scans), the prediction of the aural module aligns fairly well with the actual data (up to scan 8), as is the case for the vocal module. However, in contrast to the vocal module, the model predicts a positive BOLD response for a slightly longer time span than is actually the case.

The similarity of onset times of the vocal and aural ROI is astonishing and has interesting implications. As argued before, the vocal BOLD responses rise very early and even before the first answers are given, thus showing an unexpected anticipatory effect. Given that the aural BOLD responses rise at the same time as that of the vocal ROI, this implies that there is also an anticipatory effect in the aural ROI. Interestingly, no comparable effect was found for the aural ROI in [16] (the above-mentioned study which also experimented with the output modality), even though an anticipatory effect for the vocal ROI was found, albeit under different conditions.

7.1.5 ACC ROI / Goal Module

In general, the goal module is used in order to keep track of the current goal that is pursued, in particular when a task involves a number of different steps. The arithmetic task is composed of four major steps, since four responses have to be given. Each of these need to be calculated and then reported, resulting in eight different subgoals, which are reflected in the use of the goal buffer.

As can be seen in Figure 7.8, the model predicts fairly similar BOLD responses for all four conditions. However, the model reaches a peak at scan 9, whereas the subject data shows a dip at this point, followed by a small rise. It should be noted

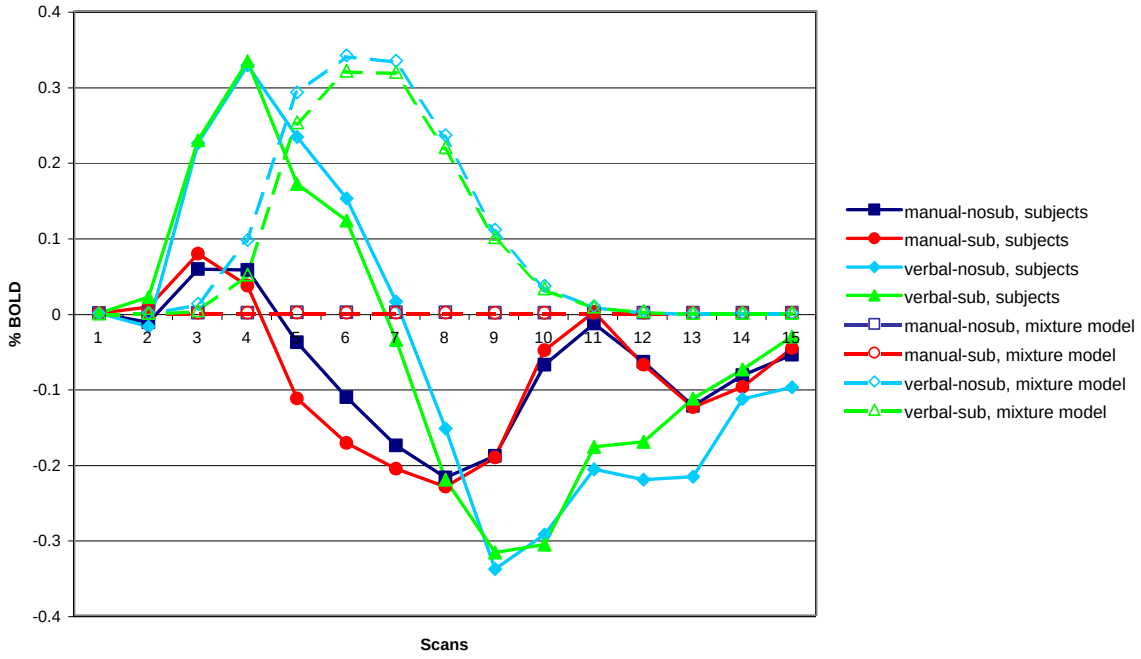

Figure 7.6: BOLD prediction for the aural module (Auditory ROI).

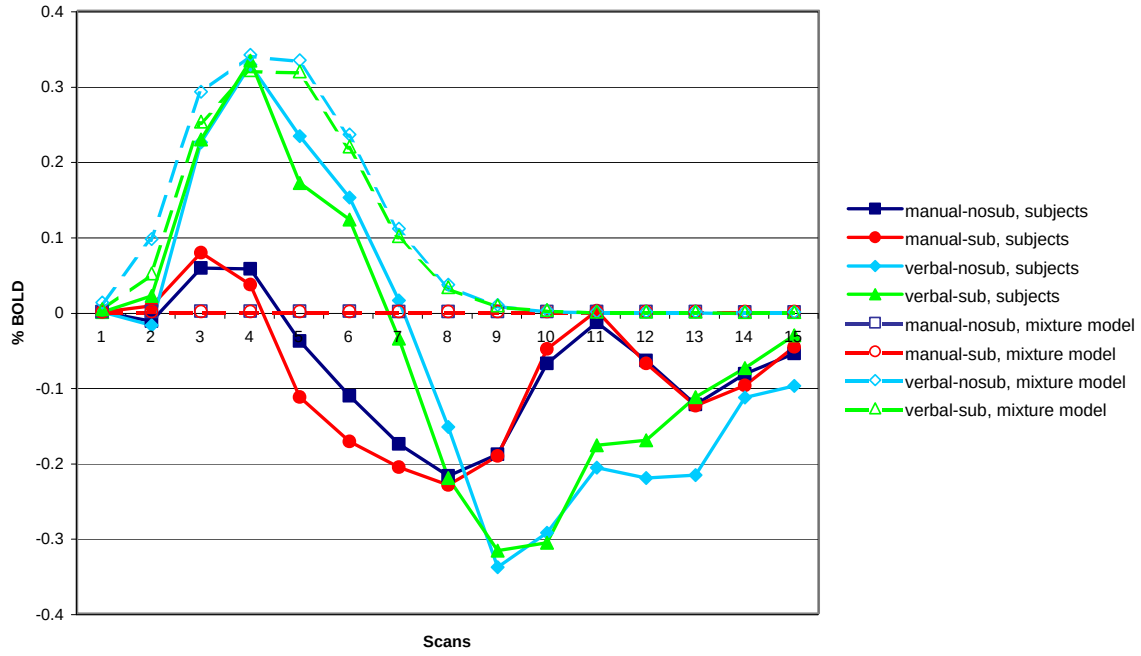

Figure 7.7: BOLD prediction for the aural module (Auditory ROI), shifted forward in time by 2 scans (4 s).

that scan 8 corresponds to the end of the trial, due to event-locked averaging that was applied to the data. However, the effects of the calculation and last response can still be seen until scan 10 because of the sluggish nature of the BOLD response. Thus, there seems to be a difference between the model and the subjects in how the last step (calculation and response of *newDenom*) is processed. Even though a small rise can be seen in the human BOLD response, suggesting that a subgoal is changed related to the output of the final response, the level of activity is lower than shown by the model. Essentially, the model handles each of the four steps in the same way, i.e. the goal module sets a new state for the calculation and the output of a response for each of them, even though setting that many control states would not have been necessary from a modeling point of view. Instead, the goal buffer could have been used less towards the end of each trial, since the first three steps are more complex than the last one. The imaging data suggests that this would have been more appropriate: it seems that the activity of the goal module or the associated ROI are not directly proportional to the number of steps taken (for this task), but rather to the complexity of these steps.

It has been reported in [38] that ongoing speech-monitoring also drives activation of the ACC, besides its other typical functions. Since speech and thus speech monitoring only occurs in the verbal conditions, this is likely to lead to different BOLD responses for the manual and the verbal conditions. This might be an explanation why the BOLD responses of the verbal conditions have a different shape than those of the manual conditions and also why their peaks occur earlier (at scan 4) and is more pronounced than those of the manual conditions. As can be seen in Figure 7.6, the auditory ROI also peaks at scan 4, even though it drops below baseline soon thereafter which is not the case for the ACC. This effect cannot predicted by the model, of course, which is one possible explanation for the discrepancy between prediction and actual imaging data for the verbal conditions.

In contrast, the prediction of the model for the manual conditions shows a fairly good match with the brain imaging data, reflecting that the activity is lower for the substitution condition (*manual-sub*) than if no substitution is involved (*manual-nosub*). Furthermore, it can be seen that the BOLD response of the model and that of the subjects rise at the same time (scan 1), but the responses of the subjects drop slightly sooner than those predicted by the model. Interestingly, the opposite can be observed when comparing the BOLD responses of the parietal to the activity of the imaginal module.

7.1.6 Parietal ROI (PPC) / Imaginal Module

In the computational model devised for modeling the task, the imaginal buffer is effectively used as working memory, that is for keeping intermediate results in mind as well as pieces of information shown on the screen. This leads to a relatively continuous, sustained activity of the imaginal module, as can be seen in Figure 7.9.

