

Neural reflections of meaning in gesture,  
language, and action

Roel M. Willems

The research reported in this thesis was supported by the Netherlands Organisation for Scientific Research (NWO) 'Cognition' program, grant number 051.02.040

Cover image: 'Hieronymus lezend in de wildernis' ['Hieronymus reading in the wild'], engraving (detail), Rembrandt van Rijn, 1634

© Rijksmuseum Amsterdam; reprinted with kind permission

ISBN 978-90-9023728-2

Printed by PrintPartners Ipskamp, Enschede, The Netherlands

© Roel Willems, 2009

Neural reflections of meaning in gesture, language, and action

Een wetenschappelijke proeve op het gebied van de  
Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus Prof. mr. S.C.J.J. Kortmann,  
volgens besluit van het College van Decanen  
in het openbaar te verdedigen op 16 januari 2009  
om 15:30 uur precies

door

Roel Mathieu Willems

geboren op 29 februari 1980  
te Limbricht

Promotor: Prof. dr. P. Hagoort  
Copromotor: Dr. A. Özyürek

Manuscript commissie:

Prof. dr. H. Bekkering  
Dr. J. J. van Berkum (Max Planck Institute for Psycholinguistics,  
Nijmegen)  
Dr. S. Kita (University of Birmingham, UK)

## Table of contents

<b>Chapter 1</b>	Introduction	7
<b>Chapter 2</b>	Online integration of semantic information from speech and gesture: Insights from event-related brain potentials	21
<b>Chapter 3</b>	When language meets action: The neural integration of gesture and speech	47
<b>Chapter 4</b>	Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context	81
<b>Chapter 5</b>	Early decreases in alpha and gamma band power distinguish linguistic from visual information during sentence comprehension	115
<b>Chapter 6</b>	The neural integration of language and action information: Co-speech gestures versus pantomimes	149
<b>Chapter 7</b>	Embodied action understanding in the motor system: Evidence from left- and right-handers	183
<b>Chapter 8</b>	Summary and discussion	215
	References	227
	Samenvatting in het Nederlands	249
	Acknowledgements	257
	Publications	259
	Curriculum Vitae	261
	Series Donders Institute for Brain, Cognition and Behaviour	263
	Appendix Colour Figures	265



## Chapter 1 Introduction\*

We move our hands when we talk. Although we might be unconsciously doing so, we do it all the time. People across cultures do it, children do it from early on in life, we do it when the other person cannot see us (like when talking on the phone), blind people do it when talking to other blind people. Conversely, this means that when we talk to someone, we often not only hear speech, but we also see movements of the hands of the speaker. Research (described below) indicates that hand movements made during speech are not mere 'hand waving', but that meaningful information is conveyed through such hand gestures. This thesis is about how meaning from hand actions is encoded and combined with language in the brain. Although co-speech gestures constitute an important part of the thesis, also other information types and actions were studied. That is, for reasons that will be outlined below, besides studies of brain correlates of the understanding of co-speech gestures and speech (Chapters 2, 3 and 6), we investigated the combination of meaning presented in visual format (pictures of objects) and speech (Chapters 4 and 5) as well as the neural implementation of actions that convey meaning without speech (Chapter 7).

In this introduction I will first briefly describe the role of co-speech gestures during language comprehension. Second, studies of neural correlates of understanding co-speech gestures will be reviewed. Third, I will give an overview of the role of the cortical motor system in coding action information. Fourth, I will give an introduction into the brain imaging methods that were used in the experiments in this thesis. Finally, I will outline the structure of the thesis. Note that the chapters were written to be understandable when read as stand-alone, without

---

\*Part of this introduction is based upon Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, 101(3), 278-289.

knowledge of the other chapters. This inevitably means that content is sometimes repeated in several chapters.

## **1.1 Co-speech gestures and their role in speech comprehension**

Co-speech gestures are the hand movements we make while we speak. McNeill (1992) classifies co-speech gestures as follows: *beats* are rhythmic hand movements that support speech but have no semantic relationship with the speech. *Deictic gestures* are hand movements that refer to some entity in space, like pointing to an object. *Metaphoric gestures* are hand movements that describe an abstract idea that is being talked about. Finally, *iconic gestures* are a reflection of the content of speech, but in a more literal, visuo-spatial way as compared to metaphoric gestures. Consider for example a speaker retelling a cartoon scene in which someone climbs up a ladder. He might say ‘and he climbed up’ while at the same time moving his hands in a climbing manner, as if holding the rungs of the ladder (see McNeill 1992). Another example is a speaker who says ‘they walked back and forth’, while repeatedly moving the hand from left to right. The gestures studied in this thesis are all iconic gestures, simply called gestures from now on. Importantly, gestures are systematically related to the speech with which they are co-expressed. This relationship exists at three levels. First, there is semantic overlap between the representation in gestures and the meaning expressed in the concurrent speech, as in the ‘climb up’ example above (e.g. McNeill 1992; Kita and Özyürek 2003). That is, speech and gesture usually convey similar or related information. Second, speech and gesture are temporally aligned to each other. The onset of the gesture usually precedes the onset of the relevant speech segment by less than a second (Butterworth and Shovelton 1978; Morrel Samuels and Krauss 1992). More importantly, in most speech-gesture pairs the stroke (semantically the most meaningful part of a gesture) coincides with the

relevant speech segment (McNeill 1992). Finally, it has been shown that the spontaneous use of gestures has a similar function as speech, namely to communicate the intended message to the addressee (e.g. Özyürek 2002; Kendon 2004; Melinger and Levelt 2004). Although they are systematically related to each other, gesture and speech are expressed in different representational format. That is, there is no form matching between gestures and speech (McNeill 1992). Consider the example described above: an upward hand movement in a climbing manner when a speaker says: “He climbed up the ladder”. Here, the gesture depicts the event as a whole, describing manner (‘climb’) and direction (‘up’) simultaneously. In speech, however, the message unfolds over time, broken up into smaller meaningful segments (i.e. the individual words ‘climb’ and ‘up’). Because of these form differences, the mapping of speech and gesture information must occur at a semantic level. However, whether gestures actually do influence speech comprehension at the level of semantics has been debated.

Whether gestures have a communicative function has been an important question in gesture research. More precisely, is information conveyed in gesture picked up by the listener and used in his / her representation of the message of the speaker? Roughly speaking, this question has been investigated in two ways. One set of studies compared comprehension of a speaker’s message when speech-and-gestures are observed to comprehension when only speech is observed. For instance, Graham and Argyle (1975) had participants describe abstract line drawings when they were either permitted or prohibited to use their hands. A panel of observers subsequently drew the figures that were described to them. It was found that reproduction of the figures was more accurate in the speech-and-gesture condition as compared to the speech alone condition. Riseborough (1981) had observers guess what object (out of three object names) an actor was describing when they could either see and hear speech and gesture, speech and facial expression, or speech alone. It was found that

recognition was fastest when the whole body was in view together with the speech and slowest when only speech was presented. However, this effect was only present for the object which was hardest to guess in general. For the other two objects, no difference was found for the different presentation modes. Riseborough also found that recall of items in a short story was better for words that were accompanied by meaningful co-speech gestures as compared to words accompanied by 'vague hand movements'. The size of this effect was influenced by the level of noise that was artificially added to the speech signal.

Meaningful gestures had their biggest influence on recall in the noisy condition, i.e. when speech was less illegible. Despite differences in details of the findings of these studies, the general picture that emerges is that gestures are helpful in comprehension of the message of the speaker. However, a remaining issue is whether gestures convey semantic information to the listener. It is possible that in the studies described above, gestures serve a facilitatory function, in some way making it easier to encode the speech (e.g. Krauss et al. 1991). A remaining question is whether the *content* of gesture can influence understanding of the message.

Another line of research has looked at the information observers glean out of the observation of only gestures, i.e. presented without the speech they were originally accompanied by. Feyereisen and colleagues (1988) had participants watch video recordings of lectures with the sound turned on or not. They found that participants were able to recognize gestures as being iconic gestures or beats, but that the meaning ascribed to the gestures most of the time did not coincide with the speech that they were originally accompanied by. It was concluded that whereas gestures may convey some semantic information, this is of a very general nature. In a similar vein, Krauss and colleagues (1991) presented gestures without speech to participants. Participants had to choose from words that had originally accompanied gestures, simply write down what they thought a gesture was about, assign

gestures to semantic categories or indicate whether they had seen a gesture before or not. Although performance was above chance in all measures that were used, it was far from perfect, even on the seemingly simple task of choosing between two words as to which one matches the observed gesture best. Beattie and Shovelton (2002) used a more elaborate scoring technique to assess the accuracy of information picked up by listeners from gestures presented without speech as compared to the original retelling of a cartoon story. Accuracy was assessed by scoring the answer of listeners or viewers to questions on several semantic categories such as the shape, size and identity of the actions or objects that were depicted in the gestures. Again, the results indicate that information picked up from only gestures is rather imprecise: the overall mean accuracy for a gesture was only 23% (100% is maximal score).

It seems that without the speech with which gestures are normally co-expressed, their meaning becomes ambiguous. Some have taken this to imply that gestures add little or no semantic information to comprehension of a message. For instance, Krauss et al. write: “[...] the gestures in our corpus in this study seem to convey relatively little information, and it is difficult to see how they could play an important role in communication.” (p. 751-752). On the contrary, in a series of experiments, Goldin-Meadow and co-workers have claimed that additional information only conveyed in gesture (and not in speech) is used by listeners. They studied the relationship between hand movements and speech in children trying to solve mathematical problems. Some children used an incorrect strategy to solve the problem, which they described verbally. Interestingly, they did indicate the correct strategy by means of their gestures. It seems as if the correct strategy is ‘present’ in the learner, but is somehow not applied. Goldin-Meadow has claimed this to be an important predictor of readiness to learn during development (Goldin-Meadow et al. 1993; Goldin Meadow 2003). More importantly for present purposes is that

observers are influenced by this additional information in gestures. For instance, it was found that children learn a new strategy for solving a math problem when speech and gesture of their teacher convey different strategies (Singer and Goldin Meadow 2005). Finally, one study has found that gestures that convey additional and even contradictory information influence the retelling of a story (McNeill et al. 1994). Participants were shown retellings of cartoon stories in which sometimes gestures conveyed additional or contradicting information. Information which was only presented through gesture was nonetheless detectable in participants' subsequent description of the retelling that they had seen. This indicates that information conveyed only through gestures is noticed by observers and can be used in their representation of the message of the speaker.

In conclusion there is evidence that information conveyed through gestures is noticed by listeners and that it can influence a listener's understanding of the speaker's message. The fact that people are relatively bad at assigning the correct meaning to a gesture presented without speech underscores the tight relationship between speech and gesture. That is, it seems that gestures cannot convey information unambiguously when presented alone. When presented together with speech, information conveyed in gesture can however be picked up by the observer and can influence the understanding of the message by the observer.

## **1.2 Co-speech gestures in the brain**

Only recently researchers have set out to investigate the neural underpinnings of understanding co-speech gestures. Kelly et al. (2004) conducted an ERP study in which subjects saw an actor make a gesture corresponding to a property of an object, like its width or height. If the gesture had been preceded by a spoken word indicating a different property of the object, a stronger negative deflection was observed in

the Electroencephalogram (EEG) signal compared to when word and gesture referred to the same property. This effect was maximal around 400 milliseconds after the gesture and is commonly known as the N400 effect. In many language studies N400 effects are found when semantic processing of an item (i.e. a word) is harder to integrate into a previous context (see Kutas and Van Petten 1994; Brown et al. 2000). Consequently, Kelly et al. argued for the N400 effect to index semantic processing as triggered by the hand gesture. In a follow-up study, Kelly and colleagues (2006) replicated the N400 effect to incongruent gestures. They furthermore showed that the effect size and its scalp distribution are modulated by whether subjects believed speech and gesture to be acted out by one person or not. That is, when subjects heard an utterance produced by one person while another person produced the accompanying hand gestures, N400 effect size and scalp distribution were different then when speech and gesture were coming from the same person. This was interpreted as reflecting the fact that semantic processing of gesture information is at least to some extent under cognitive control (see also Holle and Gunter 2007). A related ERP study looked at the effect of presenting hand gestures after a more elaborate context, that is, an excerpt of a cartoon movie. Short movie clips of an actor performing a gesture that could either match the preceding cartoon or not, were shown. It was found that hand gestures that do not match a preceding cartoon movie also lead to an increased N400 (Wu and Coulson 2005). The evidence for gestures to evoke semantic processing was further supported by a study looking at the possibility that hand gestures could disambiguate the meaning of an otherwise ambiguous word (Holle and Gunter 2007). Subjects listened to a sentence in which an ambiguous noun was accompanied by a gesture that hinted at the intended meaning of the ambiguous word. An N400 effect was observed to a word later in the sentence if the meaning of that later word did not match with the meaning indicated by the gesture earlier in the sentence. In a later study, Holle and

colleagues showed that the presence of gestures as compared to ‘self-adaptors’ such as scratching oneself leads to increased activation in posterior STS, a region known to be important for multimodal integration (Holle et al. 2008).

In short, these studies indicate that co-speech gestures evoke semantic processing, a claim which had been debated in the literature, as described above.

However, an important remaining question is how comparable the semantic processing evoked by hand gestures is to that of linguistic items such as words. In Chapter 2 we directly compared semantic processing as evoked by meaningful co-speech gestures to spoken words by using the Event-Related Potential (ERP) technique. Moreover, in Chapter 3 the neural loci involved in semantic integration of co-speech gestures and words are assessed in an experiment employing functional Magnetic Resonance Imaging (fMRI). In Chapters 4 and 5, semantic integration of words and pictures of objects are compared. The rationale for the latter studies was to see whether the effects obtained in the gesture-studies would also be present when information was conveyed in a different extra-linguistic representational format, in this case in the format of a picture. Differences may be expected because pictures, contrary to co-speech gestures, are fully recognizable without language. That is, whereas co-speech gestures need the accompanying speech to be meaningfully recognized, this is not the case for a picture of an object.

As described above, behavioural research indicates that the meaning of co-speech gestures is not unambiguously recognised when presented without speech. This clearly sets co-speech gestures apart from other types of actions. Consider for instance *pantomimes*. Pantomimes are actions in which somebody demonstrates an action without using the objects that would normally accompany the action. So if one wants to demonstrate the use of a hammer without actually having a hammer around, one can do so in a pantomimic way by

mimicking to hold a nail in the one hand and a hammer in the other hand. Despite the fact that the objects are not present in this scene, most people will recognise that the person is acting out the act of hammering. So pantomimes rely much less on accompanying language as compared to co-speech gestures. In Chapter 6 it was employed whether this inequality in the ability to signal meaning when presented without speech results in different neural correlates for the integration of speech and gestures on the one hand, and speech and pantomimes on the other hand.

### **1.3 Neural basis of action meaning**

A second issue addressed in this thesis is the neural representation of meaning conveyed through actions without language. The neural basis of action-related meaning has been mostly studied in the context of action-related *language*, such as action verbs. An important question in these studies is: Do words describing actions activate parts of the brain involved in sensori-motor processes, such as premotor cortex? The theoretical starting point for most of this research is embodied cognition. The embodied cognition viewpoint stresses the importance of bodily processes for cognition. As far as action semantics is concerned, in embodied cognition it is hypothesized that action-meaning is partially coded in brain structures involved in sensori-motor processing (e.g. Glenberg and Kaschak 2002; Gallese and Lakoff 2005; see also Pecher and Zwaan 2005).

There is some evidence that, indeed, perception of action-related language leads to activation of the cortical motor system. For instance, Hauk and colleagues (2004) took advantage of the somatotopic organization of the motor cortex to investigate the representation of action verbs. Subject read verbs describing actions performed with the feet, hands or face (e.g. 'kick', 'pick', 'lick'). Subsequently, they performed simple actions with foot, finger or tongue, which activated

primary and premotor cortex in a somatotopic fashion, as expected. Interestingly, reading action verbs led to a similar somatotopic pattern of activation. Overlap between parts of (pre)motor cortex activated by action verbs and by action production was clearly observed for two of the three effectors (see also Tettamanti et al. 2005; Aziz-Zadeh et al. 2006; Vigliocco et al. 2006). Besides these findings there is a considerable amount of other evidence for the claim that listening to action-related language activates cortical motor areas. It seems that parts of the action system such as premotor cortex are involved in the coding of meaning of action language.

Decety and colleagues however found increased activation in left inferior frontal cortex, but not premotor cortex, when they compared meaningful actions (pantomimes) and meaningless actions (sign language signs) (Decety et al. 1997). To resolve these conflicting findings, in Chapter 7 we tested whether premotor cortex is modulated by whether an action has a meaning or not. If so, this would argue for a role of premotor cortex as not passively reacting to the observation of an action, but to be involved in coding the meaning of the action, analogously as has been suggested for action language.

Furthermore, in Chapter 7 we tested the influence of hand preference on neural correlates of action understanding. In this way we tried to assess whether motor cortex is involved in action understanding in a way that is strictly tied to the motor production specifics of the observer.

#### **1.4 Some notes on methods**

In the experiments described in this thesis, two methodologies available for the measurement of neural activity were used. For a full introduction the reader is referred to two excellent handbooks (Huettel et al. 2004; Luck 2005). First, Event-Related Potentials (ERPs) are averaged time segments of the Electroencephalogram (EEG), time-

locked to a cognitive event. EEG is measured from the scalp and reflects the summed activity of electrical fields generated by large groups of neurons. ERPs provide the researcher with a high temporal resolution, in the order of milliseconds. In reaction to a cognitive event, such as the presentation of a stimulus, the ERP exhibits a characteristic shape, with 'peaks' and 'troughs' at several time points. These characteristic deflections in the signal are called components. ERP research over the years has been relatively fruitful in describing which components of the ERP react to which type of cognitive process. In the cognitive neuroscience of language for instance, the so-called N400 component has been found to be sensitive to semantic violations of words into a preceding context. The nature of this component will be discussed in much more detail in some of the chapters below. Another way of looking at the EEG signal is by decomposing the signal into several frequency bands. In so-called time-frequency analysis the time-varying power spectrum of every single trial is computed and trials are subsequently averaged. This means that the outcome reflects the changes in relation to the cognitive event of the ongoing oscillatory activity in the EEG signal. Time-frequency analysis can reveal information that goes unnoticed in traditional ERP analysis. Relating changes in power of a specific frequency band to cognitive events has been relatively commonplace and successful in the study of visual perception and attention (see e.g. Engel et al. 2001; Tallon-Baudry 2003; Jensen et al. 2007), but remains less well studied in the neurocognition of language (but see Bastiaansen and Hagoort 2006 for a recent review). In Chapters 2 and 4, the ERP method was used and in Chapter 5 time-frequency analysis of the EEG signal was employed.

It should be noted that the experimental power of ERP research lies in the possibility of looking at parts of the cognitive process under study at a relatively high temporal resolution. The spatial resolution of the EEG signal is however rather poor. Because the skull is between the cortical activity and the electrode that picks up the electrical signal,

the signal is spatially ‘smeared out’. This means that it is very hard to assess in which part of the brain the activity that was measured at the scalp originated. In functional Magnetic Resonance Imaging (fMRI) it is the other way around. In fMRI the Blood Oxygenation Level Dependent (BOLD) signal is measured while a participant is performing some cognitive task inside an MR scanner. BOLD is an indirect hemodynamic correlate of neural activity which takes advantage of the influx of oxygenated blood after neural activation. This means that the BOLD signal ‘lags behind’ neural activation, typically around 8-12 seconds. Moreover, if one want to measure the whole brain this is typically only possible every 2-2.5 seconds. Taken together, this means that the temporal resolution of fMRI is rather poor. However, localization of activation is typically possible with a resolution of ‘voxels’ (small pieces of brain tissue) of around 3x3x3 mm. In Chapters 3, 4, 6 and 7, fMRI was employed.

### **1.5 Outline of the thesis**

The topic of this thesis is the neural basis of understanding and integrating meaning conveyed through hand actions and through spoken language. First, we investigate how the meaning of iconic co-speech gestures is integrated into a language context (i.e. a sentence). Specifically, we asked whether integration of information from gestures and from spoken words follow a different or a similar neural time course (Chapter 2) and recruit overlapping neural loci (Chapter 3). Subsequently, in Chapters 4 and 5 it is assessed how integration of meaning into a preceding sentence context occurs for other non-linguistic information, i.e. for pictures of common objects. Again, an important question was whether integration of information conveyed through a picture and through a word would follow a different neural time course and / or overlapping neural loci. The combination of these studies allowed for an indirect comparison of non-linguistic information

that is by its very nature bound to language (co-speech gestures) and non-linguistic information conveyed in a format that is not strictly dependent upon a speech context (pictures). In Chapter 6 neural processing of pantomimes was compared to that of co-speech gestures. When presented without speech, the pantomimes were clearly recognizable, whereas the co-speech gestures were not. We investigated how integration of information from language and action may be different for these two classes of meaningful hand movements. In Chapter 7 the neural correlates of meaningful actions presented without language were assessed. We asked whether the cortical motor system plays a role in coding the meaning of an action, as would be suggested from some of the literature described above. Moreover, the influence of hand preference on action observation was investigated. This latter manipulation was used to test the nature of activation of the cortical motor system during action observation. The rationale for every single study is described in more detail in the introduction of each chapter.

Finally, in Chapter 8, the results of each chapter are summarized and put into a broader perspective.



## **Chapter 2** Online integration of semantic information from speech and gesture: Insights from event-related brain potentials\*

### **Abstract**

During language comprehension listeners use the global semantic representation from previous sentence or discourse context to immediately integrate the meaning of each upcoming word into the unfolding message-level representation. Here we investigate whether communicative gestures that often spontaneously co-occur with speech are processed in a similar fashion and integrated to previous sentence context in the same way as lexical meaning. Event related potentials were measured while subjects listened to spoken sentences with a critical verb (e.g., *knock*) which was accompanied by an iconic co-speech gesture (i.e., KNOCK). Verbal and / or gestural semantic content matched or mismatched the content of the preceding part of the sentence. In spite of differences in modality and in specificity of meaning conveyed by spoken words and gestures, the latency, amplitude and topographical distribution of both word and gesture mismatches are found to be similar, indicating that the brain integrates both types of information simultaneously. This provides evidence for the claim that neural processing in language comprehension involves the simultaneous incorporation of information coming from a broader domain of cognition than only verbal semantics. The neural evidence for similar integration of information from speech and gesture emphasizes the tight interconnection between speech and co-speech gestures.

---

\*This chapter is a slightly modified version of: Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19(4), 605-616.

## **Introduction**

In ordinary face-to-face conversation, language users not only hear speech but also see the speaker's hand, mouth and body movements. The listener's brain therefore continuously integrates spoken language information with several streams of visual information including information from the lips, the eyes and, crucially, semantic information from the hand gestures that accompany speech (McNeill, 1992). For example, when talking about drinking a glass of milk, speakers often perform a concomitant drink gesture (i.e., C shaped hand moved towards the mouth) as they say 'drink' in their spoken utterance. Yet, whether and how listeners integrate the semantic information from co-speech gestures on-line into the previous sentence context, and how this compares to the integration of spoken words has not been addressed.

So far, most studies on language comprehension have focused on the online processing of the acoustic and written input in isolation (but see Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy, 1995, for visual world related processing). Studies of on-line language comprehension using the event related potential (ERP) technique have shown that spoken words are integrated into a context representation in an incremental way. That is, listeners use the global semantic representation from the sentence or discourse context to integrate the meaning of each upcoming word immediately into an overall message representation (e.g. Kutas and Hillyard 1980; van Berkum et al. 1999; Hagoort 2003b; van Berkum et al. 2003; Hagoort and van Berkum 2007).

Previous studies on multi-modal processing during language comprehension have often investigated the relationship between speech and lip movements by exploiting the McGurk effect (e.g., acoustic /pa/ combined with visual /ka/ perceived as /ta/ (McGurk and MacDonald 1976). These studies using electrophysiological recordings have shown that visual information from articulation interacts with the auditory

information quite early, that is, within 200 ms during audio / visual speech observation (e.g. Sams et al. 1991; Mottonen et al. 2002; Mottonen et al. 2004). However, little is known about how other types of visual information, such as gestures, are processed in relation to speech. The relationship between lip movements and syllables is based on form matching whereas the relation between speech and gestures is based on meaning. Thus the latter might be processed in a different way, that is, at a higher, semantic level.

#### *The role of co-speech gestures in communication*

Here we focus on a ubiquitous form of communication that speakers use along with speech, namely meaningful hand movements, usually referred to as co-speech gestures (McNeill 1992, 2000; Goldin Meadow 2003; Kendon 2004). During face-to-face conversation, along with speech speakers spontaneously use different types of gestures. These can be classified as either iconic (e.g., hands represent a climbing action), deictic (e.g., pointing), or emblematic (e.g., thumbs-up, OK etc.). In this study, we focus on iconic gestures which convey information about the shape, size, motion and action characteristics of the events described in the spoken utterance. These gestures are meaningful within the speech context, but do not have conventional or unambiguous meanings in the absence of speech (Feyereisen et al. 1988; Krauss et al. 1991).

Iconic gestures have different representational properties than speech in terms of the meaning they convey. Consider for example an upward hand movement in a climbing manner when a speaker says: “the cat climbed up the tree”. Here, the gesture depicts the event as a whole, describing manner (‘climb’) and direction (‘up’) simultaneously, whereas in speech the message unfolds over time, broken up into smaller meaningful segments (i.e. different words for manner and direction). However, in spite of these differences in representational format, the information expressed in the two modalities is

systematically related to each other (McNeill 1992, 2000; Kendon 2004; Bernardis and Gentilucci 2006).

The systematic relationship between speech and gestures exists at three levels. First, there is semantic overlap between the representation in gestures and the meaning expressed in the concurrent speech, as in the 'climb up' example above (e.g. McNeill 1992; Kita and Özyürek 2003). Speech and gesture convey related and similar information. Second, speech and gesture are temporally aligned to each other. A gesture phrase has three phases: the preparation, the stroke (semantically the most meaningful part of the gesture), and the retraction or hold (McNeill, 1992). Studies have shown that the onset of the gesture phrase (i.e., preparation) usually precedes the onset of the relevant speech segment by less than a second (Butterworth and Shovelton 1978; Morrel Samuels and Krauss 1992). More importantly, in most speech-gesture pairs the stroke coincides with the relevant speech segment (McNeill, 1992). Finally, it has been shown that the spontaneous use of gestures has a similar function as speech (e.g. Özyürek 2002; Kendon 2004; Melinger and Levelt 2004), namely to communicate the intended message to the addressee.

A considerable amount of behavioural studies on speech and gesture comprehension has shown that listeners / viewers pay attention to iconic gestures and pick up the information that they encode. For example, Graham and Argyle (1975) had speakers describe abstract line drawings with and without gestures, and required listeners to make drawings on the basis of the speakers' input. Listeners were more accurate in their drawings in the speech-and-gesture condition than in the speech-alone condition. In another study, Beattie and Shovelton (1999) showed that listeners answer questions about the size and relative position of objects in a speaker's message more accurately when gestures are part of the description than when gestures are absent.

Another set of studies has investigated whether listeners pick up the information in gesture when gesture conveys different information than speech. McNeill and colleagues (1994) presented listeners with a videotaped narrative in which the semantic relationship between speech and gesture was manipulated. It was found that listeners / viewers incorporated information from the gestures in their retellings of the narratives, and attend to the information conveyed in gesture when that information complemented or even contradicted the information conveyed in speech (see also Kelly and Church 1998; Goldin Meadow and Momeni Sandhofer 1999; Singer and Goldin Meadow 2005).

Despite firm evidence that co-speech gestures contribute to comprehending the speaker's message, not much is known about the nature of the on-line cognitive processes underlying the comprehension of co-occurring multi-modal semantic information from speech and gesture. The present study investigates the integration of speech and gesture occurring simultaneously, and embedded into a sentence context. For this purpose, we exploited an Event-Related Potential (ERP) paradigm that is often used for studying the nature of on-line semantic integration in sentence and discourse contexts.

#### *ERP studies on semantic integration during comprehension*

ERPs are voltage deflections generated by the brain and recorded from electrodes placed on the scalp. One important characteristic of ERPs is their high temporal resolution, which is in the order of milliseconds. Especially the processing of semantic information has been found to influence the amplitude of a negative-going ERP component between 250-550 ms. This amplitude modulation is referred to as the N400 effect and is usually larger over posterior electrodes than over frontal sites (Kutas and Hillyard 1980).

