Abstract

Background
The irrelevant speech effect refers to a reduction in the immediate serial recall of lists of presented items when irrelevant auditory material is presented together with the items to be memorized. The ability to ignore irrelevant distractors is vital in increasingly demanding work environments and the capability to keep information in and out of mind is central in human information processing. Hence, the scope of the study of irrelevant speech effects ranges from implications for theories of attention and working memory, to practical consequences of irrelevant speech in work environments or command and control centers. Whereas the detrimental effect of irrelevant speech is well established in the behavioral literature, much less is known about its neurophysiological correlates. As evident from the different theoretical accounts of the irrelevant speech effect, there is no general agreement on which representational level is affected by the interference. The hypotheses cover different stages in the information flow, ranging across perception, attention, and memory.

Aim
The overall purpose of the present research is to provide an understanding of how the human brain functions when speech interferes with the processing of visually presented material in general, and, more specifically, to explore some possible propositions as to the representational level (e.g. perceptual, attentional, or mnemonic) of the irrelevant speech effect.

Methods
In the present thesis, behavioral experiments combined with PET methodology were used to investigate the neurophysiological mechanisms behind the irrelevant speech effect. In a network approach using structural equations modeling, we also investigated the functional and effective connectivity between different regions of the brain. In addition, we employed a new behavioral task paradigm known as the serial recognition task. This task has several potential benefits compared to a standard serial recall task. Most importantly, serial recognition requires primarily that the serial order is maintained in short-term memory while minimizing the need for a representation of the content of the presented items as well as output systems.

Results
In an initial study we used an exploratory, low level of working memory load. We found that the irrelevant speech effect correlated with a decrease of regional blood flow in
several areas previously associated with verbal working memory. These areas were located in the bilateral superior temporal and inferior/middle frontal cortices extending into Broca’s area on the left, as well as in the left inferior parietal cortex. The results suggest that areas associated with verbal working memory were broadly suppressed during irrelevant speech. In paper II, the working memory load was increased and corresponded to a conventional behavioral study of the irrelevant speech effect. The decreases in activity that were observed in the first low load study were not replicated in this study. However, in the reversed contrast, an increase in the bilateral dorsolateral prefrontal activity was observed. This area is frequently associated with executive aspects of working memory and attentional control mechanisms. In paper III we investigated the functional and effective connectivity in the dataset from study I. Initial analyses revealed an interaction between the level of irrelevant speech and the functional connectivity between the regions related to verbal working memory and the left medial temporal lobe. We investigated network interactions between the working memory network and the medial temporal lobe, including different functional sub-networks. No meaningful functional sub-network differed significantly between conditions except for the connections related to the medial temporal lobe. Based on these observations, we propose that the exposure of irrelevant speech results in a switch from a relatively pure processing of phonological representations to an engagement in processing of episodic representations and that this is indicated by a stronger interaction between the verbal working memory network and the medial temporal lobe. In the fourth and final study we observed a significant irrelevant speech effect of changing versus steady-state auditory items in a serial recognition task. This finding lends support for and extends previous empirical findings, suggesting that irrelevant speech has the potential to interfere with the coding of the order of the items to be memorized.

Conclusions

Taken together, we suggest that these studies provide findings as to the functional neuroanatomical nature of the irrelevant speech effect. Firstly, there is no single neurophysiological locus of the irrelevant speech effect. The results of the present studies do not exclude any of the hypothesized levels of representation at which the effect occurs, rather it appears that all levels are affected. Secondly, functional neuroimaging methods can provide additional information about the mechanisms underlying the irrelevant speech effect that is difficult, if not impossible, to obtain with behavioral methods alone. Finally, within the paired subtraction paradigm, changes in rCBF take place in areas that are inherently activated by the verbal working memory task itself. However, as reflected in the network analysis, dynamic changes between these areas and the medial temporal lobe, suggest that new areas are recruited according to task demands.
List of publications


Contents

1. Introduction .................................................................................................................................. 11
   1.1. The irrelevant speech effect ................................................................................................. 11
   1.2. A perceptual phenomenon? ................................................................................................ 12
   1.3. Interference in memory ....................................................................................................... 13
       1.3.1. A working memory account of the irrelevant speech effect ........................................ 13
           1.3.1.1. Short-term memory .................................................................................................. 14
           1.3.1.2. The working memory model .................................................................................. 15
           1.3.1.3. The phonological store hypothesis ........................................................................ 16
           1.3.2. The changing state hypothesis ................................................................................... 18
   1.4. Semantic influences ............................................................................................................ 21
       1.4.1. Effects of meaningfulness .......................................................................................... 21
       1.4.2. Semantic similarity effects ........................................................................................ 21
   1.5. Attentional resources ......................................................................................................... 22
   1.6. Locus of the irrelevant speech effect .................................................................................. 23

2. Cognitive neuroanatomy ........................................................................................................... 25
   2.1. Speech perception ................................................................................................................ 25
   2.2. Selective attention .............................................................................................................. 26
   2.3. The functional anatomy of working memory ....................................................................... 29
   2.4. The aim of the present thesis ............................................................................................. 30

3. Materials and methods ............................................................................................................. 33
   3.1. Study population ................................................................................................................ 33
   3.2. Behavioral paradigms ......................................................................................................... 33
       3.2.1. Serial recall ................................................................................................................... 34
       3.2.2. Serial repetition ........................................................................................................... 34
       3.2.3. Serial recognition ......................................................................................................... 34
   3.3. Auditory stimuli .................................................................................................................. 35
   3.4. Positron emission tomography ............................................................................................ 35
       3.4.1. Image preprocessing ..................................................................................................... 37
       3.4.2. Statistical modeling and inference ................................................................................ 38
   3.5. Functional connectivity and network analysis ....................................................................... 38

4. Results and discussion .............................................................................................................. 41
   4.1. Behavioral effects ................................................................................................................. 41
       4.1.1. Interaction between irrelevant speech and working memory load ................................ 41
       4.1.2. Irrelevant speech effect size and distribution ............................................................... 44
       4.1.3. Stability over time — behavioral vs. scanning environments ...................................... 47
       4.1.4. Effects of working memory load .................................................................................. 48
       4.1.5. Serial position effects .................................................................................................. 49
           4.1.5.1. Serial recall ............................................................................................................. 49
           4.1.5.2. Serial recognition ................................................................................................... 50
       4.1.6. Reaction times ............................................................................................................. 51
   4.2. Subjective reports ................................................................................................................ 52
       4.2.1. Choice of cognitive strategy ......................................................................................... 52
       4.2.2. Perceived difficulty of irrelevant speech ..................................................................... 53
       4.2.3. Perceived difficulty of load ......................................................................................... 55
   4.3. Effects of regional cerebral blood flow ............................................................................... 56
       4.3.1. Neural correlates of working memory ........................................................................ 56
       4.3.2. Neural correlates of irrelevant speech ......................................................................... 57
           4.3.2.1. Effects of speech compared to no speech .............................................................. 58
           4.3.2.2. Effects of varied speech ......................................................................................... 58
           4.3.2.3. A suppression of the verbal working memory network (study I) ....................... 58
4.3.2.4. A role for the prefrontal cortex in the irrelevant speech effect (study II) ............... 60
4.3.2.5. Effective connectivity and the finding of an MTL involvement (study III) ........... 62
5. Summary and conclusions ........................................................................................ 67
Sammanfattning av avhandlingen .............................................................................. 69
Acknowledgements ...................................................................................................... 71
References .................................................................................................................. 73
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Anterior cingulate cortex</td>
</tr>
<tr>
<td>BA</td>
<td>Brodmann’s area</td>
</tr>
<tr>
<td>Cdx</td>
<td>Cerebellum dexter</td>
</tr>
<tr>
<td>CV</td>
<td>Consonant-vowel</td>
</tr>
<tr>
<td>ERP</td>
<td>Event related potentials</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>FMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>IFC</td>
<td>Inferior frontal cortex</td>
</tr>
<tr>
<td>IPC</td>
<td>Inferior parietal cortex</td>
</tr>
<tr>
<td>LTM</td>
<td>Long-term memory</td>
</tr>
<tr>
<td>M</td>
<td>Motor cortex</td>
</tr>
<tr>
<td>MBq</td>
<td>Megabecquerel</td>
</tr>
<tr>
<td>mCi</td>
<td>Millicurie</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>MTL</td>
<td>Medial temporal lobe</td>
</tr>
<tr>
<td>O-OER</td>
<td>Object-oriented episodic record model</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PFC</td>
<td>Prefrontal cortex</td>
</tr>
<tr>
<td>rCBF</td>
<td>Regional cerebral blood flow</td>
</tr>
<tr>
<td>RT</td>
<td>Reaction time</td>
</tr>
<tr>
<td>SEM</td>
<td>Structural equations modeling</td>
</tr>
<tr>
<td>SPM</td>
<td>Statistical parametric mapping</td>
</tr>
<tr>
<td>STG</td>
<td>Superior temporal gyrus</td>
</tr>
<tr>
<td>TMS</td>
<td>Transcranial magnetic stimulation</td>
</tr>
</tbody>
</table>
1. Introduction

This thesis is about the neurophysiological effects of auditory distraction. It revolves around the so called irrelevant speech effect, which is the observation that even ignored speech or sound can cause an involuntary disruption of a visually presented working memory task. More specifically, consider any attention demanding task. In order to succeed on the task it is of vital importance to keep certain relevant information in focus, and to ignore irrelevant information. Are any of the irrelevant stimuli registered? If so, to what extent are they processed? The fate of these unattended stimuli is one of the main conceptual questions in the study of selective attention.

The irrelevant speech effect is important to study for several reasons. In general terms, the study of cognitive processes that arise without active attention may give insights in auditory cognition and lead to the further development of theories of memory and attention. From a human factors perspective (for reviews, see Banbury, Macken, Tremblay and Jones, 2001; Banbury and Berry, 2005; Beaman, 2005), the irrelevant speech effect has rendered considerable interest since the effect occurs at very low levels of sound pressure and does not appear to diminish with repeated exposure. The conventional measure taken in abatement technology of reducing the sound pressure will therefore not be effective in this case. Surrounding noise and speech are one of the most frequently mentioned sources of disturbance in the working environment (Banbury and Berry, 1998). Another concern about the interference from irrelevant sounds is related to safety issues. For example, human errors caused by distraction have been estimated to explain a substantial part of aeronautical accidents (Banbury, Fricker, Tremblay and Emery, 2003). Consequently, knowledge of how to reduce auditory distraction may lead to increased efficiency and/or safety within a range of workplace situations.

1.1. The irrelevant speech effect

The focus of the present thesis is the deleterious effect of irrelevant auditory information upon recalling a list of unrelated verbal items in correct serial order. It is known as the unattended speech effect (Colle and Welsh, 1976; Salame and Baddeley, 1982), the irrelevant speech effect (Jones and Morris, 1992), or the irrelevant sound effect (Beaman and Jones, 1997). Basically, they refer to the same phenomenon. Although the irrelevant speech effect is an empirically robust effect, the effect size can vary substantially depending on experimental design. The effect is highly replicable and applies to most individuals. A consistent finding is that the effect is most pronounced in experimental tasks involving temporary memory for serial order (Salame and Baddeley, 1990; Jones and Macken, 1993; Beaman and Jones, 1997). However, it has been demonstrated in naturalistic tasks such as mental arithmetic and memory for prose as well (Banbury and Berry, 1997).
There are two lines of research in the behavioral study of the irrelevant speech effect. Either the property of the task is manipulated, or the property of the speech is manipulated. Research has concentrated on these variables likely because they are open to direct observation, as opposed to the cognitive locus of the effect, which remains a theoretical controversy (LeCompte, 1994; LeCompte, 1996; Beaman and Jones, 1997; Beaman and Jones, 1998; Baddeley, 2000; Jones and Tremblay, 2000; Neath, 2000; Henson, Hartley, Burgess, Hitch and Flude, 2003; Larsen and Baddeley, 2003; Neath, Farley and Surprenant, 2003; Page and Norris, 2003). Theories of the irrelevant speech effect can be divided into those that explain it as an effect of distraction (Broadbent, 1983; Cowan, 1995), or as being a consequence of contamination between the visual and auditory material. The contamination approach can be further subdivided into theories of contamination by content (Salame and Baddeley, 1982; LeCompte, 1996; Neath, 2000) or by process (Jones, Madden and Miles, 1992).

Since the first observation of the irrelevant speech effect (Colle and Welsh, 1976), various possible explanations of the effect have been investigated. The hypotheses of the effect origin correspond to different levels of representation in information processing, ranging from the physical level such as intensity of the sound, across the perceptual and attentional level, to the mnemonic and language level, including phonological, lexical and semantic contributions to the effect. While sensory processing, attention and memory are interrelated phenomena, they are often studied separately and the nature of their interaction is seldom addressed in detail. As will be evident from the scope of the present thesis, the irrelevant speech effect provides an interface to the interaction between these processes. However, the purpose of the present thesis is not to attempt to entangle the subtleties of the different theoretical models of the irrelevant speech effect. Rather, the emphasis is put on which level of representation for the effect locus that is predicted from these models, and in particular, what may be derived from these predictions in terms of candidate neurophysiological correlates of the effect. To this end, a functional imaging approach to the study of the irrelevant speech effect serves to further illuminate its underlying mechanisms as hypothesized by the different theoretical accounts.

1.2. A perceptual phenomenon?
The irrelevant speech effect implies that information reaching the brain from the eye is subject to interference from information reaching the brain from the ear. Intuitively, because information is coming from different sensory modalities, it seems unlikely that there is some kind of masking at the sensory level. Still, early theoretical interest in the irrelevant speech effect advocated that the effect of irrelevant speech was essentially a perceptual phenomenon (Broadbent, 1983). The idea behind this position is that the sound impairs the initial registration or encoding of the items from the visual list. In
keeping with this hypothesis, one has to assume that the effect occurs between phonological codes, that is, between a translated code of the memory items that were visually presented and the irrelevant speech, and that this interference is somehow taking place at a perceptual level.

The disruption at encoding idea is facing several difficulties. If the irrelevant speech effect is a perceptual phenomenon, it should affect any visually presented task. In contrast, there is abundant evidence that the effect is generally not observed in non-memorial tasks (e.g. Salame and Baddeley, 1982; Baddeley and Salame, 1986; Salame and Baddeley, 1990; Jones and Macken, 1993; Beaman and Jones, 1997). These findings confirm that the visually presented stimuli are accurately perceived in these tasks, despite the exposure to irrelevant speech. Secondly, irrelevant speech that is presented after the encoding period, during the rehearsal period of that task, also produce a significant disruption in performance (Miles, Jones and Madden, 1991). A third finding that is difficult to reconcile with the perceptual approach is that if the irrelevant speech is arranged so that the irrelevant items are synchronized as to coincide with the items to-be-remembered, then the disruption is not more pronounced than if the auditory items are presented in between the visual items (Salame and Baddeley, 1982; Jones, 1994). Taken together, this pattern of results strongly suggests that the interference takes place at a later (downstream) information processing stage, presumably within (short-term) memory.