Both the predicted responses and the actual subject data show an almost immediate rise in activity. With respect to the model, this is caused by the fact that visual stimuli on the screen are perceived and need to be kept in mind for the first calculation, thus driving the activation of the imaginal module. However, in contrast to what can be observed when comparing the activity of the ACC and the goal module, the BOLD response of the parietal ROI drops later than that predicted by

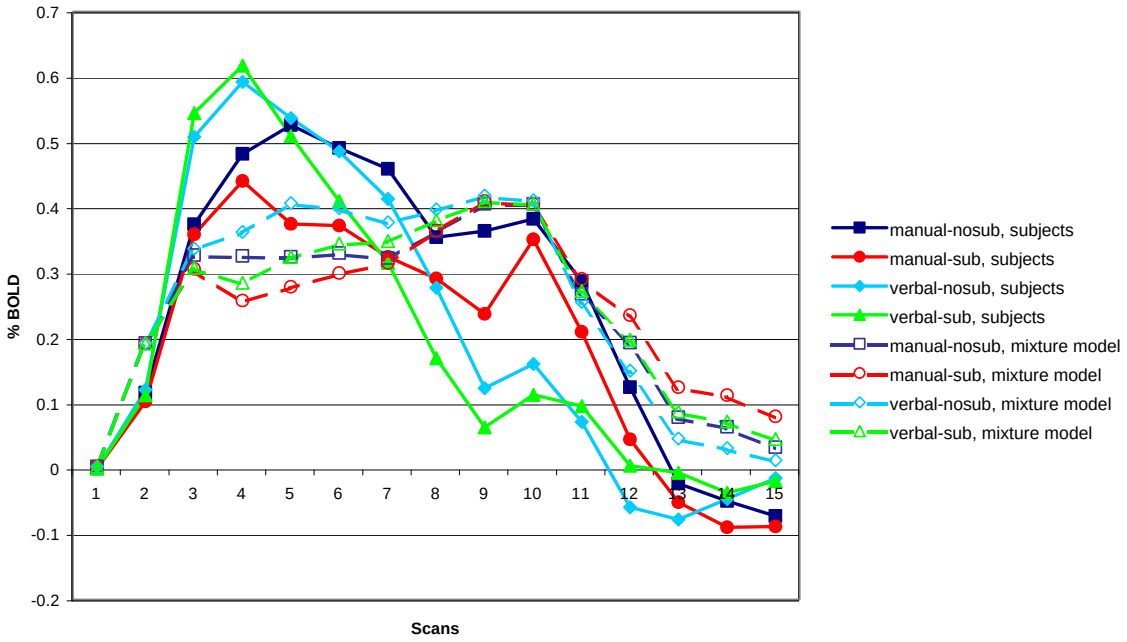

Figure 7.8: BOLD prediction for the goal module (ACC ROI).

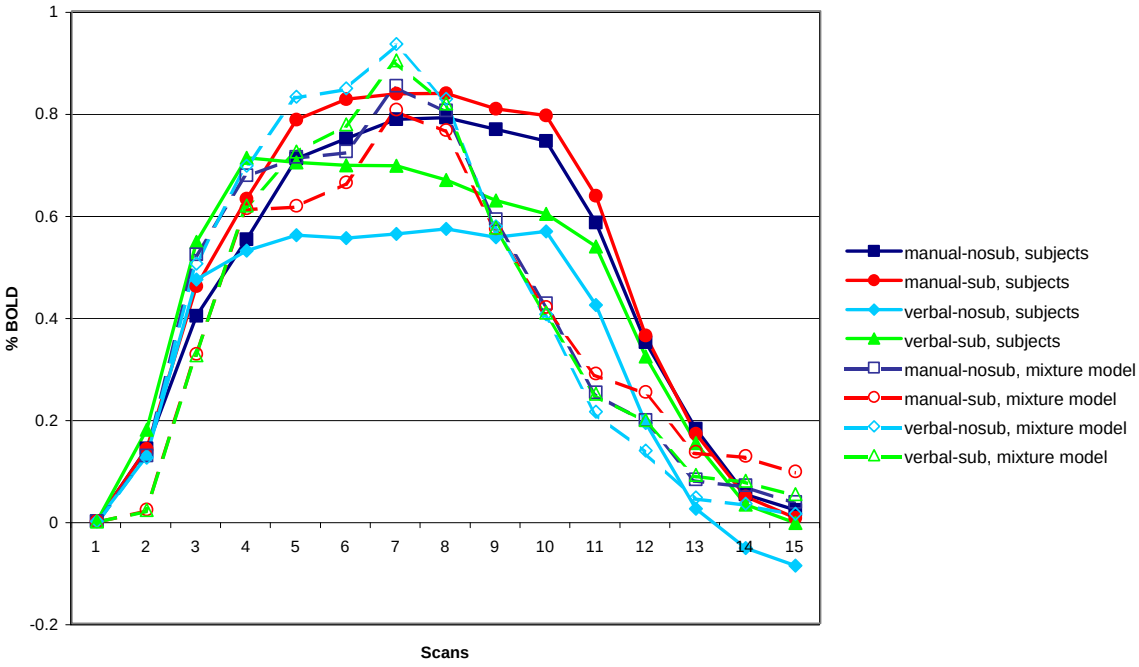

Figure 7.9: BOLD prediction for the imaginal module (Parietal ROI).

the imaginal module. Actually, the responses of the ACC and the parietal drop at about the same time (scan 11), whereas the activity of the two modules (goal and imaginal) does not align with each other.

Apart from the fact that the model predicts a shorter activity, it also suggests a peak towards the end of the trial (scan 7) which does not occur in the brain imaging data. Furthermore, it is surprising that the level of activation is completely reversed in the model's prediction: whereas the subject data shows the highest activation for *manual-sub*, followed by *manual-nosub*, *verbal-sub* and *verbal-nosub*, the exact opposite is the case for the model. However, it should be noted that the predictions of the model for the different conditions are fairly similar in amplitude, as the use of the imaginal buffer by the model does not differ much across conditions. Thus, the different order of the predictions should not be given too much weight. In contrast to that, the actual BOLD responses for the different conditions are spread further apart from each other. This might be related to the artifacts caused by the verbal protocols, leading to a lower magnitude of the BOLD response in the verbal conditions (refer to section 7.2.1 for more details).

7.1.7 Prefrontal ROI (LIPFC) / Declarative Module

The declarative module is used whenever a fact containing previous knowledge such as a multiplication-fact or mapping needs to be retrieved from declarative memory.

Just as for the prediction based on the activity of the imaginal module, the predicted order of the different conditions does not reflect the subject data well, as can be seen in Figure 7.10. However, it should be noted that this region proved to be highly variable over subjects, such that averaging might have blurred the order of the amplitudes. Even though the overall order is not predicted correctly by the model, one important aspect shows both in the subject data and the prediction: As hypothesized in section 3.3.4, a higher amplitude can be observed in the substitution condition compared to the no-substitution condition. This is the case for both output modalities. This result supports the assumed role of the prefrontal ROI for retrieval of declarative knowledge, as it reflects that two more retrievals are necessary in the substitution condition.

For the manual conditions, the model manages to predict the times for rise and fall of the BOLD responses fairly well. It should be noted, however, that the model predicts a plateau or at least just a gradual rise between scan 5 and scan 9, while a peak can be found in the imaging data at scan 6.

The situation for the verbal conditions is different. Although the timing of the rise of the response shows a good match, the activity starts dropping during scan 7 already for the subject data, whereas the model predicts that the response falls at about the same time as seen for the manual conditions. As for the imaginal module, this can be probably attributed to artifacts caused by the verbal protocols (see section 7.2.1).

7.1.8 Caudate ROI / Procedural Module

The procedural module is the core of the ACT-R production system, matching the patterns of buffer use and buffer content shown in a given situation with appropriate production rules. Although it is argued theoretically that this function

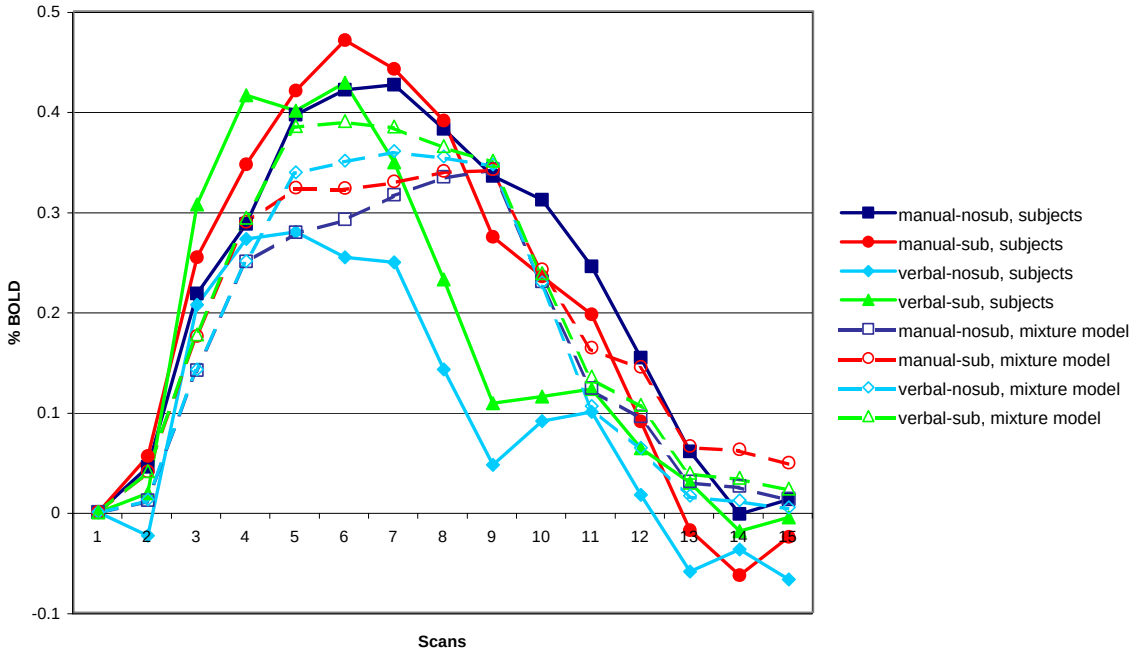

Figure 7.10: BOLD prediction for the declarative module (Prefrontal ROI).