N400 studies have typically employed a paradigm, in which the semantic integration load of a word in relation to the preceding

sentence context is manipulated. Kutas and Hillyard (1980) were the first to observe that relative to a semantically acceptable control word, a sentence-final word that is semantically anomalous in the sentence context, as in “He spread the warm bread with *socks*”, elicits an N400 effect. Additional studies have shown that it does not require a semantic violation to elicit an N400 effect. In general, N400 effects are triggered by more or less subtle differences in the semantic fit between the meaning of a word and its context, where the context can be a single word, a sentence or a discourse (e.g. Kutas and Hillyard 1984; Hagoort and Brown 1994; van Berkum et al. 1999; van Berkum et al. 2003)

More recent studies on semantic processing have investigated how extralinguistic information such as world knowledge or pictorial information is integrated into previous context. Hagoort and colleagues (2004) showed their subjects sentences that contained either a semantically anomalous word (e.g., “Dutch trains are *sour* and very crowded”) or a world knowledge violation (e.g., “Dutch trains are *white* and very crowded”). The N400 effects to the semantic and to the world knowledge violations were identical in their latency and topography. These results indicate that even in the case of extralinguistic information such as world knowledge, the brain integrates this information immediately, that is with the same temporal profile as lexical-semantic information (see Hagoort and van Berkum 2007).

Processing of extralinguistic information has also been investigated in terms of integrating information from pictures to previous context (Barrett and Rugg 1990; Ganis et al. 1996; Federmeier and Kutas 1999; McPherson and Holcomb 1999; Federmeier and Kutas 2001, 2002; West and Holcomb 2002). In picture priming studies, an N300 has been reported that is more negative for unrelated than for related pictures (Barrett and Rugg 1990; Holcomb and McPherson 1994; McPherson and Holcomb 1999). This N300 has a frontal distribution and is not reported in ERP studies that used only

linguistic stimuli. The N300 was followed by a more widely distributed N400 effect. However, in other studies in which either anomalous words or pictures were presented in a sentence context, only N400 effects were found (Nigam et al. 1992; Ganis et al. 1996). In these studies, the pictures elicited an N400 effect with a more frontal distribution than is usually observed for language stimuli. Finally, studies investigating the semantic integration of pictures to a scene or event without using any linguistic context sometimes (West and Holcomb 2002), but not always (Ganis and Kutas 2003; Sitnikova et al. 2003), found a frontal N300 preceding an N400. In the light of these findings it is especially interesting to see how iconic gestures compare to semantic integration of pictures. Gestures can be claimed to share certain visual characteristics with pictures. However, they do not have the exact semantic specificity of pictures, since unlike pictures the full interpretation of gestures depends on the semantic content of the accompanying speech.

Finally, two recent priming studies have investigated the modulation of ERPs to words preceded by gestures, or to gestures preceded by cartoon images. Kelly, Kravitz and Hopkins (2004) found that ERPs to spoken words (targets) are modulated when these words are preceded by gestures (primes) that contained information about the size and shape of objects that the target words referred to. Compared to matching target words, mismatching words evoked an early P1 / N2 effect, followed by an N400 effect. On the basis of these findings Kelly et al. (2004) claimed that the gesture primes influenced word comprehension, first at the level of 'sensory / phonological' processing and later at the level of semantic processing. In a recent study by Wu and Coulson (2005) it was found that congruous and incongruous gestures shown without speech and following cartoon images elicit a negative-going ERP effect around 450 ms. In addition, it was observed that congruous or incongruous words following the cartoon-gesture pairs elicited an N400 effect. However, neither of these studies has

investigated speech and gesture comprehension within sentence context and when they occur simultaneously as in everyday conversations.

### *The present study*

The present study investigates the integration of speech and gesture presented simultaneously and embedded into a sentence context. Using a similar ERP paradigm as for investigating the semantic integration of words, we aim to compare the latency, the amplitude, and the topography of gesture integration to sentence context with the integration of spoken words. Our main focus is on understanding how the integration of conceptual information from gestures into a previous sentence context (i.e., global integration) compares to integration of semantic information from spoken words. Second, we also investigate how listeners / viewers comprehend and integrate the information from the temporally overlapping speech and gesture segments (i.e., local integration). For example, when a listener hears “the cat climbed up the tree and caught the bird” and sees a CATCH gesture as he / she hears the word ‘caught’, the comprehension of gesture in relation to previous sentence would be ‘global integration’ and its relation to the verb ‘catch’ is referred to as ‘local integration’. Thus we aim to reveal the underlying nature and time course of these two types of multi-modal integration processes.

The particular questions that we investigated are: (i) Are gestures and speech integrated to previous sentence context simultaneously, or is speech integrated first and gesture later? (ii) How does the integration of gesture information to previous sentence context (i.e., global integration) compare to integration of gesture information to the temporally overlapping word (i.e., local integration)?

In order to determine the nature of the integration of verbal and gestural semantic information, we manipulated the semantic fit of speech (i.e., a critical verb) and / or gesture in relation to the preceding

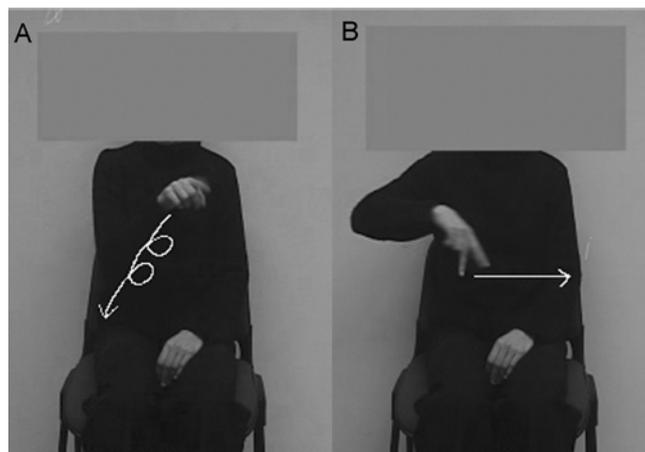
A) Correct condition (L+G+):	He slips on the roof and <u>rolls</u> down [roll down]
B) Language mismatch (G+L-):	He slips on the roof and <b>walks</b> to the other side [roll down ]
C) Gesture mismatch (G-L+):	He slips on the roof and <u>rolls</u> down [walk across]
D) Double mismatch (G-L-):	He slips on the roof and <b>walks</b> to the other side [walk across]

**Table 2.1.** An example of the Materials. In brackets [ ] is a verbal description of the iconic gesture. Gestures were time-locked to the onset of the critical verb (underlined). ERPs were time-locked to the beginning of the critical word and the gesture in each sentence. The condition coding (G+L+; G+L-, etc.) refers to the match / mismatch of either the verb (Language: L) or the gesture (Gesture: G) to the preceding sentence context, with a minus sign indicating a mismatch. Mismatches to the preceding context are indicated in bold. Conditions B and C also contain local mismatches where the concurrent speech and gesture are different. All stimuli were in Dutch.

part of the sentence (global integration) as well as the semantic relations between the temporally overlapping gesture and speech (local integration) (see Table 2.1).

Movie clips of twelve iconic gestures were temporally aligned to the critical verbs in the sentences. This manipulation resulted in four conditions (see Table 2.1, Fig. 2.1 Correct condition (Gesture (G) +, Language (L) +); Language mismatch condition (G+L-); Gesture mismatch condition (G-L+); Double mismatch condition (G-L-). In the Language mismatch the critical verb was harder to fit semantically to the preceding context while the co-occurring gesture matched the

sentence context. In the Gesture mismatch condition the gesture was harder to integrate to previous context, while the critical verb matched the spoken sentence context. In the Double mismatch condition both the gesture and the word were difficult to integrate to previous sentence context. Note that in the Language and Gesture mismatch conditions the critical verb and the overlapping gesture locally mismatched (i.e., Speech: ‘Roll’; Gesture: WALK, and vice-versa), while in the Double mismatch condition they locally matched (i.e., both ‘Walk’). This extra manipulation allowed us to investigate and compare the effects of local and global integration of speech and gesture in sentence context.



**Fig. 2.1.** Examples of the gesture movies. Stills from two gestures that were used as stimuli: **A)** Roll down; **B)** Walk across.

In our materials, an increase of semantic integration load does not necessarily involve a semantic violation. The meaning of the critical verb in the mismatch condition, however, always fits the previous sentence context less well than the meaning of its counterpart in the correct condition. ERP studies in language processing have found that semantically less expected critical words elicit an increase in the amplitude of the N400, just as semantic violations do (Hagoort and

Brown, 1994; Kutas and Hillyard, 1984). For reasons of simplicity we will refer to conditions in which speech and / or gesture is harder to integrate as ‘mismatches’.

If the brain uses an incremental and parallel processing of linguistic and extralinguistic information as found in previous studies (Hagoort et al., 2004), we expect a similar latency and amplitude of the N400 effect for all types of mismatches (i.e., Language, Gesture and Double) revealing that the brain integrates information from both speech and gesture at the same time. These results would also be in line with the claims that speech and gestures are tightly linked systems of communication (Kendon, 2004; Goldin-Meadow, 2003; Özyürek, 2002; Clark, 1996; McNeill, 1992, 2000). However, if the latency of the N400 effect was found to be later for the Gesture mismatch than for the Language mismatch, this would support a speech-first-gesture-later model of comprehension. This model is compatible with the view that the semantic interpretation of sentences precedes the integration of pragmatic, extralinguistic information (Forster 1979). It would be also in line with the view of Krauss, et al. (1991) that the meaning we assign to gestures is mostly constructed from the meaning of concurrent speech, and that gestures do not add any information to what the listener picks up from the concurrent speech. Accordingly, gestural information will have to be integrated after the relevant speech segment has been interpreted (if it is integrated at all).

Furthermore, according to the incremental processing principle, we do not expect differences across conditions with local mismatches (Language and Gesture mismatches) and the condition with the local match (Double mismatch), since integration takes place immediately in relation to a discourse model and not in multiple steps from lower to higher levels of semantic organization (van Berkum et al. 1999; van Berkum et al. 2003). According to this view, the gesture and the concurrent speech segment (i.e. the verb) are integrated in parallel into

the preceding context, and not after they first formed a common semantic object. Alternatively, it might be argued that the local conflict between speech and gesture has to be resolved first, before the global integration can take place in the local mismatch conditions. In this case, the double mismatch effect should precede the effects for the single language and gesture mismatches, since in this condition a local integration problem is absent.

### **Materials and Methods**

*Participants* Sixteen healthy subjects (12 female; mean age=22.4, range=19-27) with normal or corrected to normal vision and no hearing complaints took part in the study. All subjects were right-handed and had Dutch as their mother tongue. Subjects gave written informed consent and were paid for participation.

*Materials* The materials consisted of 320 spoken Dutch sentences. The sentences were spoken by a female native speaker of Dutch and digitized at a sample frequency of 44.1 kHz. The sentences formed 160 sentence pairs. The members of the pair were identical up until the critical verb. Half of the sentences contained a critical verb that matched the preceding context. In the other half, the critical verb was semantically anomalous in relation to the prior sentence context. Overall, twelve different critical verbs were used (see Appendix with this chapter). For each sentence, the onset of the critical verb was determined by using the speech analysis software package Praat (version 4.0; <http://www.praat.org>). The sentences had an average duration of 3720 ms (SD=81), and the critical verbs had an average duration of 322 ms (SD=85 ms).

The spoken sentences were combined with twelve iconic gestures (see Appendix with this chapter). Iconic gestures are a class of gestures that speakers spontaneously use as they talk about spatial and activity related aspects of events (e.g., using wiggling fingers moving

horizontally while talking about someone walking). The iconic gestures used in this study were based on a larger database collected to investigate speakers' natural and spontaneous use of speech and gestures in narratives of spatial events (Özyürek 2002; Kita and Özyürek 2003). For the purposes of this study, twelve of these gestures were selected and modelled by a native Dutch speaker with the requirement that they resembled spontaneous gestures in this database. Modelled gestures were preferred over natural ones from the database to make each gesture comparable across the conditions in terms of gesture space that was used, the handedness, and the gesturing person. In order to match the speed and length of the gestures as closely as possible to naturally occurring ones, we asked our model to produce concurrent sentences as she was performing the gestures. The gestures were filmed by using a digital camera (Sony, TCR-TRV950, PAL). During editing the audio was removed from the movie. Movies were edited using Adobe Premier (version 6.0; Adobe Systems Inc., San Jose, USA; <http://www.adobe.com>). The preparation and the retraction phase of each gesture were removed, leaving the stroke. Previous research has shown that especially the stroke phase conveys the meaning of a gesture (McNeill, 1992). By isolating the gesture stroke phase, we eliminated differences among gestures that were due to the fact that for some gestures hand shape might reveal information before the stroke began, and / or that some gestures might have longer preparation time than others. The average length of the strokes was 767 ms (SD=284 ms). Finally, the face of the model was blocked to eliminate the contribution of information coming from the lips.

The gestures corresponded to the meaning of the critical verbs. They were combined with the sentence pairs in such a way that in half of the items the gesture matched the preceding sentence context, and in the other half it mismatched the preceding sentence context. This resulted in a total of 160 stimulus quartets (see Table 2.1).

The gesture movies and the sentence files were combined using the Adobe Premier (version 6.0) and After Effects software (version 5.5; Adobe Systems Inc., San Jose, USA, <http://www.adobe.com>). For each movie file, the onset of the gesture stroke was temporally aligned with the onset of the critical verb, since in 90% of natural speech-gesture pairs the stroke coincides with the relevant speech segment (McNeill, 1992). For verbs with a separable prefix, the alignment point was not word-onset, but the body of the verb following the prefix. The latter was the case for 44 sentences. Additional still frames with the hand resting on the lap were added to the part of the sentence before the critical verb, and the last frame of the stroke was elongated until the end of the sentence.

Four different stimulus lists were created, to distribute the four versions of each item equally over the four lists (see Table 2.1). This was done in such a way that all four lists contained an equal number of items (40) per condition. Each list was presented to one quarter of the participants. As a result, none of the participants were presented with more than one item of a stimulus quartet as in Table 2.1.

*Experimental Procedure* The stimuli were presented using the Nijmegen Experiment Setup programme (NESU, MPI for Psycholinguistics). The visual content of the movies was presented via a computer screen. The subjects watched the movies at a distance of 80 cm from the screen. The size of the movie frame was 10 cm in height and 11.8 cm in width. The movies were presented at 25 frames per second. Speech was presented to the subjects through headphones.

Subjects were instructed to carefully listen to the sentences and watch the movies without a specific task. They were given the instruction that they could blink or move their eyes only during the inter-stimulus-intervals when a fixation cross was shown. The fixation cross was presented between the movies for a duration of 3600 ms. Finally, they were told that they would receive general questions about

the items after the experiment to make sure that they would attend to the items.

The test session started with a practice block of 30 practice items to familiarize the subjects with the procedure. The whole test session lasted approximately 40 minutes.

*EEG recording and analysis* The electroencephalogram (EEG) was recorded from 26 electrode sites across the scalp using an Electrocap with 26 Ag / AgCl electrodes, each referred to the left mastoid and off-line re-referenced to average mastoids. Electrodes were placed on midline (Fz, FCz, Cz, Pz), frontal and fronto-central (F3, F4, F8, F7, FC5, FC1, FC2, FC6), temporal (T7, T8), central (C3, C4), centroparietal (CP5, CP1, CP2, CP6), parietal (P7, P3, P4, P8) and occipital (O1, O2) sites. Vertical and horizontal eye movements were monitored via a supra- to suborbital bipolar montage and a right-to-left canthal bipolar montage, respectively. Activity over the right mastoid was recorded on an additional channel to determine if there were additional contributions of the experimental variables to the two presumably neutral mastoid sites. No such differences were observed.

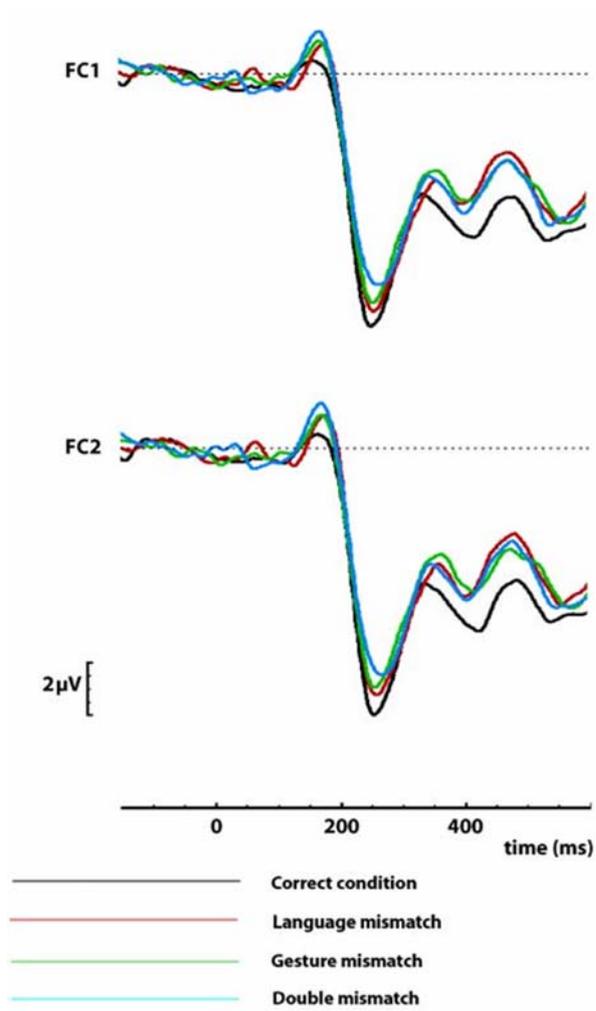
The EEG and the electrooculogram (EOG) recordings were amplified with BrainAmp DC amplifiers. A bandpass filter was applied from 10 s to 70 Hz. Impedances were kept below 5 kOhm for all channels. The EEG and EOG signals were recorded and digitized using Brain Vision Recorder Software (version 1.03), with a sampling frequency of 500 Hz.

Prior to off-line averaging, all single-trial waveforms were screened for eye movements, electrode drifting, amplifier blocking and muscle (EMG) artefacts in a critical window that ranged from 150 ms before to 1000 ms after the onset of the critical verb and the gesture stroke. Trials containing such artefacts were rejected (7.7 %). Rejected trials were equally distributed across conditions.

ERPs time-locked to the onset of the critical verb and the gesture were averaged after baseline-correction by subtracting the mean amplitude in the -150 to 0 ms pre-stimulus interval, for each condition (Correct, Gesture mismatch, Language mismatch, Double mismatch) for each subject at each electrode site. Repeated measures analyses of variance (ANOVAs) with the factors Match (Correct, Gesture mismatch, Language mismatch and Double mismatch) and Quadrant (Left Anterior: F3, F7, FC1, FC5, C3; Right Anterior: F4, F8, FC2, FC6, C4; Left Posterior: CP1, CP5, P3, P7, O1; and Right Posterior: CP2, CP6, P4, P8, O2) were conducted for three time windows. Separate ANOVAs were conducted for the midline electrodes. Huynh-Feldt correction for violation of sphericity was applied when appropriate (Huynh and Feldt 1976).

## Results

Figure 2.2 displays the grand-average waveforms time-locked to the onset of critical verbs and gesture strokes. A visual inspection of the waveforms (see Fig. 2.2) shows an N1 followed by a P2, and a negativity with a bimodal morphology peaking at about 380 ms and 480 ms respectively. Apart from a slightly smaller N1 in the correct condition, the waveforms suggest that the mismatch conditions started to deviate from the correct condition in the latency window of the P2 component around 225-275 ms. However, this effect seems especially strong for the correct condition, which could be a carry-over from the reduced N1 amplitude in this condition. Next, around 350 ms the mismatch conditions deviate from the correct condition. This effect is followed by a similar modulation between 410 and 550 ms, with a peak latency that is slightly later than is usually seen for the N400. For the P2, especially the Double mismatch seemed to show a reduced amplitude. A repeated measures ANOVA on the mean amplitudes in the 225-275 ms latency range, with the factors Match and Quadrant failed to show a significant main effect of Match ( $F(3,45)=1.90$ ;



**Fig. 2.2.** (For colour version see Appendix, p. 265). Grand-average waveforms for ERPs elicited in the three mismatch conditions and the correct condition at two representative electrode sites (FC1 and FC2). Negativity is plotted upwards. Waveforms are time locked to the onset of spoken verb and gesture (0 ms).

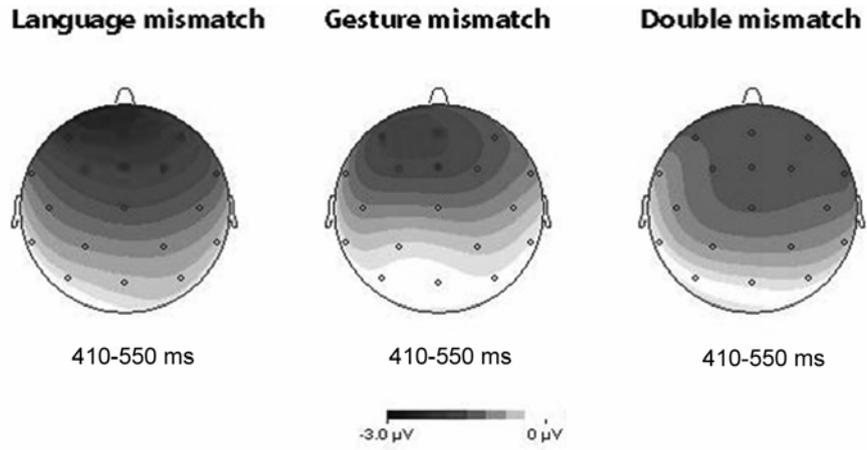
MSe=22.1;  $p=0.14$ ) . There was also no significant Match by Quadrant interaction ( $F < 1$ ). In addition, the ANOVA over the midline sites failed to reach significance ( $F(3,45)=2.29$ ; MSe=10.29;  $p=0.09$ ). However, in a planned comparison, a significant difference between the Double mismatch and the Correct condition was found ( $F(1,15)=4.69$ ; MSe=5.63;  $p < 0.05$ ). Also over the midline electrodes, this planned comparison showed that ERPs to the Double mismatch were less positive than those to the Correct condition ( $F(1,15)=5.28$ ; MSe=26.17;  $p < 0.05$ ). In addition, in a planned comparison it was found that over the midline sites also the Language mismatch was significantly less positive than the Correct condition ( $F(1,15)=4.7$ ; MSe=11.4;  $p < 0.05$ ). However, these effects are qualified by the fact that for the N1, the Correct condition shows a smaller amplitude than the other conditions. If we take this unexplained early difference into account by using another baseline (100-200 ms), no significant differences remain. In short, the P2 effect observed for the Double mismatch does not seem to be a stable effect. The conclusion that there is an earlier effect for the Double mismatch than for the Language and Gesture mismatches would therefore be premature.

The next window in which effects were tested was in the latency range of 350-410 ms. This is the window around the first negative peak (approximately at 380 ms) in the waveforms following the P2. For this latency window, repeated measures ANOVAs with the factors Match and Quadrant did not show a significant main effect for Match ( $F(3,45)=1.22$ ; MSe=18.1;  $p=0.32$ ), nor a significant Match by Quadrant interaction ( $F(9,135)=1.40$ ; MSe=5.43;  $p=0.23$ ). Additional planned comparisons resulted in significant differences between the Gesture mismatch and the Correct condition in the Left Anterior quadrant ( $F(1,15)=7.92$ ; MSe=20.49;  $p < 0.05$ ), the Right Anterior quadrant ( $F(1,15)=14.77$ ; MSe=12.19;  $p < 0.005$ ) and over the midline sites ( $F(1,15)=6.18$ ; MSe=12.5;  $p < 0.05$ ). In addition, the Language mismatch condition showed a marginally significant effect in the Left Anterior

quadrant ( $F(1,15)=4.48$ ;  $MSe=13.18$ ;  $p=0.051$ ), and just failed to reach significance over the midline sites ( $F(1,15)=3.40$ ;  $MSe=21.64$ ;  $p=0.085$ ). For the Double mismatch no significant effects were found in this latency window. However, in contrasts testing the differences between the three mismatching conditions, no significant effects were obtained.

Finally, the average waveforms were tested in the time window of 410-550 ms. As can be seen in Figure 2.2, all three types of mismatch elicit a clear negative deflection that all peak around the same time. Moreover, the topographic distribution shows that for all three mismatches, the effects are maximal over anterior sites, without a clear hemispheric dominance (see Fig. 2.3). The results of repeated measures ANOVAs in this latency window are summarized in Table 2.2. The main effect of Match is modulated by an interaction between Match and Quadrant, due to the clear anterior distribution of the condition effects. Planned comparisons conducted in separate quadrants revealed significant differences between all three mismatch conditions and the Correct condition for the anterior electrode sites (both left and right hemisphere sites), as well as for the midline sites (with the exception of the Double mismatch). Further planned comparisons between the three mismatch conditions did not reveal any significant differences. No significant effects were obtained over posterior quadrants.

Thus, the results show that Language, Gesture and Double Mismatch conditions modulated the N400 in a similar way, in terms of N400 latency and amplitude. In all conditions, the N400 component reached its peak around 480 ms. Furthermore all conditions showed a similar topographical distribution.



**Fig. 2.3.** Spline-interpolated isovoltage maps displaying the topographic distributions of the mean differences from 410-550 ms between the three types of mismatches and the correct condition.

Source	df	F	MSe	p
Omnibus ANOVA				
Match	3, 45	2.84	14.90	0.048*
Match x Quadr.	9, 135	3.12	4.91	0.013*
Left Anterior Quadr.				
Match	3, 45	4.01	8.02	0.013*
Planned comp.:				
L-G+	1,15	16.8	10.08	0.001**
L+G-	1,15	9.38	11.54	0.008**
L-G-	1,15	4.46	19.07	0.052
Right Anterior Quadr.				
Match	3, 45	6.03	5.31	0.002**
Planned comp.:				
L-G+	1,15	16.77	9.38	0.001**
L+G-	1,15	8.60	10.46	0.01*
L-G-	1,15	9.35	13.14	0.008**
Midline Sites				
Match	3, 45	3.15	8.10	0.034*
Planned comp.:				
L-G+	1,15	7.92	18.03	0.013*
L+G-	1,15	4.37	16.19	0.054
L-G-	1,15	2.73	23.39	0.12

**Table 2.2.** Repeated measures ANOVAs on mean ERP amplitudes for the four experimental conditions in the 410-550 ms latency range. Huynh-Feldt correction is applied when appropriate. The original degrees of freedom are reported. Planned comparisons are always against the Correct condition. L-G+: Language mismatch; L+G-: Gesture mismatch; L-G-: Double mismatch.

## **Discussion**

This study investigated the semantic integration of words and iconic gestures into a sentence context when they both occur simultaneously as in natural speech and gesture production. The most important finding of the study is that co-occurring speech and gestures are integrated simultaneously into a preceding sentence context. That is, semantic information provided by both spoken words and visual gestures is integrated within 350-550 ms after word and gesture onset. The time course of the observed N400 effects testifies to the immediacy of contextual integration, since in many cases they occur well before the end of the acoustic word token or the visual gesture. As the topographic distributions of the gesture and word integration effects are identical, it is most parsimonious to assume that the nature of the semantic integration process is very similar in both cases.

No solid evidence was obtained that the effect for the Double mismatch came earlier than the single mismatch effects (i.e., Gesture and Language mismatches). In the Double mismatch condition, the co-occurring critical verb and the gesture provided compatible semantic information (i.e., local match). This was different in the Language and Gesture mismatch conditions. In these conditions, the co-occurring verb and gesture were mutually inconsistent (i.e., local mismatch). This local mismatch, however, did not seem to modulate the global mismatch effect, which is the effect triggered by the mismatch in relation to the preceding sentence context. More in particular, no evidence was obtained that the effect for the Double mismatch (i.e. the local match) preceded the effects of the Language and Gesture mismatches (i.e. the local mismatch). This suggests that verb and gesture are not first integrated together to form a common semantic object, before integration into the preceding context takes place. Instead, verb and gesture seem to be integrated in parallel. This is in line with the view supported by N400 data in Van Berkum et al. (1999, 2003) that semantic integration takes place immediately in relation to

a discourse model rather than in a series of sequential steps from lower to higher levels of semantic organization.