1.3. Interference in memory

When assuming that the contamination between the relevant and irrelevant material is primarily caused by its content, the most influential approach is the working memory account. It is based on the assumption that there is confusion between the actual items held in memory and the items presented auditorily. It should be noted that there are other content-based approaches to the irrelevant speech effect, such as the Feature model (Nairne, 1990; Neath, 2000), which differ in many respects pertaining to the assumptions about the mechanisms of the irrelevant speech effect as well as the view in which information is stored and maintained in memory. However, they share the common perspective that the irrelevant speech effect is primarily due to interference by content, not process. Consequently, regardless of theoretical differences, content-based theories of the irrelevant speech effect would seem to predict that the interference takes place at a single locus where both the relevant as well as the irrelevant items are stored in memory.

1.3.1. A working memory account of the irrelevant speech effect

The working memory model by Baddeley and colleagues (Baddeley and Hitch, 1974; Baddeley, 1986; Baddeley, 1992; Baddeley, 2000; Baddeley, 2003) is no doubt the most
widely used framework for studies investigating the temporary on-line monitoring, storage and manipulation of information. This has been true also for the irrelevant speech phenomenon. However still defended, during the past few years, the working memory model has been competing with several other theoretical views of the irrelevant speech effect. In order to comprehend the arguments for and against the working memory position, we now turn to a brief overview of the model itself and its development.

1.3.1.1. Short-term memory

Before the emergence of the information processing approach to human cognition in the fifties, memory was viewed as a unitary system. Although there had been some early suggestions of a possible division between a short-term and a long-term memory (Hebb, 1949; Miller, 1956; Broadbent, 1958), it was not until Brown (Brown, 1958) and Petersons (Peterson and Peterson, 1959) published their results that such a division received a broader scientific acceptance. Independently of each other, they showed that even for small amounts of information, active rehearsal was necessary for successful immediate recall. These demonstrations provided strong support for the existence of a temporary short-term memory system that was limited in storage capacity. Already at that time, there were several behavioral findings that were consistent with this suggestion. For example, it was observed that when subjects are presented with a list of items to be remembered in no particular order, recall is typically improved for the first and last items compared to the items in the middle part of the list. This is known as the primacy and recency effects in free recall (Atkinson and Shiffrin, 1968) and was taken to indicate two difference memory systems, one short-term system for the last items presented (still in a temporary buffer), and one long-term system for the earlier presented items (since the items had been rehearsed, memory for these items lasted longer). Other experimental findings in evidence for two separate memory systems were the phonological and semantic similarity effects. In immediate serial recall tasks, recall of sequences of items that are similar in sound is more prone to error than recall of phonologically dissimilar sequences. When the items are similar in meaning (semantically similar) however, there is no difference in error rates. When promoting the utilization of long-term memory by lengthening the sequence to about 10 items and inserting a delay period between presentation and recall, the pattern is the converse. This observation was consistent with the view that short-term and long-term memory operated with different codes, an acoustic code and a semantic code respectively.

During the early seventies, there was a controversy about the role of short-term coding in long-term learning. Three different viewpoints were proposed. The ‘modal model’ (Atkinson and Shiffrin, 1968) comprised of three components: a sensory register, a short-term store and a long-term store. Incoming information is first processed in
parallel in different sensory buffers in the sensory register. It is then passed through the short-term store and depending on the time it stays in this capacity limited store, some information is subsequently transferred into the long-term system. Note that information has to enter the short-term system before it can be stored for long-term purposes and therefore this is represents a serial processing system. The modal model was not able to explain a range of empirical evidence, one of the most critical being the observation of patients with pronounced short-term memory deficits who still showed normal long-term memory learning capacities (Shallice and Warrington, 1970). One way to circumvent this inconsistency is to suppose that short-term and long-term memory operates as parallel and separate processes (Shallice and Warrington, 1970). Still another approach to the relation between short-term memory and long-term memory is represented by the 'levels of processing model' (Craik and Lockhart, 1972). An assumption of this model is that long-term learning involves a shift from a surface based (shallow), visual or phonological code, to a semantic (deep), code. The deeper and more elaborate information processing, the better the memory for it.

1.3.1.2. The working memory model

It should be noted that the concept of working memory is not exclusive to a particular theory, although the term has become intimately linked to the model by Baddeley and colleagues (Baddeley and Hitch, 1974; Baddeley, 1986). In their view, working memory is a system for the temporary maintenance of information that is necessary for the performance of higher cognitive abilities such as learning and comprehending language, reasoning and problem solving. As already mentioned, the modal model of Atkinson and Shiffrin (1968) had difficulties to account for several important neuropsychological (Shallice and Warrington, 1970) as well as other experimental findings (Craik and Watkins, 1973). These shortcomings of the modal model spoke against a unitary short-term memory model. Baddeley and Hitch (1974) used secondary tasks to tax the availability of short-term memory that were assumed to rely on working memory. In order to account for the pattern of data, they suggested a model of a working memory system that constituted a supportive role for higher cognitive abilities. As opposed to a unitary short-term memory system such as that put forward by Atkinson and Shiffrin, this was the first multicomponent model.

The first version of the working memory model included a central executive for attentional control and two modality-specific slave systems for storage – the visuospatial sketchpad for the processing of visuo-spatial information and the phonological loop for the processing of verbally coded information (Figure 1, Baddeley and Hitch, 1974). The central executive system is the least defined part of the model, although its designated role is to allocate attentional resources and to co-ordinate the two buffers.
The structure of the phonological loop was prompted by neuropsychological evidence as well as a number of behavioral phenomena; the word-length effect, the phonological similarity effect, the irrelevant speech effect and effects of articulatory suppression. In consequence, the phonological loop was further divided into two subcomponents – a store and a rehearsal system (Baddeley, 1986). The phonological store is able to passively store information for a few seconds. Articulatory rehearsal is comparable to the process of subvocalization or inner speech. It allows for the maintenance of information through prevention of temporal decay and also support the recoding of visually presented information into a phonological code. In order to account for the finding that noise did not affect serial recall performance (Salame and Baddeley, 1987), an important assumption of the phonological loop is that auditory information has obligatory access to the store, whereas non-speech sounds are filtered out. It should be further noted that the slave systems utilize a phonological or visual code independently of the sensory modality in which the stimuli are presented.

As outlined above, Baddeley initially viewed working memory and long-term memory as quite separate systems. An important reason behind this position was that patients with clear short-term memory deficits appeared to have intact long-term memory. However, subsequent research has shown that these patients did indeed have specific deficits in long-term phonological learning, for example, learning the vocabulary of a new language or in tests of non-word learning (Baddeley, Papagno and Vallar, 1988). As a consequence, the working memory model was modified so that the two buffers provided an interface to long-term memory (Figure 2).

1.3.1.3. The phonological store hypothesis

That immediate serial recall of verbal material relies principally on phonological coding has been a core finding since the cognitive revolution during the fifties. The working memory model incorporated a phonological loop as the primary carrier of verbal
information. The loop consisted of two subcomponents, a phonological store and an articulatory rehearsal process and if the irrelevant speech effect was to be explained within this model, one had to assume that the locus of the effect was associated with either one of these components. The central executive was considered to be an unlikely candidate since it was not defined as having storage capacities and the visuo-spatial sketchpad was assigned to deal with visuo-spatial material (but see Morris and Jones, 1990; and Jones, Farrand, Stuart and Morris, 1995, for investigations of interference from irrelevant speech for executive and visuo-spatial tasks). Besides, it was found that tasks which did not involve phonological storage were not disrupted by irrelevant speech. In a study by Salamé and Baddeley (1982) several experiments were conducted in order to investigate the implications of the irrelevant speech effect for the structure of working memory. When articulation were minimized, but the phonological store was still in use (through articulatory suppression and auditory presentation of the to-be-remembered items), the irrelevant speech effect was still evident (Experiment 3). It was further found that the length of the irrelevant words was not important to the irrelevant speech effect (Experiment 4), which would be the case in which the articulatory rehearsal process was the locus of the effect. In Experiment 5 it was observed that the degree of disruption by irrelevant speech was a function of the phonological similarity of the irrelevant words to the visually presented digits. In other words, for serial recall of digits (one, two, three … nine) phonological similar irrelevant words (tun, gnu, tee … sign) were more disruptive than phonological dissimilar words (tennis, jelly, tipple … wicket). This finding formed the basis of the phonological store hypothesis in Baddeley’s (1986) working memory model. It was hypothesized that visually presented items enter the phonological store via an active translation and rehearsal process, while heard speech-like sounds has obligatory access to this store. The notion of a speech/noise filter is supported by a number of
studies in which noise show no or negligible reductions in serial recall performance (Salame and Baddeley, 1982, Experiment 2; Salame and Baddeley, 1987; Jones, Miles and Page, 1990; see Smith, 1990, for a review). The hypothesis further holds that, for the irrelevant speech effect, there is confusion in this store between the irrelevant sounds and the phonological memory trace that has been translated from the visually presented material. Since this finding has proven not to replicate (Jones and Macken, 1995) and subsequent research has shown that the similarity between the two streams of information is largely unimportant to the effect (Jones and Macken, 1995; Bridges and Jones, 1996; LeCompte and Shaibe, 1997; Larsen, Baddeley and Andrade, 2000; but see Tolan and Tehan, 2002), this particular hypothesis was subsequently abandoned (Salame and Baddeley, 1986). Nevertheless, recent computational approaches (Neath, 2000) seem to suggest that even though the actual process of interference might be different from that of the original phonological store hypothesis, the locus of interference may still be attributed to the phonological store (Baddeley, 2000; Baddeley and Larsen, 2003; Larsen and Baddeley, 2003; Norris, Baddeley and Page, 2004). However, the phonological store hypothesis is facing further challenges. A finding that is particularly problematic for this hypothesis is that non-speech sounds such as pure tones, instrumental music, or even various types of clicks and bangs, also produce a marked disruption (Salame and Baddeley, 1989; Jones and Macken, 1993; Jones, Macken and Murray, 1993; Klatte, Kilcher and Hellbruck, 1995). This observation lies at the core of a competing explanation of the irrelevant speech effect, termed the changing state hypothesis (Jones et al., 1992; Macken, Mosdell and Jones, 1999). In contrast to the phonological store hypothesis, which suggests that the between stream similarity is the principal determinant, the changing state hypothesis argues that the within stream dissimilarity is the critical factor behind the effect.

1.3.2. The changing state hypothesis

In an experiment by Jones and Macken (1993, Experiment 5), four tones produced an effect that was a little less marked than for speech but significant compared to a quiet condition. This finding suggests that speech is not necessary for the irrelevant speech effect to occur since non-speech material also proved to be disruptive. The term “irrelevant speech effect” is in light of these findings somewhat misleading and in the recent literature, it is frequently replaced with the term ’irrelevant sound effect' (Beaman and Jones, 1997). Conversely, speech did not seem to be a mandatory constituent for the effect. Jones, Madden and Miles (1992) found that a single repeated syllable was not significantly disruptive compared to a quiet condition in a serial recall task. However, when four different syllables were used, it proved to be significantly more disruptive than the single repeated syllable. A conclusion that can be drawn from these observations is that any
irrelevant speech sound can interfere with serial recall performance, and that the primary cause of disruption is whether there is a change between successive irrelevant items in the sound sequence. Recall that the speech/non-speech filter mechanism was central to the original working memory account of the irrelevant speech effect. Changing state effects could be accommodated by the phonological store hypothesis, provided that they were attributed to a filter that is dynamic and sensitive to changing sequences of segmentable sounds rather than a static speech/non-speech filter (Jones and Morris, 1992). However, the assumption that the effect depended on the phonological characteristic of the sounds had difficulties explaining disruptive effects of non-speech sounds such as pure tones (Jones and Macken, 1993). The observation that reversed and forward speech is equally disruptive (Jones et al., 1990) gives further support for the supposition that the irrelevant speech is not interfering with a speech-specific phonetic code. What seems to be more important is the acoustic distinctiveness between the successive items in the speech sequence. If pink noise is added to a speech signal, performance is linearly improved with decreasing signal-to-noise ratios (Ellermeier and Hellbruck, 1998). Similar results have been reported with procedures in which the speech is subjected to degradation by filtering (Jones, Alford, Macken, Banbury and Tremblay, 2000).

The many difficulties that the phonological store hypothesis has encountered as an explanation for the irrelevant speech effect has led to criticism of the entire phonological loop model (Macken and Jones, 2003; Jones, Macken and Nicholls, 2004). In consequence, Jones and his colleagues (Jones, 1993; Jones, Beaman and Macken, 1996) have proposed an alternative working memory account in the Object-Oriented Episodic Record (O-OER) model. The O-OER model is a unitary model in the sense that it posits a unitary store that is used for the short-term maintenance of information. In this mnemonic store, incoming events are temporarily represented by abstract objects. These objects are amodal, which means that they do not contain any information about whether they were presented in a visual, auditory or other modality. This assumption clearly departs from the perspective of the multi-component model of Baddeley, in which visuo-spatial and verbal information are stored in different systems. According to the O-OER framework (1993), two sets of order cues, ‘episodic pointers’, are generated in the irrelevant sound paradigm. One is propagated from the articulation of visual items as they appear by their temporal presentation. This order information is maintained through rehearsal. The importance of a serial component to the visually presented task is emphasized by the finding that the effect occurs primarily in tasks that require, or is subject to, maintenance of serial order information (e.g. Salame and Baddeley, 1990; Jones and Macken, 1993; Beaman and Jones, 1997). For example, in an experiment by Henson and colleagues (2003), a list probe task was compared to an item probe task. Item probe tasks call for the recognition of a single item rather than a sequence of items as does a list probe task. Consequently,
the item probe task makes no explicit demand for serial order memory. Consistent with the changing state prediction, the list probe task did produce a significant irrelevant speech effect, whereas the item probe task did not.

The second source of serial information is produced by the auditory stream. This stream carries information about pitch, prosody and timing and emerges from the automatic perceptual organization of the sound. The O-OER model presumes that the auditory scene is organized into perceptual objects which are temporally structured into sequences and that the original order of the auditory events is preserved. It is further assumed, and recently supported by ERP data (Winkler, Teder-Salejarvi, Horváth, Näätänen and Sussman, 2003), that this organization occurs obligatory and without deliberate attention. The relation between perceptual organization and order information was investigated in a study by Jones and Macken (1995). When presenting a fixed sequence of three syllables to both ears at once, the auditory stream will appear to be coming from the same location in auditory space. However, if each syllable is assigned to a separate location in this space by means of stereophony, then phenomenally, three separate streams corresponding to three different spatial locations are heard. Notably, disruption was significantly reduced in this condition. This finding was predicted by the changing state hypothesis, since the streams, when played in each ear, will appear as repeated items which does not produce an irrelevant speech effect (Jones et al., 1992).

What constitutes a ‘change’ in the irrelevant stream? Obviously, if the changes become very small, the information about order is reduced. But what if the change is increased? In a study by Jones and colleagues (1999, Experiment 3), the irrelevant speech effect was manipulated by frequency differences between the auditory items. When there was only a small tone distance between the items, a slight effect was observed. As the differences increased, the disruption increased but beyond a certain point of separation (10 semitones), the disruption decreased again. This result can also be understood within the framework of perceptual organization. When the items are separated enough for auditory streaming to occur, the two streams form separate objects composed of repeated items with the same pitch (‘perceptual fission’) which no longer contain information about order.