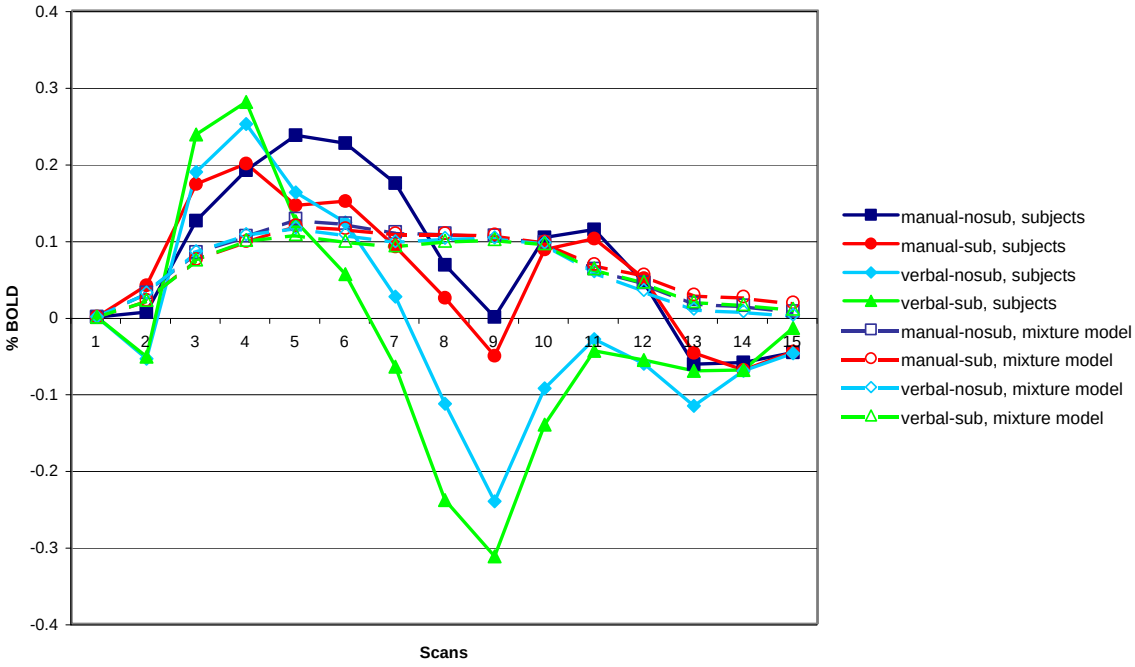

Figure 7.11: BOLD prediction for the procedural module (Caudate ROI).

should be associated with the basal ganglia (caudate ROI), most of the previous ACT-R fMRI studies found huge discrepancies between the predicted and the actual activity. It is well-known that activity in the caudate is not a pure measure for involvement of the procedural system, since it also shows a reinforcement response. As the thalamus does not show reinforcement effects while having a similar role in information-processing, this might be a promising region to look into as an alternative to the caudate. This could be investigated using the fMRI data collected in both the current and previous experiments.

The prediction for the caudate does not show a good fit as expected and is only included for the sake of completeness. It is interesting to note, however, that the two peak structure that could be seen in many of the other regions of interest, can also be observed in the caudate and even fits into the general picture timewise, with the first peak at scan 4-5 and the second at scan 11. As argued in [15], this might relate to visual activity, as it has been shown that the caudate responds to eye movements [58].

7.1.9 Summary

Both the manual and vocal ROI show a very early onset of activation, indicating anticipatory effects. The model predicts their respective peaks to occur about 7 seconds after the stimulus, thus lagging 4 seconds behind the brain imaging data of the subjects. In previous studies, anticipatory activity was found as well, but the time shift had been about 2 seconds (see [16], among others). Also, it can be observed that the vocal ROI shows activation contemporaneously with activity in the manual ROI during the end of the trial, even though subjects do not speak then. This can be explained by the fact that the activity of the manual and vocal ROI are related, as both involve parts of the homunculus. Hence, one activation may drive the other, as seems to be the case here. Neither this effect nor the anticipatory activation could be predicted by the model, which fits the data well otherwise.

The predefined fusiform and auditory ROIs reflect perception of the environment; to be more precise, they reflect advanced processing of visual and audio stimuli. Surprisingly, speaking and hearing are processed similarly fast, even though the verbal response has not been given yet at the time at which the BOLD response of the auditory ROI rises. The early onset of activity in the auditory ROI occurs 4 seconds earlier than would be expected, thus aligning with the onset of the vocal ROI. In a previous study that involved verbal responses, it could be shown that the speech and auditory regions were only affected when they should be [16], in contrast to our findings here. Probably, this is not related to a synchronization issue, since the timing of the other regions is about right (with only the manual ROI showing anticipatory activation). Thus, one could argue that anticipatory vocal activity produces anticipatory auditory activity, i.e. subjects prepare to move their facial muscles to speak and thus also expect to hear an audio signal upon preparation.

For the ACC, the model predicts the activity well for the manual conditions, even though the responses are predicted to be too similar over conditions. The latter is also the case for the parietal and the prefrontal ROIs. However, it should be noted that the verbal conditions are likely to be affected by artifacts, such that the differences between the output conditions might not only reflect differences in brain activity, but also artifacts. With the exception of the caudate, which is off by

far, as expected, the timing of the predictions for the cognitive ROIs matches the subject data well. Only for the parietal ROI, the activity is predicted to drop too early. However, for both the ACC and the prefrontal, the activity should be higher in the first half of the trial, whereas the opposite is the case for the parietal. These discrepancies point towards differences in the timing of the cognitive steps taken by the model and the subjects, respectively, thus providing us with insights about the strategies taken.

The brain imaging data seems to indicate that the strategy to keep the two denominators in mind for the calculation of *newDenom* instead of retrieving them from the screen is correct: the visual module shows too much activity towards the end, while the imaginal module shows too little activity compared to the subject data, even though the model already assumes that the two denominators are memorized. Had the model assumed that this was not the case, this would have lead to an even bigger discrepancy between prediction and subject data.

Concerning the use of the convolution versus the basic strategy, the shape of the BOLD responses for the manual ROI seem to suggest that the subjects rather use the convolution strategy in the easier *nosub* condition and the more straight-forward basic strategy in the harder *sub* condition. Finally, the activity pattern of the ACC shows a higher activation during the beginning of the trial which involves more complex steps than towards the end. In contrast to this, the model assumes that the activity would be distributed relatively uniformly over the length of the trial. Thus, the data suggests a different use of the goal buffer and the corresponding control sequences, respectively.

7.2 Assessment of the Remaining Hypotheses

7.2.1 Hypothesis 1

Hypothesis 1 stated that it was possible and sensible to use verbal protocols in the given fMRI study. As part of the larger project (refer to section 3.2), other project members developed the methodology necessary for both running fMRI experiments while recording verbal protocols and for processing the resulting data, neither of which was part of this thesis work. Thus, the principal technical challenges involved in recording audio data in the scanner and synchronizing it with the log files containing the remaining behavioral data were solved.

However, we have reason to assume that the imaging data recorded in the verbal conditions is affected by artifacts. As can be seen in section 7.1, the BOLD signal drops sharply as soon as the second response is given in the verbal conditions, i.e. at the end of scan 3 and 4 on average (*verbal-nosub* and *verbal-sub*, respectively, as shown in Table 7.2). This as can be observed for all the ROIs, except the parietal. It could be argued that this drop of activity (compared to baseline) was simply part of the actual signal. In this case, we would expect this drop only to show in regions in which activity is triggered by the task. However, it could be shown by Yulin Qin, a member of the project group, that this is not the case: The BOLD responses also drop in regions that typically show almost no activity related to tasks that are similar in nature to our arithmetic task. Thus, we conjecture that this drop of activity is likely caused by artifacts.

This points towards a fundamental issue of using verbal protocols which has yet to be solved. The drop of activity is probably related to the continuous changes in volume of air in the lungs of the subjects during articulation, as well as the tongue and jaw movement related to uttering a word. As pointed out in section 2.5.3, this leads to inhomogeneities in the magnetic field and can thus cause artifacts.

Techniques exist to handle artifacts induced by using having subjects use overt speech in an fMRI scanner (see [25]). One option might be to examine the effect that can be seen in brain regions that are unaffected by the task and then derive a method for subtracting this effect from the actual regions of interest. However, it is argued in [74], that it is probably not possible to detect real activation reliably by subtracting off the effect of an artifact, as the artifacts are not distributed uniformly over the images. A more promising approach seems to be to introduce pauses in volume acquisition, i.e. interrupting the scanning process during overt speech. It has been reported in the literature, that motion-related artifacts can be considerably reduced by this technique (see [25, 61], for instance). This might be a step to take for future research. However, as pointed out by [25], this techniques is substantially limited with regard to regions that are not affected by speech-related motion. Alternatively, using arterial spin labeling (ASL) might be an option, as it was shown that this technique offers some advantages over BOLD for studying language [74].

Thus, the methodological challenge of using verbal protocols in the given fMRI study is not solved yet, as the issue of artifacts remains to be solved. For this purpose, the above-mentioned techniques could be tested and evaluated.

7.2.2 Hypothesis 2

Hypothesis 2 assumed that the output modality would not affect the cognitive steps taken by the subjects to solve a given arithmetic problem. More precisely, we conjectured that the activity shown by the pre-defined prefrontal, parietal, anterior cingulate and caudate ROI would be independent of the output modality, which is in accordance with previous ACT-R fMRI research on the modality-specificity of cognition [16]. In the light of the discussion of Hypothesis 1 in the above section, it obviously becomes difficult to judge this hypothesis, given that the brain imaging data for the verbal conditions is likely affected by artifacts.