In terms of their latency and amplitude characteristics, the effects are similar to the well-known N400 effect that is observed if word meaning violates the semantic context (Kutas and Hillyard 1980). However, the waveforms show a clearly biphasic morphology, and the effects have a more anterior distribution than is reported for the classical N400 effect. The first negative peak in the biphasic negativity is reminiscent of the N300 that has been reported before for visual materials, and which has been found to be more negative for unrelated than for related pictures (Barrett and Rugg 1990; Holcomb and McPherson 1994; McPherson and Holcomb 1999). The N300 effect might be related to the presence of the visual-gestural information.

For the N400 an anterior distribution has been observed before for visual information such as pictures (e.g. Ganis et al. 1996; Federmeier and Kutas 2001; West and Holcomb 2002). In the current study the visual characteristics of the gestures might have elicited a frontal distribution. It is interesting here to note that even the Language mismatch condition elicited an anterior effect, which suggests that the mere presence of a simultaneous gesture is responsible for the anterior distribution, even when the integration problem is located in the speech channel. The finding that all mismatch conditions have similar topographic distributions suggests that semantic integration of information from both modalities might be instantiated by overlapping neuronal sources. Interestingly, it suggests that with respect to contextual integration, there is no reason to distinguish between visual semantics and verbal semantics.

As a cautionary note, we want to point out that the N300 and N400 effects are descriptive labels. There is no evidence that both effects are independently modulated, or generated by non-overlapping neural generators. Earlier studies involving visual materials have reported both N300 and N400 effects. We have chosen our descriptive

terms here in connection to these earlier studies. However, our main conclusions do not depend on the question whether or not N300 and N400 effects are one and the same extended negativity.

The present study and also the studies by Kelly et al. (2004) and Wu and Coulson (2005) point to the fact that iconic gestures trigger semantic processing, as is indicated by the presence of N400 effects. However, the current study differs from these earlier studies in crucial ways. In these studies, words and gestures were presented sequentially, and ERPs were measured to either word targets preceded by gestures or to gesture targets preceded by cartoon images. In the present study, the gestures and the relevant speech segments were presented simultaneously as they naturally occur, and furthermore in a sentence context by which the integration of gestural information to speech context beyond single word and gesture levels could be investigated.

It is also important to note that we found an N400 effect instead of the earlier negativities normally reported to speech-lip movement mismatches in the McGurk effect (e.g. Mottonen et al., 2002; Sams et al., 1991). This provides evidence that speech and gesture integration occurs at a higher semantic level than the integration of information from lip movements and speech sounds. That is, different types of multi-modal information are processed in different ways in the brain, even though both concern processing relations between speech and visual movements.

Finally, our results parallel those of Chapter 3 (Willems et al. 2007), using the same stimuli in a design with the same conditions. In the fMRI study it was found that all mismatch conditions activated a common area, namely the left inferior frontal cortex. This area has been claimed to be crucial for the integration of semantic information into previous context (Hagoort 2003a; Hagoort et al. 2004; Hagoort 2005b; Hagoort and van Berkum 2007). Together with the ERP results of the current study, the fMRI data suggest that the semantic

integration of both speech and gesture semantics to sentence context involves very similar processes.

In conclusion, when understanding an utterance, the brain does not restrict itself to language information alone, but also integrates semantic information conveyed through other modalities, such as co-speech gestures. Furthermore, the neural sources and the time course of the integration processes seem to be similar across gesture and language semantics. Both constrain the interpretation domain simultaneously during on-line processing. This opens the interesting possibility that language comprehension involves the incorporation of information in a ‘single unification space’ (Hagoort 2003a; Hagoort et al. 2004; Hagoort 2005b; Hagoort and van Berkum 2007), coming from a broader range of cognitive domains than is usually thought. The neural evidence for the tight link between speech and gesture that we observed underscores the fact that in natural conversation speech and gesture are often tightly interconnected (McNeill 1992; Clark 1996; McNeill 2000; Özyürek 2002; Goldin Meadow 2003; Kita and Özyürek 2003; Kelly et al. 2004; Kendon 2004; Bernardis and Gentilucci 2006). Further research has to reveal if in this sense co-speech gestures are special, or representative of a broad domain of visual information constraining on-line sentence interpretation (Tanenhaus et al., 1995).

### **Acknowledgements**

Supported by a grant from the Netherlands Organization for Scientific Research (NWO), 051.02.040, and Biotechnology and Biological Sciences Research Council, UK (BBS / B / 08906). We thank Femke Deckers, Miriam Kos, Nienke Weder, and Tineke Snijders for their assistance during the running of this experiment, and Jos van Berkum for his comments on an earlier version of this article.

**Appendix Chapter 2** List of critical verbs (originals in Dutch) and gestures used as stimuli within sentence context

<b>Critical verb</b>	<b>Gesture</b>	<b>Gesture description</b>
Break	BREAK	Fist hands make a break motion from the middle to the sides and down
Give	GIVE	Hand opens up as it moves forward
Knock	KNOCK	Fist hand moves back and forth
Punch	PUNCH	Fist hand make a punching motion away from body
Push	PUSH	Both flat hands move away from body
Roll away	ROLL_AWAY	Index finger pointing to the right makes circles as it moves away from the body
Roll down	ROLL_DOWN	Index finger pointing away from body makes circles as it moves down and left
Swing across	SWING_ACROSS	Index finger pointing away from body moves left making an arc
Swing away	SWING_AWAY	Index finger pointing towards right moves away from body making an arc
Walk away	WALK_AWAY	V handshape with wiggling fingers moves forward away from self
Walk across	WALK_ACROSS	V handshape with wiggling fingers moves left horizontally
Write	WRITE	One hand makes a writing gesture moving to the right

## **Chapter 3** When language meets action: The neural integration of gesture and speech\*

### **Abstract**

Although generally studied in isolation, language and action often co-occur in everyday life. Here we investigated one particular form of simultaneous language and action, namely speech and gestures that speakers use in everyday communication. In an fMRI study, we identified the neural networks involved in the integration of semantic information from speech and gestures. Verbal and / or gestural content could be integrated easily or less easily with the content of the preceding part of speech. Premotor areas involved in action observation (BA 6) were found to be specifically modulated by action information ‘mismatching’ to a language context. Importantly, an increase in integration load of both verbal and gestural information into prior speech context activated Broca’s area and adjacent cortex (BA 45 / 47). A classical language area, Broca’s area, is not only recruited for language-internal processing, but also when action observation is integrated with speech. These findings provide direct evidence that action and language processing share a high-level neural integration system.

---

\*This chapter is a slightly modified version of: Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex* 17(10):2322-33.

## **Introduction**

Language and action are two core systems of human cognition. Moreover, they are often used together, as in pointing towards an object while producing its name. Despite this common co-occurrence, language and action are usually studied and conceived as separate domains within cognitive neuroscience. Consequently, very little is known about the neural circuitry underlying the integration of meaning from simultaneously perceived speech and action. Nevertheless, recent findings on the neurocognition of language semantics on the one hand (e.g. Pulvermuller 2005), and human action observation systems on the other (Decety et al. 1997; Rizzolatti and Arbib 1998; Rizzolatti et al. 2001; Rizzolatti and Craighero 2004; Molnar-Szakacs et al. 2005), suggest that the two systems recruit partly overlapping neural networks. In this study, we investigate the commonalities between language comprehension and action observation directly by presenting action and language related stimuli simultaneously. To do so, we focus on one particular form of action that often co-occurs with language, namely co-speech gestures.

When someone talks to us, we not only hear speech but also see the speaker's hand, mouth and body movements. In conversational settings, the brain therefore continuously integrates several streams of language and action related information that contribute to the listener's understanding of a speaker's message. Among those sources of information, co-speech gestures constitute a particular form of action. That is, they have communicative content and are naturally produced together with speech, contrary to, for instance, goal-directed object manipulations. As such, they are a prime example of actions that are recruited in the context of another domain of cognition (i.e. language). The present functional magnetic resonance imaging (fMRI) study investigates the neural locus of the integration of speech and action semantics as they co-occur simultaneously.

Previous studies investigating multimodal integration during communication have mostly focused on the relationship between lip movements and speech (Calvert 2001). Although both gestures and lip movements are examples of the natural co-occurrence of auditory and visual information during communication, they are fundamentally different with respect to their relationship to the speech they accompany. Whereas speech sounds and lip movements match with respect to form properties of language, there is no form matching between gestures and speech (McNeill 1992). Consider for example an upward hand movement in a climbing manner when a speaker says: “He climbed up the ladder”. Here, the gesture depicts the event as a whole, describing manner (‘climb’) and direction (‘up’) simultaneously. In speech, however, the message unfolds over time, broken up into smaller meaningful segments (i.e. the individual words climb and up). Because of these form differences (McNeill 1992), the mapping of speech and gesture information must occur at a higher, semantic level. Nevertheless, despite the fact that gestures express information in a different representational format than speech, the two modalities are systematically related in jointly conveying the speaker’s overall meaning (Clark 1996; Goldin Meadow 2003; Kendon 2004; Kita and Özyürek 2003; McNeill 1992; McNeill 2000). Thus it has been claimed that speech and gesture are part of the same system of communication (Bernardis and Gentilucci 2006; Kendon 2004; McNeill 1992).

The systematic relationship between speech and gestures exists at three levels. First, there is semantic overlap between the representation in gestures and the meaning expressed in the concurrent speech, as in the ‘climb up’ example above (e.g. Kita and Özyürek 2003; McNeill 1992). That is, speech and gesture usually convey similar or related information. Second, speech and gesture are temporally aligned to each other. A gesture phrase has three phases: the preparation, the stroke (semantically the most meaningful part of the gesture), and the retraction or hold (McNeill 1992). Studies have

shown that the onset of the gesture (i.e., preparation) usually precedes the onset of the relevant speech segment by less than a second (Butterworth and Shovelton 1978; Morrel Samuels and Krauss 1992). More importantly, in most speech-gesture pairs the stroke coincides with the relevant speech segment (McNeill 1992). Finally, it has been shown that the spontaneous use of gestures has a similar function as speech, namely to communicate the intended message to the addressee (e.g. Kendon 2004; Melinger and Levelt 2004; Özyürek 2002).

Furthermore, behavioural studies on speech and gesture comprehension have shown that listeners / viewers integrate information from gesture into their semantic interpretation of the speech input (Thompson and Massaro 1986, 1994; Beattie and Shovelton 1999; Kelly et al. 1999). Listeners / viewers pick up information coming from gestures in naturally occurring situations when information is expressed only in gesture but not in the concurrent speech (Church and Goldin-Meadow 1986; Goldin Meadow et al. 1999; Goldin Meadow and Momeni Sandhofer 1999; Singer and Goldin Meadow 2005) and even in cases when gestures contradict the information simultaneously conveyed in speech (McNeill et al. 1994).

Recently, few studies that have investigated brain responses with electrophysiological recordings (ERPs) during speech and gesture comprehension show that gestures evoke semantic processing. Kelly et al. (2004) found that ERPs to spoken words (targets) are modulated when the words are preceded by gestures (primes) containing information about the size and shape of objects that the target words referred to. Compared to words that matched the gesture primes, mismatching words evoked an early P1 / N2 effect, followed by an N400 effect. On the basis of these findings Kelly et al. (2004) claimed that the gesture primes influenced word comprehension, first at the level of 'sensory or phonological' processing and later at the level of semantic processing. In a study by Wu and Coulson (2005), it was found that incongruous gestures shown without speech and following cartoon

images elicited a negative-going ERP effect around 450 ms compared to congruous gestures. In addition, it was observed that incongruous words following the cartoon-gesture pairs elicited an N400 effect. Holle and Gunter (2007) presented spoken sentences in which an ambiguous word was combined with a pantomimic gesture that cued one meaning of the ambiguous word. An N400 effect was found to a word later in the sentence when that word was incongruous to the meaning of the ambiguous word cued by the gesture. The authors concluded that a gesture cue can disambiguate the meaning of an ambiguous word.

With respect to neuroimaging, many studies have investigated types of actions other than co-speech gestures. These include the observation of pantomimes (Decety et al. 1997; Gallagher and Frith 2004), of hand emblems (Nakamura et al. 2004), of simple finger movements (Iacoboni et al. 1999; Koski et al. 2002; Molnar-Szakacs et al. 2005), and of actions towards objects (Hari et al. 1998; Nishitani and Hari 2000; Buccino et al. 2001; Grezes et al. 2003; Hamzei et al. 2003). Crucially, in all these studies actions were presented in isolation. Different neural responses in areas involved in action observation have been reported in different task settings (e.g. 'passive observation' versus 'observe to imitate'). However, it is unknown to what extent these areas can be modulated by a language context. Nevertheless, a direct link between the language and action domains has been proposed, mainly inspired by neural findings in the monkey. When a monkey observes an action, neurons in areas that are thought to be homologous to human language areas are activated (Rizzolatti and Arbib 1998; Arbib 2005). Since classical language areas (Broca's area) are also found activated in human action observation, some have speculated about gestural communication as an immediate precursor of language in evolution (Arbib 2005; Nishitani et al. 2005; Rizzolatti and Arbib 1998, cf Aboitiz and Garcia 1997; Aboitiz et al. 2006).

With respect to language, a number of fMRI studies on language processing beyond the single word level are available. The language

studies that have examined the neural networks underlying the semantic integration of word meaning into a representation of the preceding part of the utterance mostly used a mismatch paradigm. In this paradigm the semantic integration load of a word's meaning in relation to the preceding speech context is manipulated. fMRI studies using this approach found that language-internal semantic violations result in stronger activation in left superior temporal and left inferior frontal areas compared to a matching (i.e. semantically correct) control condition (Ni et al. 2000; Kuperberg et al. 2003).

Finally, there have been recent neuroimaging studies investigating sign language comprehension. Even though sign languages also use actions for communicative expressions like co-speech gestures, they differ in important ways from co-speech gestures, based on the fact that signs are lexicalized and produced in hierarchic combinations (Goldin Meadow 2003; McNeill 1992). A few fMRI studies investigated sentence comprehension in deaf signers (Neville et al. 1998; MacSweeney et al. 2002a; MacSweeney et al. 2002b; Newman et al. 2002; MacSweeney et al. 2004; MacSweeney et al. 2006). These have shown that processing sentences in sign language activates a network of inferior frontal and temporal areas, which strongly overlaps with areas involved in sentence comprehension in hearing non-signing individuals (see Corina and Knapp 2006; Emmorey 2006 for review). However, despite using the same visuo-spatial domain of expression, it is unknown and unclear whether co-speech gestures will activate the same areas.

In order to bridge the gap between these separate lines of research in the action and language domains, we investigated if similar neural systems are involved when semantic information conveyed through action or language needs to be integrated into the preceding context. We addressed this question by investigating which brain regions are responsive to variations of the semantic relationship between a gesture and / or a spoken word and the preceding part of a spoken sentence. In

our study we presented participants with spoken sentences in which we manipulated the semantic ‘fit’ of a verb (language) and / or a gesture (action) to the preceding sentence context (Table 3.1, Fig. 3.1). As found in previous language studies (Hagoort and Brown 1994; Kutas and Hillyard 1980; Kutas and Hillyard 1984), semantic integration load was expected to vary with this manipulation, which is commonly employed in neuroimaging studies of language (e.g. Kuperberg et al. 2000; Ni et al. 2000; Friederici et al. 2003; Kuperberg et al. 2003; Hagoort et al. 2004; Ruschemeyer et al. 2006). In this way, regions specific for speech and gesture processing, as well as areas common to the integration of both information types into the prior sentence context could be identified. If integrating semantic information from both gesture and language into a broader sentence context activates the same areas, this would be direct evidence that the two systems recruit overlapping neural networks.

We call the critical verb or gesture which is semantically less fitting the previous sentence context ‘mismatch’. It is important to note that the term mismatch is used here in a different sense than in other studies in the speech and gesture literature (e.g. Church and Goldin Meadow 1986), where it is called a mismatch when gesture conveys additional - not incongruent - information compared to speech.

The manipulation of the semantic fit in our materials resulted in four conditions (see Table 3.1, Fig. 3.1): Correct condition, Language mismatch condition, Gesture mismatch condition, Double mismatch condition. In the Language mismatch the critical verb was harder to fit semantically to the preceding context while the co-occurring gesture matched the sentence context. In the Gesture mismatch condition the gesture was harder to integrate to previous context, while the critical verb matched the spoken sentence context. In the Double mismatch condition both the gesture and the word were difficult to integrate to previous sentence context. Note that in the Language and Gesture mismatch conditions the critical verb and the overlapping gesture were

locally incompatible (e.g., Speech: Write; Gesture: Hit, and vice-versa), while in the Double mismatch condition they were locally consistent (e.g., both Hit). Even though our study mainly targeted the effects of global sentence-level integration, this extra manipulation allowed us to check if our findings could be attributed to locally incompatible information of speech and gesture instead of to the context effects that we intended to study. The Double mismatch condition should elicit similar effects as the Language and Gesture mismatch conditions, if what we are testing is really a global, sentence-level effect. Note that in our materials an increase of semantic integration load does not necessarily involve a strict semantic violation. That is, the critical word always fits the preceding sentence context less well in the mismatch conditions compared to the correct condition, but is often not impossible as a continuation of the sentence. The condition label ‘mismatch’ thus refers to cases where the continuation is pragmatically less plausible than in the correct condition, but not necessarily impossible.

In particular, in this study we test the following specific hypotheses. The first concerns theories about the relation between speech and gesture systems. If speech and gesture are part of the same system of communication or interact at a high level of cognitive processing as is claimed on the basis of behavioural findings (Goldin Meadow 2003; Kendon 2004; Kita and Özyürek 2003; McNeill 1992; McNeill 2000), we expect the Gesture and the Language mismatch conditions to activate overlapping areas. If however co-speech gestures are considered not to have a communicative function (see Krauss et al. 1991), the mismatching gesture will not elicit similar effects as a mismatch in the language domain. Furthermore, neural overlap would provide further evidence for the claim that in the human brain there is a strong link between action and language systems (Nishitani et al. 2005; Rizzolatti and Arbib 1998).

Our second prediction concerns the fact that we expect the overlapping area of activation of Language and Gesture mismatch to

include Broca's area and adjacent cortex. This is in relation to a recent proposal (Hagoort 2003a; Hagoort et al. 2004; Hagoort 2005b), in which Broca's area and adjacent cortex (including Brodmann Area, BA 47, 45, 44 and the ventral part of BA 6) in the left hemisphere serves as a unification space for language, with a focus in BA 45 / 47 for the unification of semantic information. During unification, lexical information retrieved from memory (i.e. from the mental lexicon) is integrated into a unified representation of a multi-word utterance, such as a sentence. It is still an open question as to whether this unification space is specific for language or whether it integrates information across different domains of cognition. If Broca's area and adjacent cortex acts as the general (not domain-specific) unification space for language and action, we predict left inferior frontal cortex to be activated stronger with higher semantic integration load of speech and gesture information. Specifically, based upon previous research we predict BA 45 and 47 to show increased activation with an increase in semantic integration load (Bookheimer 2002; Hagoort 2005b).

Third, we investigate whether and how regions of the human action recognition network are modulated by a language context. We focus on the neural processing of hand actions in premotor cortex (BA 6) and parietal cortex. Previous work has shown that part of the motor system 'resonates' in a mirror-like fashion in response to the observation of actions (Nishitani et al. 2005; Rizzolatti et al. 2001). That is, the observation of an action triggers similar neural activity as executing an action. This 'neural simulation' of actions may underlie the understanding of actions performed by others (Jeannerod 2001; Rizzolatti et al. 2001; Nishitani et al. 2005). Studies in which actions were presented in isolation found modulations of the premotor cortex depending on the task, i.e. whether participants observed actions with the intention to imitate versus passive observation of actions (Grezes et al. 1999; Molnar-Szakacs et al. 2005). Parietal regions are reported to be specifically modulated by object-related versus non object-related

actions (Buccino et al. 2001), and by biologically impossible versus possible actions (Costantini et al. 2005). In this study, we seek to answer whether motor related areas, besides being modulated by task setting and type of actions, can also be influenced by a language context. We hypothesize that part of the action recognition system will be more strongly activated when the semantic content of an action cannot be easily integrated into a broader context, that is, in the gesture mismatch condition. This would provide evidence that the action recognition system not only automatically codes features of observed actions, but that it is also influenced by a previous semantic context provided by the language system.

### **Materials and Methods**

*Participants* Sixteen healthy volunteers (N=16; 8 female; mean age=24.1 years, range: 18-33) with normal or corrected to normal vision and normal hearing participated in the study. All participants were right-handed (Oldfield 1971) and had Dutch as their native language. None of the participants had any known neurological impairment. Participants gave written informed consent in accordance with the declaration of Helsinki. The participants were paid for participation. The study was approved by the local ethics committee.

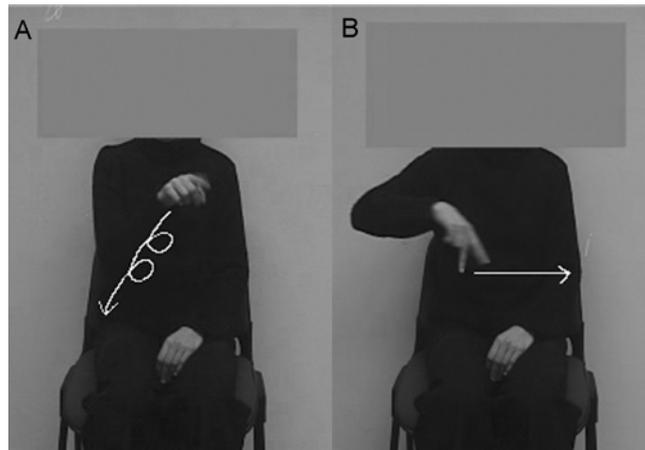
*Materials* Please note that the stimuli used in this experiment were the same as described in Chapter 2. The materials consisted of 640 items of spoken sentences that were accompanied by co-speech gestures. The sentences formed 160 sentence pairs. The members of a pair were identical up until the critical verb. Half of the sentences contained a critical verb that matched the preceding context. In the other half, the critical verb was semantically anomalous to the prior sentence context. Overall, twelve different critical verbs were used (see Appendix Chapter 2). The sentences had an average duration



(Fig. 3.1). Iconic gestures are a class of gestures that speakers spontaneously use as they talk about spatial and activity related aspects of events (e.g., using wiggling fingers moving horizontally while talking about someone walking (Kita and Özyürek 2003; McNeill 1992)). The iconic gestures used in this study were based on a larger database collected to investigate speakers' natural and spontaneous use of speech and gestures in narratives of spatial events (Kita and Özyürek 2003). For the purposes of this study, twelve of these gestures were selected and modelled by one native female Dutch speaker with the requirement that they resembled spontaneous gestures in this database. The purpose behind using modelled gestures instead of natural ones was to be able to keep external factors constant across different gestures. In order to match the speed and length of the gesture phases (e.g., the stroke) as closely as possible to naturally occurring iconic gestures, we asked our model to produce concurrent sentences originally used in the narrative database as she was performing the gestures. During editing the audio was removed from the movie. Movies were edited using Adobe Premier (version 6.0; Adobe Systems Inc., San Jose, USA; <http://www.adobe.com>). The preparation and the retraction phase of each gesture were removed, leaving the stroke. Previous research has shown that especially the stroke conveys the meaning of a gesture (Goldin Meadow 2003; Kendon 2004; Kita and Özyürek 2003; McNeill 1992; McNeill 2000). By isolating the gesture stroke phase, we eliminated differences among gestures that were due to the fact that for some gestures hand shape might reveal information before the stroke began, and that some gestures might have longer preparation time than others. The average length of the strokes was 767 ms (SD=284 ms). Finally, the face of the model was blocked to eliminate the contribution of information coming from the lips.

The gestures corresponded to the meaning of the critical verbs. They were combined with the sentences in such a way, that in half of the items the gesture matched the preceding sentence context, and in

the other half it ‘mismatched’ the preceding sentence context. This resulted in a total of 160 stimulus quartets (Table 3.1). In sum, there were four experimental conditions (Table 3.1): Correct condition (Gesture (G) +, Language (L) +); Language mismatch condition (G+L-); Gesture mismatch condition (G-L+); Double mismatch condition (G-L-).



**Fig. 3.1.** Two examples of the iconic gestures that were used. Depicted is one frame of the **A)** ‘Roll down’ gesture, and one frame of the **B)** ‘Walk across’ gesture. The line and arrow indicate the movement made by the hand.

The gesture movies and the sentence files were combined using the Adobe Premier (version 6.0) and After Effects software (version 5.5; Adobe Systems Inc., San Jose, USA, <http://www.adobe.com>). For each movie file, the onset of the gesture stroke was temporally aligned with the onset of the critical verb, since in 90% of natural speech-gesture pairs the stroke coincides with the relevant speech segment (McNeill 1992). For verbs with a separable prefix, the alignment point was not word-onset, but the body of the verb following the prefix. The latter was the case for 44 sentences. Additional still frames with the hand resting on the lap were added to the part of the sentence before the critical verb, and the last frame of the stroke was elongated until the end of the sentence.

Four different stimulus lists were created, to distribute the four versions of each item equally over the four lists. This was done in such a way that all four lists contained an equal number of items per condition. Each list was presented to a quarter of the participants. As a result, none of the participants were presented with more than one version of the stimulus items, i.e. every participant was presented with only one item from a quartet as in Table 3.1.

*Experimental Design and Procedure* Forty items per condition were presented, resulting in a total of 160 items per participant. The items were presented in an event-related design, in a pseudo-randomized order with the constraints that no more than two items of the same condition were presented after each other. The four stimulus lists were presented in normal or reversed order, resulting in eight stimulus lists that were evenly distributed across male and female participants.

Stimuli were presented using the Nijmegen Experiment Setup software (NESU, MPI for Psycholinguistics; <http://www.mpi.nl/world/tg/experiments/nesu.html>). The visual content of the movies was presented through an Eiki LC-X986 TFT-LCD projector outside the scanner room at a refresh rate of 60 Hz. Participants watched the screen via a non-magnetic mirror mounted to the head coil. The movies subtended 10 cm (height) x 11.8 cm (width) and were shown at a viewing distance of 80 cm. Speech was presented to the participants through non-magnetic headphones (Commander XG, Resonance Technology Inc.; <http://www.mrvideo.com>), which dampened scanner noise.

Participants were instructed to carefully listen to the sentences and watch the movies. They were told that they would receive questions about the items after the experiment. Before the beginning of a run, each participant received two practice runs consisting of 5 practice items each. These items were also used to adjust the volume level of the sentences. Therefore, the scanner was switched on during

the practice items and participants were asked to indicate whether the volume should go up or down. The volume level that suited each participant best was used in the following experimental run. The functional data acquired during the practice runs were not used in the analysis.

*fMRI data acquisition* MR imaging was performed on a Siemens Magnetom Trio scanner (Siemens Medical Systems, Erlangen, Germany) with 3 Tesla magnetic field strength. Functional data were acquired with echo planar whole brain images in 32 transversal slices (TR=2230 ms; TE=30 ms; flip angle=80°; slice thickness=4 mm; FoV=224 mm, voxel resolution=3.5 mm x 3.5 mm). Slices were positioned to cover the participant's whole brain. Intertrial interval was two or three scanner volumes (TRs) and the onset of each trial was synchronized to a scanner pulse. Sentence onset was effectively jittered by adding either 0, 500 or 1000 ms (mean=500 ms) to the trial onset (Josephs et al. 1997; Dale 1999; Miezin et al. 2000).

After the functional run, for each participant an anatomical scan was made using a high resolution T1 weighted 3D-MPRAGE sequence consisting of 192 sagittal slices (TR=2300 ms; TE=3.93 ms; FoV=256 mm; slice thickness=1 mm).

*Data analysis* Data were analyzed using Brainvoyager QX (Brain Innovation, Maastricht, The Netherlands; <http://www.brainvoyager.com>). The first five volumes of every functional run were discarded from the analysis to minimize T1-saturation effects. Preprocessing involved rigid body transformations of all volumes to the first volume to correct for small head movements, slice scan time correction, linear trend removal and high pass temporal filtering of 3 or fewer cycles per time course. The functional data of each run were co-registered to the anatomical data and were interpolated to a 1x1x1 mm voxel size. Subsequently, anatomical and

functional data were transformed into stereotaxic space as defined by Talairach and Tournoux (Talairach and Tournoux 1988). The functional data were spatially smoothed with a Gaussian filter kernel of 12 mm FWHM (Xiong et al. 2000).