The perceptual discrimination can also be reduced by increasing the number of sources of sounds, so that they produce a relatively homogeneous background. It has been shown that individuals are rather poor at extracting serial order information from sequences of sounds that come from a variety of sources (Broadbent and Ladefoged, 1959). Interestingly, if an irrelevant sound consists of many speakers that talk simultaneously, no irrelevant speech effect is observed (Hellbruck & Kilcher, 1993; Jones & Macken 1995; Kilcher & Hellbruck, 1993). Optimal changes for serial order processing then, seems to be modest variation on a common ground, that is, events coming from
the same source. This thinking represents a modification of the changing state hypothesis; both the presence of change as well of the amount of change will determine the interference by irrelevant speech (Tremblay and Jones, 1998). An interesting consequence of relating the irrelevant speech effect to the perceptual organization of sounds is that it allows prediction of the extent of disruption of a sound. In other words, the conditions in which irrelevant sound is most disruptive are those in which, under different conditions of testing, the order of events is better preserved. To conclude, the O-OER model suggests that the degree of interference is determined by the degree of seriality in the two streams of information. By reducing the level of serial information in one or both of these streams, it follows that the amount of disruption will be diminished as well.

1.4. Semantic influences

When examining semantic influences of irrelevant sound, a distinction can be made between the effect of meaningfulness of the irrelevant sound, as opposed to the semantic similarity between the sound and the to-be-remembered items (Buchner, Irmen and Erdfelder, 1996). Note that this division either focuses on effects within the auditory stream of information, or on effects between the visual and the auditory material.

1.4.1. Effects of meaningfulness

When comparing effects of narrative speech in a native language with a foreign language, thus producing quite large differences in meaning, there are no considerable differences in disruption (Colle and Welsh, 1976; Salame and Baddeley, 1986; Jones et al., 1990). Meaningful words and non-words produce similar amounts of disruption (Salame and Baddeley, 1982), as does reversed speech compared to forward speech (Jones et al., 1990). Contrary to these findings, LeCompte, Neely and Wilson (1997, Experiment 3) reported a difference between meaningful and meaningless speech. However, the effect size was small (approximately 2%) with a relatively large sample. In addition, the comparison was made between words and nonsense syllables, which is a rather crude measure of meaningfulness. Taken together, the evidence suggest that the meaningfulness of the irrelevant speech has only a small (if any) effect on the serial recall performance of visually presented material.

1.4.2. Semantic similarity effects

Semantic similarity effects can be manipulated by using auditory material with a similar meaning to that of the visually presented material. The most obvious case would be to use spoken digits as irrelevant speech together with a digit span task (e.g. Salame and Baddeley, 1982, Experiment 5; Jones, 1994, Experiment 1). In a study by Buchner, Irmen
and Erdfelder (1996), the serial recall of digit sequences was exposed to either two-digit numbers, non-words comprised of phonemes of the numbers, or word combinations with phonemes that were similar to those of the memorized digits. Compared to a quiet condition, all three auditory conditions produced substantial decrements in serial recall. However, semantically similar auditory information did not affect the recall of the visual material more than pseudonumbers (e.g. 'fentytwive'), suggesting that semantic similarity does not play a role in the irrelevant speech effect. Similar results have been obtained by other researchers (Salame and Baddeley, 1982, Experiment 5; LeCompte, 1994, Experiment 3; Bridges and Jones, 1996, Experiment 4). A potential weakness of these studies is however that digits are essentially meaningless tokens. An alternative would be to use more meaningful words as to-be-remembered stimuli. In a study by Neely and LeCompte (1999, Experiment 1) background speech comprising words known to be strong associates to the visually presented material was compared to background speech comprising words that were relatively unrelated to the to-be-remembered stimuli. There was a small (2 %), although highly significant difference between these conditions, suggesting that there is some support to the claim that semantic similarity may, to some degree, influence the magnitude of the irrelevant speech effect.

1.5. Attentional resources

Empirical evidence show no effect of phonological similarity (Jones and Macken, 1995; Bridges and Jones, 1996; LeCompte and Shaibe, 1997; Larsen et al., 2000) nor any appreciable effect of semantic similarity (Colle and Welsh, 1976; Salame and Baddeley, 1982; Salame and Baddeley, 1986; Jones et al., 1990; LeCompte, 1994; Buchner, Irmem and Erdfelder, 1996) between the two processing streams. These findings implicate that the processing of the irrelevant speech is automatic and pre-attentive, rather than attentive and controlled. In Cowan's (1995) habituation-based theory, selective attention is viewed as being embodied within a general model of long-term memory. This theory suggests that the irrelevant speech effect arises from the diversion of attentional resources away from the process of attentional scanning of the representations of the to-be-remembered items. Put in another way, the irrelevant speech induces an orienting response towards the sound and away from serial recall. Changing state effects are thought to depend on habituation of these orienting responses. Habituation occurs with repetition of the sequence while the changing sequences continuously evoke the orienting responses. Multiple orienting responses draw upon attentional resources that are necessary for the serial recall task to be successfully executed. However, there is little empirical support for this idea. In keeping with the orienting response approach, there should be at least some disruption even from the repeated sequence before this is habituated. Typically, no such observations are found. Instead, the differences in performance between varied sequences
and repeated sequences of a sound, remain the same both within (Hellbruck, Kuwano and Namba, 1996; Jones, 1997; Tremblay and Jones, 1998) and between (Hellbruck et al., 1996) experimental sessions. Only two studies show some habituation over trials, after 20 min of pre-exposure to the sound (Morris and Jones, 1990; Banbury and Berry, 1997). A possible explanation for the discrepancy of results is that the interference can be restored after a relatively short period of quiet (Banbury and Berry, 1997).

Cowan’s orienting response framework also predicts that habituation should decrease as the number of items in the irrelevant sound sequence increases, and that this relation should be fairly linear. In a study by Bridges and Jones (1996), using ‘word-doses’ of 1, 2, 5 and 7 items with a fixed presentation rate, it was found that the most prominent effect was between 1 and 2 items. From 3 items and beyond, there was no appreciable difference, an observation that is at variance with the habituation approach. A similar result has been demonstrated with ‘token-set-sizes’. By sampling from a larger pool of irrelevant items, the number of different irrelevant items can be increased. Again, the greatest increase in disruption of serial recall is between one and two tokens. Beyond two, there are only minor non-significant additional disruption related to enlargements of set-size (Tremblay and Jones, 1998). The habituation theory would predict that token-set-sizes beyond two should produce additional disruptive effects on serial recall.

1.6. Locus of the irrelevant speech effect

Whereas the detrimental effect of irrelevant speech is well established in the behavioral literature, much less is known about its neurophysiological correlates. However, some attempts have been made to correlate the irrelevant speech effect with specific brain processes. For example, in a MEG study by Valtonen and colleagues (2003) it is suggested that the effect is related to alterations in signal amplitudes in auditory cortex. Another research group speculate that the effect is caused by a disruption of left-lateralized fronto-central networks, as measured by an EEG signal coherence approach suggested to reflect the functional cooperation between different brain regions (Kopp, Schroger and Lipka, 2004). EEG has also been used to record ERPs related to irrelevant sound (Campbell, Winkler, Kujala and Näätänen, 2003), in which it was found that the N1 component (a long-latency vertex negativity) of the auditory ERP accompanied the effect. Still another approach has been to present the irrelevant sounds to either the left ear, right ear or both ears (Hadlington, Bridges and Darby, 2004). The authors suggest that the mechanisms behind the interference are lateralized to the right hemisphere of the brain, as indicated by the finding that irrelevant speech presented to the left ear produced the greatest disruption (cf. Colle, 1980). Evidently, several investigations have proposed different brain mechanisms behind the disruption from irrelevant speech, apparently yielding conflicting results. As evident from the different theoretical accounts of the irrelevant speech effect,
there is no general agreement on which representational level is affected by the interference. The hypotheses cover different stages in the information flow, ranging across perception, attention, and memory. Chapter 2 outlines the functional anatomic architecture of these cognitive functions in turn. In specific, some important findings from functional imaging studies within the fields of speech perception, selective attention and working memory will be discussed.
2. Cognitive neuroanatomy

For several decades, since the beginning of the study of cognition, the vast majority of research on cognitive functions was carried out almost exclusively using behavioral experiments. For the past 20 years, however, neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (FMRI) have emerged that allow an (in)direct observation of the neural activity in specific brain regions. Certain domains of human cognition have received particularly intense interest from cognitive neuroscientists. For example, language is unique to the human brain and as a consequence, animal models are probably not suitable to the study of the neurophysiology of the comprehension and production of speech (this constraint would seem to include the study of irrelevant speech effects). To this end, neuroimaging methods can complement the findings from neuropsychological patients by providing an approach to study the healthy, intact, and normally functioning brain. Still, the contribution of functional neuroimaging techniques is perhaps most apparent in the study of attentional mechanisms. One reason behind this is that we have adequate knowledge about where in the brain sensory signals are processed and as a consequence we are able to investigate how signal processing in these regions is modified by attention. Coupled with behavioral data, functional imaging methods show great potential to improve our understanding of how cognitive processes are implemented in the brain (Ingvar and Petersson, 2000). A condition for a successful outcome of this endeavor is that we have a sufficiently detailed knowledge of a specific cognitive task and a consistent theoretical model. While this might not always be the case (for a critical review, see Poeppel, 1996), Baddeley’s working memory model provides a well-studied cognitive framework that has been extensively used to interpret neuroimaging data. Numerous studies have provided a fairly consistent mapping between neural substrates and the proposed cognitive model. In the present chapter, some important neuroimaging findings from the study of speech recognition, selective attention and working memory, will be considered.

2.1. Speech perception

A sound is not something we can physically prevent from entering the nervous system as we can for visual stimuli. The sense of hearing is omnidirectional in nature and is on at almost all times. Speech, music and most other forms of environmental sounds are segmented in nature. PET studies using phonetic, musical and tonal auditory stimuli reveal activation of both left and right auditory cortices (Brodmann’s area [BA] 41, 42, 22), indicating that early auditory sensory areas support the acoustical and sequential processing of sounds in general (Tervaniemi, Medvedev, Alho, Pakhomov, Roudas, van
Neuropsychological data, following the ground-breaking work of Broca and Wernicke, strongly suggest that speech recognition and language comprehension is largely lateralized to the left hemisphere. However, functional imaging studies suggest that the laterality of language comprehension might be less pronounced than what would be expected from lesion data. Several studies have demonstrated that speech sounds compared to quiet, consistently produce activations in bilateral superior temporal cortices (e.g. Petersen, Fox, Posner, Mintun and Raichle, 1988; Howard, Patterson, Wise, Brown, Friston and Weiller, 1992; Mazoyer, Tzourio, Frak, Syrota, Murayama, Levrier, Salamin, Dehaene, Cohen and Mehler, 1993). Furthermore, it has been shown that the response of the left planum temporale is as sensitive for tones as it is for speech (Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman and Possing, 2000). This finding indicates that this classic speech region is concerned with the processing of acoustic rather than linguistic features of a sound. Binder and colleagues (2000) suggest that bilateral activations in more ventral and temporo-parietal regions (in the mid-portion of the superior temporal sulci) are more likely associated with speech-specific processing. Another hypothesis, based on comparisons between speech and rotated speech (Scott, Blank, Rosen and Wise, 2000), is that a posterior superior area of the superior temporal sulcus is associated with phonetic and sequential analysis, whereas a more anterior area mediates post-phonetic processing, such as word recognition and beyond.

2.2. Selective attention

In order to effectively deal with even the phenomenally simplest mental activity, such as to name an object in front of us, attentional mechanisms are required. This particular object has to be selected to be the focus of attention, rather than any other object in its surrounding. Some parts of the object might be more relevant than others, or perhaps some details of the objects have to be ignored in order to recognize the nature of it? In essence, attention implies that some features of the external (and internal) world must be selected, whereas others must be ignored. There are different kinds of attentional processes, such as selective, divided and sustained attention. What is most relevant to the irrelevant speech paradigm, however, is the voluntary control of selective attention. In the irrelevant speech task, subjects are required to selectively attend to the visually presented material, and to ignore any material presented auditorily.

Several psychological models of attention have been proposed. A common distinction is made between those that emphasize early or late selection of unattended stimuli. Broadbent’s filter theory (Broadbent, 1958) represents one of the most influential advocates of early selection. This position held that stimuli were filtered at the level of
their physical characteristics and before stimulus identification had occurred. Unattended stimuli were not supposed to be processed beyond this point. An opposite view would be to assume that all stimuli are processed non-selectively and involuntarily, leading to successful object identification regardless of relevance. This represents a late selection view on attention (Deutsch and Deutsch, 1963; Norman, 1968) and was supported by behavioral phenomena, such as the ‘cocktail party effect’ (Cherry, 1953; Moray, 1959), indicating that even unattended information can under certain circumstances be correctly identified. A weakness of the late selection position, is that it regards attentional capacity to be unlimited, which is less likely from an information processing point of view. Subsequent theories of attention can be regarded as compromises between the early-late selection distinction. For example, according to Cowan’s habituation-based theory (Cowan, 1995) discussed earlier, all stimuli are processed to some extent, possibly including some semantic processing, but they would recruit attention (by eliciting an orienting response) only if they deviated from the subject’s current neural model of the environment. In Desimone and Duncan’s ‘biased competition model’ (1995), a basic assumption is that there is a competition between stimuli for processing by the nervous system. This competition can be biased, through endogenous (internally driven, voluntary, ‘top-down’) as well as exogenous (stimulus-driven, automatic, ‘bottom-up’) mechanisms, depending on their relevance to the individual at any given moment. In an attempt to investigate the constraints on selective attention and to resolve the early and late selection debate, Lavie (1995) suggested that the perceptual load of relevant information will determine the selective processing of irrelevant information. An instruction to subjects to focus attention on a certain task is not sufficient to prevent interference from distractors. What is also needed is a high perceptual load that engages the full attention from the participants. In other words, when there exists available processing capacity, irrelevant distractors will automatically be processed. Consequently, tasks that require a low perceptual load are compatible with late selection of unattended stimuli, whereas tasks with high perceptual load are associated with early selection of unattended stimuli.