In accordance with Hypothesis 2, the computational model predicted BOLD responses that are fairly similar across conditions (for the ROIs associated with the cognitive modules). In contrast to the predictions, the actual brain imaging data shows clear differences between the manual and the verbal conditions, with the exception of the parietal. For the latter, the BOLD responses are fairly similar in shape and timing, independent of the condition. However, the amplitude of the BOLD responses is lower in the verbal conditions than it is in the manual conditions. For ACC, prefrontal and caudate, both the shape and the timing of the BOLD responses clearly differ between manual and verbal conditions. As pointed out in section 7.1.8, the results for the caudate should be handled with care, since most previous ACT-R fMRI studies showed mismatches for this region. For ACC and prefrontal, the BOLD response drops much earlier for the verbal conditions compared to that of the manual conditions, while the BOLD responses start rising at the same time (beginning of the trial), independent of the condition. On the one hand, the early drop of activity might be attributed to potential artifacts due to having

subjects speak in the scanner, as explained in the previous discussion of Hypothesis 1 (refer to section 7.2.1). If this is the case, we cannot evaluate Hypothesis 2. On the other hand, the difference between manual and verbal condition might be a result of the fact that the cognitive steps that subjects take are indeed affected by the output modality; thus, the results would refute Hypothesis 2. Given the high likelihood of artifacts due to the use of verbal protocols (see section 7.2.1), the first option is most probably correct. This would be in accordance with the results of a previous ACT-R fMRI study concerned with the modality-specificity of cognition ([16], for details refer to section 2.5.1). However, we cannot make a definite statement concerning Hypothesis 2. This will only be possible once the issue of artifacts caused by verbal protocols is solved.

Nonetheless, it should be pointed out that the parietal ROI does show similar brain activity over all conditions. As mentioned before, artifacts due to the susceptibility effect are not distributed uniformly over the images. Thus, the parietal might actually be a region that is unaffected by artifacts. In this case, the imaging data from the parietal would be in accordance with Hypothesis 2 by showing independency of output modality for this ROI.

7.2.3 Hypothesis 3

Hypothesis 3 stated that raising the level of difficulty by introducing substitutions in the task would have a similar impact for both the verbal and the manual conditions. Unfortunately, it seems likely that the brain imaging data recorded in the verbal conditions is affected by artifacts, as discussed before (refer to section 7.2.1). Thus, the situation for Hypothesis 3 is similar to that of Hypothesis 2, as neither of them can be evaluated completely. Nonetheless, some conjectures about Hypothesis 3 are provided in the following.

First of all, it is important to note that in most cases, the output modality has a higher impact on the BOLD response than does the level of difficulty. As can be seen in Table 7.3, BOLD responses with the same output modality have a higher correlation than those that share the same difficulty level. Nonetheless, difficulty level and output modality have a similar impact on the activity of the motor, fusiform and parietal ROI (associated with the manual, visual and imaginal module). In those cases, the activation pattern is generally fairly similar across conditions.

Table 7.3: Correlations between BOLD responses recorded in different conditions for each ROI (subject data). The column 'Common' denotes the common factor, i.e. 'manual' signifies that the correlation between *manual-nosub* and *manual-sub* is provided.

	Region of Interest							
Common	Fusiform	Auditory	Motor	Vocal	Parietal	Prefrontal	ACC	Caudate
manual	0.99	0.96	1.00	0.94	1.00	0.97	0.98	0.91
verbal	0.98	0.97	0.99	0.99	0.99	0.98	0.97	0.95
nosub	0.94	0.55	0.98	0.52	0.96	0.84	0.87	0.73
sub	0.94	0.45	0.99	0.46	0.97	0.91	0.84	0.70

For the most part, the model predicts that there was only a relatively small difference in brain activation between the substitution condition (*sub*) and the no-substitution condition (*nosub*). This is the case for both output modalities, i.e. the model fulfills Hypothesis 3. For all ROIs, the BOLD response is predicted to be slightly higher in the *nosub* condition compared to the *sub* condition, with the exception of the prefrontal ROI where this is reversed. As explained in section 7.1.7, the order predicted for the prefrontal ROI makes sense and is in accordance to the assumed role of this ROI for retrieval and to the first 9 scans of the subject data.

For the model, the effect of the two difficulty levels is fairly clear, as the BOLD responses keep their relative position to each other in most cases. In contrast, the subject data shows a more complex pattern, as the BOLD responses for the different conditions switch places more often (prefrontal, ACC, caudate). Whereas the model predicts the *nosub* condition to result in a higher BOLD response, this is only the case for a few ROIs (motor, ACC, caudate). Also, for half of the ROIs (fusiform, vocal, prefrontal, ACC), the effect of a raised difficulty level seems to depend on the output modality. Finally, it should be noted, that the differences in the BOLD signal between the *sub* and the *nosub* condition are fairly small for the fusiform, vocal and auditory ROIs.

Thus, even though we have reason to assume that the data collected in the verbal condition is affected by artifacts, the subject data seems to indicate that the effect of a raised level of difficulty depends on the output modality for some regions of interest. Furthermore, for most ROIs, there is almost no difference to be found between the BOLD responses recorded in the *sub* and *nosub* condition. Instead, the differences are mainly driven by the output condition. However, the differences that are shown by subjects suggest that the additional cognitive load of having to retrieve substitutions influences them beyond the mere retrieval of mappings.

8. Summary and Future Work

8.1 Summary

The main purpose of this thesis was to build an ACT-R model for a given arithmetic task and predict the BOLD response of subjects solving this task. In total, the data of 16 subjects was used, each of whom solved 100 problems in the fMRI scanner. They were asked to do so under four different conditions, varying the output (verbal versus manual) and the level of difficulty (including substitutions or not). Thus, the data that was used to evaluate the model included both behavioral data (response times) and brain imaging data (BOLD responses).

The main novelty of this study was to use verbal protocols in the scanner which enabled us to compare the activation of the vocal and the aural module of ACT-R to brain imaging data, as well as to test the modality-specificity of cognitive modules. This aspect had been investigated in ACT-R fMRI research once before, but on a different domain [16]. The study was devised in order to investigate the following hypotheses:

Hypothesis 1. It is possible to use verbal protocols in the given fMRI study.
Hypothesis 2. The output modality (verbal or manual) does not affect the BOLD responses of the predefined prefrontal, parietal, anterior cingulate or caudate ROI, i.e. the brain regions that are associated with the ACT-R modules of cognitive control (declarative, imaginal, goal and procedural module).
Hypothesis 3. The BOLD responses for the verbal and the manual output modality are affected in a similar way by the level of difficulty (manipulated by including substitutions or not).
Hypothesis 4. The BOLD responses shown by the subjects during the arithmetic task can be predicted by an ACT-R model for the eight predefined regions of interest.

Unfortunately, it seems likely that the brain imaging data recorded in the verbal conditions is affected by artifacts, probably due to the susceptibility effect. Solving this issue involves most probably a different recording method, as suggested in section 7.2.1. Thus, Hypothesis 1 cannot be answered for the time being. As a result,

Hypotheses 2 and 3 can hardly be evaluated either, as both assessments depend on the data recorded in the verbal condition. Nonetheless, it is important to note that the output modality seems to have a higher impact on the BOLD response than the level of difficulty.

The focus of this thesis was on Hypothesis 4, that is on building a model that is capable of predicting the BOLD responses for eight pre-defined regions of interest, shown by subjects while they are solving the task. As a first step, a deterministic model was developed based on what seemed reasonable assumptions and without consulting the subject data. A number of different strategies were implemented, most notably the so-called basic and convolution strategy as well as a mixture model that choses between these two with a given probability. Experiments showed that a mixture model that utilizes the basic strategy $\frac{2}{3}$ of the time and the convolution strategy $\frac{1}{3}$ of the time is capable of capturing the behavioral data of the subjects best. In order to reflect the variability in problem solving time shown by the subject population, a variability component was added to the previously deterministic model, based on a range of experiments.

This probabilistic model was then used to generate predictions of the BOLD responses for eight pre-defined regions of interests. The model was capable of predicting the shape of the BOLD responses for the four perceptual-motor regions very well; however, three of the predictions were shifted backwards in time compared to the actual brain imaging data. The prediction of the activity in the fusiform gyrus (visual activity) provided a remarkably good match with the subject data, whereas the regions defined for predicting motor, vocal and auditory activity consistently showed activity four seconds earlier than the model predicted. We conjecture that this shift is caused by anticipatory activity in these regions, which comes expected for the manual and vocal region, but surprisingly also includes the auditory ROI. This effect was not observed in the only previous ACT-R fMRI study for which subjects were asked to give overt verbal responses in the scanner [16].

While the model predicts the patterns of the BOLD responses well for the perceptual-motor regions, but with an offset in timing, the situation is almost reversed for the four regions involved in cognitive control (anterior cingulate cortex, parietal, prefrontal, caudate). For these, the time course of the predictions aligns well with the actual brain activity, but shows partly different patterns. The caudate turns out to be the only region where the prediction of the model is off by far which comes expected as this was the case in a number of previous studies relating ACT-R to brain imaging data.

To draw a conclusion, the model is capable of capturing different aspects for seven out of the eight regions of interests in its predictions. This result confirms Hypothesis 4 which states that the BOLD responses shown by the subjects during the arithmetic task can be predicted with an ACT-R model. It should be noted that this does not imply that the model proposed in this thesis is the only ACT-R model that could have been written for the given task; however, it is one of a restricted number of models that are possible and plausible for the given arithmetic task within the constraints of the computational framework of ACT-R.

8.2 Future Work

Experiments with the Model

The aim of this research was to arrive at a prediction of the BOLD responses, based on a computational model of cognition. However, it would also be informative to use the knowledge gained from the comparison of this prediction to the actual data in order to try to adapt the model correspondingly. Some novel effects that were found in this study could not be expected and were thus not modeled. Other results were due to implementation choices that could be changed. It might lead to some interesting insights whether these changes can lead to a more accurate estimate of the BOLD response. It should be noted, however, that this is a different approach to the one presented here, as the aim would not be to predict previously unseen BOLD responses, but to match well-known BOLD responses.

First, experiments could be run with the parameters of the gamma function that is used to estimate the Hemodynamic Response Function (HRF), as described by equation (2.12)). For the work presented here, the values of the scaling factor s for the time and the shape of the curve (determined by parameter a) was kept the same throughout the experiments. A better fit might be achieved by changing these parameters, potentially following insights gained by research on the estimation of HRFs (refer to section 2.5.2).