*Regions of interest analyses* As described in the introduction, we had specific hypotheses for the anterior part of the left inferior frontal cortex (BA 45 / 47), the premotor cortex (BA 6) and the (inferior) parietal cortex. We, therefore, performed region of interest (ROI) analyses in these regions. A meta-analysis by Bookheimer (Bookheimer 2002) showed that semantic processing is centered around [x y z] [-42 25 4] (Talairach and Tournoux 1988), with a mean distance of the local maxima to this centre coordinate of 15 mm (Petersson et al. 2004). Accordingly, a spherical ROI around [x y z] [-42 25 4] (Talairach and Tournoux 1988), with a radius of 15 mm was created<sup>1</sup>. To avoid including air-tissue boundaries in our ROI, the inferior 3 mm of the sphere were not taken into the ROI.

The ROI for the premotor cortex (left and right BA 6) was defined on the basis of an observer independent cytoarchitectonic probability map (Eickhoff et al. 2005), by including voxels that had a probability of 50% or higher to fall within the borders of BA 6. This region was converted from anatomical MNI space into stereotaxic Talairach space (Talairach and Tournoux 1988) by applying a non-linear transformation (<http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml>).

For the parietal cortex an ROI was constructed by averaging the local maxima of studies in which passive action observation was contrasted to a low level baseline (Grezes et al. 1999; Iacoboni et al. 1999; Buccino et al. 2001; Hamzei et al. 2003; Buccino et al. 2004; Costantini et al. 2005). This average was [x y z] [-35 -43 49] (Talairach and Tournoux 1988) for the left hemisphere and [x y z] [41 -38 52] (Talairach and Tournoux 1988) for the right hemisphere, both in the

vicinity of the intraparietal sulcus. The mean distances of the local maxima to these centre coordinates were 16 mm (left hemisphere) and 9 mm (right hemisphere). Accordingly, two spherical regions of interest were created around these averaged maxima with a radius of 16 mm and 9 mm, respectively.

Statistical analysis was done in the context of the General Linear Model (GLM). A model with the four experimental conditions (Table 3.1), in which events were modelled as the duration of the whole sentence convolved with a canonical, two gamma hemodynamic response function (Friston et al. 1998) was tested in each participant's data separately. First, the average activation levels (beta weights) were estimated separately for each participant and condition in the a priori defined ROIs. Subsequently paired t-tests ( $df=15$ ) to test for significant differences between conditions were applied to the estimated activation levels. Tested contrasts were language mismatch versus correct condition (G+L- vs. G+L+), gesture mismatch versus correct condition (G-L+ vs. G+L+) and double mismatch versus correct condition (G-L- vs. G+L+).

*Whole brain analysis* In addition to testing condition effects in the ROIs, we also tested for the presence of other areas that were differentially activated by the experimental conditions. For this purpose, we performed a whole brain random effects analysis, with the four conditions convolved with a canonical, two gamma hemodynamic response function (Friston et al. 1998) as our model. Individual contrast maps were taken to a second level analysis in which for each voxel the mean value of a contrast was tested against zero using the student's t-distribution with  $df=15$  ( $n-1$ ). To control for the multiple comparison problem introduced by the massive univariate approach taken, a voxel-wise intensity threshold ( $p<0.003$ ) was combined with a cluster extend threshold of  $R>41$  contiguous  $3\times 3\times 3$  mm voxels, to

control for false positives at an alpha level of  $p < 0.05$  (Forman et al. 1995).

## **Results**

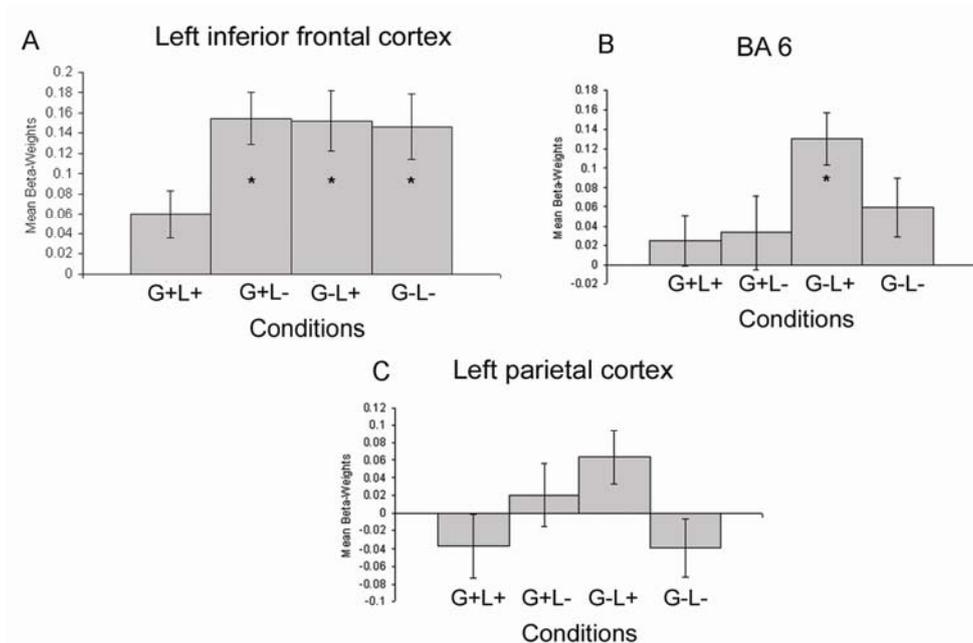
*Regions of interest analyses* First we explored if left inferior frontal cortex responds differently to action or language information that fits less easily into a sentence context than in the correct baseline condition. In the ROI in left inferior frontal cortex (BA 45 / 47) all tested contrasts revealed significant differences (Table 3.2, Fig. 3.2A): the language mismatch versus correct condition (G+L- vs. G+L+:  $t(15)=2.59$ ,  $p < 0.02$ ), the gesture mismatch versus correct condition (G-L+ vs. G+L+:  $t(15)=2.53$ ,  $p < 0.02$ ), and the double mismatch versus correct condition (G-L- vs. G+L+:  $t(15)=2.19$ ,  $p < 0.04$ )<sup>2</sup>.

Second, the response of the motor system known to be involved in action observation was tested. In the ROI in premotor cortex (left and right BA 6), no significant differences were present between the language mismatch and the correct condition (G+L- vs. G+L+:  $t(15)=0.20$ ,  $p < 0.84$ ), nor between the double mismatch and the correct condition (G-L- vs. G+L+:  $t(15)=0.79$ ,  $p < 0.44$ ). In contrast, the gesture mismatch differed significantly from the correct condition (G-L+ vs. G+L+:  $t(15)=2.64$ ,  $p < 0.02$ ; Table 3.2, Fig. 3.2B). In the left and right parietal ROIs, there was only a marginally significant effect in the left hemisphere for the contrast gesture mismatch versus correct condition (G-L+ vs. G+L+:  $t(15)=1.88$ ,  $p < 0.08$ ; Table 2.2, Fig. 3.2C).

*Whole brain analysis* Subsequently, a more exploratory analysis was performed over the whole brain by testing for areas differentially activated by the contrasts of interest. In this whole brain analysis, the comparison between the language mismatch condition and the correct condition resulted in significant activations in the left inferior frontal sulcus extending into the precentral sulcus, in the posterior part of the left superior temporal sulcus, and in the superior part of the left

Region	Coordinates			Contrast	T	df	p
	x	y	z				
Left Inferior Frontal Cortex (BA 45 / 47)	-42	25	9	Language mismatch vs. correct	<b>2.59</b>	15	<b>0.02</b>
				Gesture mismatch vs. correct	<b>2.53</b>	15	<b>0.02</b>
				Double mismatch vs. correct	<b>2.19</b>	15	<b>0.04</b>
Premotor cortex (left and right BA 6)	0	-11	57	Language mismatch vs. correct	0.20	15	0.84
				Gesture mismatch vs. correct	<b>2.64</b>	15	<b>0.02</b>
				Double mismatch vs. correct	0.79	15	0.44
Left Parietal	-35	-43	49	Language mismatch vs. correct	1.67	15	0.12
				Gesture mismatch vs. correct	1.88	15	0.08
				Double mismatch vs. correct	0.37	15	0.72
Right Parietal	41	-38	52	Language mismatch vs. correct	-1.10	15	0.29
				Gesture mismatch vs. correct	1.50	15	0.16
				Double mismatch vs. correct	-1.13	15	0.28

**Table 3.2.** Activations in regions of interest. The T values reflect differences between the averaged activation levels elicited by the various conditions. Regions were defined on the basis of previous functional imaging results (left inferior frontal cortex and left and right parietal regions) or on the basis of a cytoarchitectonic probability map (left and right BA 6). Centre coordinates are in stereotaxic space (Talairach and Tournoux 1988). Significant T and P values are in bold.



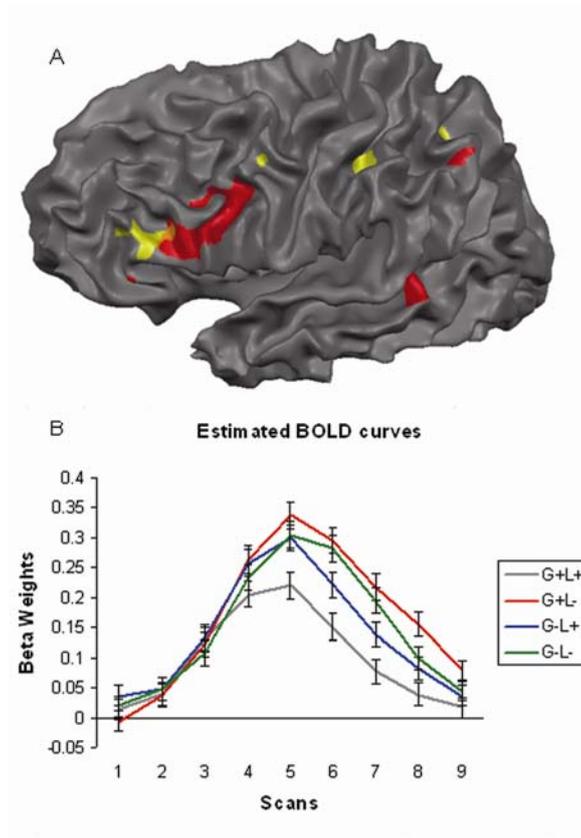
**Fig. 3.2.** Gesture and speech in a sentence context. Mean activation levels (beta weights) for the four experimental conditions in **A)** left inferior frontal cortex **B)** left and right BA 6, and **C)** left inferior parietal cortex. The activation levels are averaged over participants. An asterisk indicates a significant difference of the activation level of that condition compared to the correct condition (G+L+), at an alpha level of  $p < 0.05$ . See Table 3.2 for specific statistics. Error bars are standard error of the mean (s.e.m). G+L+: correct condition, G+L-: language mismatch, G-L+, gesture mismatch, G-L-: double mismatch.

intraparietal sulcus (Table 3.3, Fig. 3.3A). For the gesture mismatch condition compared to the correct condition, we found significant activations in the left inferior frontal sulcus and in two areas in left intraparietal sulcus, one anterior and one posterior (Table 3.3, Fig. 3.3A).

The double mismatch versus correct condition contrast showed an area in the left inferior frontal cortex and an area in the precentral cortex to be differentially activated (Table 3.3; Fig. 3.4). The fact that precentral cortex is also activated in this comparison, may seem to be in contrast to the findings from the region of interest analysis in BA 6, reported above. In the ROI analysis, only the gesture mismatch condition led to significantly increased activation compared to the correct condition. The small region found activated to the double mismatch condition in the whole brain analysis is part of the ROI used to test differential activation of BA6. The fact that no difference is found in the ROI analysis of BA 6 between double mismatch and correct condition is probably because the extent of the reaction of premotor cortex is more limited to the double mismatch condition than to the gesture mismatch condition. In other words, modulation of premotor cortex seems to be much more robust in the gesture mismatch condition than in the double mismatch condition.

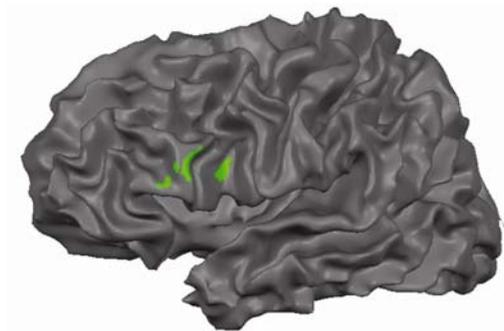
In none of the contrasts significant differences in the opposite direction (i.e. correct>mismatch) were found.

Next, a conjunction analysis, performed by a test for independently significant effects as in a logical AND (Nichols et al. 2005), was performed to test for regions involved in both the language mismatch as well as the gesture mismatch conditions. In this way, common (overlapping) activations for both processes could be defined. This analysis tested for areas activated significantly stronger in both the language mismatch versus correct condition contrast and the



**Fig. 3.3.** (For colour version see Appendix, p. 266). Gesture and speech in a sentence context. **A)** Significant activations in the whole brain analysis for the language mismatch versus correct (red) and the gesture mismatch versus correct (yellow) comparisons. Note the overlap in inferior frontal cortex (BA 45, [x y z] [-46 23 25]). Maps are thresholded at  $t(15) > 3.5$ ,  $p < 0.05$  (corr.). No activations were found in the right hemisphere. **B)** Blood Oxygen Level Dependent (BOLD) curves from the activated regions in left inferior frontal cortex (centre coordinates [x y z] [-43 11 26]). The curves were created by estimating each time point after modelling an event by nine subsequent stick functions (Dale 1999; Miezin et al. 2000). This region is also activated in the correct condition (grey line), but more so in reaction to a semantic mismatch (red, blue and green lines). The time scale on the x-axis is in repetition times ('scans'), one scan is 2230 ms. G+L+: correct condition, G+L-: language mismatch, G-L+, gesture mismatch, G-L-: double mismatch.

gesture mismatch versus correct condition contrast (G+L- vs. G+L+  $\cap$  G-L+ vs. G+L+). One area in the left inferior frontal cortex was significantly activated in this conjunction (overlap in Fig. 3.3A).



**Fig. 3.4.** (For colour version see Appendix, p. 266). Significant activations in the whole brain random effects analysis in the double mismatch versus correct condition contrast. The inferior frontal area and the precentral area in purple were more strongly activated by the double mismatch (G-L-) condition than by the correct condition (G+L+). Map is thresholded at  $t(15) > 3.5$ ,  $p < 0.05$  (corrected) and projected onto the cortical sheet of one of the participants. No activations were found in the right hemisphere.

Comparison to a cytoarchitectonic probability map of BA 45 (Amunts et al. 1999; Eickhoff et al. 2005), showed that 82% of the voxels in this region was part of BA 45 with its centre of gravity ( $[x\ y\ z]$  [-46 23 25] (Talairach and Tournoux 1988) having a probability of 60% to be part of BA 45. When the statistical threshold was lowered for informal visual inspection, the area of overlap became much bigger. This confirms that the activations in left inferior frontal cortex displayed in Figure 3.3A are not two distinct areas slightly overlapping, but are part of the same region being activated in both conditions.

To test for activations specifically obtained for the local mismatch in the Gesture and Language mismatch conditions, a conjunction analysis (Nichols et al. 2005) was performed, testing for regions activated in both the G+L- vs. G-L- and the G-L+ vs. G-L- contrasts. This analysis was preferred over an analysis contrasting the two conditions with a local mismatch versus the two conditions with a local match, since the latter analysis would contain a confound between overall sentence anomaly (absent for the correct condition) and local match (present for the correct and double mismatch conditions). To avoid such a confound, we contrasted the local mismatch conditions (Language mismatch, G+L-; Gesture mismatch, G-L+) against the double mismatch condition (G+L- vs. G-L-  $\cap$  G-L+ vs. G-L-). In this way all conditions were semantically anomalous, so that the only remaining difference was a local mismatch (G+L-; G-L+) versus a local match (G-L-). One area in the right fusiform gyrus (Table 3.3) was found to be significantly more activated in the conditions with a local mismatch (G+L- and G-L+) compared to the condition with a local match (G-L-)<sup>3</sup>.

At the end of the scanning session, participants were extensively debriefed. All participants were able to describe the manipulations in the materials and could provide examples of specific trials. All participants were aware of the cases in which language and / or gestures did not fit well into the preceding sentence context. Moreover, they were aware of both language and gesture ‘mismatches’ to an approximately equal extent, indicating that both language and gesture information had been in the focus of attention.

Contrast	Coordinates			Region	T	Number of voxels
	x	y	z			
Language mismatch vs. correct	-43	12	24	L Inferior Frontal Sulcus	5.19	7326
	-33	-65	35	L Intraparietal Sulcus (Posterior)	4.33	1591
	-52	-50	4	LSuperior Temporal Sulcus	4.83	2374
Gesture mismatch vs. correct	-46	29	23	L Inferior Frontal Gyrus / Sulcus	4.48	1651
	-32	-46	31	L Intraparietal Sulcus (Anterior)	5.93	2531
	-19	-63	32	L Intraparietal Sulcus (Posterior)	5.37	2290
Double mismatch vs. correct	-55	17	24	L Inferior Frontal Sulcus	5.96	1586
	-52	-6	49	L precentral sulcus	4.36	1145
Conjunction of G+L- vs. G-L- and G-L+ vs. G-L-	38	-54	-13	R Fusiform Gyrus	4.82	1946

**Table 3.3.** Activations in the whole brain analysis. Regions that were significantly activated in the whole brain random effects group analysis ( $t(15) > 3.5$ ,  $p < 0.05$ , corrected). Displayed are the contrasts, the centre coordinates in stereotaxic space (Talairach and Tournoux 1988), a description of the region, the T value of the maximally activated voxel and the number of significant voxels (1x1x1 mm voxel size).

## **Discussion**

The main goal of this study was to investigate the neural integration of semantic information conveyed through language (speech) and action (gestures) within a sentence context. The results show that both action and language recruit overlapping parts of left inferior frontal cortex, specifically BA 45. That is, this region is modulated by an increase in the semantic load of simultaneously presented information from the speech and action domains. Additionally, premotor cortex is modulated by the semantic processing of actions within a language context. These results are in line with accounts hypothesizing a link between language and action systems. Furthermore, the fact that we found overlapping areas provides neural evidence for claims that speech and gesture are closely linked in language comprehension (Goldin Meadow 2003; Kendon 2004; McNeill 1992; McNeill 2000).

The involvement of left inferior frontal cortex in integrating semantic information from both the action and language domains is consistent with a theory of language comprehension in which the left inferior frontal cortex serves as the general (i.e. not domain-specific) unification site for language comprehension (Hagoort 2003a, 2005b). During unification, current information is integrated into an unfolding representation of multi-word utterances. In the case of a semantic mismatch, integration of the mismatching information is harder, resulting in an activation increase. The Blood Oxygen Level Dependent (BOLD) curves of the left inferior frontal cortex (Fig. 3.3B) show that this increased activation does not reflect a reaction to a semantic mismatch as such, but that the area is also activated in the processing of correct sentences. This lends credibility to the idea that Broca's area and adjacent cortex, more in particular BA 45 and 47, is the semantic unification site for language comprehension. Its activation in studies using the mismatch paradigm is no artefact of the materials used, but truly reflects increased semantic load. Moreover, most of our items in the 'mismatching' conditions were not straightforward semantic

violations, but contained only semantically less expected verbs and / or gestures given the semantics of the preceding sentence context. This is in line with a recent study (Rodd et al. 2005) in which Materials contained no semantic violations whatsoever. Yet, sentences containing semantically ambiguous words activated left inferior frontal cortex stronger than sentences containing non-ambiguous words. Our results are also in accordance with the established finding in the ERP literature that both straightforward semantic anomalies as well as subtle manipulations in semantic integration processes lead to the same ERP effect, namely the N400 effect (Hagoort and Brown 1994; Kutas and Hillyard 1980; Kutas and Hillyard 1984; Kutas and Van Petten 1994). Moreover, in an ERP study conducted in our lab (Özyürek et al. 2007) with the same materials as used in this study (Chapter 2) it was found that ERPs time locked to the critical verbs and gestures elicited an N400 effect in all three mismatch conditions. Furthermore, the latency and amplitude of these N400 effects were similar. This is independent evidence for the claim that the effects reported here reflect semantic unification as indexed by the N400 effects. It furthermore shows that processing of gestures versus critical verbs in relation to previous context do not involve different processing strategies, which is consistent with the debriefing reports of the participants in our study.

Based on the results for the language and gesture mismatch conditions, one could argue that the activation of the left inferior frontal cortex is not due to the increased semantic integration load with respect to the preceding sentence context, but instead to the local mismatch between the simultaneously occurring critical verb and the gesture (e.g., the verb knock and the gesture roll down). However, this interpretation is challenged by the fact that we observed the same area activated in the double mismatch condition (Fig. 3.4, Table 3.3), in which verb and gesture match with respect to each other and the mismatch is solely in relation to the preceding context.

These findings are in one important aspect different from what has been reported for the audiovisual integration of speech and lip movements, where mismatching information has been found to result in a reduced activation compared to matching information (Calvert 2001). This difference might be due to the fact that for the integration of speech and lip movements, in the matching condition the auditory and visual input converge on a common form representation in memory (e.g., a particular syllable), that as a result gets more strongly activated than in a mismatching condition. In our study, the gesture and speech signal had to be integrated in a sentence-level semantic representation that is not available in memory, but has to be constructed on-line. This semantic integration process is more strongly taxed when the integration load increases, leading to higher activation levels in the mismatch conditions.

Overall, our findings are compatible with other studies on language-related integration processes. It was recently shown that an area in left inferior frontal cortex partly overlapping with the region reported here, is involved in the integration of both semantic and world knowledge information during reading (Hagoort et al. 2004). The present study replicates the role of this area in integrating verbal information into a prior sentence context, but this time with spoken input. Within the domain of language, this area seems to operate independent of input modality (reading vs. speech). Importantly, the data presented here convincingly demonstrate that unification in the left inferior frontal cortex during language comprehension is not domain-specific. The integration of semantic information conveyed through the action domain also recruits this area. A dorsal to ventral parcellation of inferior frontal cortex into distinct subregions performing different core functions within the language domain, such as phonological, semantic and syntactic processing, has been proposed (Poldrack et al. 1999; Bookheimer 2002; Vigneau et al. 2006). Note that the location of the activation to both gesture and language conditions in

this study is in line with the location of the semantic component within left inferior frontal cortex.

Finally, our results also show overlap with sign language comprehension with regard to the involvement of left inferior frontal and temporal cortices (MacSweeney et al. 2006; MacSweeney et al. 2004; MacSweeney et al. 2002; Neville et al. 1998), despite the differences in linguistic properties of signs compared to gestures.

This study also sheds light on the semantic modulation of the action recognition system. The fact that we found a context dependent modulation of premotor cortex (BA 6) has important implications for the role of this area in action observation. A large number of studies show activation of premotor areas by action observation (Buccino et al. 2001; Costantini et al. 2005; Grezes et al. 2003; Hari et al. 1998; Jeannerod 2001; Nishitani and Hari 2000; Rizzolatti et al. 2001). This has been interpreted as evidence for the existence of an action recognition system in humans, comparable to the mirror neuron system in monkeys, in which similar neural activations exist during action observation and action execution (Jeannerod 2001; Nishitani et al. 2005; Rizzolatti et al. 2001). Premotor activation is also found when stimuli are meaningless actions (Fadiga et al. 1995; Decety et al. 1997), point light displays (Saygin et al. 2004), biologically impossible actions (Costantini et al. 2005) and in motor imagery (Schubotz and von Cramon 2004; de Lange et al. 2005). Together, these findings suggest that the activation of the premotor cortex (BA 6) is automatic and occurs to the observation of any type of action. Here we show that although possibly automatic, activation of premotor cortex is influenced by semantic information from speech. That is, premotor cortex is directly sensitive to the semantic context in which an action occurs, possibly through top-down modulations of motor representations by higher order cortical areas. Future work is needed to investigate the specific neural dynamics of this interaction.

One interesting finding with regard to the modulation of premotor cortex is that this area was modulated in a more robust way in the gesture mismatch than in the double mismatch condition, even though gesture information was harder to integrate to previous context in both conditions. In the ROI analysis, the premotor cortex was found to be activated only in the gesture mismatch condition. However, in the whole brain analysis, a small region of precentral gyrus was found activated to the double mismatch condition as well. Apparently, in the double mismatch condition premotor areas are activated to some extent, but much less robustly than for the gesture mismatch condition. One could speculate that the more robust activation in the gesture mismatch is due to the fact that in this condition there is an additional local mismatch between the co-occurring verb and gesture, which is not present in the double mismatch condition.

Given their commonly observed role in action observation, it is tempting to interpret the activations of intraparietal regions in the gesture mismatch condition in the whole brain analysis in a similar vein as the findings in BA6. However, the activation of intraparietal areas was not specific to the gesture condition but was also found in reaction to the language mismatch condition, albeit in a slightly different location. Therefore, we interpret these activations in another, more parsimonious way. That is, both these conditions might lead to increased spatial attention, a process in which (intra)parietal regions are known to also be involved (Corbetta and Shulman 2002). The non-specific nature of these activations (not in response to one particular condition) strengthens this explanation.

Finally, a left superior temporal activation was seen in the language mismatch condition. This finding is compatible with previous studies of semantic aspects of sentence processing (Kuperberg et al. 2003; Ni et al. 2000). Presumably this activation reflects the interaction between context and the retrieval of lexical-semantic information.

In summary, our results reveal two important aspects of the relations between language and action systems. One is that high-level neural integration of semantic information into a context is not domain specific and takes place in Broca's area. When understanding a sentence, the brain does not restrict itself to language information alone, but also integrates semantic action information conveyed through co-speech gestures into the preceding message context. Both action and language semantics constrain the interpretation domain simultaneously, and by recruiting the neural contribution of left inferior frontal cortex. This opens the interesting possibility that neural processing in language comprehension involves the incorporation of information in a single unification space coming from a broader range of cognitive domains than thought so far.

Different proposals on the role of inferior frontal cortex have been put forward in recent years, such as selection among competing alternatives (Thompson-Schill et al. 1997), controlled semantic retrieval (Wagner et al. 2001) or both (Badre et al. 2005). Such views are not inconsistent with our account, since selection is a necessary aspect of unification, as we have argued elsewhere and as is specified in explicit computational models of unification (Vosse and Kempen 2000; Hagoort 2005a). How the role of left inferior frontal cortex is best characterized if one wants to cover all findings available in the literature, is an open question. One possibility is that seemingly conflicting findings regarding inferior frontal cortex functioning can be subsumed under the heading of one underlying common process, such as the 'regulation of mental activity' (Thompson-Schill et al. 2005). However from the perspective of the cortex as a dynamically changing system of large-scale distributed functional networks (Mesulam 1990, 1998; Fuster 2003), it is conceivable that higher order cortices do not perform one function, but play different roles in different networks, depending upon the nature of input and task. Along these lines, we do not claim that the sole function of this cortical area is to be involved in

semantic unification. Whether different functional accounts of inferior frontal cortex can be grouped under one heading (Thompson-Schill et al. 2005) or that qualitatively different functions can co-exist within one part of cortex is an important question for future research. In any case, accounts of inferior frontal cortex as playing a role in selection (Badre et al. 2005; Thompson-Schill et al. 1997) are compatible with our interpretation that this region contributes to unification processes.

The second important finding of our study is that contextual information from the language domain can influence parts of the motor system. This adds to the growing insight that cognitive modulation of areas at a lower level in the cortical hierarchy appears to be an important principle in the neural architecture of human cognition (see also de Araujo et al. 2005).

In conclusion, we have shown that a classical language area, Broca's area, can be modulated by action processing as well as that a classical action area, premotor cortex, can be modulated by the language context in which actions are embedded. These findings provide support for the claims that in real life speech and action are often tightly interconnected (Goldin Meadow 2003; Kendon 2004; Kita and Özyürek 2003; McNeill 1992; McNeill 2000), and that there are close links between the action and language systems (Nishitani et al. 2005; Rizzolatti and Arbib 1998). Many aspects of their neural interplay remain to be unravelled, but this study provides a first insight in the neural integration of language and action information.