It is generally recognized that there is a need to relate cognitive processes to brain functions. During the 1980’s, the most influential models of attention were based on neuropsychological data and single-cell recordings in monkeys (Mesulam, 1981; Posner and Petersen, 1990), and since the development of functional neuroimaging techniques, much effort has been focused on the uncovering of the neural mechanisms by which attentional selection is achieved. An important finding from these experiments is that an attended stimulus elicits more neural activity in early sensory areas (e.g. Corbetta, Miezin, Dobmeyer, Shulman and Petersen, 1990; Mangun, Buonocore, Girelli and Jha, 1998; Rees, Russell, Frith and Driver, 1999), so even in the absence of such stimuli (Chawla, Rees and Friston, 1999). However, it is not clear from these studies whether the effects of
attention are mainly due to enhanced perception of attended stimuli or attenuated perception of unattended stimuli (cf. Smith, Singh and Greenlee, 2000). Although the research within the field of attentional selection has predominantly investigated attentional modulation in the visual modality, there is some support for similar modulatory effects in the auditory sensory cortices (Grady, Van Meter, Maisog, Pietrini, Krasuski and Rauschecker, 1997; Jäncke, Mirzazade and Shah, 1999). Several studies also suggest that decreases of activity as measured by functional imaging methods, may reflect suppression of neural activity in task-irrelevant modalities (e.g. Haxby, Horwitz, Ungerleider, Maisog, Pietrini and Grady, 1994; Shulman, Corbetta, Buckner, Raichle, Fiez, Miezin and Petersen, 1997). Notably, in one of the few studies that have used irrelevant speech together with a visually presented (spatial) working memory task, a corresponding suppression of the bilateral auditory cortices was observed (Ghatan, Hsieh, Petersson, Stone-Elander and Ingvar, 1998). Taken together, these findings are consistent with the notion of early selection processes. There is also evidence of late selection from neuroimaging studies (e.g. Downar, Crawley, Mikulis and Davis, 2000; Rees, Wojciulik, Clarke, Husain, Frith and Driver, 2002).

As predicted by Lavie’s load theory of attention (Lavie, 1995), the perceptual load has been demonstrated to modulate the neural activity in areas associated with irrelevant stimuli. In a study by Rees and colleagues (1997), participants were asked to perform a linguistic task of either low (letter-case judgment) or high (bisyllabic judgment) perceptual load, while at the same time ignoring irrelevant visual motion in the periphery of the display. Both behavioral measures of motion after-effect as well as functional imaging measures of motion perception in known motion sensitive areas (V5/MT), showed reduced motion processing during the high load condition. When the relevant task is auditory however, motion areas are equally affected, regardless of perceptual load (Rees, Frith and Lavie, 2001), suggesting that the attentional load theory is valid within but not between sensory modalities. If increasing the perceptual load is leading to a decreased processing of irrelevant distractors, the opposite relation seems to be the case with working memory load (see Lavie, 2005, for a review). This suggestion is supported both by behavioral (Lavie, Hirst, De Fockert and Viding, 2004; Lavie and De Fockert, 2005) and functional imaging studies (De Fockert, Rees, Frith and Lavie, 2001). In the study by De Fockert and colleagues (2001), participants were required to ignore distractor faces while engaged in low and high load working memory tasks. A high working memory load resulted in increased interference effects on performance measures as well as significant increases of activity in visual areas known to be selective for face processing (Fusiform face area, BA 19/37). The authors suggest that working memory serves to control visual selective attention and that the extent to which distractor interference takes place, is determined by the availability of working memory resources. In keeping with this
suggestion, it is not surprising to find that the most consistent activation pattern for attentional control is the same as that for working memory (for reviews, see Cabeza and Nyberg, 2000; Naghavi and Nyberg, 2005).

2.3. The functional anatomy of working memory

Among the different theoretical accounts of the irrelevant speech effect, it is the multiple-component model by Baddeley (1986) that has been most widely used for the purpose of correlating working memory functions to neural substrates. Early neuroimaging studies provided some neurophysiological support for the division between verbal and visuospatial processing (e.g. Jonides, Smith, Koepppe, Awh, Minoshima and Mintun, 1993; Smith, Jonides and Koepppe, 1996). These studies suggested that phonologically coded and visuo-spatially coded material was related to the left and right hemisphere of the brain, respectively. There were also neuroimaging evidence for the fractionation of the phonological loop into a phonological store and an articulatory rehearsal process (e.g. Paulesu, Frith and Frackowiak, 1993; Awh, Jonides, Smith, Schumacher, Koepppe and Katz, 1996). Storage mechanisms were associated with posterior areas, such as the left inferior parietal cortex (BA 39/40) along with parts of the superior temporal cortex (Paulesu et al., 1993; Becker, MacAndrew and Fiez, 1999), while rehearsal mechanisms were thought to be mediated by areas in the frontal cortex, specifically in the left inferior frontal cortex, including Broca’s area (BA 44) and the premotor cortex (BA 6, Smith, Jonides, Marshuetz and Koepppe, 1998; Smith and Jonides, 1999). The verbal-spatial dissociation is an example of a material-specific distinction whereas a separation between storage and rehearsal is an example of a process-specific dissociation. Although these functional-anatomical distinctions have shown to be less robust than expected (e.g. D’Esposito, Aguirre, Zarahn, Ballard, Shin and Lease, 1998; Nyström, Braver, Sabb, Delgado, Noll and Cohen, 2000; Hautzel, Mottaghy, Schmidt, Zemb, Shah, Muller-Gartner and Krause, 2002; Veltman, Rombouts and Dolan, 2003), the bulk of research has shown a remarkable consistent pattern of neuroanatomical regions associated with a wide variety of working memory tasks (Smith and Jonides, 1999; Cabeza and Nyberg, 2000). Typically, working memory tasks, as well as tasks that require attention or episodic memory (see Results and discussion section on potential medial temporal lobe contributions to the irrelevant speech effect, pp. 62-66), involve activity in a fronto-parietal network, in particular the dorsolateral prefrontal (BA 6 and BA 9) and parietal cortices (BA 7 and BA 40, Naghavi and Nyberg, 2005).

The central executive system of Baddeley’s working memory model has also received a great deal of attention from neuroscientists. According to Baddeley (2003), this component is the most important but least understood part of the model. In the original
version of the model (Baddeley and Hitch, 1974), the central executive was thought of as a general processing unit, designated to complex mental activities that were not directly related to the two sub-systems. In the modified version (Baddeley, 1986), the central executive was likened to the supervisory activating system, a concept put forward in the Norman and Shallice model of attentional control (Norman and Shallice, 1986). In a later attempt to define the capacities of the central executive system, Baddeley (1996) proposed that the attentional capacities included to focus, to divide and to switch attention. Dual-task performance and random generation were also capabilities that were attributed to central executive function. In the neuropsychological discipline, executive functions have become intimately linked to the prefrontal cortex (Luria, 1966). Neuroimaging experiments have studied a range of executive processes, such as monitoring the contents of working memory (e.g. Petrides, Alivisatos, Meyer and Evans, 1993), switching of attention between two tasks (e.g. D’Esposito and Grossman, 1996; Sylvester, Wager, Lacey, Hernandez, Nichols, Smith and Jonides, 2003), response inhibition (e.g. Konishi, Nakajima, Uchida, Kikyo, Kameyama and Miyashita, 1999; Bunge, Ochsner, Desmond, Glover and Gabrieli, 2001; Sylvester et al., 2003), dual-task performance (e.g. Klingberg and Roland, 1997; Bunge, Klingberg, Jacobsen and Gabrieli, 2000) and random number generation (e.g. Jahanshahi, Dirnberger, Fuller and Frith, 2000). Despite the attempts to locate a specific anatomical area or pattern of areas linked to central executive processing, there has been little success in pointing out any exclusively executive functional anatomical regions. A general finding is that the before mentioned prefrontal and parietal network is recruited by diverse cognitive problems, including those involving executive demands (Naghavi and Nyberg, 2005). Hence, the functions ascribed to the activation of this network depend on theoretical perspective. In an attentional framework, the fronto-parietal circuitry serves to generate top-down biasing signals that modulate the activity in early sensory cortices. In terms of the working memory model, the same network is associated to the operation of the central executive.

2.4. The aim of the present thesis

The aim of the present thesis is to locate cortical regions in the brain associated with the cognitive process of sustaining concurrent irrelevant information during a working memory task. In particular, we were interested in whether there is a single area or a specific pattern of areas of the brain that corresponds to the irrelevant speech effect. Studies I, II, and III were devoted to this exploration. Investigating the neurophysiological mechanisms of interference from irrelevant speech may elevate our understanding of what parts of the brain are influenced by auditory distraction. A further possibility with functional imaging is that of obtaining a different type of data that may help to adjudicate
between conflicting theoretical accounts of the origins of the irrelevant speech effect. A common assumption among the different models is that irrelevant speech effects derive from competition for cognitive resources between the working memory task and the interference condition. In neuroanatomical terms, it would be reasonable to hypothesize that the locus of competition would be revealed as a functional anatomical overlap between the working memory network and the set of areas engaged by irrelevant speech alone. Different theoretical positions seem to predict different neural correlates (Table 1).

For example, the phonological loop account of the irrelevant speech effect pointed out a subcomponent of the loop, termed the phonological store, as the locus of the effect. In an attempt to segregate the neural substrates to the subcomponents of the phonological loop, Paulesu, Frith and Frackowiak (1993), showed that the operation of the phonological store corresponded to the left inferior parietal region (BA 40) of the brain in particular. In keeping with this finding, it can be hypothesized that the irrelevant speech effect might be associated with the activation of the inferior parietal cortex. While it is difficult to derive from the other theories what would be predicted in terms of candidate neuroanatomical correlates to the irrelevant speech effect, at least some ideas can be inferred. The O-OER model is advocating a single unitary mechanism for the processing of serial information. This process could either be supported by a single area or by several areas, but it should be universally activated whenever seriation is involved. Since both a quiet working memory task as well as any task involving irrelevant speech generates pointers (that become entangled), those areas commonly activated in both tasks might serve as candidates for the seriation process. The habituation theory is proposing that interference of attentional mechanisms underlie the irrelevant speech effect, and so the effect should be associated with activity in areas of the brain typically associated with voluntary attentional control, or show evidence of modulatory effects of auditory sensory areas.

Table 1

<table>
<thead>
<tr>
<th>Theory</th>
<th>Locus of ISE</th>
<th>Predicted areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological store</td>
<td>The phonological store</td>
<td>Inferior parietal</td>
</tr>
<tr>
<td>Changing-state</td>
<td>Seriation</td>
<td>Inferior frontal</td>
</tr>
<tr>
<td>Embedded processes</td>
<td>Attention</td>
<td>Fronto-parietal and/or auditory</td>
</tr>
</tbody>
</table>

Competing theories of the irrelevant speech effect and their proposed locus of the effect. In addition, some brain areas that might serve as candidate neural correlates of the effect are suggested on the basis of neuroimaging evidence.
3. Materials and methods

3.1. Study population

A recruited sample of an undergraduate population is no doubt the most common study group in cognitive neuroscience. A reasonable question is therefore if the findings cover children and the elderly as well. In the case of the irrelevant speech effect, there are no obvious reasons to think otherwise. On the contrary, one might hypothesize that since these groups are known to show less attentional control than young adults, their short-term memory performance is in fact more vulnerable to the effects of irrelevant speech. Although there is some evidence in support of this suggestion when comparing children at different ages (Elliott, 2002), there are no difference in irrelevant speech effects between older adults and younger controls (Rouleau and Belleville, 1996; Beaman, 2005).

Participants in the present studies were recruited from Universities in Stockholm and the mean age across studies were about 25 years. All subjects gave informed consent to take part in the experiments. General inclusion criteria for the neuroimaging studies (Study I–III) were male sex and right-handedness. Participants were excluded from a PET experiment if they used any medication or had experienced a severe head trauma. In addition they were not included if they had a history of drug abuse, neurological or major psychiatric illness or family history there of. Specific selection criteria are reported in each study. Behavioral data for an FMRI experiment that has been executed but is to be reported in a coming manuscript is also included in some parts of the Results and discussion section. In the FMRI study, a subset (n = 15) of the participant in study IV were selected on the basis of their individual irrelevant speech effects. Selection criteria were determined both with regard to scanning regulations as well as to behavioral criteria. Those subjects showing the greatest irrelevant speech effect, expressed as the mean score of performance at multiple auditory items across all working memory loads subtracted from the mean score at the single item level across load, were included in the imaging study. The experimental design in the FMRI study was nearly identical to that in study IV. An important difference was that, in order to avoid neuronal ceiling effects, participants performed the task at lower levels of working memory load.

3.2. Behavioral paradigms

The task most frequently employed in irrelevant speech studies is a standard serial recall task\(^1\) (e.g. Colle and Welsh, 1976; Salame and Baddeley, 1982; Jones et al., 1992; LeCompte, 1995; Ellermeier and Zimmer, 1997). With increasing interest in the irrelevant speech

\(^1\) In fact, our knowledge of short-term memory is based almost entirely on the serial recall paradigm.
effect however, it has been investigated in a range of other tasks, such as free recall (Salame and Baddeley, 1990; LeCompte, 1994; Beaman and Jones, 1998), paired associate cued recall (LeCompte, 1994; Beaman and Jones, 1997), a missing item task (Beaman and Jones, 1997), and probe recall or recognition tasks (LeCompte, 1994; Beaman and Jones, 1997; Henson et al., 2003). While the pattern of results from these studies is rather mixed it is generally consistent with the notion that when serial rehearsal is employed by the subjects, an irrelevant speech effect is observed (Salame and Baddeley, 1990; Jones and Macken, 1993; Beaman and Jones, 1997). The experiments included in the present thesis employed a serial recall (study I-III), a serial repetition (study I–II) as well as a serial recognition (study IV) task.

3.2.1. Serial recall
In Study I–III, we employed a standard serial recall task with minor modifications. Basically, a list of digits was displayed on a computer screen and participants were instructed to memorize the order of the digits for auditory recall.

3.2.2. Serial repetition
For the PET studies (Study I–II), a serial repetition task served as a control condition to the serial recall task. The stimulus presentation and interval times were the same between the two tasks, but in serial repetition subjects were shown a list of digits in a straightforward order. As a consequence, subjects were not required to remember the order of the digits. As required by the ‘paired subtraction paradigm’ (see below), the tasks were different with respect to one component (working memory) but matched in every other respect, such as the perceptual and motor components of the task.

3.2.3. Serial recognition
In study IV, a new task paradigm was employed known as the serial recognition task or matching span (Allport, 1984; Gathercole, Service, Hitch, Adams and Martin, 1999). This experiment was a forerunner to an FMRI experiment on the effect, allowing higher spatial and temporal resolution than PET. Generally speaking, this task has several potential benefits compared to a standard serial recall task. In a typical serial recognition task, a list of items is presented and then probed by the same list in which the order of two adjacent items has been transposed. Due to the re-presentation of the encoding string, this task only claims memory for order while de-emphasizing item information, whereas in serial recall both item and order information is required. It is also the case that this task does not depend on the preparation and organization of articulatory output compared to a spoken serial recall task (Gathercole, Pickering, Hall and Peaker, 2001). An interesting
consequence is that serial recognition can prove to be a useful test of phonological short-term memory in experimental environments in which verbal response modes are not applicable, such as in an FMRI setting (e.g. Henson, Burgess and Frith, 2000). While several previous studies examine irrelevant speech effects in experiments using standard recognition procedures, very few studies have used the serial recognition paradigm together with irrelevant speech (but see Henson et al., 2003, Experiment 1). The present study allows a direct comparison between changing-state and steady-state sounds within the serial recognition paradigm. An effect of varied speech sounds is consistent with the changing state prediction lending support for the claim that changing-state speech has the potential to interfere with the order of items maintained in short-term memory. Participants performed the task at three different levels of working memory load. In study IV, lists of 6, 8 and 10 digits were presented, whereas in the FMRI study lists of 5, 7 and 9 digits were used.