Second, the model itself could be changed. One step worthwhile trying would be to reduce the use of the goal buffer towards the end of the task when *newNum* and *newDenom* are reported. In order to stay consistent, each of the eight steps (calculation and perception of the four responses) were modeled such that they involved the same number of changes of the control states. However, this does not necessarily have to be the case for the information flow when the last two responses are caculated and reported.

Moreover, the imaginal module could be used less towards the end of the task. The model stores the two denominators in the imaginal buffer and retrieves them from there in order to calculate the last response. However, the brain imaging data suggests that subjects rather look at the screen to re-encode the two denominators, instead of keeping them in mind. Hence, this change would probably eliminate the peak of activity shown by the imaginal module towards the end of the trial, which does not occur in the subject data.

Related to this, the declarative module could be used differently. Instead of storing the results for the manual response in the imaginal buffer, they could be retrieved via the declarative module. Since the required chunks would have the highest activation in memory (as they were used last), they would be retrieved first. Thus, the timing would remain more or less the same, but the activity would be shifted from the imaginal to the declarative module, which is what the brain imaging data suggests.

Furthermore, the model could be changed in order to capture the manual output more accurately by having the model use a mouse as the subjects do. This would also allow to reflect anticipatory mouse movements which seems to be shown by the subjects, given their early activation of the motor region in the verbal conditions.

Changes in Methodology

Apart from the model, the methodology could be refined. Event-locked averaging (ELA) anchors the data around observable events to allow for averaging over trials without losing too much of the structure of the data. For the simulations presented in this thesis, only three landmarks were used (trial onset, problem solving time, last scan). Thus, the data was separated in two intervals, trial proper (8 scans) and cool down (7 scans). However, the method is not restricted to a certain number of landmarks. Therefore, it would be an interesting question to investigate whether better results could be achieved using more landmarks, i.e. other response times.

Moreover, we used event-locked averaging in the same way for all four conditions, although the mean problem solving time differs between conditions. Instead of defining 8 scans as the trial proper, independent of the condition, we could treat the *verbal-nosub* and *manual-sub* condition separately since their mean times are lower and higher than average, respectively (using 7 and 9 scans). This would most likely preserve more of the structure in the BOLD responses.

Additionally, it would make sense to re-evaluate which data points are chosen for the database. This concerns both outliers and the selection of data points in general. Although we did exclude outliers in the behavioral data, we did not exclude them in the brain imaging data obtained. Instead, we left two subjects out who showed a high amount of outliers. Event-locked averaging will shorten these long trial to the right length; however, the subject most likely followed different steps than those implemented in the model. Thus, it would most likely lead to an improvement of the fit of the prediction to exclude this 'cognitive noise'.

Furthermore, it should be taken into account that the subjects learned the number to letter mappings on the day before the actual scanning took place. Thus, it seems reasonable to assume that they were re-learning them by trial and error during the first block or two, leading both to a different response behavior and probably also to variance in the brain regions of interest. This aspect could be analyzed, both in order to obtain cleaner data (i.e. data that is more comparable) and to investigate the neural representation of this re-learning phase.

Follow-up Experiment

Beyond the steps described here, the results of this study suggest that it would be worthwhile to investigate the effect of anticipatory auditory activation further and to confirm or possibly refute the findings presented here. For this purpose, it would be valuable to record subjects in the scanner that receive two different types of auditory cues: one from an external signal and one produced by themselves by having them speak in the scanner. This would allow us to compare the timing of the activation driven by these two types of signals. The experiment would need to be designed in a way that subjects could not anticipate the external auditory cue, resulting in a delayed activity of the auditory region. In contrast, we would expect to see an anticipatory activation in this region if a subject prepares to speak. If we identified the correct region of interest and if the assumptions underlying the aural module in the ACT-R architecture are appropriate for external cues, we would expect to be able to predict the timing of the first activity, but not the latter. This could then also be understood as a confirmation of the findings presented here.

Comparison and Combination with Hidden Process Models

Apart from changes of the model and a follow-up experiment, we could gain new insights by combining the approach presented here with statistical machine learning methods, specifically a new probabilistic time series model called Hidden Process Models [68, 69]. As mentioned in section 2.5.2, Hidden Process Models (HPMs) are a probabilistic framework that can be used for estimating the onset times of overlapping hemodynamic responses, among others.

Modeling data with ACT-R allows us both to incorporate a priori knowledge about human cognition and to learn more about the types of processes that we assume to happen. In contrast, HPMs provide a powerful tool to learn the timings of these processes. Thus, ACT-R and HPMs could inform each other and provide us with a more thorough understanding of the processes in the subjects' brains. This could be particularly helpful in dealing with the challenge of finding out which different strategies humans pursue while solving a task.

As a first step, the data of the study could be modeled with HPMs to come up with predictions for the timing of mental processes. It should be noted that these should be based on the processes defined for the model presented here, but would need to be much less fine-grained. For instance, the eight different goal changes could be used (encoding and calculation for each of the four responses). Then, the outcome of the HPM estimation could be compared with the insights gained with the ACT-R model.

Subsequently, it would only be natural to extend this research project by developing a framework that combines these two general approaches, cognitive architectures and statistical machine learning methods. This framework could be used as a tool to interpret fMRI data in a novel way, since it would combine knowledge about human cognition with the ability of finding certain parameters of processes automatically.

A. Data

In this part of the appendix, a summary of the behavioral and BOLD data is given for both the subjects and the models.

				Percent correct	
Subject ID	**Age**	**Gender**	**Set**	**Training session**	**Scanning session**
21	19	Female	dev	80.00	96.00
23	21	Male	dev	95.00	88.00
24	23	Male	dev	87.50	98.00
25	25	Male	dev	95.00	95.00
26	22	Female	dev	82.50	89.00
27	22	Female	dev	95.00	96.00
28	19	Male	dev	95.00	99.00
29	27	Female	dev	87.50	92.00
30	28	Male	dev	95.00	96.00
32	21	Female	test	70.00	93.00
34	20	Female	test	90.00	94.00
35	23	Male	test	75.00	92.00
36	20	Male	test	82.50	93.00
37	missing	Male	test	82.50	88.00
39	19	Male	test	92.50	98.00
40	22	Female	test	80.00	93.00
development set	22.89			90.28	94.33
test set	20.83			81.79	93.00
complete database	22.07			86.56	93.75

Table A.1: Empirical data of the subjects in the database.

	Number of Correct Trials				
Subject ID	**Total**	**Verbal-nosub**	**Verbal-sub**	**Manual-nosub**	**Manual-sub**
21	96	25	24	24	23
23	88	24	17	22	25
24	98	25	23	25	25
25	95	24	22	24	25
26	89	22	20	23	24
27	96	23	24	25	24
28	99	24	25	25	25
29	92	24	23	21	24
30	96	24	25	22	25
32	93	24	24	24	21
34	94	22	24	24	24
35	92	24	20	23	25
36	93	22	23	26	22
37	88	21	20	23	24
39	98	25	24	25	24
40	93	22	24	23	24
development set	849	215	203	211	220
test set	651	160	159	168	164
complete database	1500	375	362	379	384

Table A.2: Number of correct trials per subject and condition. In total, each subject solved 100 problems (25 per condition).

	Number of Outliers				
Subject ID	**Total**	**Verbal-nosub**	**Verbal-sub**	**Manual-nosub**	**Manual-sub**
21	22	4	7	3	8
23	1	1	-	-	-
24	1	1	-	-	-
25	-	-	-	-	-
26	9	4	2	2	1
27	4	-	-	2	2
28	-	-	-	-	-
29	-	-	-	-	-
30	3	1	-	1	1
32	-	-	-	-	-
34	1	1	-	-	-
35	3	-	-	3	-
36	1	-	-	1	-
37	13	2	1	7	3
39	2	-	1	1	-
40	2	1	1	-	-
development set	40	11	9	8	12
test set	22	4	3	12	3
complete database	62	15	12	20	15

Table A.3: Outliers in subject database per subject and condition.

			Absolute Times				Relative Times			
Data Set	**Measure**	**Condition**	**num1**	**num2**	**newNum**	**newDenom**	**num1**	**num2**	**newNum**	**newDenom**
complete database	Mean	verbal-nosub	4.18	5.83	9.56	12.25	4.18	1.72	3.70	2.69
		verbal-sub	5.38	7.81	12.04	14.63	5.38	2.38	4.27	2.59
		manual-nosub	4.74	7.88	11.47	14.01	4.74	3.14	3.60	2.54
		manual-sub	6.32	10.39	14.07	16.50	6.32	4.08	4.48	2.43
	Variance	verbal-nosub	1.60	1.66	3.45	3.86	1.60	0.81	2.82	1.12
		verbal-sub	2.04	2.66	4.05	4.23	2.04	1.44	3.02	0.92
		manual-nosub	1.59	2.50	3.60	3.94	1.59	1.47	2.27	0.97
		manual-sub	2.45	3.76	4.74	4.98	2.45	2.27	2.45	0.80
development set	Mean	verbal-nosub	4.60	6.21	10.39	13.13	4.60	1.68	4.13	2.74
		verbal-sub	5.83	8.34	12.84	15.40	5.83	2.46	4.60	2.56
		manual-nosub	5.25	8.65	12.49	15.15	5.25	3.40	3.84	2.66
		manual-sub	6.70	11.19	15.17	17.63	6.70	4.48	4.48	2.46
	Variance	verbal-nosub	1.78	1.81	3.31	3.66	1.78	0.74	2.66	1.02
		verbal-sub	2.22	2.89	3.90	3.98	2.22	1.58	2.79	0.81
		manual-nosub	1.60	2.33	3.30	3.62	1.60	1.32	2.27	1.01
		manual-sub	2.42	3.79	4.77	4.87	2.42	2.42	2.57	0.71
test set	Mean	verbal-nosub	3.61	5.33	8.48	11.10	3.61	1.77	3.11	2.61
		verbal-sub	4.83	7.10	11.04	13.66	4.83	2.28	3.82	2.62
		manual-nosub	4.07	6.87	10.15	12.53	4.07	2.81	3.28	2.38
		manual-sub	5.81	9.36	12.63	15.02	5.81	3.55	4.48	2.39
	Variance	verbal-nosub	1.08	1.25	3.32	3.82	1.08	0.90	2.93	1.23
		verbal-sub	1.63	2.11	4.02	4.34	1.63	1.22	3.25	1.04
		manual-nosub	1.32	2.35	3.54	3.85	1.32	1.57	2.23	0.88
		manual-sub	2.40	3.45	4.30	4.72	2.40	1.93	2.22	0.90

Table A.4: Behavioral data of the subjects (mean and variance) for the complete dataset as well as for the two subsets in which it was split (development set and test set). For each of the conditions, the absolute and relative times are given.