## Notes

1) In addition, we defined an ROI for left BA 45 based on an observer-independent cytoarchitectonic probability map (Amunts et al. 1999; Eickhoff et al. 2005). We included the voxels that had a probability of 50% or higher to fall within the borders of left BA 45. This region was converted from anatomical MNI space into stereotaxic Talairach space (Talairach and Tournoux 1988) by applying a non-linear

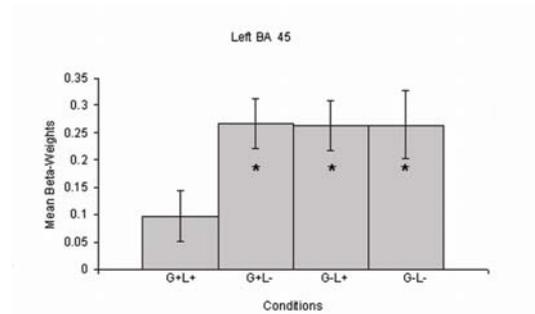
transformation (<http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml>). A similar probability map for left BA 47 does not yet exist. By including this anatomically defined ROI for BA 45 we have an additional check on the validity of the ROI analysis that is solely based upon functional data from previous studies.

2) The effects for the anatomically defined left BA 45 (see note 1) had a highly similar pattern. Again, significant differences were obtained between the language mismatch and correct condition (G+L- vs. G+L+:  $t(15)=3.24$ ,  $p<0.005$ ), the gesture mismatch and correct condition (G-L+ vs. G+L+:  $t(15)=3.65$ ,  $p<0.002$ ), and the double mismatch and correct condition (G-L- vs. G+L+:  $t(15)=3.11$ ,  $p<0.007$ ). See Supplementary Figure S3.1.

3) However, at the more liberal voxel-threshold, intraparietal areas were activated bilaterally, as is already suggested by Figure 3.2C.

### **Acknowledgements**

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO), 051.02.040. We thank Ivan Toni, Michael Coles, Sotaro Kita, Floris de Lange and Marieke Schölvinc for their comments on an earlier version of our paper and Paul Gaalman for assistance during the scanning sessions. Petra van Alphen is acknowledged for lending her voice to the sentence materials.



**Fig. S3.1.** Activation levels per condition in the anatomically defined ROI in left BA45. This ROI was used as an additional check on the outcome of the ROI in left inferior frontal cortex reported in the paper, which was constructed based upon functional studies alone. The pattern of results is highly similar to the results of the functionally defined ROI (Fig. 3.2A).

## **Chapter 4** Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context\*

### **Abstract**

Understanding language always occurs within a situational context. Therefore, understanding language often implies combining streams of information from different domains and modalities. One such combination is that of spoken language and visual information which are perceived together in a variety of ways during everyday communication. Here we investigate whether and how words and pictures differ in neural correlates when they are integrated into a previously built-up sentence context. This is assessed in two experiments looking at the time course (measuring Event-Related Potentials, ERPs) and the locus (using functional Magnetic Resonance Imaging, fMRI) of this integration process. We manipulated the ease of semantic integration of word and / or picture to a previous sentence context to increase the semantic load of processing. In the ERP study, an increased semantic load led to an N400 effect similar for pictures and words in terms of latency and amplitude. In the fMRI study we found overlapping activations to both picture and word integration in left inferior frontal cortex. Specific activations for the integration of a word were observed in left superior temporal cortex. We conclude that despite obvious differences in representational format, semantic information coming from pictures and words is integrated into a sentence context in similar ways in the brain. This study adds to the growing insight that the language system incorporates (semantic) information coming from linguistic and extra-linguistic domains with the same neural time course and by recruitment of overlapping brain areas.

---

\*This chapter is a slightly modified version of: Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, 20:7.

## **Introduction**

Understanding language always occurs within a situational context, such as knowledge about the person you are talking to or the location one is in (Clark 1996). Therefore, understanding language often implies combining streams of information from different modalities. For instance, consider a biology teacher describing the properties of an animal while at the same time showing a slide with a picture of the animal. In such a case, auditory and visual information do not bear a direct physical connection. That is, the string of sounds describing a concept never directly maps onto the visual appearance of this concept. This raises the question how verbal (linguistic) and visual (extra-linguistic) semantic information combine to form a coherent interpretation of a speaker's message in relation to the overall context. The current study investigates this question by assessing the neural integration of semantic information from words and pictures embedded within a spoken sentence context. Our main aim was to investigate if, despite differences in representational format, semantic information from pictures and words is integrated into an overall representation of an utterance in the same way as unimodal semantic information from a word is. Within the broader context of language comprehension, we wanted to investigate differences and commonalities between linguistic and extra-linguistic information processing during sentence comprehension.

We investigated both the neural time course (measuring Event-Related Potentials, ERPs) and the locus (using functional Magnetic Resonance Imaging, fMRI) of this multimodal integration process. Combining temporal and spatial neural information in this way provides a more complete picture of the integration process under study.

### *Lexical semantic information in a sentence context*

The seminal ERP studies by Kutas and Hillyard (Kutas and Hillyard 1980, 1984) showed that words that are semantically anomalous in relation to the preceding sentence context lead to a more negative deflection in the ERP waveform than words that are semantically congruent. For example, the sentence-final word of the sentence ‘She spread her bread with socks’ leads to a negative deflection in the ERP waveform in comparison to the ERP for a congruous ending as in ‘She spread her bread with butter’. This N400 effect occurs between 250-500 ms after the anomalous word, and is usually maximal at central-posterior electrode sites. N400 effects are also observed when a critical word is a possible but unexpected continuation of a sentence, without being a semantic violation (Kutas and Hillyard 1984; Hagoort and Brown 1994). The N400 has become an established ERP component which is thought to reflect the ease of integration of a word into a preceding context (see Kutas and Van Petten 1994; Brown et al. 2000 for reviews).

fMRI studies of sentences with semantic anomalies comparable to those used in N400 studies, have reported increased activation in left inferior frontal and / or left temporal areas (Kuperberg et al. 2000; Ni et al. 2000; Baumgaertner et al. 2002; Friederici et al. 2003; Kuperberg et al. 2003; Hagoort et al. 2004; Willems et al. 2007). Based on these findings it has been claimed that these areas are involved in semantic integration since they respond to a higher load of integration elicited by the difficulty of semantic processing. Recent work has indeed shown that increased semantic ambiguity without semantic anomalies also leads to increased activations in left inferior frontal and left temporal regions (Rodd et al. 2005; Davis et al. 2007).

### *Extra-linguistic information in a sentence context*

The integration of extra-linguistic information into a preceding context has been explored in a variety of ways in ERP or fMRI studies (Van

Petten and Rheinfelder 1995; Hagoort et al. 2004; Koelsch et al. 2004). A complete review of how extra-linguistic information influences sentence comprehension is beyond the scope of this paper. We therefore restrict our focus on studies investigating the integration of visual information into a preceding (sentence) context.

A small number of studies have looked into the integration of picture information into a sentence context. In an ERP study, Ganis, Kutas and Sereno (1996) presented sentences that either ended with a word or a picture that could be anomalous or not. Similar N400 effects were found to anomalous words and pictures. However, the scalp distribution for the anomalous pictures was more frontal than for the anomalous words. Nigam and colleagues (Nigam et al. 1992) also found similar N400 effects for pictures and words, but did not find a difference in scalp distribution. However, this might be due to the limited number of electrodes that they recorded from, which did not cover the frontal part of the brain. Federmeier and Kutas (2001) found a correlation between the amplitude of the N400 effect and the semantic fit of a picture with respect to the preceding part of a sentence. Again, there was a frontal scalp distribution for the effects. Additionally, they observed an N300 effect to the anomalous pictures. Some other ERP studies have investigated the processing of visual information following a visual context instead of a language context. West and Holcomb (2002) for instance presented a series of pictures forming a simple story. The last picture was either a congruous or an incongruous ending of the story. Incongruous pictures elicited increased N300 and N400 effects, with a maximal distribution over centro-frontal electrodes. Sitnikova and colleagues (Sitnikova et al. 2003) had congruous or incongruous objects appear in video clips of real world events. They observed an N400 effect for the incongruous objects with a fronto-central maximum in the scalp distribution. Finally, Ganis and Kutas (2003) had congruent or incongruent objects appear in still images of real-world events. An increased negativity strongly

resembling the N400 was observed for the incongruous as compared to the congruous objects.

Several studies report similar findings when pictures and words are presented outside of a sentence context. That is, N300 and N400 effects are reported to incongruous picture pairs, with a more frontal scalp distribution than is normally seen for word-word priming studies (Barrett and Rugg 1990; Holcomb and McPherson 1994; McPherson and Holcomb 1999).

In summary, ERP studies manipulating the semantic fit of pictures in relation to a (sentence) context report similar N400 amplitudes and onset latencies as found for integration of semantic information conveyed through a word. Differences are reported however in scalp distribution which is more frontal for pictures than for words, and in the finding of an earlier separate negativity, the N300. The latter component has been suggested to reflect the degree of effort needed to integrate an object-specific / imagistic representation into a preceding context (e.g. McPherson and Holcomb 1999).

From neuroimaging studies little is known about the neural localization of sentence-level processing of visual extra-linguistic information. In an earlier study (Chapter 3), we looked at how meaningful co-speech gestures compare to spoken words when anomalous in a sentence context (Willems et al. 2007). Overlap between lexical violations and gesture violations was found in left inferior frontal cortex. There is a considerable literature on the neural correlates of the semantic representation of visual objects, however. Studies investigating the processing of visual objects mostly find that ventral temporal cortex is activated to the perception of a large variety of objects (Schacter and Buckner 1998; Martin and Chao 2001). More important for the present study is that many of these studies also report inferior frontal cortex to be sensitive to the repeated presentation of an object (Schacter and Buckner 1998; Martin and Chao 2001) or of a word and an object (Lebreton et al. 2001). A

commonly held view derived from these and other studies is that ventral temporal activation is related to semantic knowledge of an object, whereas the inferior frontal activation is related to processes of semantic selection or retrieval (Thompson-Schill et al. 1997; Wagner et al. 1997; Martin and Chao 2001; Wagner et al. 2001). We would like to point out that although related to our study, the priming studies of objects differ in important aspects from the present study. In the present paradigm there is a relatively rich linguistic context to which picture or word can be integrated. In the repeated presentation of exemplars of object categories this is arguably not the case.

#### *The present study*

Within the study of the cognition of language, the issue of how linguistic and extra-linguistic information are integrated into a sentence context is reflected in the distinction between one-step and two-step models of language comprehension. The implication of two-step models is that, first, the meaning of a sentence is computed and second, the sentence meaning is integrated with extra-linguistic information such as information about the speaker's identity (e.g. Cutler and Clifton 1999; Lattner and Friederici 2003). This position is a consequence of Fregean compositionality, which states that the meaning of an utterance is a function of the meaning of its parts and of the syntactic rules by which these parts are combined (see Culicover and Jackendoff 2006). Since the domain of syntactic rules is the sentence, the implication of this idea is that language interpretation takes place in a two-step fashion. It is important for the present study, that the two-step model at least implies that linguistic computation should precede the integration of nonlinguistic information in time (see Hagoort and van Berkum 2007 for further discussion). The standard two-step model prohibits immediate contextualization of meaning since in this model computation of sentence level meaning has to precede the effects of contextual influences. Adherents of a one-step model, in

contrast, take as their starting point the ‘immediacy assumption’, i.e., the idea that every source of information that constrains the interpretation of an utterance (syntax, prosody, word-level semantics, prior discourse, world knowledge, knowledge about the speaker, gestures, etc.) can in principle do so immediately (Taraban and McClelland 1990; Spivey Knowlton and Sedivy 1995; Tanenhaus et al. 1995; Tanenhaus and Trueswell 1995; Hagoort and van Berkum 2007). Summarized, proponents of a two-step model would expect indicators of semantic integration in the ERP to be manifested earlier when a word has to be integrated as compared to when a picture has to be integrated into the previous sentence context<sup>1</sup>.

In terms of cortical areas important for language comprehension a recent neurobiological account of language comprehension has argued for left inferior frontal cortex to be a general (i.e. not domain-specific) unification site (Hagoort 2005b, a). Unification entails integration of information into a built-up representation of the previous sentence context as well as selection of appropriate candidates for integration (Hagoort 2005b, a). When unification is more difficult, more resources are needed to integrate linguistic as well as extra-linguistic information, resulting in increased activation levels in left inferior frontal cortex. If this is indeed the case, we should observe increased activation both when a picture and when a word is harder to integrate. However, if this area’s role is restricted to integrating language information, no such increase should be observed when information conveyed through a picture has to be integrated.

On the basis of previous studies investigating sentence level integration of co-speech gestures compared to words (Chapters 2 and 3), we have argued for linguistic and extra-linguistic information to be integrated in the same way into a linguistic context (Özyürek et al. 2007). In these studies, the semantic fit of a word or of a co-occurring co-speech gesture to the preceding sentence context was manipulated. Mismatching spoken words and co-speech gestures elicited N400 effects

with similar onset latencies. Although co-speech gestures and the pictures that we investigate here, are both extra-linguistic information, clear differences exist as well. One important characteristic of co-speech gestures is that their meaning is not recognized unambiguously when presented outside of a language context (Krauss et al. 1991). In contrast, pictures can stand on their own. Therefore, the present study is a stronger test for the claim that semantic integration at the sentence level is not domain (i.e. language) specific. If neural correlates of integration of pictures and words are similar, it follows that also information that is not necessarily bound to a language context is integrated with the same spatio-temporal profile in the brain as linguistic information is.

To address these questions we presented participants with spoken sentences in which a critical word was manipulated to either fit the sentence context or not. The critical words were accompanied by pictures (i.e., line drawings) that could also either match or mismatch with regard to the previous part of the sentence. This manipulation resulted in four conditions (see Table 4.1): Correct condition (Picture (P) +, Language (L) +); Language mismatch condition (P+L-); Picture mismatch condition (P-L+); Double mismatch condition (P-L-). In the Language mismatch the critical word was harder to integrate semantically into the preceding sentence context, while the co-occurring picture matched the sentence context. In the Picture mismatch condition the picture was harder to integrate to previous context, while the critical word matched the spoken sentence context. In the Double mismatch condition both the picture and the word were difficult to integrate to the previous sentence context. Note that in the Language and Picture mismatch conditions the critical word and the overlapping picture locally mismatched (e.g., Picture CHERRY, word 'flower', and vice-versa), while in the Double mismatch condition they locally matched (e.g., both 'cherry'). This manipulation enabled us to distinguish integration at the 'local' level of simultaneously occurring

word and picture from integration at the ‘global’ sentence level; that is, integration into a higher-level representation built-up on the basis of the preceding context information.

We had three specific hypotheses. First, for the ERP data we hypothesized that manipulating the match of both picture and word would lead to an N400 effect comparable in size and onset latency. Moreover, we were curious to see if an N300 effect would be apparent and if so, whether it would be specific to the picture mismatch condition. Previous ERP studies have mostly compared the presence of an N300 effect in reaction to pictures to the absence of an N300 to words indirectly. That is, in most studies either words or pictures were presented. Our design allows for assessing the functional relevance of the N300 in the sense that if it is sensitive to semantic load of a picture it should occur in the picture mismatch condition but not in the language mismatch condition. Second, for the fMRI study we predict a stronger involvement of inferior frontal cortex in both the picture and word mismatch conditions. If so, this would be evidence that this region, besides its well-established role in spoken and written language comprehension (e.g. Bookheimer 2002; Vigneau et al. 2006) also takes extra-linguistic visual information into account during language comprehension. Third, on the basis of earlier findings (Özyürek et al. 2007) we hypothesized our findings to reflect semantic processing at the ‘global’ sentence level but not at the ‘local’ level of the simultaneous picture and word.

Overall, our main question regards the similarity or dissimilarity of integrating linguistic and extra-linguistic information into a sentence context. Differences in neural indicators of semantic processing would favour an account in which linguistic information has a preferred status in sentence integration (Forster 1979; Fodor 1983), whereas findings of similar neural correlates would support the idea that linguistic and extra-linguistic information are integrated with a similar neural time course and by recruiting overlapping cortical areas

(Hagoort and van Berkum 2007). Moreover, an earlier effect to words than to pictures would be in favour of two step models of language comprehension (e.g. Cutler and Clifton 1999; Lattner and Friederici 2003), whereas similar neural time courses would favour accounts of immediacy in which a broad range of information types is immediately incorporated into a discourse model (Taraban and McClelland 1990; Spivey Knowlton and Sedivy 1995; Tanenhaus et al. 1995; Tanenhaus and Trueswell 1995; Hagoort and van Berkum 2007).

### **General Materials and Methods**

*Materials and Experimental Procedure* A total of 328 sentences (mean duration 3196 ms, range 2164 – 4184 ms) were recorded in a sound attenuated room at 44.1 KHz, spoken at a normal rate by a native Dutch female speaker. Half of these sentences differed in one critical word, which was never in sentence final position. In each sentence a short context was introduced to which the critical word could fit more or less easily. Critical words were nouns that corresponded to the names given by a separate group of participants (n=32) to a large set of black and white line drawings. All critical words had a picture equivalent with a naming consistency of 85% or higher. In total there were 26 critical words with their picture-equivalents. All words were one syllable long and started with a plosive consonant. Every critical word occurred equally often in a matching and a mismatching sentence context. The critical word in the mismatching sentence always had a different onset consonant than the critical word in the semantically correct sentence. Sentences were pretested in a cloze probability test that was given to a separate participant group (n=16). The percentage of participants that gave the target word as response was taken as a measure of its cloze probability. Overall, the mean cloze probability was 16% for the matching critical words (range 0 – 69%), and 0% for the semantically anomalous critical words. We choose for critical words

with low cloze probabilities to avoid confounding effects of prediction (e.g. Van Berkum et al. 2005).

Our manipulation resulted in four conditions: (i) Correct condition (Picture + Language +); (ii) Language mismatch (P+L-); (iii) Picture mismatch (P-L+); (iv) Double mismatch (P-L-). Note that mismatch in these materials is always defined relative to the preceding sentence context.

Four stimulus lists of 164 trials each were created in which only one item of every stimulus quartet (as in Table 4.1) was presented. Sentences were pseudo-randomized with the constraint that the same condition occurred maximally two times in a row. Every list contained an equal amount of stimuli from the four conditions (41 per condition). Every target word and picture was repeated on average 6.3 times (range 5-8, modus=6, median=6 repetitions) in every stimulus list. Pictures were presented from the onset of the critical word to the end of the sentence.

### **Experiment 1: EEG**

*Participants* Twenty-four healthy right-handed (Oldfield 1971) participants with Dutch as their mother tongue took part in the EEG study. None had any known neurological history or hearing complaints, and all had normal or corrected-to-normal vision. Eight participants' data had to be discarded because of an excessive number of blinks and eye movements, leaving datasets from sixteen participants (mean age=22.4 years, range 18-34, 11 female). Participants were paid for participation. The local ethics committee approved the study and all participants signed informed consent in accordance with the declaration of Helsinki.

<i>Dutch:</i>	
“De man gaf zijn vrouw een mooie <u>bloem</u> / <u>kers</u> die avond”	
<i>English:</i>	
“The man gave his wife a nice <u>flower</u> / <u>cherry</u> that evening”	
<u>Correct condition</u>	
P+L+:	The man gave his wife a nice <u>flower</u> that evening
	
<u>Language mismatch</u>	
P+L-:	The man gave his wife a nice <b>cherry</b> that evening
	
<u>Picture mismatch</u>	
P-L+:	The man gave his wife a nice <u>flower</u> that evening
	
<u>Double mismatch</u>	
P-L-:	The man gave his wife a nice <b>cherry</b> that evening
	

**Table 4.1.** An example of the Materials. Pictures were displayed time-locked to the onset of the noun (underlined). Note that the condition coding (P+L+, P+L-, etc.) refers to the match / mismatch of either the noun (Language: L) or the Picture (Picture: P) to the part of the sentence preceding the word that is underlined, with a minus sign indicating a mismatch. That is, in the correct condition (P+L+), both the word ‘flower’ as well as the picture [FLOWER] fit the preceding sentence context. In the language mismatch condition (P+L-), the word ‘cherry’ does fit the preceding sentence context less well, whereas the picture [FLOWER] does fit. Conversely, in the picture mismatch condition (P-L+) the picture [CHERRY] does not fit the preceding sentence context, whereas the word ‘flower’ does fit. Finally, in the double mismatch condition (P-L-) both the word ‘cherry’ and the picture [CHERRY] do not fit the preceding sentence context. Mismatching words are indicated in bold. All stimuli were in Dutch.

*Procedure* Stimuli were presented using Presentation software (version 9.13, <http://www.neuro-bs.com/>). Pictures had varying sizes depending upon the object they represented and were maximally 8 x 8 cm, shown at a viewing distance of 90 cm (5° x 5° visual angle). A trial started with 600 ms blank screen, followed by the spoken sentence and the picture, 1000 ms blank screen and 2500 ms with a fixation cross on the screen. Participants were instructed to sit still in a comfortable position and to blink only when a fixation cross was presented. The test session started with eight trials which contained different critical words than used in the main part of the experiment. Participants were told to attentively listen to and watch the stimuli about which they would receive questions afterwards. At the end of the test session, general questions about the stimuli were asked. All participants had understood the manipulation in the materials and could provide examples of stimuli.

*Recording and analysis* The electroencephalogram (EEG) was recorded from 27 electrode sites across the scalp using an Electrocap with Ag / AgCl electrodes, each referred to the left mastoid. Electrodes were placed on standard electrode sites (Fz, FCz, Cz, Pz, FP2, F3, F4, F8, F7, FC5, FC1, FC2, FC6, T7, T8, C3, C4, CP5, CP1, CP2, CP6, P7, P3, P4, P8, O1, O2). Vertical eye movements and blinks were monitored by means of two electrodes, one placed beneath and one above the left eye. Horizontal eye movements were monitored by means of a left to right bicanthal montage. Activity over the right mastoid was recorded to determine if there were additional contributions of the experimental variables to the two presumably neutral mastoid sites. No such differences were observed. The EEG and Electrooculogram (EOG) recordings were amplified with BrainAmp DC amplifiers, using a band pass filter from 10 s to 100 Hz. Impedances were kept below 5 kOhm for all channels. The EEG and EOG signals were recorded and digitized

using Brain Vision Recorder software (version 1.03), with a sampling frequency of 500 Hz.

The data were filtered off-line with a 30 Hz low pass filter, re-referenced to the mean of the two mastoids and segmented from 150 ms before to 1000 ms after the critical word. Segments were normalized to the mean amplitude of a baseline period 150 ms before the critical word (baseline correction). All segments were screened for eye movements, electrode drifting, amplifier blocking and muscle artefacts. Trials containing such artefacts were rejected (mean=8.6 %, SD=5.2 %, range 0-18 %). Rejected trials were equally distributed across conditions ( $F < 1$ ). Segments were averaged for each condition for each participant at each electrode site. Repeated measures analysis of variance (ANOVA) was applied to the mean activity in four time-windows (see results) with factors condition (P+L+, P+L-, P-L+, P-L-) and quadrant (Left Anterior, Right Anterior, Left Posterior, Right Posterior).

Electrodes were assigned to quadrants as follows: Left Anterior (F3, F7, FC1, FC5, C3); Right Anterior (F4, F8, FC2, FC6, C4); Left Posterior (CP1, CP5, P3, P7, O1) and Right Posterior (CP2, CP6, P4, P8, O2). A separate ANOVA was performed for the midline electrodes (Fz, FCz, Cz, Pz). Huynh-Feldt correction for violation of sphericity assumption was applied when appropriate (Huynh and Feldt 1976). Differences in N400 effect onset latencies were tested by calculating the time bin (bins of 10 ms) at which 20% of the total area of the difference waves of the experimental conditions with the correct condition in the 200-500 ms latency window was reached (fractional area latency analysis).

Statistical significance of these differences was assessed by using the jackknifing procedure described by Miller and colleagues (1998).

### **Results EEG Experiment**

Visual inspection of the grand average waveforms (Fig. 4.1) showed clear N1 and P2 components followed by a negativity starting from 350 ms resembling the N400. The correct condition showed a slightly

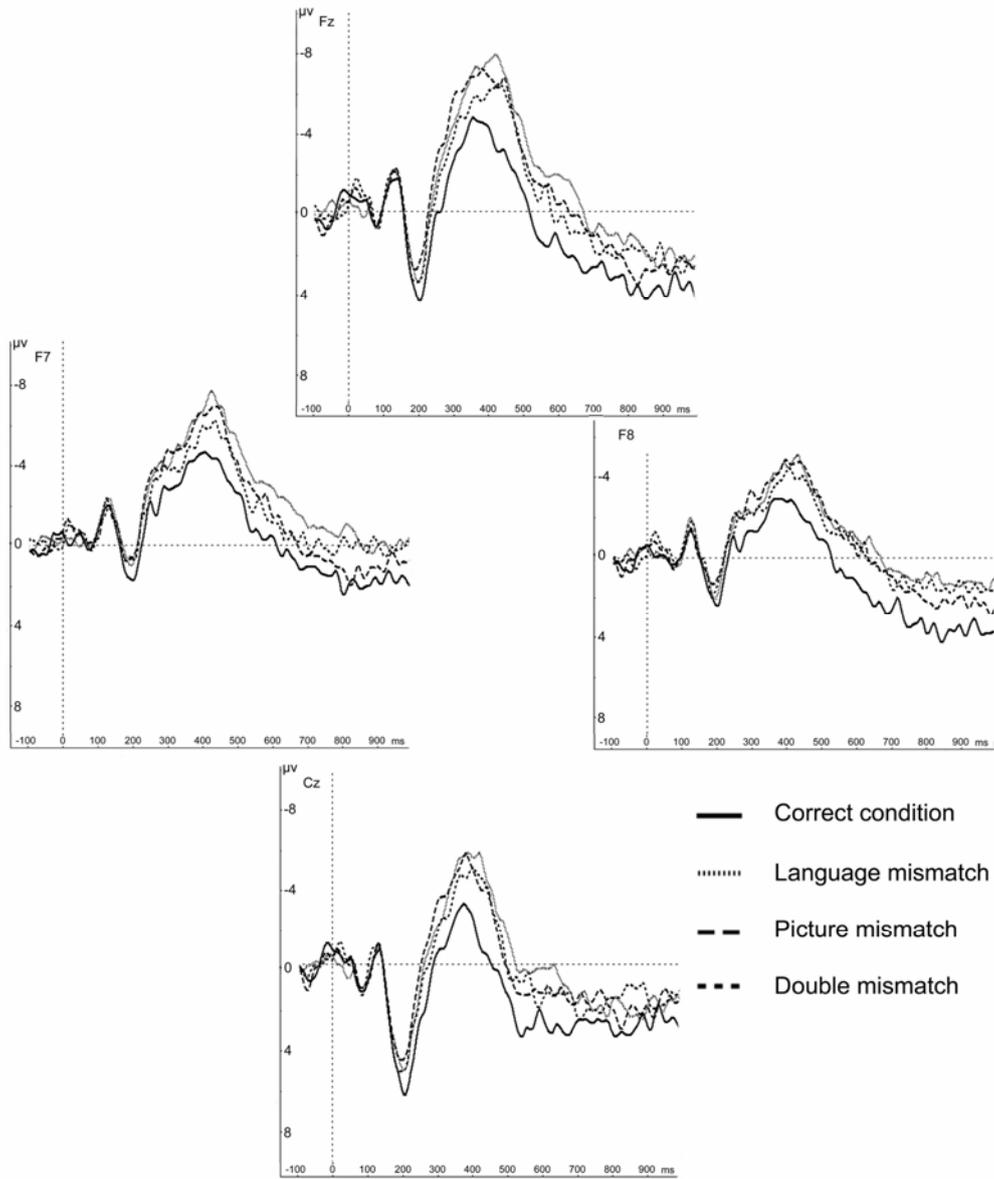
decreased peak at the N1 and a higher positivity at the P2 peak compared to the other three conditions (Fig. 4.1 and 4.2). In the N400 time window, the three mismatch conditions (P+L-, P-L+, P-L-) were more negative than the correct condition. The mismatch conditions stayed more negative than the correct condition until the end of the segment. Consequently, ANOVAs were done on the mean amplitudes in the latency ranges 170-250 ms, 300-550 ms and 600-900 ms. Additional tests were performed in the N300 time window (225-325 ms) given that previous literature (e.g. McPherson and Holcomb 1999) reports picture specific effects in this time window.