3.3. Auditory stimuli

The irrelevant speech consisted of multiple auditory items and a single repeated item (CV-pseudowords). The speech sounds were recorded and edited digitally in a sound editing software. If a white noise condition was included in the study, it was generated by the software and created with fade-in and fade-out effects as to resemblance the individual sound envelope used in the pseudoword conditions. The irrelevant speech at the single level consisted of repetitions of the identical CV-pseudoword “da” [daː], while the CV-pseudowords ne [ne], li [li], to [tuː], vu [vu], py [py], bå [bɔː], nö [nœː] were used at the multiple level in a randomized order. Subjects were instructed to ignore any sounds presented in the headphones. There were only slight modifications made between studies. In study I–III, the auditory items were presented at an approximately even pitch, while in study IV, the multiple items were digitally shifted in pitch with one semitone between items, giving each item a unique pitch in an approximate semitone scale with seven tones. In study I–III, the individual items were presented at a frequency of 1 item/s, while in study IV, the frequency was increased to about 1.7 items/s.

3.4. Positron emission tomography

In Study I–III, the neural correlates of working memory and irrelevant speech were investigated. Hemodynamically based functional neuroimaging methods, such as PET and FMRI, have developed rapidly over the last two decades. They are extensively used to access neuronal processing in the living human brain and to investigate how this processing relates to different behavioral or cognitive tasks. The basis of PET is that brain activity causes local changes in blood flow (see Raichle, 2001, for a brief overview). The exact
mechanisms by which changes in neuronal activity lead to alterations in regional cerebral blood flow and oxidative metabolism are still not known, however, and are the topic of much discussion within the functional imaging community (e.g. Horwitz and Sporns, 1994; Jeuptner and Weiller, 1995; Buxton and Frank, 1997; Logothetis, Pauls, Augath, Trinath and Oeltermann, 2001).

The emission tomography produces an image of the distribution of a previously administered radioactive tracer in the brain or in any other section of the body. The recorded tissue radioactivity is a nearly linear function of local blood flow in the tissue (Frith and Friston 1997). One of the most widely used radioactive tracers for measuring regional cerebral blood flow (rCBF) in humans is \[^{15}\text{O}]\text{H}_2\text{O}.\[^{15}\text{O}\] is an oxygen atom where one electron has been removed from the normal atom. The generation of the tracer is done by the use of a cyclotron, a particle accelerator in adjunction to the camera. \[^{15}\text{O}\] is an unstable isotope that eventually will decay to ordinary \[^{16}\text{O}\] and emit one positron. The proton travels a few millimeters before it interacts with an electron, producing two annihilation photons, which travel in opposite directions (in-plane at 180° relative to each other). Radiation detectors connected to electronic coincidence detectors record an event only when two photons simultaneously arrive at a pair of radiation detectors. The coincidence detectors determine the line along which the photons traveled and provide an accurate map of the radiation distribution (density) throughout the brain.

The major advantage of the \[^{15}\text{O}]\text{H}_2\text{O} tracer is that it has a very short half-life (~2 min). Since the tracer has almost completely vanished in about 10 min, several scans can be performed on the same volunteer, allowing direct comparisons of conditions within subjects. Because relatively few scans (12 scans, corresponding to about 400–500 MBq [10–15 mCi] each, are the maximum allowed under current radiation safety regulations in Sweden) can be performed on each volunteer, it is necessary to combine data from several subjects in order to detect activation. As the sensitivity of the newest PET scanners is high, allowing lower doses of radioactivity to be employed, more scans can be performed on each subject or even make it possible with similar-subject studies.

Despite the obvious advantages of the PET technique many problems still remain in studying the underlying neural activity during different components of a cognitive task. The spatial and temporal resolutions are still not even close to the size and response of neurons or cortical columns, being in the order of millimeters and milliseconds. Moreover, hemodynamic measurement techniques such as PET may reflect both excitatory and inhibitory synaptic activity (Jeuptner and Weiller, 1995). Another problem is that the reflected synaptic activity not only originates from the local synapses in a brain region but from afferent activity coming from all the regions that project to the region being examined (Horwitz, 1998). Noteworthy, different techniques, such as PET, FMRI,
EEG/MEG, lesion analysis, TMS and single-unit recordings, all have different spatial and temporal properties and consequently do not allow direct comparisons.

### 3.4.1. Image preprocessing

The primary PET data underwent a series of image preprocessing and statistical analyses steps using the statistical parametric mapping (SPM) software (http://www.fil.ion.ucl.ac.uk/spm). The data is analyzed by group, as most functional imaging studies are, which reveal effects that are common across participants and thus are generalizable, at least to the specific group in question. A standard procedure is to realign, normalize, and to filter the data before any statistical analysis can take place. The realignment step accounts for the movement of the subject’s head across scans. In the present studies, 12 scans were obtained with approximately 10 min of rest in between. Although plastic helmets prevented subjects from repositioning their head, small head movements still occur (on the order of 1–3 mm). To compensate for this movement, image volumes are spatially adjusted to the position of the first. Normalization deals with the problem of anatomical variability between subjects. The brains of different individuals are anatomically different. A necessary requirement for group studies in functional neuroimaging is to represent data in a standardized anatomical space. This requires a method to transform, or warp, individual data into the standardized space, so-called anatomical normalization. Anatomical normalization aims to adjust for anatomical differences in order to allow data to be averaged across subjects. Anatomical normalization transforms the image time-series of the individual participant into a standardized anatomical space and in our studies we have commonly used the stereotactic space as defined by the SPM template. The Montreal Neurological Institute in Canada has created templates for this purpose. SPM99 uses the ICBM152 template, which is based on 152 normal MRI scans. This anatomical space is approximately matched to the Talairach brain as photographed and depicted in the famous Talairach and Tournoux atlas (Talairach and Tournoux, 1988). In general, using a voxel-based approach, it is important to reduce the impact of misregistration and inter-individual residual functional-anatomical variability. A common strategy is to spatially low-pass filter (spatial smoothing) the data either at reconstruction or with a suitably chosen convolution kernel (e.g. an isotropic 3D Gaussian kernel). Spatial filtering, which in effect is a local weighted averaging procedure, also increase the local equivalence of the voxel data across measurements and individuals and thus the validity of voxel-based statistical models. Thus, this procedure is done in order to minimize individual functional-anatomical differences between subjects. Another potential benefit with filtering data spatially is that it may increase the signal to noise ratio.
3.4.2. Statistical modeling and inference

Subsequent to the image preprocessing, a statistical model and a test statistic are chosen. Typically, a general linear model is used in which (regression) parameters are estimated. This implies that some effects of interest are included in the model, together with confounds of no interest as well as an error term. Specific null-hypotheses can then be tested with the model. A P-value is generated by comparing the observed test statistic to its distribution under the null hypothesis. A significance test is carried out in every voxel, generating a statistic image of the brain. Usually, a pre-determined level of significance is used to accept or reject the null hypothesis for each voxel. The problem of multiple comparisons can be handled by several different statistical approaches within the SPM software package.

The goal with functional imaging is often to locate and isolate a region in the brain that is implicated in a specific mental activity. However, during the minute of image acquisition in the scanner, the subject typically engages in numerous mental activities, so taking just one ‘picture’ is not enough to pin down a particular cognitive process. A common approach to this problem is to record a second image in the same subject, but from a carefully selected control condition. The optimally chosen tasks are minimally different from each other and the cognitive task should encompass all of the components used in the control task. Ideally, by subtracting the image from the control task from the image of the cognitive task, the resulting activity is associated with the cognitive process of interest. This technique is known as paired-image subtraction (see Posner and Raichle, 1994, for an introduction), and has become a dominant experimental paradigm for imaging studies. However, the subtraction method has been criticized for several reasons, and as a consequence, alternative ways of analyzing imaging data have emerged. One complementary approach is to examine the functional connectivity between activations in different regions, or to use various network approaches, such as structural equations modeling.

3.5. Functional connectivity and network analysis

The nervous system of the brain no doubt represents an extremely complex system. This is reflected in its function, its evolutionary history, its structure, as well as the coding schemes it uses to represent information (see Koch and Laurent, 1999, for a brief overview). The combinatorial possibilities in a system consisting of about 100 billion neuronal components having hundreds of trillions of interconnections are overwhelming. For this reason, it can be argued that continued reductionism and atomization might not elevate our understanding of brain function. Higher cognitive functions are, most likely, the result of interactions between different regions of the brain and as a consequence, an integrative approach to imaging data is potentially beneficial. In Study III, the aim was to
analyze the functional and effective connectivity between functionally specialized areas. Functional connectivity is defined in terms of temporal correlations or covariance of activity between different neurophysiological events, while effective connectivity refers to the influence one neural system exerts over another (Friston, 1994). The basic hypothesis is that the variability in the neural response to a specific cognitive task will correspond to relevant functional interactions and that these interactions will be reflected by the covariance structure. While functional connectivity is an operational definition; it is merely a statement about the observed correlations, effective connectivity depends on some model of the influence one neuronal system exerts over another. In Study III, based on theoretical and empirical considerations, we defined a simple network for verbal working memory to investigate interactions with the medial temporal lobe system. In order to examine such interactions, a structural equation modeling (Hayduk, 1987; Bollen, 1989) approach was used. In this context, structural equations modeling can be viewed as a linear regression model for effective connectivity. With this tool, different models can be compared statistically, and constraint can be put on the model by specifying the error variance and the connections between the components. With an appropriately selected model, it is then possible to compare differences between conditions in a stacked models approach (Bollen, 1989).
4. Results and discussion

In the following section, results from studies I–IV are separated into behavioral and rCBF effects. For the behavioral results, data from study I, II and IV are on some occasions presented across different studies in the same figure. The intention with the composite graphs is to illustrate relevant similarities and differences, and does not reflect any further statistical analyses across studies. The rCBF effects for study I and II are reported and discussed within a common framework, whereas the network data from study III are presented separately.

4.1. Behavioral effects

The main effect of irrelevant speech is well established in the behavioral literature. Unfortunately however, the area has suffered from being studied by a sequence of different investigators. All have concentrated on a somewhat different question and used their own techniques. There is no doubt about the robustness of the phenomenon, but cross-laboratory replication of more subtle results is certainly needed. Below, the behavioral results from study I, II and IV will be reported and briefly commented upon.

4.1.1. Interaction between irrelevant speech and working memory load

In study I and II, the irrelevant speech effect was investigated under two different levels of working memory load. Consequently, these findings do not provide any direct evidence of load effects and interaction with irrelevant speech. In study IV, we included three levels of working memory load and two levels of irrelevant speech (single and multiple items). As evident in Figure 3, across the group of participants (n = 62), no significant interaction is seen between irrelevant speech and working memory load. It is interesting to note that there was no significant simple effect of speech at the highest load.

Previous findings show that subjects are likely to abandon phonological coding when performance rates drop below some critical level. In a study by Salame and Baddeley (1986) the interaction between the irrelevant speech effect and phonological similarity effects was investigated across a range of list lengths. For sequences of five, six and seven items, marked effects of similarity were observed under both control and irrelevant speech conditions, while at a list length of eight, the similarity effect disappeared. This pattern was interpreted in terms of a tendency of the subjects to shift from pure phonological processing towards engaging semantic processes as well when performance dropped below some critical level. This is also supported by additional behavioral results (Baddeley, 1966; Baddeley, 1966; Salame and Baddeley, 1986; Hanley and Bakopoulou,
suggesting that when working memory load is high, the use of non-phonological strategies will decrease the interference from irrelevant speech. Interestingly, participants in study IV report to have engaged in a number of additional strategies besides subvocal rehearsal. Nevertheless, while there was a slight tendency for such a decrement at the highest load in study IV, any conclusions would be premature.

Examining main effects of irrelevant speech and working memory load at a group level might have little to say about the relationship between individual working memory capacity and the ability to sustain irrelevant distractors. Working memory serves higher cognitive functions, such as reasoning, problem solving and language among many other. Indeed, correlations have been found between measures of working memory span and performance on a range of diverse tasks, such as reading comprehension (Daneman and Carpenter, 1980), following directions (Engle, Carullo and Collins, 1991), and reasoning (Kyllonen and Christal, 1990). In order for this to succeed, it is likely that an important capacity for working memory is to sustain the processing of relevant information, while inhibiting processing of irrelevant information. The ability to suppress distraction is assigned to the central executive system according to Baddeley (1996). The implication would seem to be that when working memory is occupied, this ability is weakened. Some evidence in support of this suggestion is provided by the finding that

Figure 3. The probability of correct responses in the presence of single and multiple auditory items in study IV (* p < 0.05, *** p < 0.001).
low-span individuals are more likely than high-span individuals to notice their own names in an unattended auditory stream (Conway, Cowan and Bunting, 2001). Also consistent with this suggestion, Lavie (2005) recently proposed that while high perceptual load prevents distractor interference, high working memory load increases interference by irrelevant distractors. Although Lavie has not explicitly addressed the irrelevant speech phenomenon, what appears to be predicted by his theory is that the individual working memory capacity scores of the subjects in the present study should correlate negatively with their sensitivity to irrelevant speech, such that the better they perform on the serial recognition task, the better they cope with the irrelevant speech. While there was some support for this prediction at the low level of working memory load ($r = -0.29, p < 0.05$), the medium and high levels of load showed virtually no correlation (Figure 4). As a comparison, the correlation between working memory span and other cognitive abilities, such as reading comprehension (Daneman and Carpenter, 1980), arithmetic calculation (Turner and Engle, 1989) or logic (Kyllonen and Stephens, 1990), typically correlate in the order of 0.5. In a study by Beaman (2004) it was concluded that individuals showing greater general cognitive capabilities; 'high-span individuals', are as likely to be as affected by irrelevant sound as low-span individuals. In another study by Ellermeier and Zimmer (1997), replicated by Neath and colleagues (2003), there was no correlation between short-term
memory span and susceptibility to irrelevant sound. \(r = 0.01, \text{ns}\). Overall, the results of study IV are consistent with previous findings (Ellermeier and Zimmer, 1997; Neath et al., 2003; Beaman, 2004) in suggesting that there is no linear relation between working memory capacity and the ability to sustain irrelevant speech.

We conclude that there are no appreciable interactions between load and sound in study IV, neither between conditions over group, nor within subjects as reflected by a correlational approach. The absence of a correlation between measures of working memory performance and the degree of interference from irrelevant speech is consistent with the suggestion that the effect occurs at an automatic level, beyond the control of the individual.

4.1.2. Irrelevant speech effect size and distribution

For present purposes, the term ‘effect size’ will be referring to differences between scores on multiple auditory item conditions subtracted from scores on single auditory item conditions. In Experiment 1 of study I, the effect size was 9 %, while in Experiment 2, a reversed effect of 1 % was observed. Study II demonstrated no effect of changing states, whereas in study IV, the pooled effect size across working memory load was 4 %. Mean scores between auditory conditions for the present studies are displayed in Figure 5.