Model	Strategy	Condition	Absolute Times num1	num2	newNum	newDenom	Relative Times num1	num2	newNum	newDenom
deterministic	basic	verbal-nosub	4.01	7.36	11.18	14.54	4.01	3.35	3.82	3.36
		verbal-sub	5.18	9.71	13.52	16.88	5.18	4.53	3.82	3.36
		manual-nosub	4.02	8.56	12.84	16.20	4.02	4.54	4.29	3.36
		manual-sub	5.19	10.90	15.19	18.55	5.19	5.71	4.29	3.36
	convolution	verbal-nosub	7.43	7.91	11.18	14.54	7.43	0.48	3.27	3.36
		verbal-sub	9.77	10.25	13.52	16.88	9.77	0.48	3.27	3.36
		manual-nosub	7.38	8.42	12.71	16.07	7.38	1.04	4.29	3.36
		manual-sub	9.73	10.77	15.06	18.42	9.73	1.04	4.29	3.36
	mixture	verbal-nosub	5.15	7.55	11.18	14.54	5.15	2.40	3.63	3.36
		verbal-sub	6.71	9.89	13.52	16.88	6.71	3.18	3.63	3.36
		manual-nosub	5.14	8.51	12.80	16.16	5.14	3.37	4.29	3.36
		manual-sub	6.70	10.86	15.14	18.50	6.70	4.15	4.29	3.36
probabilistic	basic	verbal-nosub	3.38	6.39	10.02	12.76	3.38	3.01	3.63	2.74
		verbal-sub	4.13	7.89	11.52	14.26	4.13	3.76	3.63	2.74
		manual-nosub	3.72	7.95	11.73	14.47	3.72	4.23	3.78	2.74
		manual-sub	4.47	9.45	13.23	15.97	4.47	4.98	3.78	2.74
	convolution	verbal-nosub	6.47	6.98	10.09	12.83	6.47	0.51	3.11	2.74
		verbal-sub	7.96	8.48	11.59	14.33	7.96	0.51	3.11	2.74
		manual-nosub	6.75	7.78	11.56	14.30	6.75	1.03	3.78	2.74
		manual-sub	8.25	9.28	13.06	15.80	8.25	1.03	3.78	2.74
	mixture	verbal-nosub	4.41	6.59	10.05	12.79	4.41	2.18	3.46	2.74
		verbal-sub	5.41	8.09	11.54	14.28	5.41	2.68	3.46	2.74
		manual-nosub	4.73	7.89	11.68	14.42	4.73	3.17	3.78	2.74
		manual-sub	5.73	9.39	13.17	15.91	5.73	3.66	3.78	2.74

Table A.5: Mean response times of the deterministic and the probabilistic model, using the basic strategy, the convolution strategy or the mixture model ($M(\frac{2}{3}, \frac{1}{3})$, i.e. the basic strategy is employed 2/3 of the time, the convolution 1/3 of the time). For each of the conditions, the absolute and relative times are given.

Bibliography

[1] Aguirre, G.K., E. Zarahn and M. D'Esposito: *The Variability of Human, BOLD Hemodynamic Responses.* NeuroImage, 8:360–369, 1998.

[2] Anderson, J.R.: *Language, memory, and thought.* Erlbaum, Hillsdale, NJ, 1976.

[3] Anderson, J.R.: *The Adaptive Character of Thought.* Erlbaum, Hillsdale, NJ, 1990.

[4] Anderson, J.R.: *Human symbol manipulation within an integrated cognitive architecture.* Cognitive Science, 29(3):313–341, 2005.

[5] Anderson, J.R.: *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press, 2007.

[6] Anderson, J.R., M.V. Albert and J.M. Fincham: *Tracing Problem Solving in Real Time: fMRI Analysis of the Subject-Paced Tower of Hanoi.* Journal of Cognitive Neuroscience, 17:1261–1274, 2005.

[7] Anderson, J.R., J.F. Anderson, J.L. Ferris, J.M. Fincham and K.-J. Jung: *The Lateral Inferior Prefrontal Cortex and Anterior Cingulate Cortex are Engaged at Different Stages in the Solution of Insight Problems.* In *Proceedings of the National Academy Sciences*, in press.

[8] Anderson, J.R., S.A. Betts, J.L. Ferris and J.M. Fincham: *Can neural imaging investigate learning in an educational task?* In *Proceedings of the 36th Annual Carnegie Symposium on Cognition*, Carnegie Mellon University, June 2-3 2009.

[9] Anderson, J.R., D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere and Y. Qin: *An integrated theory of the mind.* Psychological Review, 111, (4):1036–1060, 2004.

[10] Anderson, J.R. and G.H. Bower: *Human associative memory.* Washington: Winston and Sons, 1973.

[11] Anderson, J.R., D. Byrne, J. Fincham and P. Gunn: *Role of Prefrontal and Parietal Cortices in Associative Learning.* Cerebral Cortex, 18(4):904–914, 2008.

[12] Anderson, J.R., C.S. Carter, J.M. Fincham, Y. Qin, S.M. Ravizza and M. Rosenberg-Lee: *Using fMRI to Test Models of Complex Cognition.* Cognitive Science (submitted), 32:1323–1348.

[13] Anderson, J.R., J.M. Fincham, Y. Qin and A. Stocco: *A central circuit of the mind.* Trends in Cognitive Science, 12(4):136–143, 2008.

[14] Anderson, J.R., B. John, M.A. Just, P.A. Carpenter, D.E. Kieras and D.E. Meyer: *Production system models of complex cognition.* In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, pages 9–12, Hillsdale, NJ, 1995. Lawrence Erlbaum Associates.

[15] Anderson, J.R. and Y. Qin: *Using Brain Imaging to Extract the Structure of Complex Events at the Rational Time Band.* Journal of Cognitive Neuroscience, 20(9):1624–1636, 2008.

[16] Anderson, J.R., Y. Qin, K.-W. Jung and C.S. Carter: *Information-Processing Modules and their Relative Modality Specificity.* Cognitive Psychology, 54(3):185–217, 2007.

[17] Anderson, J.R., Y. Qin, M.-H. Sohn, V.A. Stenger and C. S. Carter: *An information-processing model of the BOLD response in symbol manipulation tasks.* Psychonomic Bulletin & Review, 10:241–261, 2003.

[18] Anderson, J.R., Y. Qin, V.A. Stenger and C.S. Carter: *The relationship of three cortical regions to an information-processing model.* Cognitive Neuroscience, 16:637–653, 2004.

[19] Anderson, J.R., L.M. Reder and C. Lebiere: *Working memory: Activation limitations on retrieval.* Cognitive Psychology, 30(3):221–256, 1996.

[20] Anderson, J.R., N.A. Taatgen and M.D. Byrne: *Learning to achieve perfect time sharing: Architectural implications of Hazeltine, Teague, & Ivry (2002).* Journal of Experimental Psychology: Human Perception and Performance, 31:749–761, 2005.

[21] Ashby, F.G. and E.M. Waldron: *The neuropsychological bases of category learning.* Current Directions in Psychological Science, 9:10–14, 2000.

[22] Bandettini, P.A., A. Jesmanowitz, E.C. Wong and J.S. Hyde: *Processing strategies for time-course data sets in functional MRI of the human brain.* Magnetic Resonance in Medicine, 30:161–173, 1993.

[23] Bandettini, P.A., E.C. Wong, R.S. Hinks, R.S. Tikofsky and J.S. Hyde: *Time course EPI of human brain function during task activation.* Magnetic Resonance in Medicine, 25:390–97, 1992.

[24] Barch, D.M., F.W. Sabb, C.S. Carter, T.S. Braver, D.C. Noll and J.D. Cohen: *Overt Verbal Responding during fMRI Scanning: Empirical Investigations of Problems and Potential Solutions.* NeuroImage, 10 (6):642–657, 1999.

[25] Birn, R.A., R.W. Cox and P.A. Bandettini: *Experimental designs and processing strategies for fMRI studies involving overt verbal responses.* NeuroImage, 23:1046–85, 2004.

[26] Birn, R.M., R.A. Bandettini, R.W. Cox and R. Shaker: *Event-related fMRI of tasks involving brief motion.* Human Brain Mapping, 7(2):106–114, 1999.

[27] Blessing, S. and J.R. Anderson: *How people learn to skip steps.* Journal of Experimental Psychology: Learning, Memory and Cognition, 22:576–598, 1996.

[28] Bothell, D.: *ACT-R 6.0 Reference Manual.* Technical Report, Retrieved April 15, 2009, from Carnegie Mellon University, Department of Psychology Website: http://act-r.psy.cmu.edu/actr6/reference-manual.pdf, 2008.

[29] Botvinick, M.M., T. S. Braver, D.M. Barch, C.S. Carter and J.D. Cohen: *Conflict monitoring and cognitive control.* Psychological Review, 108(3):624–652, 2001.

[30] Boynton, G.M., S.A. Engel, G.H. Glover and D.J. Heeger: *Linear systems analysis of functional magnetic resonance imaging in human V1.* Journal of Neuroscience, 16:4207–4221, 1996.