#### *P2 time window (170-250 ms)*

Statistical analyses in this time window failed to reveal a significant effect of Condition ( $F(3, 45)=2.34, p=0.09$ ). There was also no Condition x Quadrant interaction ( $F<1$ ). However, in the ANOVA over midline electrodes, a main effect of Condition was found ( $F(3, 45)=2.87, MSe=9.971, p=0.047$ ). Planned comparisons over the midline electrodes, of every experimental condition versus the correct condition, showed this effect to be strongest in the picture mismatch condition ( $F(1,15)=5.82, MSe=26.96, p=0.029$ ), although there were also marginally significant effects in the language mismatch versus correct condition ( $F(1,15)=4.29, MSe=21.27, p=0.056$ ) and in the double mismatch versus correct condition comparisons ( $F(1,15)=4.51, MSe=10.745, p=0.051$ ).

#### *N300 time window (225-325 ms)*

The morphology of the grand average waveforms does not clearly indicate the presence of a separate N300 component. Given previous findings of the N300 for mismatching pictures we tested effects in the 225-325 ms time window (e.g. McPherson and Holcomb 1999). A main effect of condition was observed ( $F(3, 45)=3.17, MSe=42.96, p=0.040$ ), but no Condition x Quadrant interaction ( $F(9, 135) = 1.26, MSe=10.22$ ,



**Fig. 3.1.** Grand average ERPs for the four conditions at electrodes Fz, F7, F8 and Cz. ERPs were time-locked to the onset of the critical word and picture. Negativity is plotted upwards.

p=0.29). Pairwise comparisons between all conditions were performed and p-values were corrected for the number of tests accordingly. Only the double mismatch versus correct condition differed significantly from each other ( $F(1,15)=9.82$ ,  $MSe=17.3$ ,  $p=0.042$ ). Since previous studies found the distribution of the N300 effect to be frontal, we separately tested in left and right anterior quadrants. Again, there was a main effect of condition (left:  $F(3, 45)=3.47$ ,  $MSe=22.54$ ,  $p=0.04$ ; right:  $F(3, 45)=4.22$ ,  $MSe=11.88$ ,  $p=0.01$ ). Pairwise comparisons revealed only the double mismatch condition to be significantly different from the correct condition in the left anterior quadrant ( $F(1,15)=11.33$ ,  $MSe=10.17$ ,  $p=0.024$ ). The picture mismatch versus correct condition was marginally significant in the left anterior quadrant only ( $F(1,15)=8.20$ ,  $MSe=34.96$ ,  $p=0.07$ ). No other comparisons revealed significant differences between conditions. Summarized, although there is a main effect of condition in this time window, this effect is not specific to the picture and / or double mismatch conditions.

#### *N400 time window (300-550 ms)*

Table 4.2 summarizes the results in this time window. There was a main effect of Condition ( $F(3, 45)=11.46$ ,  $p<0.001$ ), but no Condition x Quadrant interaction ( $F(9, 135)=1.56$ ,  $p=0.16$ ). To explore specific differences between conditions, pairwise comparisons were performed. Accordingly, the p-values are corrected for the number of tests performed (see Table 4.2). Pairwise comparisons revealed that all mismatch conditions differed significantly from the correct condition. No other comparisons showed significant differences between conditions (Table 4.2). These effects are spatially smeared out over the scalp, with a tendency for fronto-central electrodes to show the greatest effect size (Fig. 4.3). To formally test the onset latencies of the N400 effects, a fractional area latency measure was computed in the 200-500 ms time window (see above). The time point at which 20% of the grand average difference waveform was reached was 305 ms for the Language

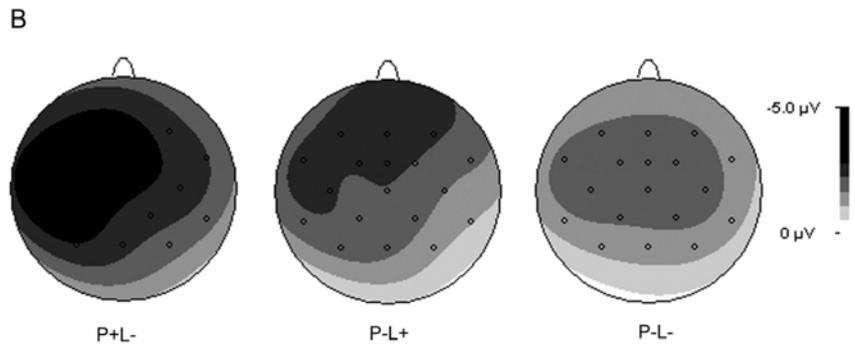
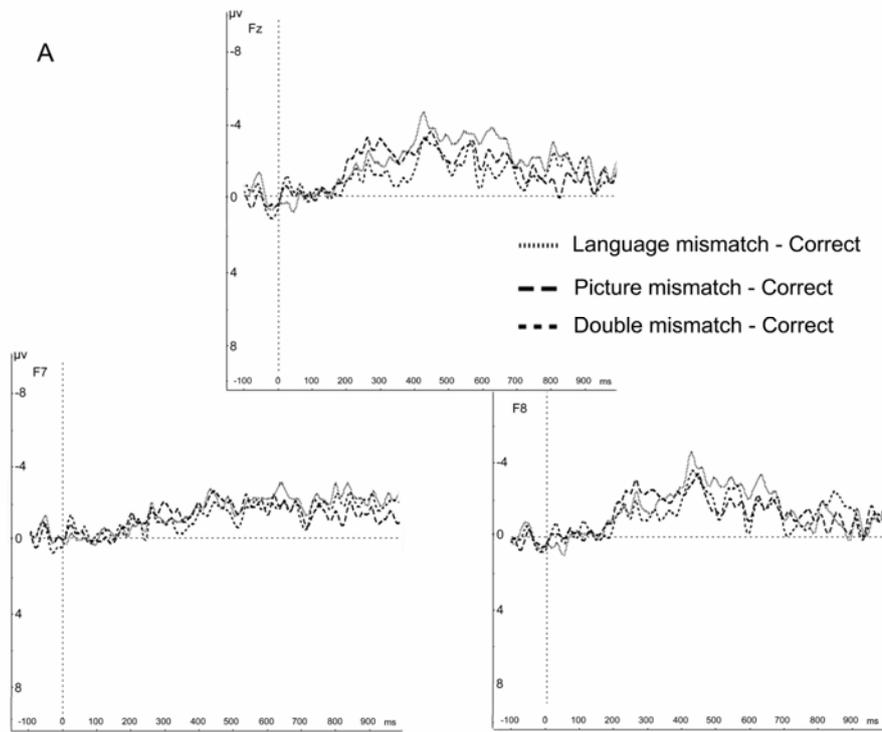
mismatch–Correct condition; 285 ms for the Picture mismatch–Correct condition and 295 ms for the Double mismatch condition–Correct condition. Statistical testing of these differences (see Miller et al., 1998) revealed that onset latencies did not differ from each other (all  $t < 1$ ).

Finally the scalp distributions of the difference waves (Fig. 4.3) were tested in a two-way ANOVA with factors Difference (Language mismatch–Correct condition, Picture mismatch–Correct condition, Double mismatch–Correct condition) and Quadrant. The lack of an interaction effect ( $F < 1$ ) indicates that the scalp distributions were similar for the N400 effects elicited by the mismatch conditions.

#### *Late time window (600-900 ms)*

In the late time window (600-900 ms) there was a significant main effect of Condition ( $F(3, 45)=3.71$ ,  $MSe=31.55$ ,  $p=0.018$ ) and a significant Condition x Quadrant interaction ( $F(9, 135)=3.62$ ,  $MSe=6.259$ ,  $p=0.003$ ). Subsequent tests in specific quadrants revealed main effects of Condition only in left anterior ( $F(3, 45)=5.816$ ,  $MSe=13.914$ ,  $p=0.002$ ) and right anterior quadrants ( $F(3, 45)=5.795$ ,  $MSe=11.974$ ,  $p=0.002$ ). Again, we tested pairwise comparisons between all conditions and we report adjusted p-values accordingly. In the left anterior quadrant the language mismatch differed significantly from the correct condition ( $F(1, 15)=30.744$ ,  $MSe=15.100$ ,  $p=0.0003$ ). The double mismatch condition and the picture mismatch condition did not

**Fig. 4.2.** (opposite page) **A**) Difference waves of the experimental conditions minus the correct condition (Language mismatch–Correct condition; Picture mismatch–Correct condition; Double mismatch–Correct condition) at electrodes Fz, F7 and F8. Difference waves are time-locked to the onset of the critical word and picture. Negativity is plotted upwards. **B**) Spline interpolated isovoltage maps of the mean difference wave 300-550 ms after the critical word. Displayed are the difference of the Language mismatch condition (left), Picture mismatch condition (middle) and Double mismatch condition (right) with the Correct condition.



differ significantly from the correct condition ( $F(1, 15)=4.313$ ,  $MSe=40.538$ ,  $p=0.33$  and  $F(1,15)=2.433$ ,  $MSe=30.470$ ,  $p>0.5$  respectively). In the right anterior quadrant a qualitatively similar pattern of results was observed with language mismatch being significantly different from the correct condition ( $F(1, 15)= 11.535$ ,  $MSe=32.543$ ,  $p=0.024$ ), while the picture mismatch and double mismatch conditions were not different from the correct conditions ( $F(1, 15)=5.295$ ,  $MSe=29.862$ ,  $p=0.216$ ; P-L- vs. P+L+:  $F(1, 15)=8.313$ ,  $MSe=27.528$ ,  $p>0.5$ ). Other comparisons did not reveal significant differences between conditions. Over the midline electrodes, there was only a trend for a main effect of condition ( $F(3, 45)=2.45$ ,  $MSe=12.41$ ,  $p=0.076$ ).

### **Discussion EEG Experiment**

Highly similar N400 effects were found for all experimental conditions compared to the correct condition. Comparing our ERP findings to earlier studies investigating semantic processing of pictures, a few differences are readily apparent. In contrast to some earlier findings (McPherson and Holcomb 1999; Federmeier and Kutas 2001, 2002; West and Holcomb 2002), we failed to observe a separate N300 effect, which has been claimed to be specific to the processing of pictorial stimuli. However, some other studies investigating pictures in a sentence context have also failed to observe a separate N300 effect (Nigam et al. 1992; Ganis et al. 1996). Therefore, we argue for the N300 as not being specific to the processing of pictures, at least not when presented within a sentence context. The absence of a picture specific effect in the ERP waveforms and the similar time course of the N400 suggest that at this level of processing, no differentiation is made between verbal and visual semantic information.

The double mismatch condition, in which both the word and the picture did fit the previous context less well, evoked an N400 similar in latency and amplitude than the other mismatch conditions, in which either

word or picture were in discordance with the previous context. Conflicting information coming from the visual or verbal domain does not add up linearly to increase the effect size in the double mismatch. Furthermore, the fact that the double mismatch N400 starts at the same latency as the other two mismatch conditions speaks in favour of the ‘immediacy assumption’ which predicts that information is used by the language comprehension system as soon as it is available. In other words, it suggests that picture and word are not first integrated at a lower level of processing before being integrated into the sentence context. If this were the case, a delay in the N400 response to the language mismatch condition and the picture mismatch condition would have been expected.

Interestingly, an earlier, marginally significant difference between mismatch conditions and the correct condition could be observed in the time window of the P2 component. This is most parsimoniously explained as a lead-in effect of the subsequent N400, which was much more negative for the three mismatch conditions than for the correct condition.

In line with earlier ERP studies, the scalp distribution of the N400 effect was more frontal than the centro-posterior distribution that is normally observed in studies of spoken or written language. The frontal distribution was however not specific to the picture mismatch condition. Therefore, although the presence of visual information might shift the N400 distribution to a more frontal maximum, the fact that this holds even when the anomaly is language-internal argues against a picture-specific integration process that is different from semantic integration of written or spoken words. Together with results in Chapter 2 (Özyürek et al. 2007) it seems that the mere presence of a visual stimulus (other than a written word) makes the scalp distribution ‘shift’ to a more frontal maximum compared to when only linguistic information is presented.

The stronger negativities to the mismatch conditions in the later

Source	df	F	MSe	p
	ANOVA (4 cond. X 4 quadr.)			
Condition	3, 45	11.46	26.72	<0.001***
Condition x Quadrant	9, 135	1.56	8.86	0.162
Pairwise comparisons	1,15	42.34	39.18	p(corr.) <0.001***
P+L- vs. P+L+	1,15	11.42	73.08	0.025*
P-L+ vs. P+L+	1,15	15.39	37.05	0.008**
P-L- vs. P+L+	1,15	4.66	60.93	0.285
P+L- vs. P-L-	1,15	0.67	37.53	ns
P-L+ vs. P-L-	1,15	2.38	58.83	ns
P-L+ vs. P+L-				
	Midline (4 cond. X 4 electr.)			
Condition	3, 45	9.81	10.58	<0.001***
Pairwise comparisons	1,15	37.72	13.76	p(corr.) <0.001***
P+L- vs. P+L+	1,15	10.82	28.82	0.029*
P-L+ vs. P+L+	1,15	10.82	19.31	0.030*
P-L- vs. P+L+	1,15	2.92	23.76	ns
P+L- vs. P-L-	1,15	0.75	13.79	ns
P-L+ vs. P-L-	1,15	1.50	17.42	ns
P-L+ vs. P+L-				

**Table 4.2.** ERP results in the 300-550 ms time window. Amplitudes of the ERPs were averaged over this time window for every participant separately and entered into repeated measures ANOVA with factors Condition (4 levels) and Quadrant (4 levels). A separate ANOVA was performed for the Midline electrodes with factors Condition (4 levels) and Electrode (4 levels). Huynh-Feldt correction for violation of sphericity assumption was applied (Huynh and Feldt 1976), but the original degrees of freedom are reported. The significance levels for the pairwise comparisons were corrected for the number of tests performed by Bonferroni correction. The corrected p-levels are reported; effects with p-values >0.5 are reported as not significant (ns). P+L+: correct condition; P+L-: language mismatch; P-L+: picture mismatch; P-L-: double mismatch. \*p<0.05, \*\*p<0.01, \*\*\*p<0.001.

time window resemble the findings in the N400 time window, although only the language mismatch condition differed significantly from the correct condition. However, no differences were observed between the experimental conditions. Therefore, these late effects are best explained as a carry over effect of the strong N400 effects.

### **Experiment 2: fMRI**

*Participants* Nineteen healthy right-handed (Oldfield, 1971) participants with Dutch as their mother tongue took part in the fMRI study. None had any known neurological history or hearing complaints, and all had normal or corrected-to-normal vision. Three data sets in the fMRI study had to be discarded, two because of inattentive participants (see below) and one because of excessive head motion. Data from the sixteen remaining participants (mean age=22.3 years, range 20-28, 8 female) were entered into the analysis. Participants were paid for participation. The local ethics committee approved the study and all participants signed informed consent in accordance with the declaration of Helsinki.

*Procedure* Stimuli were the same as in the EEG experiment. Pictures were projected from outside of the scanner room onto a screen at the end of the patient table. The screen was visible through a mirror mounted to the head coil, at a viewing distance of 80 cm (pictures subtended a 5.7° x 5.7° visual angle). Speech was presented through non-magnetic headphones (Commander XG, <http://www.mrvideo.com>), which dampened scanner noise. Intertrial interval was 6, 7 or 8 seconds. During the scanning session eye movements were recorded using an infrared IviewX eyetracker (<http://www.smi.de>), to formally control participant's vigilance during scanning.

The scanner was switched on during the practice trials and participants had to indicate whether the volume should go up or down. No participant asked for the volume to be increased to the maximally

possible level. Participants were told to attentively listen to and watch the stimuli about which they would receive questions afterwards. All participants indicated they were able to hear and understand the sentences well. At the end of the scanning session, general questions about the stimuli were asked. All participants had understood the manipulation in the materials and could provide examples of stimuli.

*Recording and analysis* MR imaging was performed on a 3T Siemens Magnetom Trio scanner (Siemens, Erlangen, Germany). Per participant, approximately 800 echo planar whole brain images were acquired (TR=2230 ms; TE=30 ms; flip angle=80°; 32 slices; slice thickness=4 mm; FoV=224 mm, voxel resolution=3.5 mm x 3.5 mm x 4 mm). Additionally, a T1 weighted anatomical scan (3D-MPRAGE, 192 slices, TR=2300 ms; TE=3.93 ms; FoV=256 mm; slice thickness=1 mm) was made. Data were analyzed using Brainvoyager QX (Brain Innovation, <http://www.brainvoyager.com>). The first five volumes of a session were discarded to avoid T1 saturation effects. Preprocessing involved rigid body transformations of all volumes to the first volume, slice scan time correction, linear trend removal and high pass temporal filtering (cut-off 3 cycles over the time course) and spatial smoothing with a Gaussian filter kernel of 8 mm FWHM. Data were transformed into stereotaxic space (Talairach and Tournoux 1988). A whole brain analysis was performed in the context of the General Linear Model, with the conditions as factors of interest and the six parameters from the motion correction as nuisance factors. Experimental factors were modelled for the duration of each sentence and convolved with a canonical 2 gamma hemodynamic response function. Parameters were estimated for every voxel's time course. Effect sizes were estimated by constructing contrast (t) maps consisting of differences between the parameter estimates in every voxel and participant separately for contrasts of interest. Subsequently, contrast maps were taken to a second level analysis, testing for differences from zero in a one-sample

t-test in a random effects analysis. Every contrast was tested two-sided. The multiple comparisons problem was addressed by thresholding the activation maps at  $t(15)=3.9$ ,  $p<0.001$  at the voxel level and taking the cluster sizes into account, leading to a correction at an alpha level of  $p<0.05$  (Forman et al. 1995). The eyetracking data were used to control for the vigilance (i.e. wakefulness) of the participant. Two datasets had to be discarded because participants had their eyes closed in more than 10% of the trials.

### **Results fMRI Experiment**

To see effects specific for the language condition, the language mismatch condition (P+L-) was contrasted against the correct condition (P+L+). The correct condition served as a high-level baseline in this way. An extensive region in left inferior frontal cortex, stretching into premotor cortex and an area in left superior temporal sulcus were found activated (Table 4.3; Fig. 4.4A). One area in the right middle frontal sulcus was activated in the reversed contrast (i.e. correct condition versus language mismatch). Second, effects to the picture condition were assessed, again by comparing it to the correct condition (P-L+ vs. P+L+). Part of left inferior frontal sulcus showed significant activation to this contrast (Table 4.3, Fig. 4.4B). Finally, the double mismatch condition compared to correct condition tested the effect of both picture and word being in discordance with the sentence context. This comparison (P-L- vs. P+L+) led to increased activity in an extensive part of the inferior frontal cortex stretching into premotor cortex, an area in left superior temporal sulcus, an area in the left temporo-parietal junction, and a small area of activation in the right cerebellum (Table 4.3, Fig. 4.4C).

Figure 4.4 displays the results of a conjunction analysis (conjunction as in a logical AND, see Nichols et al. 2005) testing for overlap between the comparisons described above (P-L+ vs. P+L+  $\cap$  P+L- vs. P+L+  $\cap$  P-L- vs. P+L+). One region in left inferior frontal

cortex (max [-40 11 31]) was found activated in this contrast. Compared to a cytoarchitectonic probability map, 39% of this region overlapped with BA 44 (with a mean probability per overlapping voxel of 40% to be part of BA 44) in contrast to only 3% of the voxels that were classified as being part of BA 45 (Eickhoff et al. 2005).

Finally, we tested for differential effects to the local and global match or mismatch effects. This was done by comparing the language mismatch and picture mismatch to the double mismatch condition (P+L- vs. P-L-  $\cap$  P-L+ vs. P-L-). In this way, all conditions involve a sentence level mismatch, but only the language and picture mismatch conditions had an additional local mismatch. No areas were found activated in this contrast.

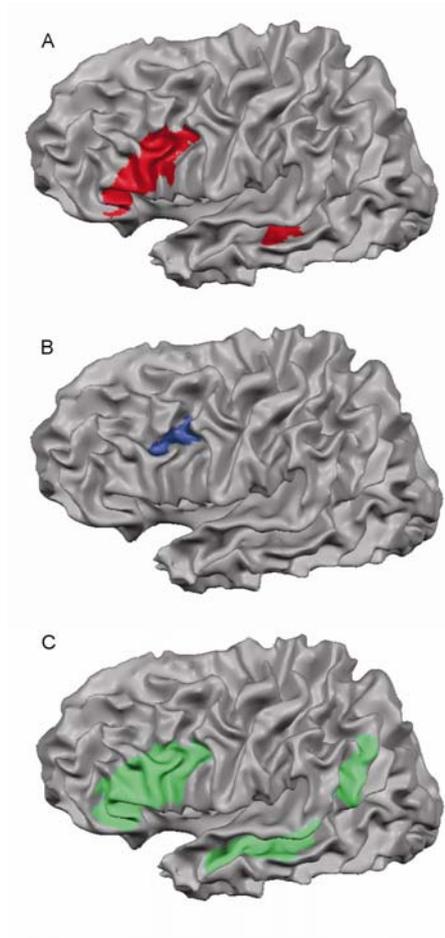
### **Discussion fMRI Experiment**

We observed increased activation levels in all three mismatch conditions compared to the correct condition in left inferior frontal cortex. This study adds to a large number of studies showing that left inferior frontal cortex is an important node in the speech comprehension network (for reviews see Bookheimer, 2002; Vigneau et al., 2006). We interpret our findings as reflecting unification processing in left inferior frontal cortex. This entails integration of information into a built-up representation of the previous sentence context as well as selection of appropriate candidates for integration (Hagoort 2005b, a). We show here that also integration of extra-linguistic information such as a visual picture recruits this area.

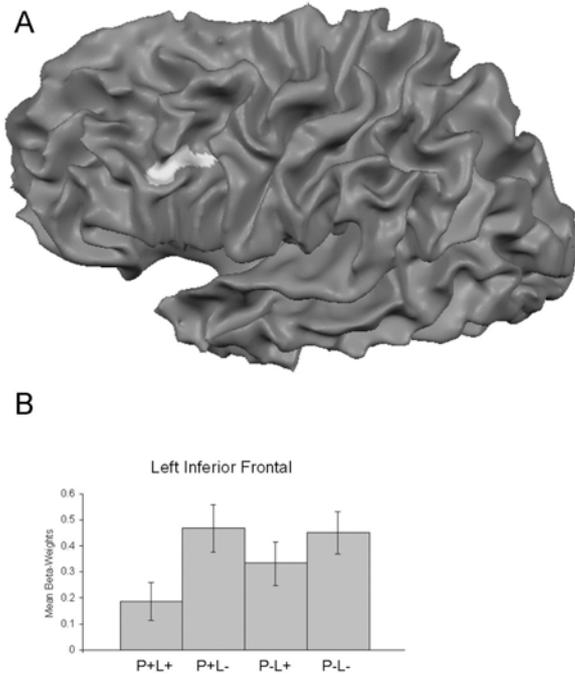
The peak of overlapping activation in inferior frontal cortex was in BA 44, which is at odds with proposals of a gradient of linguistic sub processes (such as semantics, syntax and phonology) within left inferior frontal cortex (Bookheimer, 2002; Vigneau et al., 2006). Semantic processing is centered more ventrally, around BA 45 and 47 in these proposals (Bookheimer, 2002; Vigneau et al., 2006). However, as can be

Contrast	Coordinates			Region	T	Number of voxels
	x	y	z			
Language mismatch vs. correct	-45	15	24	Left Inferior Frontal Sulcus / Premotor cortex	8.25	8655
	-52	-34	-1	Left Superior Temporal Sulcus	5.42	2507
Picture mismatch vs. correct	-38	8	32	Left Inferior Frontal Sulcus	6.23	2404
Double mismatch vs. correct	-43	17	23	Left Inferior Frontal Sulcus / Premotor cortex	11.44	17304
	-50	-29	-6	Left Superior Temporal Sulcus	11.20	9370
	-38	-60	14	Left Temporo-Parietal junction	5.84	3103
	17	-38	-27	Right Cerebellum	5.46	906
Correct vs. language mismatch	28	30	44	Right Middle Frontal Sulcus	7.15	1965

**Table 4.3.** Results from the fMRI experiment. Regions are presented that were significantly activated in the whole brain random effects group analysis ( $t(15) > 3.9$ ,  $p < 0.05$ , corrected) by contrasting each mismatch condition against the correct condition. Displayed are the contrasts, the centre coordinates in stereotaxic space (Talairach and Tournoux 1988), a description of the region, the T value of the maximally activated voxel and the number of significant voxels (1x1x1 mm voxel size).



**Fig. 4.3.** (For colour version see Appendix, p. 267). Results from the fMRI whole brain random effects group analysis ( $t(15) > 3.9$ ,  $p < 0.05$ , corrected). Areas significantly activated in the **A**) Language mismatch versus Correct condition contrast (red), **B**) Picture mismatch versus Correct condition contrast (blue), **C**) Double mismatch versus Correct condition contrast (green). Results are overlain on a cortical sheet segmented along the grey-white matter border in stereotaxic (Talairach) space.



**Fig. 4.4. A)** Part of the left inferior frontal cortex commonly activated by all mismatch conditions. This area was the result of a conjunction analysis (Nichols et al., 2005) of each mismatch condition to the correct condition (P-L+ vs. P+L+  $\cap$  P+L- vs. P+L+  $\cap$  P-L- vs. P+L+). **B)** Parameter estimates for all conditions from the area in A. Although this area is more strongly activated in all mismatch conditions than in the correct condition, the parameter estimate to the correct condition was significantly different from zero ( $t(15)=2.43$ ,  $p<0.03$ ), indicating that the area is also activated in the correct condition.

seen in Figure 4.4, two of the mismatch conditions did activate a more ventral part of left inferior frontal cortex (language mismatch condition and double mismatch condition). More in general, it should be pointed out that what is most striking from meta-analyses (such as Bookheimer, 2002; Vigneau et al., 2006) is the large spread around the mean centre coordinates. Because of this variance across studies, one single study such as the present one cannot be taken as evidence in favour of or against a gradient of linguistic processes in left inferior frontal cortex.

A potential worry is that the activation of left inferior frontal cortex could be a by-product of using the mismatch paradigm, bearing little relevance to general language processing. As is clear in Figure 4.5B, however, also the correct condition resulted in an activation increase in this region. The activation levels of the correct condition were found to be significantly different from zero (see figure caption). We show that inferior frontal cortex activation is also involved in the processing of semantically correct sentences and that its activation in this study cannot be attributed to the use of the mismatch paradigm (see also Hagoort 2005b; Rodd et al. 2005; Davis et al. 2007; Hasson et al. 2007; Willems et al. 2007).

In relation to the object priming literature cited in the introduction we want to point out that our results cannot be explained in terms of increased conceptual priming in the correct condition. That is, both in the correct condition as well as in the double mismatch condition word and picture were conceptually the same. However, increased inferior frontal cortex activation was nevertheless observed in the double mismatch condition. The reason that no priming effects are observed is probably because picture and word are not presented after another as is usually done in priming paradigms. That is, a picture or a word does not form the context for the other item; rather, the preceding sentence is the crucial context.

Finally, apart from overlapping areas across conditions we also found an increase in activation in the left superior temporal sulcus specific to the language and double mismatch conditions, but not to the picture mismatch condition<sup>2</sup>. A similar result was obtained in Chapter 3 investigating the processing of co-speech gestures (Willems et al. 2007). This suggests that superior temporal regions might be specifically involved in verbal semantics<sup>3</sup>. Interestingly, no specific effect was observed for the picture mismatch condition. Given the role of ventral temporal cortex in object representations, this area might have been expected to be more activated in the picture mismatch condition as compared to the other conditions. Such effects were however not observed.

### **General Discussion**

In this study, we compared the integration of semantic information conveyed through spoken language (words) and visual information (pictures) at the sentence level. Overall, our results provide strong evidence for both processes to tax the same neural processes. That is, neural indicators of semantic integration react the same to both a higher integration load when information is conveyed through a word than when it is conveyed through a picture. A same neural time course is indicated by same onset latencies and effect sizes of the N400 effects. The processing at this level of comprehension does not give temporal precedence to linguistic information over extra-linguistic information as indicated by the N400 effects. In terms of neural locus, part of left inferior frontal cortex was commonly activated by all mismatch conditions. A recent neurobiological account of sentence comprehension has interpreted increased activation in left inferior frontal cortex as being the neural indicator of increased integration load of a word's meaning into a built-up (sentence) context (Hagoort 2005b). Here we provide evidence for this region not to be domain-specific since the integration of information presented in a non-linguistic modality also

taxes this region. This is in line with an earlier study in which we found left inferior frontal cortex to be activated more strongly to both spoken words and co-speech gestures in a sentence context. However, there is an important difference between our previous and the present study. Co-speech gestures are necessarily bound to a language context; that is, they do not clearly represent their meaning when presented on their own (Krauss et al. 1991; McNeill 1992). Pictures, on the contrary, are fully meaningful outside of a language context. In this way the present study provides more convincing evidence for the claim that the role of left inferior frontal cortex in language comprehension is not domain-specific. Left inferior frontal cortex plays an important role in integration and selection operations that combine linguistic and extra-linguistic visual information into a coherent overall interpretation of an expression.