Taken together, the observed effects sizes in the present studies are quite small. In

![Figure 5. Average scores across auditory conditions for each study.](image-url)
experiments that carefully manipulate the acoustic character of the irrelevant sound typically demonstrate disruptive effects of about 5–10% (e.g. Jones, 1994; Tremblay and Jones, 1999; Campbell et al., 2003). Using naturalistic irrelevant speech compared to a quiet condition, effects can be anywhere between about 10% (e.g. Colle and Welsh, 1976) and 30% (e.g. Ellermeier and Zimmer, 1997). Given the closely matched auditory conditions used in the present studies, relatively subtle effects can be expected. The absence of changing state effects in study I (Experiment 1) is probably reflecting ceiling effects given the low working memory load (6 digits). However, the lack of an effect in study II, using a conventional working memory load (8 digits) is more surprising. Possible factors are low degree of exposure. The frequency of presentation of the auditory items in study II was about 1/s. It has been demonstrated that as the amount of changing state information increases (up to a certain point), so does the degree of disruption from irrelevant speech (Bridges and Jones, 1996; Jones et al., 1999). Accordingly, in order to increase the exposure of changing state information, the presentation frequency of the auditory items was increased in study IV. For the same purpose, pitch differences were generated between the items in study IV. It is difficult to derive the significant irrelevant speech effect in study IV to the implementation of these, rather subtle changes in auditory stimuli. The experimental paradigm was different and we conducted the study on a different subject population. Considering the relatively large variability of the effect, as illustrated in Figure 6 (study IV), the absence of an effect in study II might be explained in terms of the small sample size (n = 9). Study IV was designed in order to isolate the effect of varied speech attributed to a disruption of serial order information by contrasting changing state with steady state auditory items in a serial recognition task. The results demonstrated a highly significant changing state effect of multiple versus single auditory items. In the serial recognition paradigm memory for the order between adjacent items is crucial, while minimal demands are put on representation of content due to the re-presentation of the items at test. This contrasts with serial recall, which requires the storage of both item and order information. Consequently, the results of study IV adds strong support for and extends previous empirical findings (Jones and Macken, 1993; Beaman and Jones, 1997; Henson et al., 2003) claiming that irrelevant varied speech has the potential to interfere with the coding of the order of the items to be memorized. The observed effect of varied speech in study IV was relatively small (4%). In comparison, unspecific speech vs. quiet comparisons typically produce smaller effects. For example, there is a minor effect in the item probe task in the study of Henson and colleagues (Henson et al., 2003), consistent with similar findings of irrelevant speech effects in tasks where serial order is not required and in which a quiet control is used (Beaman and Jones, 1997). In keeping with this observation, there is typically a small task unspecific effect of steady-state sounds as well. Notably, in both Experiment 3 and 4 in a study by Jones and colleagues (1992), the effect of steady-
state sounds compared to quiet also disrupted performance by 3 % and in Experiment 4 this effect was significant. This and similar observations (Jones, 1994, Experiment 1; LeCompte, 1995, Experiment 1–4) have led to the suggestion that the irrelevant speech effect consists of two effects, one specific to tasks that require memory for serial order, and one more subtle effect that is unspecific in the sense that it affects any task requiring attention and memory (cf. LeCompte, 1996; Neath, 2000). The present results show a difference between multiple and single auditory conditions, and is therefore not a general speech versus quiet effect.

As evident in Figure 6, the distribution of individual irrelevant speech effects in study IV ranged from positive effects of about 20–25 %, to negative effects of 10–15 % with an average effect of 4 %. In this context, ‘negative effects’ refers to a reverse irrelevant speech effect, that is, an enhancement of performance due to multiple auditory items compared to single items. Thus, some participants actually benefited from the exposure to irrelevant speech. Eleven (18 %) of the subjects, increased their performance by 5 % or more while exposed to the changing auditory material compared to the repeated items. Similar results have been reported in other irrelevant speech studies with large numbers of participants (Ellermeier and Zimmer, 1997; Neath et al., 2003). Apparently, people show great variability in their susceptibility to irrelevant sound interference. The individual characteristics that determine the degree of disruption is not known.
4.1.3. Stability over time – behavioral vs. scanning environments

A subset of participants from the behavioral study IV was selected to participate in an FMRI experiment that took place within the next 2–3 months after completing study IV. In order to narrow the variability in the selected sample and to increase the possibility of detecting an irrelevant speech effect, the subjects showing the largest irrelevant speech effects were asked to return for the FMRI study. A comparison between the behavioral results from the same group of participants in study IV and the FMRI study reveals that the effect size dropped from about 10% in study IV to about 4% in the FMRI study. As is evident in Figure 7, all subjects except one exhibited a reduced irrelevant speech effect in the FMRI study compared to the behavioral study. The effect size showed to be approximately evenly distributed across different levels of working memory load, ruling out the possibility that the difference was due to the relatively lower load in the FMRI study\(^2\). One possible explanation would be in terms of practice related effects. However, this seems unlikely given the reported high test-retest reliability \((r = 0.82)\) of the irrelevant speech effect (Ellermeier and Zimmer, 1997) and that it has been demonstrated to remain stable both within (Hellbruck \textit{et al.}, 1996; Jones, 1997; Tremblay and Jones, 1998) and between (Hellbruck \textit{et al.}, 1996) experimental sessions.

\footnotesize
\begin{figure}
\centering
\includegraphics[width=0.6\textwidth]{figure7.png}
\caption{Individual irrelevant speech effects (S–M) in study IV compared to the FMRI study for the same subjects \((n = 15)\).}
\end{figure}

\(^2\) Note that, at the group level, there is also a small reduction of the varied speech effect in the PET experiment (Experiment 2) compared to the behavioral experiment (Experiment 1) in Study I (Figure 5). However, this difference possibly reflects ceiling effects due to the low level of working memory load in the PET experiment.
Another obvious difference between studies was the experimental environment. The MR scanner produces a substantial amount of ambient acoustic noise. Thus, the decrement in effect size might be explained by unspecific effects due to the fact that the subjects are placed inside the bore of the scanner. Possibly, the loud noise associated with FMRI recording activates the auditory system and reduces the perception of other sounds. Nevertheless, a comparison between study IV and the FMRI experiment suggests that the behavioral main effect of irrelevant speech subsists (although slightly diminished) despite the noise. This finding is consistent with behavioral pilot studies obtaining irrelevant speech effects when recorded scanner noise was played aloud throughout the session (Chein, 2003, Appendix A).

4.1.4. Effects of working memory load

Average probability scores across auditory conditions are given in Table 2. In the lowest load study (Study 1, Experiment 2), performance was close to faultless. When working memory load is increased (Study 1, Experiment 1; Study II), performance drops considerably.

Table 2

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Load</th>
<th>Probability correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>I, Exp 1</td>
<td>16</td>
<td>8</td>
<td>0.69</td>
</tr>
<tr>
<td>I, Exp 2</td>
<td>9</td>
<td>6</td>
<td>0.93</td>
</tr>
<tr>
<td>II</td>
<td>14</td>
<td>8</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Note that the probability of recall was slightly higher in study 2 than in study 1 (Experiment 1), despite that the level of working memory load is the same. Possible explanations are that the subjects in study 2 consisted of a selected subset from a pilot in which they performed the same serial recall task under a white noise background condition. Consequently, they were familiar with the experimental settings and requirements, which might have improved their overall performance. This assumption receives some support from the results (Figure 8) in the serial recognition task (study IV), showing that performance improves when subjects are retested in the scanner (FMRI study). Data from a pilot to study IV is also depicted in Figure 8. In the pilot study, a similar serial recognition task was used in a quiet background at five levels of working memory load. Taken together, serial recognition performance displays an approximately linear pattern across different levels of working memory load. These results harmonize well with empirical evidence. Previous research has shown that, within the range of about 4 to 11
items, serial recall performance exhibit a highly linear relation to the number of items in a stimulus set (Cowan, 2001). Below and beyond that range, ceiling effects and floor effects are typically observed. The present results are also consistent with the notion of working memory as a limited capacity system (Miller, 1956; Baddeley, 1986; Cowan, 2001). Considering that the chance level is 50%, the low mean score on the high level of working memory load suggests that the working memory capacity was reaching its limits.

4.1.5. Serial position effects

A common procedure when analyzing data from tasks in which the serial order of events is required, is to produce serial position curves. A linear slope with primacy and recency effects are thought to reflect the serial processing of the to-be-remembered items. In free recall tasks, primacy and recency effects are thought to reflect long-term and short-term memory processing, respectively (Atkinson and Shiffrin, 1968). However, it is less clear as to what primacy and recency effects signify in serial recall or serial recognition tasks.

4.1.5.1. Serial recall

Figure 9 displays serial position plots for the serial recall studies. Overall, the results show the u-shaped serial position curves typically produced by serial and free recall tasks.
Serious recognition

The serial recognition task is far less extensively used as the serial recall task and as a consequence, serial position plots are intended as exploratory. However, it would be fair to speculate that if serial rehearsal mechanisms are at play, they should be reflected in a serial position slope in the serial recognition task as well. Figure 10 displays the probability correct as a function of probe position for each level of working memory load in study IV. Note that since the first and last pair of digits in each string was not transposed, serial position data only reflects \( k - 3 \) possible transpositions, where \( k \) is the number of items in the string. In a study by Henson and colleagues (2003), a list probe task was used that was similar to the one in study IV. In general, performance in the list probe task used in Henson’s study decreased with probe position. However, there are several differences between studies that potentially explain the observed inconsistency in serial position results. In Henson’s study, subjects were presented shorter list lengths of 5, 6 and 7 items whereas in study IV, list lengths of 6, 8 and 10 items were used. Previous studies have indicated that a high working memory load may be associated with shifts in rehearsal strategy (Baddeley, 1966; Baddeley, 1966a; Salame and Baddeley, 1986; Hanley and Bakopoulou, 2003; Larsen and Baddeley, 2003), suggesting that subjects were prone to use non-phonological/non-serial rehearsal in the highest level of load in study IV. A post-experimental interview confirmed that subjects were engaged in a range of complementary cognitive strategies, besides phonological rehearsal. It was not clear, however, if these strategies were more intensively employed at higher levels of working memory load.
An additional difference between studies was that in study IV, the first and last pair of positions was not transposed, while in the study of Henson, all positions were transposed. Allowing each neighboring pair in the list to function as probe increases the range of possible transpositions, and thereby the sensitivity in a serial position analysis. By restricting the extreme positions from being transposed in study IV, the intention was to discourage atypical cognitive strategies (i.e. checking the first and last digits before rehearsing the sequence). Taken together, the results from the serial position analysis in the serial recognition task show no simple pattern of interactions between accuracy and position. The line plot at the medium level of working memory load bears some resemblance to the serial position curve typically found in serial recall tasks. However, the 'slope' is rather mixed for the different working memory loads, probably in part reflecting floor effects and/or choked rehearsal strategies at the highest level of working memory load.

4.1.6. Reaction times

In study IV, reaction times (RT) were obtained for different levels of working memory load as well as in response to different auditory background conditions. Subjects used significantly more time to respond during the multiple item sound compared to the single item sound at the low and medium working memory load (Table 3). The general increase in RT due to working memory load observed in the present experiment is consistent with the earlier findings that RT increase with list length, approximated by a
linear function (Sternberg, 1969). The best linear fit gave a slope of 250 ms/item increased in working memory load. This is a slope close to other tasks thought to require serial rehearsal and too slow to merely reflect a visual search task (Henson et al., 2003).

4.2. Subjective reports

Subjective data are scarcely collected in functional imaging experiments. Purely behavioral studies typically use a series of experiments to rule out possible strategic interpretations of the data, whereas for imaging studies this approach might be seen as impractically expensive. In the present imaging studies, a post-experimental interview followed each experiment in order to provide introspective data that potentially could show to be useful when interpreting the behavioral as well as functional imaging data.

4.2.1. Choice of cognitive strategy

The post-experimental debriefing concerning the cognitive strategies in study I (Experiment 2) indicated that most of the subjects recalled the digits in triplets as suggested to them during the practice session. However, about half of the subjects explicitly reported to use a visual strategy for the last few digits of the sequence. The most common mnemonic strategy in study II was to rehearse the digits in pairs or triplets. About a third of the subjects specifically reported to use a visual strategy for the last two or three digits of the sequence. All subjects in study IV reported that they had mainly used the suggested mnemonic strategy, where digits most commonly were memorized in pairs or triplets. About 40% of the subjects specifically stated that they had also used one or several complementary strategies. These strategies could be classified as being either associative (for example, associating numbers with years of historic or personal relevance), trend sensitive (upward or downward tendency of the numbers), visual (some subjects stated they created an inner vision of a part of the digit sequence), or melodic (a few subjects reported that they remembered the digits by mapping them onto pitch sequences). Some subjects also reported that the digit zero was left out of the remembered sequence as a 'blank space'.

<table>
<thead>
<tr>
<th>Irrelevant stimuli</th>
<th>Working memory load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low M</td>
</tr>
<tr>
<td>Single items</td>
<td>1.51</td>
</tr>
<tr>
<td>Multiple items</td>
<td>1.65</td>
</tr>
</tbody>
</table>
Overall, the subjective reports from study I, II and IV have two important implications. First, the introspective data suggest that participants spontaneously engage in alternative mnemonic strategies beside phonological rehearsal. In study IV in particular, subjects were creative in finding various means to cope with the serial recognition task. This indirect evidence implies that subjects sometimes shift strategy according to task demands. For example, in a serial recall task with a list length of five items, a phonological similarity effect but no semantic similarity effect has been demonstrated (Baddeley, 1966a). At a list length of ten items, the reverse pattern was observed (Baddeley, 1966b), indicating that subjects had shifted rehearsal strategy from a phonological coding to a semantic coding. Second, participants do not always perform the task according to the given task instructions. Previous irrelevant speech studies have demonstrated that task instructions can have a significant impact on the behavioral results. Recall that tasks that are most sensitive to interference from irrelevant speech effect are those involving a serial component. For free recall tasks, however, this component should be minimal. Still, in a study by Beaman and Jones (1998), an irrelevant speech effect was observed in a free-recall task for both free-recall and serial recall instructions. Although somewhat circular, this finding suggests that giving free-recall instructions does not mean that serial strategies will not be used. Furthermore, it has been shown that strikingly different patterns of performance can be obtained depending on the instruction given. A study by Hanley and Bakopoulou (2003, Experiment 2) demonstrated that when subjects are specifically instructed to use a phonological rehearsal strategy, the phonological similarity effect and the irrelevant speech effect were additive, as predicted by the phonological loop model (Baddeley, 1986). However, when no instructions were given concerning rehearsal, the effect of phonological similarity was abolished in the irrelevant speech conditions. As for the studies by Baddeley (1966a; 1966b), the results can be taken to indicate that the phonological similarity effect disappears in some experiments because participants sometimes employ an alternative (semantic) rehearsal strategy.