[31] Brown, J.W. and T.S. Braver: *Learned predictions of error likelihood in the anterior cingulate cortex.* Science, 307:1118–1121, 2005.

[32] Buckner, R.L., J. Goodman, M. Burock, M. Rotte, W. Koutstaal, D.L. Schachter, B.R. Rosen and A.M. Dale: *Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI.* Neuron, 20:285–296, 1998.

[33] Buckner, R.L., W.M. Kelley and S.E. Petersen: *Frontal cortex contributes to human memory formation.* Nature Neuroscience, 2:311–314, 1999.

[34] Cabeza, R., F. Dolcos, R. Graham and L. Nyberg: *Similarities and differences in the neural correlates of episodic memory retrieval and working memory.* NeuroImage, 16:317–330, 2002.

[35] Cabeza, R., F. Dolcos, S.E. Prince, H.J. Rice, D.H. Weissman and L. Nyberg: *Attention-related activity during episodic memory retrieval: a cross-function fMRI study.* Neuropsychologia, 41:390–99, 2003.

[36] Carbonell, J.G., C.A. Knoblock and S. Minton: *Architectures for intelligence*, chapter PRODIGY: An integrated architecture for planning and learning. Lawrence Erlbaum, Hillsdale, NJ, 1990.

[37] Carter, C.S., A.M. MacDonald, M. Botvinick, L.L. Ross, V.A. Stenger and D. Noll: *Parsing executive processes: Strategic versus evaluative functions of the anterior cingulate cortex.* Proceedings of the National Academy of Sciences of the USA, 97:1944–1948, 2000.

[38] Christoffels, I.K., E. Formisano and N.O. Schiller: *Neural correlates of verbal feedback processing: an fMRI study employing overt speech.* Human Brain Mapping, 28:868–879, 2007.

[39] Ciuciu, P., J.-B. Poline, G. Marrelec, J. Idier, C. Pallier and H. Benali: *Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment.* IEEE Transactions on Medical Imaging, 22(10):1235–1251, 2003.

[40] Clark, D and A.D. Wagner: *Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control.* Neuropsychologia 41, pages 304–317, 2003.

[41] Cohen, M.S.: *Parametric analysis of fMRI data using linear systems methods.* NeuroImage, 6:93–103, 1997.

[42] Dale, A.M. and R.L. Buckner: *Selective averaging of rapidly presented individual trials using fMRI.* Human Brain Mapping, 5:329–340, 1997.

[43] Danker, J.F.: *A Bayesian model of the left prefrontal cortex accounts for both previous experience and current context during declarative retrieval.* Cerebral Cortex, (in press).

[44] Deborah, A.H., A.Q. Summerfield, M.S. Goncalves, J.R. Foster, A.R. Palmer and R.W. Bowtell: *Time-course of the auditory BOLD response to scanner noise.* Magnetic Resonance in Medicine, 43(4):601–606, 2000.

[45] Dehaene, S.: *Sources of mathematical thinking: behavioral and brain-imaging evidence.* Science, 284:970–974, 1999.

[46] Dehaene, S., M. Piazza, P. Pinel and L. Cohen: *Three parietal circuits for number processing.* Cognitive Neuropsychology, 20 (3):487–506, 2003.

[47] Dehaene, S., M.I. Posner and D.M. Tucker: *Localization of a neural system for error detection and compensation.* Psychological Science, 5:303–305, 1994.

[48] Denes, G. and L. Pizzamiglio (editors): *Handbook of clinical and experimental neuropsychology.* Psychology Press, Hove, East Sussex, UK, 1999.

[49] D'Esposito, M., J.A. Detre, D.C. Alsop, R.K. Shin, S. Atlas and M. Grossman: *The neural basis of the central executive system of working memory.* Nature, 378:279–281, 1995.

[50] Doessel, O.: *Bildgebende Verfahren in der Medizin: von der Technik zur medizinischen Anwendung.* Springer, Berlin, Heidelberg, 2000.

[51] Falkenstein, M., J. Hohnbein and J. Hoorman: *Perspectives of Event-Related Potentials Research*, chapter Event related potential correlates of errors in reaction tasks, pages 287–296. Elsevier, 1995.

[52] Fletcher, P.C. and R.N.A. Henson: *Frontal lobes and human memory: insights from functional neuroimaging.* Brain, 124:849–881, 2001.

[53] Frank, M.J., B. Loughry and R.C. O'Reilly: *Interactions between the frontal cortex and basal ganglia in working memory: A computational model.* Cognitive, Affective, and Behavioral Neuroscience, 1:137–160, 2001.

[54] Friston, K.: *Human brain function*, chapter Introduction: experimental design and statistical parametric mapping, page 599ff. Elsevier Science & Technology Books, 2nd edition, 2003.

[55] Friston, K.J.: *Functional Neuroimaging: Technical Foundations*, chapter Statistical parametric mapping, pages 79–93. 1994.

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

[57] Geary, D.C. and J.G. Wiley: *Cognitive Addition: Strategy Choice and Speed-of-Processing Differences in Young and Elderly Adults.* Psychology and Aging, 6(3):474–483, 1991.

[58] Gerardin, E., S. Lehericy, J.B. Pochon, S. Tezenas du Montcel, J.F. Mangin, F. Poupon, Y. Agid, D. Le Bihan and C. Marsault: *Foot, hand, face and eye representation in the human striatum.* Cerebral Cortex, 13:162–169, 2003.

[59] Glover, G.H.: *Deconvolution of impulse response in event-related BOLD fMRI.* NeuroImage, 9:416–429, 1999.

[60] Goutte, C., F.A. Nielsen and L.K. Hansen: *Modeling the hemodynamic response in fMRI using smooth filters.* IEEE Transactions on Medical Imaging, 19(12):1188–1201, 2000.

[61] Gracco, Z.L., P. Tremblay and B. Pike: *Imaging speech production using fMRI.* NeuroImage, 26:294–301, 2005.

[62] Grill-Spector, K., N. Knouf and N. Kanwisher: *The fusiform face area subserves face perception, not generic within-category identification.* Nature Neuroscience, 7:555–562, 2004.

[63] Hecht, S. A.: *Individual solution processes while solving addition and multiplication math facts in adults.* Memory & Cognition, 27(6):1097–1107, 1999.

[64] Hikosaka, O., H. Nakahara, M.K. Rand, K. Sakai, Z. Lu and K. Nakamura: *Parallel neural networks for learning sequential procedures.* Trends in Neuroscience, 22:464–471, 1999.

[65] Houk, J.C. and S.P. Wise: *Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: Their role in planning and controlling action.* Cerebral Cortex, 5(2):95–110, 1995.

[66] Huang, J., T.H. Carr and Y. Cao: *Comparing cortical activation for silent and overt speech using event-related fMRI.* Human Brain Mapping, 15:39–53, 2001.

[67] Huettel, S.A., A.W. Song and G. McCarthy: *Functional Magnetic Resonance Imaging.* Sinauer Associates, Sunderland, Mass., 2004.

[68] Hutchinson, R.A., T.M. Mitchell and I. Rustandi: *Hidden Process Models.* In *Proceedings International Conference for Machine Leaning*, 2006.

[69] Hutchinson, R.A., R.S. Niculescu, T.A. Keller, I. Rustandi and T.M. Mitchell: *Modeling fMRI data generated by overlapping cognitive processes with unknown onsets using Hidden Process Models.* NeuroImage, 46:1:87–104, 2009.

[70] Jung, K.-J.: *Voice Elicitor v1.1.* Technical Report, Retrieved May 13, 2009, from Brain Imaging Research Center Website (University of Pittsburgh, Carnegie Mellon University): http://www.birc.pitt.edu/technotes/BIRC_TN_VERecording.pdf, 2006.

[71] Jung, K.-J., P. Prasad, Y. Qin and J.R. Anderson: *Extraction of overt verbal response from acoustic noise from the scanner in fMRI by use of segmented active noise cancellation.* Magnetic Resonance Imaging, 53:739–744, 2005.

[72] Kao, Y.S., S.A. Douglass, J.M. Fincham and J.R. Anderson: *Traveling the Second Bridge: Using fMRI to Assess an ACT-R Model of Geometry Proof.* (in revision).

[73] Kay, K.N., T. Naselaris, R.J. Prenger and J.L. Gallant: *Identifying natural images from human brain activity.* Nature, 452:352–55, 2008.

[74] Kemeny, S., F.Q. Ye, R. Birn and A.R. Braun: *Comparison of Continuous Overt Speech fMRI Using BOLD and Arterial Spin Labeling.* Human Brain Mapping, 24:173–183, 2005.

[75] Kieras, D.E. and D.E. Meyer: *The EPIC architecture: Principles of operation.* Technical Report, Retrieved April 15, 2009, from University of Michigan, Department of Electrical Engineering and Computer Science Website: ftp://www.eecs.umich.edu/people/kieras/EPICtutorial/EPICPrinOp.pdf, 1996.

[76] Kieras, D.E. and D.E. Meyer: *An overview of the EPIC architecture for cognition and performance with application to human-computer interaction.* Human-Computer Interaction, 12(4):391–438, 1997.

[77] Kwong, K.K., J.W. Belliveau, D.A. Chesler, I.E. Goldberg and R.M. Weisskoff: *Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation.* Proceedings of the National Academy of Sciences of the USA, 89:5675–79, 1992.

[78] Laird, J.E., A. Newell and P.S. Rosenbloom: *Soar: An architecture for general intelligence.* Artificial Intelligence, 33:1–64, 1987.

[79] Langley, P., J. E. Laird and S. Rogers: *Cognitive architectures: Research issues and challenges.* Cognitive Systems Research, 10(2):141–160, June 2009.

[80] Lebiere, C.: *The dynamics of cognition: An ACT-R model of cognitive arithmetic.* Kognitionswissenschaft, 8(1):5–19, 1999.