The current study adds to an understanding of the language comprehension system as taking several types of information into account in the same way when understanding a message (see also Taraban and McClelland 1990; Trueswell and Tanenhaus 1994; Spivey Knowlton and Sedivy 1995; Tanenhaus and Trueswell 1995). That is, the system does not restrict itself to one source of information (speech), but seems to use a rich variety of sources of meaningful information in a qualitatively similar way when understanding a message. Note that the visual information in our study was rather simple, consisting of pictures of single objects. In contrast, in the eye movement literature cited above, visual context often involves several objects (e.g. Tanenhaus et al., 1995). For reasons of comparability with Chapter 2 (e.g. Özyürek et al. 2007) as well as for reasons of experimental control we restricted ourselves to using pictures of single objects as stimuli. However, the few ERP studies that did use a richer visual context seem to suggest that similar findings would be obtained if the visual stream of information had been richer (e.g. West and Holcomb 2002; Ganis and

Kutas 2003; Sitnikova et al. 2003). This is however an issue that is open for empirical investigation.

Importantly, our results support a theory of language processing that goes against the classical two-step model of interpretation (e.g. Cutler and Clifton 1999; Lattner and Friederici 2003). Instead, in line with the immediacy assumption, all available information is used directly to co-determine the interpretation of linguistic expressions. Moreover, we show that the role of inferior frontal cortex in the language comprehension network is not restricted to linguistic information. Rather, also an increased semantic integration load conveyed by a picture activates this area.

### **Notes**

- 1) Note that this distinction between two-step and one-step models is different from the distinction between syntax-first models (e.g. Frazier 1987) versus constraint-based models of sentence comprehension (see Hagoort and van Berkum 2007).
- 2) Informal visual inspection of the contrast map at a lower threshold confirmed that in the picture mismatch condition superior temporal cortex was not activated.
- 3) This could relate to the debated issue of the semantic system to be organized in multiple semantic codes or in one common code (McCarthy and Warrington 1988; Shallice 1988; Caramazza et al. 1990). Our study was designed to investigate sentence level integration of semantic information and not targeted at revealing the neural representations of words and pictures. Therefore we are reluctant to interpret our results in terms of multiple or modality specific codes.

### **Acknowledgements**

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO), 051.02.040. Petra van Alphen is acknowledged for expertly voicing the sentences. We thank

Heidi Koppenhagen and Niels Schiller for providing the naming consistency information of the line drawings and Tineke Snijders, Giosuè Baggio and Tessa van Leeuwen, as well as two anonymous reviewers for helpful comments. We thank Paul Gaalman for assistance during the scanning sessions and Miriam Kos for help in EEG data collection.

## **Chapter 5** Early decreases in alpha and gamma band power distinguish linguistic from visual information during sentence comprehension\*

### **Abstract**

Language is often perceived together with visual information. This raises the question how the brain integrates information conveyed in visual and / or linguistic format during spoken language comprehension. In this study we investigated the dynamics of semantic integration of visual and linguistic information by means of time-frequency analysis of the EEG signal. A modified version of the N400 paradigm with either a word or a picture of an object being semantically incongruous with respect to the preceding sentence context was employed. Event-Related Potential (ERP) analysis showed qualitatively similar N400 effects for integration of either word or picture. Time-frequency analysis revealed early specific decreases in alpha and gamma band power for linguistic and visual information respectively. We argue that these reflect a rapid context-based analysis of acoustic (word) or visual (picture) form information. We conclude that although full semantic integration of linguistic and visual information occurs through a common mechanism, early differences in oscillations in specific frequency bands reflect the format of the incoming information and, importantly, an early context-based detection of its congruity with respect to the preceding language context.

---

\*This chapter is a slightly modified version of Willems, R. M., Oostenveld, R., & Hagoort, P. (2008). Early decreases in alpha and gamma band power distinguish linguistic from visual information during sentence comprehension. *Brain Research: 1219*, 78-90.

## **Introduction**

Language is often perceived in the presence of concomitant semantic information. For instance, linguistic utterances often take place with reference to objects in the environment. Consider someone showing his friend the features of a new car. The speaker will perhaps talk about improvements to the engine of the vehicle, while at the same time showing the engine to the listener. Here, the co-occurrence of language and visual information is an important feature of the way the message is conveyed by the speaker as well as how it is understood by the listener. A consequence of this common co-occurrence is that the brain continuously has to integrate streams of information conveyed through different modalities during language comprehension. Importantly, as in the example above, such integration has to happen at a semantic level. That is, there is no way in which the form properties of the visual information and of the spoken language overlap. Here we investigated the possibility that such integration can be distinguished neurally in terms of differences in changes in power in specific frequency bands. Previous research indicates that integration of visual and linguistic information at this level of processing taxes overlapping neural correlates. Several studies employing the event-related potential (ERP) technique show that an incongruous picture of an object evokes a qualitatively similar N400 effect as compared to an anomalous word (the N400 is thought to reflect semantic integration load of an item with respect to a previous context; see below). For instance, Ganis, Kutas and Sereno (1996) presented sentences that either ended with a word or a picture that could be anomalous or not. Similar N400 effects were found to anomalous words and pictures. The scalp distribution for the anomalous pictures was more frontal than for the anomalous words. Nigam and colleagues (Nigam et al. 1992) also found similar N400 effects for pictures and words, but did not find a difference in scalp distribution. However, this might be due to the limited number of electrodes that they recorded from, which did not cover the frontal part

of the brain. Federmeier and Kutas (2001) found a correlation between the amplitude of the N400 effect and the semantic fit of a picture with respect to the preceding part of a sentence. Again, there was a frontal scalp distribution for the effects. Additionally, they observed an N300 effect to the anomalous pictures. Some other ERP studies have investigated the processing of visual information following a *visual* context instead of a language context. West and Holcomb (2002) for instance presented a series of pictures forming a simple story. The last picture was either a congruous or an incongruous ending of the story. Incongruous pictures elicited increased N300 and N400 effects, with a maximal distribution over centro-frontal electrodes. Sitnikova and colleagues (Sitnikova et al. 2003) had congruous or incongruous objects appear in video clips of real world events. They observed an N400 effect for the incongruous objects with a fronto-central maximum in the scalp distribution. Ganis and Kutas (2003) had congruent or incongruent objects appear in still images of real-world events. An increased negativity strongly resembling the N400 was observed for the incongruous as compared to the congruous objects. Finally, in an earlier report of the ERP analysis of part of the data presented in this paper, we found that incongruent pictures and words evoke similar N400 effects and lead to overlapping activations in left inferior frontal cortex (Willems et al. 2008b); (see also Özyürek et al. 2007; Willems et al. 2007).

In summary, ERP studies manipulating the semantic fit of pictures in relation to a (sentence) context report similar N400 amplitudes and onset latencies as found for integration of semantic information conveyed through a word. Moreover, integration of information from pictures and words into a sentence context leads to overlapping activations in left inferior frontal cortex.

From these and other findings it has been claimed that despite differences in representational format, integration of linguistic and non-linguistic semantic information with language recruits the same

neural mechanisms. In line with this, Hagoort and colleagues showed that integration of two types of knowledge (lexical semantic knowledge and general knowledge of the world) during sentence comprehension follows the same neural time course and recruits an overlapping neural locus (Hagoort et al. 2004; Hagoort and van Berkum 2007). However, it was also found that integration of these information types can be distinguished in terms of differences in frequency band power of the EEG (Hagoort et al. 2004). Specifically, the world knowledge violations led to an increase in gamma band power that was not observed for semantic violations. Here we investigated whether analogous to Hagoort et al., visual and linguistic information also elicit different responses in the frequency domain.

Time-frequency analysis can reveal effects that go unnoticed in the time-locked ERP, due to the averaging of the signal in ERP analysis. In several domains of cognitive neuroscience it has proven to be fruitful to study frequency-specific changes in power to specific cognitive events (see e.g. Engel et al. 2001; Tallon-Baudry 2003; Herrmann et al. 2004b; Jensen et al. 2007). However, analysis in the time-frequency domain remains less well studied in the neurocognition of language (but see Bastiaansen and Hagoort 2006 for a recent review).

To investigate the issue of frequency-specific effects related to linguistic and visual semantic processing during sentence comprehension, we employed the N400 paradigm. A word with a meaning that is incongruous with respect to a preceding part of the sentence leads to a more negative deflection in the ERP around 400 milliseconds after word onset (Kutas and Hillyard 1980). This effect is labelled the N400 effect and has become a well-established indicator of semantic integration of for instance a word into a preceding context (see Kutas and Van Petten 1994 for review; Brown et al. 2000). In contrast, the oscillatory correlates of semantic processing are not well established. Semantic processing has been linked to increases in power in the theta band (around 4-6 Hz) by some (Bastiaansen et al. 2005)

and by decreases in power in the alpha band (around 10 Hz) by others (Rohm et al. 2001).

Relevant for the present paper are two recent studies in which an N400 paradigm was used to assess oscillatory correlates of semantic processing during sentence comprehension (Hald et al. 2006; Davidson and Indefrey 2007). Hald and colleagues observed an increase in theta band (3-5 Hz) power after a semantic incongruity. This was interpreted as reflecting an increased difficulty in lexical selection in the case of a semantically incongruous word (Hald et al. 2006). Interestingly, also an early (50-200 ms after presentation of the critical word) decrease in power in the gamma band (35-45 Hz) was observed. This effect was tentatively linked to the absence of integration or 'unification' at the sentence level in the incongruous condition. That is, in the case of a semantic incongruity unification of all words of the sentence into a coherent whole is rendered impossible, leading to the gradual built-up of gamma power to be halted (Hald et al. 2006).

Davidson and Indefrey (2007) observed a similarly late increase in theta power (3-7 Hz) after a semantic violation. No other differences were observed, but it should be noted that the gamma band was not analyzed in that study.

Here we investigated similarities or differences between oscillatory correlates of integration of information conveyed linguistically (words) or visually (pictures) during spoken sentence comprehension. To do this, we adapted the N400 paradigm to modulate semantic load of either a spoken word, a picture or of both. Participants listened to spoken sentences in which a critical word was presented which could be either semantically congruous or incongruous with respect to the preceding part of the sentence. Together with the critical word, a picture was presented which could also be congruous or incongruous (Table 5.1). There were four conditions: 1) Correct condition (Picture congruous, Word congruous) 2) Language mismatch (Picture congruous, Word incongruous) 3) Picture mismatch (Picture

incongruous, Word congruous) and 4) Double mismatch (Picture incongruous, Word incongruous). The Double mismatch condition was added to test whether effects are a reflection of increased semantic load with regard to the preceding sentence context or of mismatching co-occurring picture and word. That is, in the Language mismatch condition and the Picture mismatch condition one can argue that possible effects are driven by the fact that picture and word in these conditions convey a different meaning. If so, the effects would not be a reflection of sentence-level semantic integration. Since in the Double mismatch condition picture and word convey the same meaning (but are incongruous with respect to the preceding sentence context), this cannot be the case in this condition.

We hypothesized that all mismatch conditions would evoke an N400 effect in the ERP analysis, corroborating earlier findings as described above and as we have reported before (see Chapter 4 (Willems et al. 2008b)). Since the results of the ERP analysis of almost the same data set have been published and discussed elsewhere (Willems et al. 2008b), our focus will be on the outcome of analysis in the frequency domain. We hypothesized a relatively late theta band power increase to be a reflection of a general (that is, not language-specific) integration mechanism, analogous to the N400. If this is indeed the case, it should be obtained in all three mismatch conditions. Furthermore we expected decreases in the alpha (Rohm et al. 2001) and / or gamma (Hald et al. 2006) frequency bands in response to the Language mismatch condition. A crucial question was whether similar effects would be observed when the picture, but not the word was in discordance with the previous sentence context. Alternatively, effects specifically related to increased semantic load as conveyed through a visual stimulus may be observed. One candidate frequency band to manifest such specific effect is the gamma frequency band, which has been implicated in successful recognition of objects (see e.g. Rodriguez et al. 1999; Tallon-Baudry 2003).

## **Materials and Methods**

*Participants* Data of three participants in Chapter 4 (Willems et al. 2008b) had to be discarded because of excessive (muscle) artefacts in high frequency bands. These were replaced by three novel data sets, such that data of 16 participants went into the analysis (mean age=22.8 years, range 18-34, 13 female). All participants were healthy, right-handed (Oldfield 1971), and had Dutch as their mother tongue. None had any known neurological history, hearing complaints and all had normal or corrected-to-normal vision. Participants were paid for participation. The local ethics committee approved the study and all participants signed informed consent in accordance with the declaration of Helsinki.

*Materials* Note that the same stimuli were used as in Chapter 4. A total of 328 sentences (mean duration 3196 ms, range 2164 – 4184 ms) were recorded in a sound attenuated room at 44.1 KHz, spoken at a normal rate by a native Dutch female speaker. Half of these sentences differed in one critical word, which was never in sentence final position. In each sentence a short context was introduced to which the critical word was congruous or not. Critical words were nouns that corresponded to names given by a separate group of participants (n=32) to a large set of black and white line drawings. All critical words had a picture equivalent with a naming consistency of 85% or higher. In total there were 26 critical words with their picture equivalents. All words were one syllable long and started with a plosive consonant. Every critical word occurred equally often in a matching and in a mismatching sentence context. The critical word in the mismatching sentence always had a different onset consonant than the critical word in the semantically correct sentence. Sentences were pretested in a cloze probability test by a separate participant group (n=16). The percentage of participants that gave the target word as response was

taken as a measure of its cloze probability. Overall, the mean cloze probability was 16% for the semantically congruous critical words (range 0 – 69%), and 0% for the semantically incongruous critical words.

*Procedure* Stimuli were presented using Presentation software (version 9.13, <http://www.neurobs.com/>). Four stimulus lists of 164 trials each were created in which only one item of every stimulus quartet (as in Table 5.1) was presented. Sentences were pseudo-randomized with the constraint that the same condition occurred maximally two times in a row. Every list contained an equal amount of stimuli from the four conditions (41 per condition). Pictures had varying sizes depending upon the object they represented and were maximally 8 x 8 cm (5° x 5° visual angle; minimum height x width: 2.5 cm x 7.5 cm and 7 cm x 3 cm), shown at a viewing distance of 90 cm. Pictures were presented from the onset of the critical word to the end of the sentence. A trial started with 600 ms blank screen, followed by a spoken sentence and a picture, 1000 ms blank screen and 2500 ms with a fixation cross on the screen. Participants were instructed to sit still in a comfortable position and to blink only when the fixation cross was presented. The session started with eight practice trials which contained different critical words than used in the main part of the experiment. Participants were told to attentively listen to and watch the stimuli about which they would receive questions afterwards. At the end of the test session, general questions about the stimuli were asked. All participants had understood the manipulation in the materials and could provide examples of stimuli.

*Recording* The electroencephalogram (EEG) was recorded from 27 electrode sites across the scalp using an Electrocap with Ag / AgCl electrodes, each referred to the left mastoid. Electrodes were placed on standard electrode sites (Fz, FCz, Cz, Pz, F3, F4, F8, F7, FP2, FC5,

FC1, FC2, FC6, T7, T8, C3, C4, CP5, CP1, CP2, CP6, P7, P3, P4, P8, O1, O2). Vertical eye movements and blinks were monitored by means of two electrodes, one placed beneath and one above the left eye. Horizontal eye movements were monitored by means of a left to right bicantonal montage. Activity over the right mastoid was recorded to determine if there were additional contributions of the experimental variables to the two presumably neutral mastoid sites. No such differences were observed. Recordings were amplified with BrainAmp DC amplifiers, using a hi-cut of 100 Hz and a time constant of 10 sec. Impedances were kept below 5 kOhm for all channels. The EEG and EOG signals were recorded and digitized using Brain Vision Recorder software (version 1.03), with a sampling frequency of 500 Hz.

*Analysis* Analyses were done using the FieldTrip software package, which is an open-source Matlab toolbox designed for EEG / MEG data analysis (<http://www.ru.nl/fcdonders/fieldtrip/>). Data were filtered off-line with a 70 Hz low pass filter, re-referenced to the mean of the two mastoids and segmented from 600 ms before to 1000 ms after the critical word. All segments were screened for eye movements, electrode drifting, amplifier blocking and muscle artefacts, leading to 32% of the trials to be rejected, equally distributed over conditions ( $F < 1$ ).

For the ERP analysis, baseline correction was applied by subtracting the mean of the pre-stimulus period from 150 ms to the onset of the critical word. ERPs were created by averaging all trial segments for each condition and subject separately. Statistical analysis was performed by employing repeated measures analysis of variance (ANOVA) on the mean amplitude in the 300-600 ms time window with factors Condition (4 levels; Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels; Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2). In the case of a main effect of Condition or Condition x Electrode interaction, subsequently, planned

<i>Dutch:</i>	
“Voor in de keuken kocht zij een eenvoudige <u>kom</u> / <u>trein</u> en borden”	
<i>English:</i>	
“For (use in) the kitchen she bought a simple <u>bowl</u> / <u>train</u> and plates”	
<u>Correct condition</u>	
P+L+:	For in the kitchen she bought a simple <u>bowl</u> and plates
<u>Language mismatch</u>	
P+L-:	For in the kitchen she bought a simple <u>train</u> and plates
<u>Picture mismatch</u>	
P-L+:	For in the kitchen she bought a simple <u>bowl</u> and plates
<u>Double mismatch</u>	
P-L-:	For in the kitchen she bought a simple <u>train</u> and plates
	

**Table 5.1.** An example of the materials. Pictures were displayed time-locked to the onset of the noun (underlined). Note that the condition coding (P+L+, P+L-, etc.) refers to the match / mismatch of either the noun (Language: L) or the Picture (Picture: P) to the part of the sentence preceding the word that is underlined, with a minus sign indicating a mismatch. That is, in the correct condition (P+L+), both the word ‘bowl’ as well as the picture [BOWL] fit the preceding sentence context. In the Language mismatch condition (P+L-), the word ‘train’ does not fit the preceding sentence context, whereas the picture [BOWL] does fit. Conversely, in the Picture mismatch condition (P-L+) the picture [TRAIN] does not fit the preceding sentence context, whereas the word ‘bowl’ does fit. Finally, in the Double mismatch condition (P-L-) both the word ‘train’ and the picture [TRAIN] do not fit the preceding sentence context. All stimuli were in Dutch; the literal translation in English is ungrammatical.

comparisons were performed to test for differences between each mismatch condition and the Correct condition as well as between the Language mismatch condition and the Picture mismatch condition. Huynh-Feldt correction for violation of sphericity assumption was applied when appropriate (Huynh and Feldt 1976), but original degrees of freedom are reported (Table 5.2).

For the time-frequency analysis, the time-frequency representation (TFR) was computed for every trial using a multi-taper procedure (Mitra and Pesaran 1999). Low (4-40 Hz) and high frequencies (40-70 Hz) were analyzed separately. For the low frequencies a 500 ms sliding window with a single Hanning taper was used with no spectral smoothing. For the high frequencies, a 200 ms sliding window with three orthogonal Slepian tapers was used with 10 Hz spectral smoothing. Note that conversion into the frequency domain limited the maximal time-point of a segment which could be estimated at 750 ms (low frequencies) and 900 ms (high frequencies) after critical word onset. In the analysis, each segment was analyzed up to 750 ms after critical word onset. Average TFRs were computed by averaging single trial TFRs for every condition and subject separately.

Subsequently, statistical analysis involved repeated measures analysis of variance (ANOVA) on mean power in six a priori defined time-frequency windows with factors Condition (4 levels; Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels; Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2). The a priori defined frequency windows were based upon previous literature and included the theta band (4-6 Hz) (Bastiaansen et al. 2005; Hald et al. 2006; Davidson and Indefrey 2007), alpha band (8-12 Hz) (Rohm et al. 2001) and the lower gamma band (40-50 Hz) (Hagoort et al. 2004; Hald et al. 2006). Separate ANOVAs were conducted in early (0-300 ms) and late time windows (350-750 ms). In the case of a main effect of Condition or Condition x Electrode interaction,

subsequently, planned comparisons were performed to test for differences between each mismatch condition and the Correct condition as well as between the Language mismatch condition and the Picture mismatch condition. Huynh-Feldt correction for violation of sphericity assumption was applied when appropriate (Huynh and Feldt 1976), but original degrees of freedom are reported.

Moreover, to test for significant differences between conditions in time-frequency-electrode clusters that were outside of the a priori defined time-frequency windows, the data were analyzed using a cluster randomization procedure which identifies consistent changes between conditions in time-frequency-electrode clusters. First, single subject statistics were computed (two-sided t-test for the difference between two conditions for every electrode-time-frequency point). Consequently, group statistics involved a clustering procedure on the thresholded ( $t > 1.96$  and  $t < -1.96$ ) single subject statistics which identifies clusters of time-frequency-electrode points showing the same direction of effect (Maris 2004). To assess statistical significance of each cluster, the sum of all t-statistics in the cluster was computed. This was chosen as the cluster-level statistic (Maris and Oostenveld 2007). Second, significance of each cluster-level statistic was assessed by comparing the cluster statistic to its randomization distribution which was created by 2500 random re-assignments of the single-subject statistics and zero. That is, in each randomization the single-subject statistics were randomly re-assigned to zero or to their original value. The actual cluster statistic was then compared to the randomization distribution obtained and significance was assessed by evaluating the cluster statistic to the  $p < 0.05$  significance level. Note that the validity of the inference drawn is not dependent upon the exact statistic (in this case sum of all t-statistics) chosen (as is explained in more detail in Maris and Oostenveld, 2007). The choice for the sum of individual t-statistics is based upon theoretical / modelling considerations (Maris 2004; Maris and Oostenveld 2007) as well as by practice in other

studies employing a cluster-randomization procedure (e.g. Hald et al. 2006; Osipova et al. 2006; Davidson and Indefrey 2007; Medendorp et al. 2007; Tuladhar et al. 2007). This non-parametric procedure effectively controls the multiple comparisons problem introduced by the massive univariate approach taken (Maris and Oostenveld 2007).

We tested the hypothesis that some effects may have decreased / increased over the time course of the experiment (see Discussion section) in two ways. First, it was tested whether there was a linear correlation between effect size and item order in three time-frequency windows in which such correlation may have been expected (early time window (0-300 ms) in alpha frequency band (8-12 Hz) in Language mismatch - Correct condition and Double mismatch - Correct condition comparisons and early time window (0-300 ms) in the gamma frequency band (40-50 Hz) in the Picture mismatch - Correct condition comparison). Second, data from each of these three time-frequency windows were split into four equal parts according to their occurrence in the experiment and subjected to repeated measures analysis of variance (ANOVA) with factors Time (4 levels) and Electrode (27 levels).

## **Results**

*ERP results* Although the focus of this chapter is on the results obtained from the time-frequency analysis, we briefly report results of the ERP analysis to be able to compare the two measures. The ERP grand average waveforms (Fig.5.1) show a N1 followed by a P2, followed by a negativity resembling the N400. The latter negativity shows clear differences between the correct condition (black in Fig. 5.1) and the mismatch conditions (collared in Fig. 5.1), lasting from 300 ms to 750 ms after stimulus onset. Statistical analysis involved repeated measures analysis of variance (ANOVA) of the mean amplitude in the 300-600 ms time window with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and

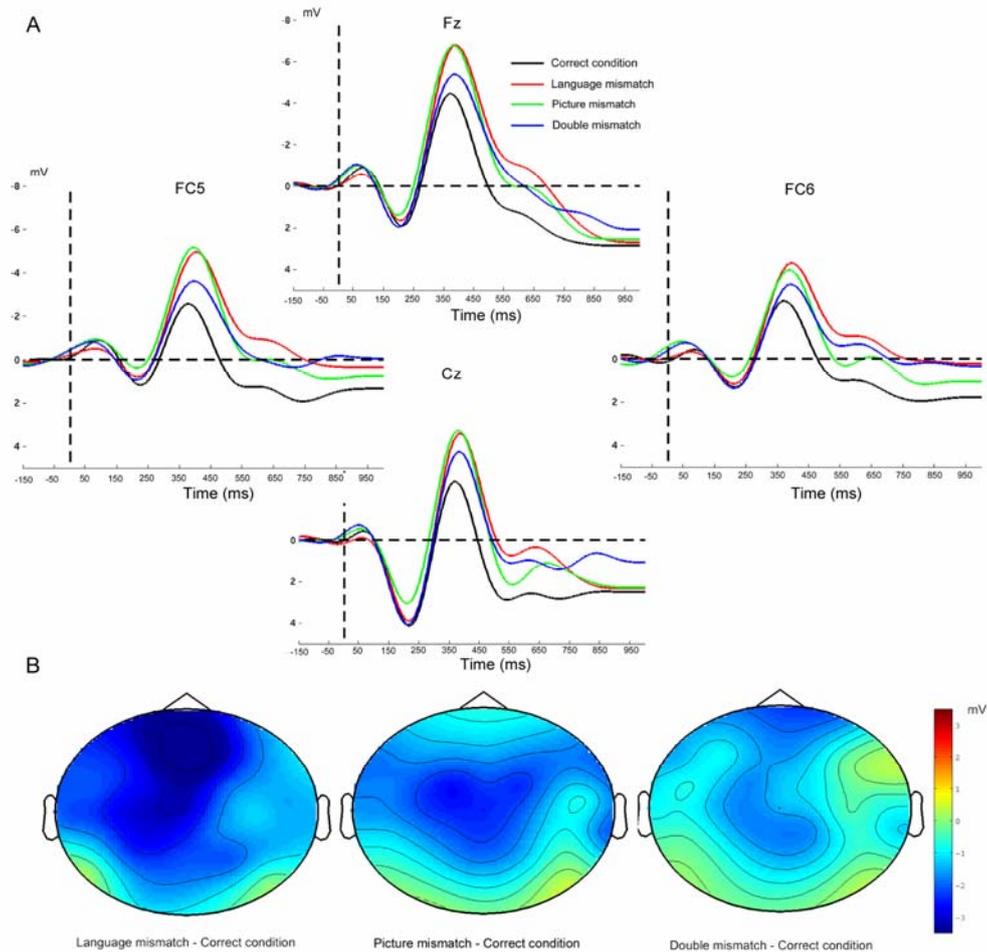
Electrode (27 scalp electrodes). There was a main effect of Condition ( $F(3,45)=6.17$ ,  $MSe=61.75$ ,  $p=0.002$ ), but no Condition x Electrode interaction ( $F(78, 1170)=1.09$ ,  $MSe=8.32$ ,  $p=0.364$ ). Planned comparisons showed that all mismatch conditions differed significantly from the Correct condition (Language mismatch vs. Correct condition:  $F(1,15)=12.98$ ,  $MSe=154.66$ ,  $p=0.003$ ; Picture mismatch vs. Correct condition:  $F(1,15)=5.61$ ,  $MSe=183.79$ ,  $p=0.032$ ; Double mismatch vs. Correct condition:  $F(1,15)=5.21$ ,  $MSe=114.92$ ,  $p=0.037$ ). The Picture mismatch and Language mismatch conditions were not significantly different from each other ( $F(1,15)=1.91$ ,  $MSe=84.52$ ,  $p=0.187$ ).

In all mismatch conditions the N400 effect was widely spread across the scalp (Fig. 5.1B), but with a more anterior distribution than is usually observed for the N400 effect to spoken or written words (e.g. Kutas and Hillyard 1980; Hagoort and Brown 2000; van den Brink et al. 2001). Such a relatively anterior distribution has been observed before in studies employing the N400 paradigm with pictures (e.g. Ganis et al. 1996; Federmeier and Kutas 2001).