4.2.2. Perceived difficulty of irrelevant speech

In study I and study II, both using the serial recall task, subjects rated the multiple item condition to be more difficult than the single item condition (Figure 11). The single items condition, in turn, was judged to be more difficult than the auditory control condition (quiet or noise). When examining the individual ratings for both serial recall studies (n = 25), only one subject (4%) rated the single items to be the more difficult one, whereas four subjects (16%) rated them as equally difficult. The remaining subjects (80%) perceived the multiple item condition to be the most demanding. Notably, there was no observed behavioral difference between the single and multiple item conditions in
study II, and only a negligible difference in study I. Turning to study IV, it is surprising to find that almost one third of the subjects (31%) stated that it was more difficult to perform during the single item background sound (Figure 12). Interestingly, the magnitude of the effect of varied speech was dependant on what sound the subjects perceived was the most difficult (Figure 13). This difference turned out to be significant only for those subjects that had stated the multiple item background to be the most demanding. Apparently, subjective ratings of difficulty were less reflective of the actual performance in the single item than in the multiple item condition. In comparison, Ellermeier and Zimmer (1997), found that a priori estimates of the extent to which subjects anticipated the irrelevant speech would disrupt their performance did not correlate with the observed susceptibility. However, when ratings were obtained subsequent the first experimental

![Figure 11. Perceived difficulty of the serial recall task for different auditory background conditions in study I and study II.](image1)

![Figure 12. Perceived difficulty of the serial recognition task for different auditory background conditions in study IV (n = 62).](image2)
session, a moderate correlation was observed suggesting that an initial lack of insight could be improved by relatively little direct experience.

### 4.2.3. Perceived difficulty of load

The perceived difficulty of load could only be obtained in study IV, in which different levels of load were included in the same experimental design. Ratings were scored on a visual analogue scale, and the results revealed a linear increase across the levels of working memory load (Figure 14). Note that the linearity in performance data in relation to
working memory load (Figure 8) is paralleled by the subjective reports of the perceived difficulty of the task at the different levels of load.

4.3. Effects of regional cerebral blood flow

4.3.1. Neural correlates of working memory

In study I and study II, we examined the main effect of working memory by combining the results from the three serial recall tasks and comparing them with the control tasks. Previous functional imaging studies lend considerable support to a working memory related network in the brain that is typically associated with prefrontal and parietal regions (Cabeza and Nyberg, 2000; Naghavi and Nyberg, 2005). Consistent with these findings, the main effect of working memory in study I included significant posterior parietal activations and in study II prefrontal and parietal activations were observed. In both studies, the anterior cingulate cortex (BA 32) was activated as well. The anterior cingulate cortex is frequently associated with working memory tasks (Cabeza and Nyberg, 2000), but this area seems to be related to task difficulty (Barch, Braver, Nyström, Forman, Noll and Cohen, 1997), rather than to any processes specific to working memory. In addition to the anterior cingulate, several activations were found in regions also previously associated with the verbal component of working memory. These included left superior temporal cortex (BA 22, Paulesu et al., 1993; Ghatan et al., 1998), the premotor (BA 6, Awh et al., 1996; Smith and Jonides, 1999) and supplementary motor areas (Schumacher, Laub, Awh, Jonides, Smith and Koepp, 1996; Smith and Jonides, 1999), as well as the cerebellum (Paulesu et al., 1993; Smith, Jonides and Koepp, 1996).

All regions that were activated in study I were also activated in the study II. In a conjunction analysis, using the contrast showing the main effect of working memory, commonalities of activations between studies included increases of blood flow in the anterior cingulate (BA 24/32), the anterior insula (BA 13/47), the inferior frontal (BA 6/44) and premotor cortex (BA 6) as well as in the posterior parietal cortex (BA 40 and BA 7). Additional activations were observed in the basal ganglia and the cerebellum. The impact of load is also related to working memory operations and was assessed by contrasting the main effect of working memory as an interaction contrast between studies. One obvious difference between studies when observing the activation images was the presence of bilateral dorsolateral prefrontal activity in study II (Figure 15a). Activations in this region were confirmed in the multi-group comparison, specifically in the right dorsolateral prefrontal cortex (right BA 46/10) as well as in the anterior prefrontal cortex bilaterally (BA 10/9). In functional imaging studies of working memory, activations in prefrontal areas are generally associated with executive processes (e.g. D’Esposito, Detre, Alsp, Shin, Atlas and Grossman, 1995; Smith and Jonides, 1999) or working memory load.
In keeping with these findings, the prefrontal activations observed in study II were possibly reflecting the relatively higher working memory load compared to study I.

4.3.2. Neural correlates of irrelevant speech

In study I, a low level of working memory was used in order to avoid potential neuronal ceiling effects. In study II, the load was increased to a level comparable to most behavioral
studies of the irrelevant speech effect. In neither of study I or II, an irrelevant speech effect were obtained between the varied speech and steady-state conditions. One might argue, that since no behavioral effect of changing-state speech occurred, the activation images do not reflect the desired effect. However, it has been demonstrated that neurophysiological effects of irrelevant speech persist both in the presence and absence of behavioral differences, as confirmed by an examination of correct-only trials in a study of Chein and colleagues (Chein, 2003, Experiment 1).

4.3.2.1. Effects of speech compared to no speech
When comparing irrelevant speech (multiple or single auditory items) to quiet (study I) or noise (study II) in conditions of serial recall, increases in rCBF were confined to the bilateral superior temporal regions (BA 22). This finding is consistent to empirical evidence, demonstrating that that early auditory sensory areas support the acoustical and sequential processing of sounds (Tervaniemi et al., 2000; Patterson et al., 2002).

4.3.2.2. Effects of varied speech
Varied-speech effects were investigated by comparing simple effects of multiple items with single items during the serial recall task. In study I, this varied-speech comparison revealed a decrease in activity in the left superior temporal cortex. In study II, an increase in activity was observed in the bilateral dorsolateral prefrontal (BA 9) and the superior-middle temporal (BA 22/21) cortices.

4.3.2.3. A suppression of the verbal working memory network (study I)
The irrelevant speech effect may also be characterized in terms of an interaction between the level of irrelevant speech (multiple and single items) and working memory (recall and repetition) in the general linear model. In study I, the interaction contrast revealed no significant increases in activity. However, relative decreases in blood flow were detected in several areas associated with verbal working memory. These decreases were significant in the left superior temporal (BA 22) and right inferior/middle frontal areas (BA 6/44), while nonsignificant decreases were also observed in homologous regions in the contralateral hemisphere, and in the left inferior parietal cortex (BA 40). Whereas increases in blood flow in a region of the brain can be interpreted as enhanced neural activity within that region, reduced activity may reflect the suppression of that activity (cf. Haxby et al., 1994; Shulman et al., 1997). In keeping with this interpretation, the changing-state sounds resulted in a suppression of verbal working memory related functions, specifically those involved in the temporary storage and rehearsal of speech-based material (Paulesu et al., 1993; Awh et al., 1996; Schumacher et al., 1996; Smith and Jonides, 1999). Interestingly, in study I, the left inferior parietal cortex was also implicated in the suppressed network
of areas. The left inferior parietal activation is of particular importance from a theoretical perspective, since this region is suggested to reflect the phonological store, the subcomponent of the working memory model where the irrelevant speech effect is assumed to take place (Baddeley and Salame, 1986; Baddeley, 1992; Baddeley, 1994). Functionally speaking, a way of attempting to sustain the interference from irrelevant sounds would be to inhibit the system as a whole while increasing the internally generated signal representing the remembered items. Some further support for such an interpretation is provided in an FMRI study of the irrelevant speech effect by Chein and colleagues (2003). Subjects were scanned while performing a delayed probed recall task under various irrelevant information conditions. Three types of irrelevant information were used: irrelevant speech, irrelevant non-speech, and articulatory suppression. Irrelevant speech sequences consisted of the spoken digits one through four, presented in a pseudo-random order. Whereas the working memory task was associated with increased activity in areas consistently observed in verbal working memory tasks, irrelevant speech tended to reduce activity in these areas. In specific, significant decreases in activity were found in the anterior cingulate, inferior frontal, and anterior insular regions, predominantly in the left hemisphere.

Activations in the superior temporal region should reflect early phonological processing and might indicate modulatory effects related to attentional aspects (i.e. ‘early selection’) of the task (Grady et al., 1997; Jäncke et al., 1999). Study II demonstrated an increase in activity in this area in the varied speech vs. unvaried speech comparison, while in study I a decrease was observed in the same contrast and region. A similar decrease in activity in auditory areas have been reported in a visuo-spatial working memory task where irrelevant speech was present compared to a quiet condition (Ghatan et al., 1998). However, in the irrelevant speech study of Chein and colleagues (2003), also using a speech-quiet comparison, an increase in activity was reported in the left middle temporal cortex (BA 21). These apparent inconsistencies are not easily incorporated by theoretical explanations. According to Lavie’s theory of attention (Lavie, 2005), early selection depends on high perceptual load. Whereas high perceptual load prevents distractor interference, low perceptual load may or may not be associated with interference depending on the working memory load. As long as active cognitive control functions are available for the selective attention task, interference may be resolved. However, resolution is only required between targets and a potent salient distractor that strongly competes with the target. In the behavioral section of the present thesis, a simple prediction derived from the Lavie’s framework of attention was that individual working memory capacity scores should correlate negatively with their sensitivity to irrelevant speech, such that the better they perform on a working memory task, the better they would cope with the irrelevant speech. This prediction was not met by the behavioral data. In terms of neuroanatomy,
however, it is not clear what to predict. Lavie’s hypothesis is based mainly on studies conducted using stimuli within a single, usually visual, modality and might not apply to the effects of irrelevant speech. A few studies have examined crossmodal effects of perceptual load, but the results from these studies are rather mixed (Rees et al., 2001; Houghton, Macken and Jones, 2003; Weissman, Warner and Woldorff, 2004). A further complication is that the level of perceptual or working memory load is a relative concept and dependant on individual differences. As a consequence, they must be qualified by some measure in order to have any explanatory value. To this end, correlational approaches, where working memory capacity scores are correlated to measures of rCBF in auditory cortex during an irrelevant speech task, might provide useful information. In general, these approaches require larger subject samples than examined in study I–II. In study III, an attempt was made to explore the interaction between the superior temporal cortex and additional areas implicated in a verbal working memory network using a structural equations modeling approach (see below). Nevertheless, it remains for further studies to investigate whether the observed pattern of temporal activations reflects a possible modulatory mechanism for task-irrelevant auditory stimuli.

4.3.2.4. A role for the prefrontal cortex in the irrelevant speech effect (study II)

In study II, the pattern of activations in the interaction contrast was quite different from the one observed in study I. First, there were no significant decreases in activity (besides the right cerebellum). Instead, the reverse contrast revealed increases in activity in the dorsolateral prefrontal cortices bilaterally (BA 9/10/46, Figure 15b). These activations presumably reflect the neural responses to varied speech that is specifically related to the context of the serial recall task. The varied irrelevant speech evidently engaged a region of the brain previously shown to be associated with working memory (Cabeza and Nyberg, 2000), and executive functions in particular (e.g. D’Esposito et al., 1995; Smith and Jonides, 1999). According to the working memory model, interference from irrelevant speech is unlikely to involve any other subcomponents of the model, such as the visuo-spatial sketchpad or the central executive (Salame and Baddeley, 1982; Hanley and Broadbent, 1987; Morris and Jones, 1990). However, the ability to suppress distractions is central to the task in irrelevant speech paradigms and this mechanism is commonly ascribed to the executive component of working memory (Baddeley, 1996; Smith and Jonides, 1999). Tasks that require this ability frequently invoke activity in the prefrontal cortex of the brain (Fuster, 1997), suggesting that this region might play an important role in the irrelevant speech effect as well.

As noted earlier, a tentative prediction from the changing-state hypothesis is that any potential neural correlates of the irrelevant speech effect should include activation from both the working memory task as well as from the changing auditory material.
itself. The common process by which interference occurs is thought to be seriation, that is, the serial maintenance of sequentially presented material. One way forward would seem to be to compare neurophysiological studies of changing auditory material to studies specifically aimed at localizing the serial component of a working memory task. Any common areas found from these studies would serve as a possible neural locus of the changing-state effect. Hemodynamic responses to infrequent ('mismatch') sound changes have frequently been observed in the superior temporal as well as the inferior and dorsolateral prefrontal cortices (Dittmann-Balcar, Juptner, Jentzen and Schall, 2001; Schall, Johnston, Todd, Ward and Michie, 2003), even when the subjects are explicitly instructed to ignore the sounds (Opitz, Rinne, Mecklinger, von Cramon and Schroger, 2002; Rinne, Degerman and Alho, 2005). The superior temporal activity is consistent with its confirmed role of general acoustical and sequential processing of sounds but the functional role of the frontal cortices remains poorly understood. Switching of attention to the sound (Nääätänen, 1990), compensatory enhancement of activity (Opitz et al., 2002) or inhibitory mechanisms (Rinne et al., 2005) have been suggested as possible functions underlying the frontal involvement in auditory change detection.

Remembering the order of events in a presented sequence is central to any serial recall task. Working memory tasks in general invoke activity in a distributed network of areas, and rehearsal mechanisms are typically associated with the left inferior frontal cortex (Paulesu et al., 1993; Awh et al., 1996; Smith and Jonides, 1997). Whereas this area (Broca’s) is suggested to support the articulatory rehearsal processes required for the phonological recoding of visual stimuli, the specific process of maintaining temporal order has been linked to the left dorsolateral premotor cortex in particular (Henson et al., 2000). Taken together, the empirical evidence from auditory change detection tasks and serial recall tasks appear to have the activation of the left inferior frontal cortex in common. Notably, the left superior temporal cortex has also occasionally been implicated in the verbal working memory network (Paulesu et al., 1993), although the functional role of this activation is probably related to speech rather than to storage mechanisms. While these findings at most provide some indirect clues as to the locus of the irrelevant speech effect according to the changing state hypothesis, a second way to deal with this question is to analyze the activations images further in study II. Interestingly, increases in activity were observed in the dorsolateral prefrontal cortices, both in response to changing-state auditory material (cf. Dittmann-Balcar et al., 2001) as well as to the working memory task. A conjunction approach, combining the contrast showing the main effect of working memory with the one reflecting multiple versus single items, confirmed that the dorsolateral prefrontal cortices were activated in both situations (Figure 15c). Arguably, a potential explanation for the effect is in terms of competition for common neuronal resources. An alternative interpretation is that the prefrontal activation is not associated with the varied
speech per se, but reflects a recruitment of compensatory processing. Since the subjects were able to fully compensate behaviorally for the irrelevant speech in the low level of working memory load used study I, we hypothesized that there would be evidence of such compensatory processing in study I as well. Note that the activation pattern for the irrelevant speech effects was quite different in study I compared to study II. One obvious interpretation for this discrepancy is that the experimental context differed between studies. Besides subtle differences in auditory conditions, the level of working memory load was considerably higher in study II which might explain the apparent inconsistencies. To examine this possibility, three different levels of load were included in study IV and a subsequent FMRI experiment. The behavioral results of study IV have been reported in the above, while the FMRI data awaits further analysis. However, another way to shed light on the mechanisms behind the irrelevant speech effect is to explore an alternative type of contextual information, that of the interaction between regions in terms of their functional and effective connectivity. This was the aim of study III.