[81] Lepage, M., O. Ghaffar, L. Nyberg and E. Tulving: *Prefrontal cortex and episodic memory retrieval mode.* Proceedings of the National Academy of Sciences, 97:506–511, 2000.

[82] Levitt, M.H.: *Spin Dynamics: Basics of Nuclear Magnetic Resonance.* John Wiley & Sons, New York, 2008.

[83] MacDonald, A.W., J.D. Cohen, V.A. Stenger and C.S. Carter: *Dissociating the role of dorsolateral prefrontal and anterior cingulate cortex in cognitive control.* Science, 288:1835–1838, 2000.

[84] McCandliss, B.D., L. Cohen and S. Dehaene: *The visual word form area: expertise for reading in the fusiform gyrus.* Trends in Cognitive Science, 7:293–299, 2003.

[85] McGonigle, D.J., A.M. Howseman, B.S. Athwal, K.J. Friston, R.S.J. Frackowiak and A.P. Holmes: *Variability in fMRI: An Examination of Intersession Differences.* NeuroImage, 11:708–734, 2000.

[86] Minton, S.N.: *Quantitative results concerning the utility of explanation-based learning.* Artificial Intelligence, 42:363–391, 1990.

[87] Mitchell, T.M., S.V. Shinkareva, A. Carlson, K.-M. Chang, V.L. Malave, R.A. Mason and M.A. Just: *Predicting Human Brain Activity Associated with the Meanings of Nouns.* Science, 320:1191–5, 2008.

[88] Moore, J.E. and G. Zouridakis (editors): *Biomedical technology and devices handbook.* Boca Raton: CRC Press, 2004.

[89] Munhall, K.G.: *Functional imaging during speech production.* Acta Psychologica, 107:95–117, 2001.

[90] Newell, A.: *Visual Information Processing*, chapter You can't play 20 questions with nature and win. Academic Press, New York, 1973.

[91] Newell, A.: *Unified theories of cognition.* Harvard University Press, Cambridge, MA, 1990.

[92] Nielsen, F. A., L.K. Hansen, P. Toft, C. Goutte, N. Lange, S.C. Stroher, N. Morch, C. Svarer, R. Savoy, B. Rosen, E. Rostrup and P. Born: *Comparison of two convolution models for fMRI time series.* NeuroImage, 5:473, 1997.

[93] Ogawa, S., T.M. Lee, A.R. Kay and D.W. Tank: *Brain magnetic resonance imaging with contrast dependent on blood oxygenation.* Proceedings of the National Academy of Sciences of the USA, 87:9868–72, 1990.

[94] Ogawa, S., D.W. Tank, R. Menon, J.M. Ellermann and S.-G. Kim: *Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging.* Proceedings of the National Academy of Sciences of the USA, 89:5951–55, 1992.

[95] Palmer, E.D., H.J. Rosen, J.G. Ojemann, R.L. Buckner, W.M. Kelley and S.E. Petersen: *An event-related fMRI study of overt and covert word stem completion.* NeuroImage, 14:182–193, 2001.

[96] Paus, T.: *Primate anterior cingulate cortex: where motor control, drive and cognition interface.* Nature Reviews Neuroscience, 2:417–424, 2001.

[97] Paus, T., L. Koski, Z. Caramanos and C. Westbury: *Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies.* Neuroreport, 9:R37–R47, 1998.

[98] Posner, M.I. and S. Dehaene: *Attentional networks.* Trends in Neurosciences, 17:75–79, 1994.

[99] Qin, Y., C.S. Carter, E. Silk, V.A. Stenger, K. Fissell, A. Goode and J.R. Anderson .: *The change of the brain activation patterns as children learn algebra equation solving.* In *Proceedings of the National Academy of Sciences*, volume 101, pages 5686–5691, 2004.

[100] Qin, Y., M.-H. Sohn, J.R. Anderson, V.A. Stenger, K. Fissell and A. Goode et al.: *Predicting the practice effects on the blood oxygenation level-dependent (BOLD) function of fMRI in a symbolic manipulation task.* In *Proceedings of the National Academy of Sciences of the U.S.A.*, volume 100, pages 4951–4956, 2003.

[101] Raichle, M.E.: *Handbook of Physiology The Nervous System*, chapter Circulatory and metabolic correlates of brain function in normal humans, pages 5:643–74. American Physiological Society, Washington, DC, 1987.

[102] Raichle, M.E. and M.A. Mintun: *Brain Work and Brain Imaging.* Annual Review of Neuroscience, 29:449–76, 2006.

[103] Rajapakse, J.C., F. Kruggel, J. M. Maisog and D. Von Cramon: *Modeling hemodynamic response for analysis of functional MRI time-series.* Human Brain Mapping, 6:283–300, 1998.

[104] Redgrave, P., T.J. Prescott and K. Gurney: *The basal ganglia: A vertebrate solution to the selection problem?* Neuroscience, 89:1009–1023, 1999.

[105] Reichle, E.D, P.A. Carpenter and M.A. Just: *The neural basis of strategy and skill in in sentence-picture verification.* Cognitive Psychology, 40:261–295, 2000.

[106] Ritter, F.E., N.R. Shadbolt, D. Elliman, R. Young, F. Gobet and G.D. Baxter: *Techniques for modeling human performance in synthetic environments: A supplementary review.* Technical Report, Wright-Patterson Air Force Base, OH: Human Systems Information Analysis Center, 2003.

[107] Rosenberg-Lee, M., M. Lovett and J. R. Anderson: *Neural correlates of arithmetic calculation strategies.* (under revision).

[108] Rosenberg-Lee, M., M. Lovett and J.R. Anderson: *Strategy variation in multi-digit mental multiplication: fMRI, behavioral and modeling perspectives.* In *Proceedings of the Annual Meeting of the Cognitive Neuroscience Society*, page 72, San Francisco, CA, 2008.

[109] Rugg, M.D., P.C. Fletcher, C.D. Frith, R.S.J. Frackowiak and R.J. Dolan: *Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval.* Brain, 119:2073–2083, 1996.

[110] Rykhlevskaia, E., G. Gratton and M. Fabiani: *Combining structural and functional neuroimaging data for studying brain connectivity: A review.* Psychophysiology, 45:173–187, 2008.

[111] Salvucci, D.D.: *A model of eye movements and visual attention.* In *Proceedings of the International Conference on Cognitive Modeling*, pages 252–259, Veenendaal, The Netherlands, 2000. Universal Press.

[112] Schneider, D.W. and J.R. Anderson: *Asymmetric Switch Costs as Sequential Difficulty Effects.* (submitted), 2009.

[113] Schneider, W. and M.W. Cole: *The cognitive control network: Integrated cortical regions with dissociable functions.* NeuroImage, 16:343–360, 2007.

[114] Shapiro, D. and P. Langley: *Controlling physical agents through reactive logic programming.* In *Proceedings of the third annual conference on Autonomous Agents*, pages 386–387, 1999.

[115] Siegler, R.S.: *Strategy choice procedures and the development of multiplication skill.* Journal of Experimental Psychology, 117:258–275, 1988.

[116] Sohn, M.-H., M.V. Albert, K. Jung, C. Carter and J.R. Anderson: *Pay now or pay later: Preparatory conflict resolution in the anterior cingulate cortex and the prefrontal cortex.* Poster session presented at the 11th Annual Meeting of the Cognitive Neuroscience Society, San Francisco, 2004.

[117] Sohn, M.-H., M.V. Albert, K. Jung, C.S. Carter and J.R. Anderson: *Anticipatory conflict monitoring in the anterior cingulate cortex and the prefrontal cortex.* In *Proceedings of the National Academy of Sciences*, volume 104, pages 10330–10334, 2007.

[118] Sohn, M.-H., A. Goode, V.A. Stenger, K.-J. Jung, C.S. Carter and J.R. Anderson: *An information-processing model of three cortical regions: Evidence in episodic memory retrieval.* NeuroImage, 25:21–33, 2005.

[119] Sohn, M.H., A. Goode, V.A. Stenger, C.S. Carter and J.R. Anderson: *Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex.* In *Proceedings of the National Academy of Sciences*, volume 100, pages 7412–7417, 2003.

[120] Stocco, A. and J.R. Anderson: *Endogenous control and task representation: An fMRI study in algebraic problem solving.* Journal of Cognitive Neuroscience, 20(7):1300–1314, 2008.

[121] Sun, R., E. Merrill and T. Peterson: *From implicit skills to explicit knowledge: A bottom-up model of skill learning.* Cognitive Science, 25:203–244, 2001.

[122] Taatgen, N.A. and J.R. Anderson: *Handbook of Computational Psychology*, chapter Constraints in Cognitive Architectures, pages 170–185. Cambridge University Press, 2008.

[123] Taatgen, N.A., C. Lebiere and J.R. Anderson: *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*, chapter Modeling Paradigms in ACT-R, pages 29–52. Cambridge University Press, 2006.

[124] Villringer, A., J. Planck, C. Hock, L. Schleinkofer and U. Dirnagl: *Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults.* Neuroscience Letters, 154:1014, 1993.

[125] Wagner, A.D., A. Maril, R.A. Bjork and D.L. Schacter: *Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex.* NeuroImage, 14:1337–1347, 2001.

[126] Wagner, A.D., E.J. Parlagoev, J. Clark and R.A. Poldrack: *Recovering meaning left prefrontal cortex guides controlled semantic retrieval.* Neuron, 31:329–338, 2001.

[127] Woods, R.P., S.T. Grafton, C.J. Holmes, S.R. Cherry and J.C. Mazziotta: *Automated image registration: I. General methods and intrasubject intramodality validation.* Journal of Computer Assisted Tomography, 22(1):139–152, 1998.

[128] Zacks, J.M., J.M. Ollinger, M.A. Sheridan and B. Tversky: *A parametric study of mental spatial transformation of bodies.* NeuroImage, 16:857–872, 2002.