4.3.2.5. Effective connectivity and the finding of an MTL involvement (study III)

Recently, Baddeley (2000) added another component, the episodic buffer, a limited capacity system that provides temporary storage of information held in a multimodal code capable of binding information from the subsidiary systems and long-term memory into an episodic representational format (Figure 16). This extension of the model was published around the time when we commenced analyzing the data in study I through a network approach. Naturally, we speculated in the neural underpinnings of the new component. The episodic buffer shares some characteristics with the concept of episodic memory encoding (Tulving, 1989) with respect to its principal mode of storing information in an episodic format as well as its integrative aspects, but differs in that it is assumed to be a capacity limited temporary store. Thus, the episodic buffer provides an interface between the original components of working memory and long-term memory. In emphasizing its short-term integrative role based on an episodic format, one may hypothesize that the episodic buffer is related to the prefrontal cortex and the medial

![Figure 16. The extended working memory model of Baddeley. The episodic buffer comprises a limited capacity system that provides temporary storage of information held in a multimodal code, capable of binding information from the subsidiary systems, and from long-term memory, into an episodic representational format. The episodic buffer is supposed to provide an interface to the other slave systems of working memory and to long-term memory.]
temporal lobe as well as the interaction between these structures. The transient early role of the medial temporal lobe system in long-term memory formation and sequence encoding in conjunction with the prefrontal cortex makes these likely candidates (cf. e.g., Eichenbaum, 2000; Simons and Spiers, 2003).

The initial motivation for study III was to further characterize the role of the superior temporal activation as observed in study I with respect to the irrelevant speech effect. To this end, we investigated the functional connectivity of this region with the rest of the brain. The most prominent finding was a correlation between the left superior temporal cortex and the left medial temporal lobe that was sensitive to the varied vs. steady-state speech manipulation during the serial recall task. The subsequent network analysis (Figure 17) indicated that this finding was part of a more general phenomenon suggesting a more intense interaction between the verbal working memory network and the medial temporal lobe as expressed in changes of the connection strengths related to the level of irrelevant speech (Figure 18). One might argue that, since the multiple items condition contains greater variability in the irrelevant speech input and perhaps to some degree relatively more novel information compared to the single item condition, the present result simply reflects an episodic encoding effect. We suggest that this is not the case, because we observed a positive correlation between the left STG and the left MTL in the single item condition, while the reverse was the case for the multiple items condition (Figure 19). Instead we suggest that the observed difference may reflect a change in the interaction between these regions, depending on the level of irrelevant speech, and related
to a processing shift between the two conditions. The results of the network analysis support this interpretation, because, given the changes in functional STG–MTL connectivity as a function of irrelevant speech level, one would predict that the other related systems should show parallel changes in interaction pattern. Since, the episodic buffer provides an interface between the support systems of working memory and long-term memory, an interesting possibility in the context of this study is that the episodic buffer is instantiated as an interaction between the prefrontal cortex and the medial temporal lobe, possibly also including the posterior parietal cortex (Figure 18). Given the close relation between episodic memory and the medial temporal lobe, an observation of interest in relation to immediate serial recall is that when task difficulty increases there is a tendency for subjects to abandon phonological coding in favor of other strategies.

Figure 18. The relative difference in connection strengths between the multiple and single item conditions. No meaningful functional sub-network differed significantly between conditions except for the connections related to the MTL.
The episodic buffer might serve one possible compensation mechanism for such a strategy switch since it would provide a potential remedy against the detrimental influence of interfering or disturbing stimuli. Thus our observations might reflect a strategy shift, from pure working memory processing, to an engagement of capacities related to episodic processing and by implication the medial temporal lobe, reflecting integrative aspects of on-going information processing. However, it should be pointed out that a similar perspective on the role of the medial temporal lobe as well as the prefrontal and posterior cortical regions in working memory, which is not dependent on the concept of an episodic buffer, can be argued and have recently been outlined by Cabeza and colleagues (2002) and Ranganath (2006). In this regard, it is worth noticing that alternative views to the episodic buffer, including long-term working memory and working-with-memory, have been suggested by Ericsson and Kintsch (1995) and Moscovitch (1994), respectively. Working memory as activated portions of long-term memory (Cantor and Engle, 1993) and the embedded processes model of working memory (the framework of the habituation hypothesis of the irrelevant speech effect) by Cowan (1995), illustrates analogous perspectives. The assumption that working memory entails the activation of representations in long-term memory receives some support from findings with calculating prodigies. In a PET study by Pesenti and colleagues (2001), it was demonstrated that prodigies and non-experts used different brain areas for calculation. Experts were able to switch between short-term effort requiring storage strategies and highly efficient episodic memory encoding and retrieval, as indicated by activations in the right prefrontal and medial temporal areas. Similar strategy-switching, but to visuo-spatial coding, has been observed in abacus experts (Tanaka, Michimata, Kaminaga, Honda and Sadato, 2002). These findings suggest that a certain amount of practice on a complex short-term memory task enables the circumvention of the limited capacity of verbal short-term memory into the reliance for alternative on-line memory representations. Nevertheless, these perspectives remain controversial, in particular with
respect to the medial temporal lobe since patients with medial temporal lesions do not
typically show dramatic short-term memory deficits. However, it should be noted that
short-term memory investigations in these patients typically have used simple short-
term memory tasks such as digit span, and a potential explanation may be that these are
not sensitive enough to detect subtle short-term memory deficits. Such deficits may be
more pronounced if an additional distracting task or disturbing input is delivered and
this hypothesis can be tested directly on patients with medial temporal lesions.

As illustrated in the above, a difference between different models of working memory
is the position on whether it constitutes a separate memory system or not. Accordingly,
in cognitive theory, a distinction can be made between structural views on working
memory as opposed to activation-based approaches. A parallel division has been made
within neuroscience between theories that emphasizes organization by domain rather
than by process (for a review, see Fletcher and Henson, 2001). Within the irrelevant
speech community as well, this debate has surfaced partly as a consequence of the
inadequacy of the phonological store hypothesis to account of the irrelevant speech
effect. The critique has in some instances come to extend to Baddeley's working memory
model as a whole, suggesting that it represents a reification (Macken and Jones, 2003).
A similar stance is taken in a recently published paper by Postle (2006) in which working
memory is seen as an emergent property of the nervous system. Rather than being a
specialized system, the functions of working memory is suggested to arise through the
coordinated recruitment, via attention, of brain systems that have evolved to accomplish
sensory-, representation-, and action-related functions. In a concluding remark, Postle
proposes that while the working memory model by Baddeley has contributed significantly
to the advancement of our understanding of working memory, it is now time to abandon
it in favor of competing frameworks that more effectively incorporate the growing body
of neuroscientific data. Arguably, while the role of the medial temporal lobe in study III
may be interpreted as the operation of a domain-specific, multisensory buffer, it may as
well be seen as simply reflecting a nervous system that is capable of representing many
kinds of information. For the purpose of the present thesis, however, the suggestion put
forward is that dynamic changes between areas inherently involved in the serial recall
task and the medial temporal lobe, indicate that new areas in the brain may be recruited
according to task demands.
5. Summary and conclusions

The aim of the present thesis was to explore some possible propositions as to the representational level of the effect. Based on neuroimaging evidence within the fields of speech perception, selective attention and working memory, some neural correlates of the effect were suggested from dominant theoretical explanations of the irrelevant speech effect (Table 1). In the present thesis, through a series of behavioral experiments in combination with the PET methodology as well as structural equations modeling, we conclude that several areas in the brain are implicated in the effect. In study I, we observed a decrease in neuronal activity associated with varied speech in several areas previously shown to be associated with verbal working memory. This finding is also broadly consistent with the irrelevant speech effects reported in an FMRI study by Chein and colleagues (2003) and provides support for an interpretation based on interference within working memory. However, fronto-parietal activations are consistently observed in both working memory tasks and tasks that require attentional control. According to the ‘biased competition model’ by Desimone and Duncun (1995), the fronto-parietal circuitry serves to generate top-down biasing signals that modulate the activity in early sensory cortices. Consistent with this hypothesis, we observed a significant decrease in activity in the left superior temporal cortex in study I. Interpreted as a suppression of early auditory cortex, this observation argues for a theoretical explanation of the effect based on a perceptual or attentional level of representation for the irrelevant speech effect. In study II, varied speech was found to induce neural activity in the dorsolateral prefrontal cortices and this area was also activated by the working memory task alone. While it has to be recognized that in general, no obvious mapping has been found in the prefrontal cortex between subregions and cognitive functions, the observed results in study II might be interpreted in terms of interference between two concurrent cognitive processes. Since this area is frequently associated with the central executive aspects of working memory, our results suggest a role previously not recognized for the central executive in the irrelevant speech effect literature. Alternatively, as formulated in the changing state hypothesis, the process of interference could as well be the confusion between different cues to serial order. Consistent with this suggestion, prefrontal activations have been reported in PET and FMRI studies of auditory change detection tasks (Dittmann-Balcar et al., 2001; Opitz et al., 2002). In study III, a network approach to the dataset in study I revealed that the irrelevant speech effect was associated with a closer interaction between a verbal working memory network and the medial temporal lobe. It was argued that this observation might reflect a strategy shift, from pure working memory processing, to an engagement of capacities related to episodic processing. Baddeley’s ‘episodic buffer’ (Baddeley, 2000) might serve one possible compensatory mechanism for such a strategy.
An alternative view is that the engagement of the medial temporal lobe reflects integrative aspects of on-going information processing related to episodic coding, as part of a general long-term memory framework in which the need to assume a separate and specialized working memory system no longer remains.

Based on the overall findings from the present studies, the following conclusions are proposed:

- There is no single neurophysiological locus of the irrelevant speech effect. The results of the present studies do not exclude any of the hypothesized levels at which the effect occurs (perceptual, attentional, or mnemonic), rather it appears that all levels are affected.

- Functional neuroimaging methods can provide additional information about the mechanisms underlying the irrelevant speech effect that is difficult, if not impossible, to obtain with behavioral methods alone.

- Within the paired subtraction paradigm, changes in rCBF take place in areas that are inherently activated by the verbal working memory task itself. However, as reflected in the network analysis, dynamic changes between these areas and the medial temporal lobe, suggest that new areas are recruited according to task demands.

Study I is, to our knowledge, the first published functional imaging study of the irrelevant speech effect. As such, the starting point was exploratory. Since then, we have completed two additional functional imaging studies of the effect, one PET study (study II) and one FMRI study still in progress (Gisselgård, Uddén, Fransson, Ingvar and Petersson, in preparation). Meanwhile, other imaging experiments have also begun to emerge in the literature (e.g. Chein, 2003) and two FMRI papers on the irrelevant speech effect will be submitted in a near future (J. Chein, personal communication, 28 July 2006; K. van Dijk, personal communication, 7 July 2006). This development, together with continued theoretical interest in the mechanisms behind the irrelevant speech effect (e.g. Neath, 2000; Page and Norris, 2003), is promising.
Sammanfattning av avhandlingen

Den så kallade irrelevanta tal- eller ljudeffekten innebär en försämrad prestationsförmåga på en visuellt presenterad arbetsminnesuppgift som orsakas av irrelevant tal eller andra ljud i omgivningen som presenteras samtidigt som uppgiften skall lösas. Beteendeeffekten är experimentellt välbelagd och visar att ovidkommande ljud under vissa omständigheter underminerar vår förmåga till selektiv uppmärksamhet med försämrad kognitiv prestationsförmåga som följd. Förmågan att ignorera omgivande ljud är avgörande i de flesta arbetsmiljöer och att kunna sortera ut vad som är relevant från det som är irrelevant är en fundamental princip för den mänskliga hjärnans informationshantering. I denna avhandling har vi använt positron emissions tomografi (PET) för att undersöka de neurofysiologiska korrelaten till den irrelevanta taleffekten.

I ett första delarbete användes en seriell återgivningsuppgift med låg arbetsminnesbelastning för att undvika neuronalna takeffekter. Tre olika bakgrundsljud presenterades, ett varierat ljud bestående av korta stavelser med olika fonologi; ett repeterat ljud bestående av samma stavelse; samt en tyst bakgrund som jämförelsebetingelse. Tidigare studier har visat att ett varierat ljud ger en större sänkning av prestationen än ett repeterat ljud. I det första delarbetet visade resultaten att den irrelevanta taleffekten motsvarades av ett relativt lägre regionalt blodflöde i flera områden i hjärnan som tidigare förknippats med verbalt arbetsminne. Dessa områden omfattade auditoriska hjärnbarken och nedre/bakre frontalloben bilateralt, inklusive Brocas område i vänster hjärnholm, samt ett mindre område i nedre delen av vänstra parietala hjärnbarken. Förändringarna i blodflöde talade för att aktiviteten i områden som generellt understödjer verbalt arbetsminne hämmades av de irrelevanta talljuden.

I det andra delarbetet användes ett identiskt experimentellt paradigm, men med en högre arbetsminnesbelastning i syfte att undersöka om fynden från första delarbetet skulle bestå även vid en för rena beteendestudier konventionell minnesbelastning. De förändringar av regionalt blodflöde som observerats i studien med låg minnesbelastning replikerades inte i denna studie. Istället observerades en relativ ökning av blodflödet i dorsolaterala prefrontala hjärnbarken bilateralt. Detta är ett område i hjärnan som är starkt associerat med en viktig komponent eller aspekt av arbetsminnet som har med den centrala exekutiven att göra. De två första studierna talar för att det inte finns något enskilt område i hjärnan som är tydligt kopplat till den irrelevanta taleffekten. Våra observationer tyder på att effekten istället bör förstås utifrån förändringar i flera områden som är funktionellt relaterade till varandra. En andra slutsats är att de områden vars blodflöde förändras som en följd av skillnaderna mellan talljuden, är områden som också aktiveras av arbetsminnesuppgifter i allmänhet. Slutligen kan det antas, mot bakgrund av de observerade skillnaderna mellan de båda studiernas resultat, att valet av arbetsminnesbelastning har stor betydelse för de observerade blodflödesförändringarna.

The studies were conducted at the PET unit and MR research Center at the Karolinska Hospital, Sweden. I wish to thank everyone who has contributed in various ways to this thesis. In particular, I wish to express my sincere gratitude to Professor Martin Ingvar and Dr Karl Magnus Petersson for a dual-head supervisory teamwork that has created a dynamic work environment during the rocky road of this research endeavor. I also want to thank: Karolinska Apoteket and the PET organization at the Karolinska Hospital, including Professor Sharon Stone Elander, Vahid Halouuli, Julio Gabriel, Göran Printz, Anna Fredriksson, Ellenor Anderson and the late Valter Pulka for technical support and tracer delivery. Jesper Andersson, Stefan Skare and Peter Fransson for being helpful with computer related problems and for sharing their technical and statistical skills. My fellow colleagues during the many years: Katrina Carlsson, Predrag Petrovic, Johan Sandblom and Christian Forkstam, for great company and mutual efforts. Special thanks are due to Dr. Jean-Luc af Geijerstam and Henrik von Horn, for invaluable companionship and many laughter and conversations at lunch time, and to Isabell, for all love and support at home. This thesis is dedicated to our beautiful daughter – Livia Gisselgård – who was born somewhere between study II and III.
References


phenomena and practical implications. 


Hadlington, L., Bridges, A. M. and Darby, R. J. (2002). Auditory location in the irrelevant sound effect: The effects of presenting auditory stimuli to either the left ear, right ear or both ears. *Brain & Cognition*, 55, 545-57.


Mazoyer, B., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamin, G., Dehaene, S., Cohen, L. and Mehler, L.


Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. and Griffiths, T. D. (2002). The