### *Time-frequency results*

#### *Analysis with a priori defined time-frequency windows*

Statistical analysis involved repeated measures analyses of variance (ANOVA) on mean power in six pre-defined frequency-time windows of interest with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 scalp electrodes). The frequency bands tested were based upon previous literature. We tested for effects in the theta band (4-6 Hz) (Bastiaansen et al. 2005; Hald et al. 2006; Davidson and Indefrey 2007), alpha band (8-12 Hz) (Rohm et al. 2001) and the lower gamma band (40-50 Hz) (Hagoort et al. 2004; Hald et al. 2006). Separate ANOVAs were conducted in early (0-300 ms) and late (350-750 ms) time windows. The results are summarized in Table 5.3. In the case of a main effect of Condition or Condition x Electrode interaction, planned comparisons



**Fig. 5.1** (For colour version see Appendix, p. 268). **A)** Averaged event-related potentials time-locked to the onset of the critical word. Presented are the waveforms from electrodes FC5 (left), Fz (upper), FC6 (right) and Cz (lower) of all four conditions. The increased negativity of the collared lines (Mismatch conditions) as compared to the black line (Correct condition) is clearly visible. Negative is plotted upwards. Waveforms are low-pass filtered for illustration purposes only. **B)** Scalp topographies of the N400 effects in the 300-600 ms range for the Language mismatch-Correct condition (left), Picture mismatch-Correct condition (middle) and Double mismatch-Correct condition (right) comparisons. Note the more anterior distribution than is normally observed for the N400 effect elicited by spoken or written words.

<b>Factor</b>	<b>F</b>	<b>MSe</b>	<b>p</b>
Condition	F(3,45)=6.17	61.75	<b>0.002</b>
Electrode	F(26,390)=15.98	353.02	<b>&lt;0.001</b>
Condition x Electrode	F(78,1170)=1.09	8.32	0.364
<i>Planned comparisons</i>			
Language mismatch vs. Correct condition	F(1,15)=12.98	154.66	<b>0.003</b>
Picture mismatch vs. Correct condition	F(1,15)=5.61	183.79	<b>0.032</b>
Double mismatch vs. Correct condition	F(1,15)=5.21	114.92	<b>0.037</b>
Picture mismatch vs. Language mismatch	F(1,15)=1.91	84.52	0.187

**Table 5.2.** Results of the ERP analysis. Analysis involved repeated measures analysis of variance (ANOVA) on the mean amplitude in the 300-600 ms time-window with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels, see Experimental Methods section). Planned comparisons involved testing for differences between each mismatch condition and the Correct condition as well as between the Picture mismatch condition and the Language mismatch condition. Huynh-Feldt correction for violation of the sphericity assumption was applied (Huynh and Feldt 1976), but original degrees of freedom are reported.

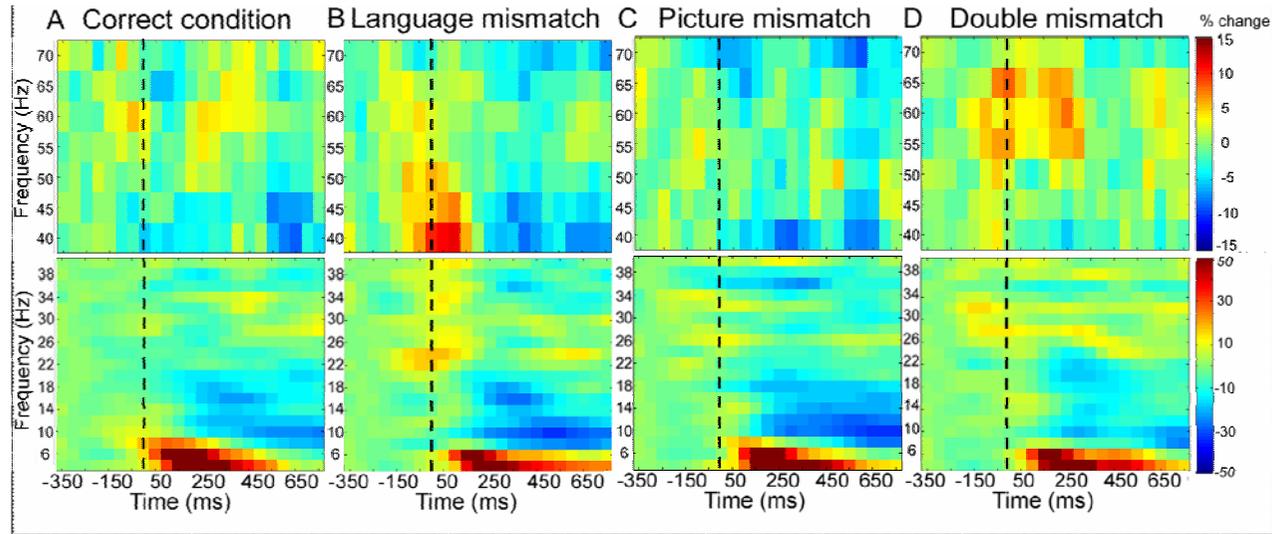
were performed to test for differences between each mismatch condition versus the Correct condition as well as between the Picture mismatch condition and the Language mismatch condition. The factor Electrode was always significant and is not reported in the text (see Table 5.3). In the early theta cluster (4-6 Hz, 0-300 ms) there was no significant main effect of Condition ( $F(3,45)=1.13$ ,  $MSe=82.39$ ,  $p=0.322$ ). The Condition x Electrode interaction was marginally significant ( $F(78,1170)=1.839$ ,  $MSe=10.89$ ,  $p=0.080$ ). Planned comparisons showed

a trend towards statistical significance for a power increase in the Double mismatch vs. Correct condition comparison only ( $F(1,15)=3.74$ ,  $MSe=34.47$ ,  $p=0.072$ ).

In the late theta cluster (4-6 Hz, 350-750 ms) there was a main effect of Condition ( $F(3,45)=4.70$ ,  $MSe=20.10$ ,  $p=0.017$ ) but no Condition x Electrode interaction ( $F(78,1170)=1.47$ ,  $MSe=104.43$ ,  $p=0.183$ ). Planned comparisons revealed that all mismatch conditions evoked significantly stronger power as compared to the correct condition (Language mismatch vs. Correct condition:  $F(1,15)=9.23$ ,  $MSe=29.99$ ,  $p=0.008$ ; Picture mismatch vs. Correct condition:  $F(1,15)=22.76$ ,  $MSe=11.93$ ,  $p<0.001$ ; Double mismatch vs. Correct condition:  $F(1,15)=9.44$ ,  $MSe=20.46$ ,  $p=0.008$ ). Effects were maximal over Frontal electrode sites for all comparisons (Fig. 5.3A, B and C). The Language mismatch and Picture mismatch condition did not differ significantly from each other ( $F<1$ ).

In the early alpha cluster (8-12 Hz, 0-300 ms), there was a marginally significant main effect of Condition ( $F(3,45)=2.90$ ,  $MSe=32.95$ ,  $p=0.071$ ) but no Condition x Electrode interaction ( $F<1$ ). Planned comparisons showed that the Double mismatch condition elicited significantly less alpha power as compared to the Correct condition ( $F(1,15)=4.54$ ,  $MSe=13.19$ ,  $p=0.050$ ). The effect was maximal over centro-posterior electrodes (Fig. 5.3C). The Language mismatch condition was not statistically different from the Correct condition ( $F(1,15)=3.28$ ,  $MSe=32.95$ ,  $p=0.093$ ), neither did the Picture mismatch condition differ from the Correct condition ( $F(1,15)=1.05$ ,  $MSe=18.80$ ,  $p=0.321$ ). Importantly, however, the Language mismatch evoked significantly less alpha power as the Picture mismatch condition ( $F(1,15)=5.08$ ,  $MSe=63.90$ ,  $p=0.040$ ). Again, the effect was maximal over centro-posterior electrodes (Fig. 5.3D).

In the late alpha cluster (8-12 Hz, 350-750 ms) there was no main effect of Condition ( $F(3, 45)=1.73$ ,  $MSe=34.60$ ,  $p=0.202$ ) or a Condition x Electrode interaction ( $F<1$ ).



**Fig. 5.2.** (For colour version see Appendix, p. 269). Time-frequency representations of the four conditions. Power is normalized with respect to power in the -350 to 0 ms time window in each frequency band by computing the relative change (percent signal change) as compared to the baseline condition for each frequency band separately. That is, baseline correction involved subtracting the mean of the baseline of that specific frequency band from the measured value and dividing this number by the mean power in the baseline ( $\text{value} - \text{baseline} / \text{baseline}$ ). Therefore, the values in the figure represent percentage power change as compared to baseline. It was made sure that no post-stimulus activation was included in the baseline period due to conversion into the frequency domain. Although instructive, this figure does not clearly illustrate the differences between conditions. These are displayed in Figure 5.3.

Frequency	Factor	Early (0-300 ms)				Late (350-750 ms)			
		F	MSe	df	p	F	MSe	p	
Theta (4-6 Hz)	Condition	1.13	82.39	3,45	0.322	4.70	20.10	<b>0.017</b>	
	Electrode	33.71	2902.45		<0.001	60.42	97.68	<b>&lt;0.001</b>	
	Cond x Electrode	1.839	10.89	78,1170	0.080	1.47	104.43	0.183	
	<i>Planned comparisons</i>								
	Language mismatch vs. Correct condition	2.21	87.86	1,15	0.158	9.23	29.99	<b>0.008</b>	
	Picture mismatch vs. Correct condition	<1		1,15	ns	22.76	11.93	<b>&lt;0.001</b>	
	Double mismatch vs. Correct condition	3.74	34.47	1,15	0.072	9.44	20.46	<b>0.008</b>	
	Picture mismatch vs. Language mismatch	<1		1,15	ns	<1		ns	
	Alpha (8-12 Hz)	Condition	2.90	32.95	3,45	0.071	1.73	34.60	0.202
		Electrode	13.78	139.92	26,390	<b>&lt;0.001</b>	15.49	85.82	<b>&lt;0.001</b>
Cond x Electrode		<1		78, 1170	ns	<1		ns	
<i>Planned comparisons</i>									
Language mismatch vs. Correct		3.28	32.95	1,15	0.093				
Picture mismatch vs. Correct condition		1.05	18.80	1,15	0.32				
Double mismatch vs. Correct condition		4.54	13.19	1,15	<b>0.050</b>				
Picture mismatch vs.	5.08	63.90	1,15	<b>0.040</b>					

Gamma (40-50 Hz)	Language mismatch Condition	3.63	0.28	3,45	<b>0.023</b>	1.07	0.27	0.365	
	Electrode	3.01	45.02	26,390	<b>0.013</b>	3.15	40.59	<b>0.008</b>	
	Cond x Electrode	<1		78, 1170	ns	1.32	0.22	0.200	
	<i>Planned comparisons</i>								
	Language mismatch vs. Correct	1.41	0.76	1,15	0.254				
	Picture mismatch vs. Correct condition	4.87	0.36	1,15	<b>0.043</b>				
	Double mismatch vs. Correct condition	<1		1,15	ns				
	Picture mismatch vs. Language mismatch	13.62	0.41	1,15	<b>0.002</b>				

**Table 5.3.** Results of time-frequency analysis in a priori defined time-frequency clusters. Analysis involved repeated measures analysis of variance (ANOVA) on mean power in six pre-defined frequency-time windows of interest, chosen on the basis of previous literature (i.e. Rohm et al. 2001; Hagoort et al. 2004; Bastiaansen et al. 2005; Hald et al. 2006; Davidson and Indefrey 2007). Data from the theta band (4-6 Hz), alpha band (8-12 Hz) and the lower gamma band (40-50 Hz) were analyzed in early (0-300 ms) and late (350-750 ms) time windows in ANOVAs with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels, see Experimental Methods section). Planned comparisons involved testing for differences between each mismatch condition and the Correct condition as well as a direct comparison between Picture mismatch and Language mismatch conditions. Huynh-Feldt correction for violation of the sphericity assumption was applied (Huynh and Feldt 1976), but original degrees of freedom are reported

In the early gamma cluster (40-50 Hz, 0-300 ms), there was a main effect of Condition ( $F(3,45)=3.63$ ,  $MSe=0.28$ ,  $p=0.023$ ) but no Condition x Electrode interaction ( $F<1$ ). Planned comparisons showed that only the Picture mismatch condition led to a significant decrease in power compared to the Correct condition ( $F(1,15)=4.87$ ,  $MSe=0.36$ ,  $p=0.043$ ). Moreover, the Picture mismatch condition was significantly different from the Language mismatch condition ( $F(1,15)=13.62$ ,  $MSe=0.41$ ,  $p=0.002$ ). The scalp distribution shows two maxima, one over left and one over right frontal electrodes, which is a rather unusual distribution (Fig. 5.3B). Inspection of the mean power differences of this comparison from the two electrodes in which the effect was maximal showed that the effect is consistent over participants in electrode FC6, and can therefore not be attributed to an artefact present in only some of the participants' data. However, the power differences at electrode T7 were much less consistent over participants and the effect is mostly due to one outlier in the data. It seems that the effect observed on electrode T7 in the Picture mismatch vs. Correct condition comparison (Fig. 5.3B) is best explained as an artefact that was not detected in the artefact detection procedure. Note that an ANOVA without the data from electrode T7 yielded similar results: a main effect of Condition ( $F(3,45)=3.76$ ,  $MSe=0.30$ ,  $p=0.026$ ) and a significant difference between Picture mismatch and Correct condition ( $F(1,15)=4.32$ ,  $MSe=0.323$ ,  $p=0.054$ ) as well as between Picture mismatch and Language mismatch ( $F(1,15)=14.00$ ,  $MSe=0.37$ ,  $p=0.002$ ) (Fig. 5.3D).

In the late gamma cluster (40-50 Hz, 350-750) there was no main effect of Condition ( $F(3,45)=1.07$ ,  $MSe=0.27$ ,  $p=0.365$ ) or a Condition x Electrode interaction ( $F(78,1140)=1.32$ ,  $MSe=0.22$ ,  $p=0.20$ ).

In summary, first, we observed late increases in theta power (4-6 Hz, 350-750 ms) in all mismatch conditions as compared to the Correct condition. Second, the Double mismatch condition differed significantly from the Correct condition in the early alpha cluster (8-12 Hz, 0-300 ms) and the Language mismatch versus Correct condition showed a

trend towards a stronger decrease in the Language mismatch condition. Importantly, a direct comparison between Language mismatch and Picture mismatch revealed that the decrease in alpha power in the early time window is significantly stronger in the Language mismatch condition as compared to the Picture mismatch condition. Finally, we observed a significant decrease in power in the early gamma cluster in the Picture mismatch condition as compared to the Correct condition as well as compared to the Language mismatch condition.

*Analysis without a priori defined time-frequency windows*

The analysis that we employed so far is restricted to the frequency bands in which we expected to find an effect based upon previous literature. To additionally test for time-frequency-electrode clusters that showed differences between conditions, but are not within these pre-defined frequency bands of interest, we employed an analysis which identifies time-frequency-electrode clusters which exhibit consistent changes between conditions. We did this by means of a non-parametric randomization procedure which is described in more detail in the Experimental Procedure (see also Maris 2004; Maris and Oostenveld, 2007). Such analysis has the potential to detect changes between conditions that go unnoticed in the analysis that we employed so far. Given the limited amount of previous research in this field of investigation, such additional effects may very well be present. A summary of the results of this analysis is provided in Table 5.4. In general, the results of this additional analysis corroborated the findings from the previous analysis. Two significant increases in power in the 4-6 Hz frequency band were observed. First, the Language mismatch vs. Correct condition led to a significant increase of power in the 4-6 Hz (theta) frequency range from 600 to 750 ms. Second, for the Double mismatch condition vs. Correct condition a similar increase in power in the 4-6 Hz (theta) frequency range from 550 to 750 ms was observed.

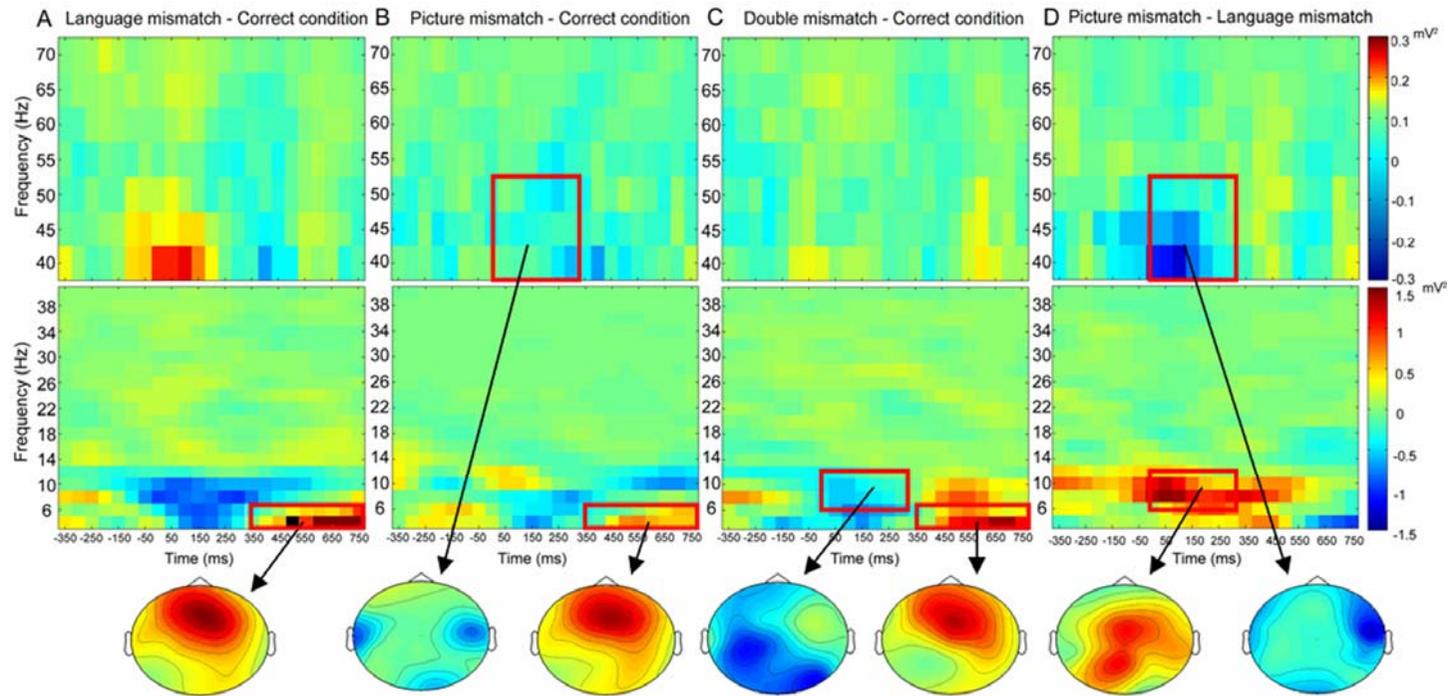


Fig. 5.3. (For colour version see Appendix, p. 270). Average time-frequency representations of **A**) Language mismatch-Correct condition, **B**) Picture mismatch-Correct condition, **C**) Double mismatch-Correct condition and **D**) Picture mismatch-Language mismatch condition. Time-frequency clusters (defined a priori based upon previous literature) in which the particular mismatch differed from the Correct condition are indicated with a red square. Displayed is the average power difference over all electrodes. Scalp topographies of significant differences between conditions are also displayed. Note the difference in scaling between lower and higher frequencies.

A significant decrease in power in the 6-10 Hz (alpha) frequency range was observed between 50 to 200 ms for the Double mismatch vs. Correct condition. A comparable decrease in the Language mismatch vs. Correct condition was not statistically significant ( $p=0.13$ ).

The comparison between the Picture mismatch condition and the Correct condition led to one significant decrease in power in the 40-65 Hz (gamma) frequency range from 0 to 250 ms.

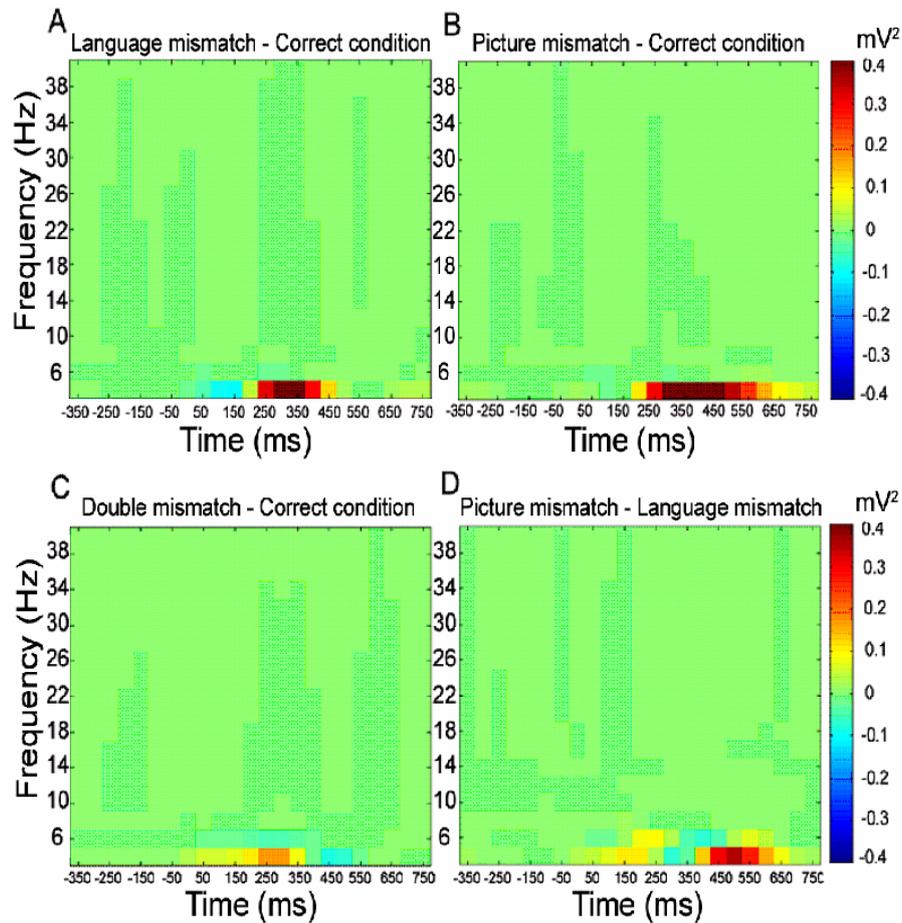
Finally, a direct comparison of the Picture mismatch condition to the Language mismatch condition revealed one significant cluster, involving a decrease in the 40-55 Hz frequency range from 0 to 250 ms to the Picture mismatch condition. No other differences were observed between the experimental conditions (Table 5.4).

Overall, these results are highly similar as compared to the analysis we performed with a priori defined time-frequency windows of interest. Importantly, no additional clusters of significant changes between conditions were observed.

As described above, we suspected that the early decreases in alpha / gamma frequency band would be modulated by the amount of prediction which develops over the time course of the experiment. We tested for a linear correlation between item order and effect size in the early (0-300 ms) gamma range (40-50 Hz) between Picture mismatch and the Correct condition, as well as in the early (0-300 ms) window in the alpha frequency band (8-12 Hz) comparing the Language mismatch versus Correct condition and the Double mismatch versus Correct condition. No statistically significant correlations between effect size and item position were observed (all  $p>0.25$ ), indicating that the effect size in the alpha and gamma clusters did not increase or decrease over the course of the experiment. Neither did analyzing the effect sizes in four separate time-bins reveal an effect of item order (see Experimental Methods for details).

<b>Comparison</b>	<b>Time (ms)</b>	<b>Freq (Hz)</b>	<b>Difference</b>	<b>p</b>	<b>Electrodes</b>
Language mismatch vs. Correct	600-750	4-6	Increase	0.016	F7, F3, Fz, F4, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2
Picture mismatch vs. Correct	0-250	40-65	Decrease	0.028	F3, Fz, F4, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, O1
Double mismatch vs. Correct	550-750	4-6	Increase	0.030	F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, C3, Cz, C4, T8, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2
	50-200	6-10	Decrease	0.043	F7, F3, Fz, FC5, FC1, FCz, T7, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, P8
Picture mismatch vs. Language mismatch	0-250	40-55	Decrease	0.021	F3, Fz, F4, F8, FC1, FCz, FC2, FC6, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, O1
Double mismatch vs. Language mismatch	-	-	-	-	
Double mismatch vs. Picture mismatch	-	-	-	-	

**Table 5.4.** Results of the time-frequency analysis with the cluster randomization analysis. This analysis identifies consistent differences between conditions over time-frequency-electrode clusters without the need to define time-frequency windows of interest a priori. Displayed are the contrasts that were tested and the significant time-window, frequency range and electrodes in which a cluster of increased or decreased activation was detected (see Experimental Procedure). ‘Decrease’ denotes a relative decrease in power, ‘increase’ denotes a relative increase in power.



**Fig. 5.4.** (For colour version see Appendix, p. 271). Time-frequency representation of the averaged Event-Related Potentials. Displayed are the TFRs of the difference waves of the **A)** Language mismatch-Correct condition, **B)** Picture mismatch-Correct condition, **C)** Double mismatch-Correct condition and the **D)** Picture mismatch-Language mismatch condition. The manifestation of the N400 as an increase in power around 4 Hz is clearly visible. TFRs were created by applying the same analysis procedure for the averaged ERP difference waves as used in the time-frequency analysis of the single trial data.

## **Discussion**

We investigated whether visual and linguistic information presented during language comprehension can be distinguished in terms of changes in power in specific frequency bands of the EEG signal. To increase semantic load of either word and / or picture during sentence comprehension we employed a modified version of the N400 paradigm in which either a word, a picture, or both word and picture could be semantically incongruous with respect to the preceding part of the sentence. Indeed, we found that early decreases in power in the alpha and gamma frequency bands were specifically related to linguistic or visual information load. Before we discuss these specific effects, it should be noted that we replicated earlier findings of similar N400 effects for all three experimental conditions. That is, qualitatively similar N400 effects were observed regardless of the format (linguistic or visual) in which incongruous semantic information was conveyed. For discussion of the theoretical implication of this finding we refer to two recent papers from our lab (Hagoort and van Berkum 2007; Willems et al. 2008b). An increase in theta power was observed for all mismatch conditions. It is tempting to interpret the theta power increase as an oscillatory counterpart of the N400 effect. Analogous to the hypothesized role of the N400, theta increases may reflect increases in semantic integration load. That is, the harder it is to integrate information into a context representation, the stronger the theta power increase. Indeed, conversion of the ERP difference waves into the time-frequency domain, shows an increase around 4 Hz, comparable to the theta increase observed in the frequency analysis (Fig. 5.4). This pattern suggests that an increase in theta power and the N400 effect may reflect at least partially overlapping cognitive processes. However, previous findings of theta power increases during language comprehension suggest that the role of theta power increases may be broader than just semantic processing. That is, although theta increases have been implicated in semantic processing (Hagoort et al.

2004; Bastiaansen et al. 2005; Hald et al. 2006), they are also reported for increased syntactic processing (Bastiaansen et al. 2002) and working memory operations in general (e.g. Klimesch 1999). This is not the case for the N400 effect which is closely linked to semantic processing (Kutas and Van Petten 1994; Brown et al. 2000). A viable possibility is that the theta increases observed here are a reflection of increased (working) memory load, as was suggested recently (Bastiaansen and Hagoort, 2006).

Specifically related to linguistic information, we observed an early (around 150 ms) decrease in alpha power in the Double mismatch condition. Moreover, a direct comparison between Language mismatch and Picture mismatch conditions showed that the decrease in alpha power was significantly stronger in the Language mismatch condition. However, this effect failed to reach significance in the Language mismatch condition versus Correct condition comparison. The alpha band has been claimed to be implicated in general levels of attention or vigilance (Klimesch 1999), in semantic processing in language tasks (Rohm et al. 2001) and in working memory processes (Jensen et al. 2002; Jokisch and Jensen 2007). It is however not directly evident how this previous literature can explain the decrease in alpha that we observed. General effects of attention / vigilance or working memory are unlikely to have been different in the Language and Double mismatch conditions compared to the Picture mismatch condition. Second, the occurrence of the effect seems to be too early to reflect a full semantic analysis of the critical word in connection to the preceding context. An alternative explanation that we entertain here is that a decrease in alpha power reflects an early detection of mismatch in the observed acoustic input based upon the preceding sentence context. That is, in the Language mismatch condition as well as in the Double mismatch condition the incoming acoustic information from the mismatching critical word is different from the 'correct' or matching critical word from the first phoneme on. It is hypothesized that a rapid,

context-based analysis of the acoustic information that was not in accordance with the preceding context is at the basis of the alpha power decrease. This effect is reminiscent of the N200 effect that has been reported when the onset of a spoken word differs from the word onset of words that form a congruent completion (Hagoort and Brown 2000; van den Brink et al. 2001).

Interestingly, Weiss and Rappelsberger (Weiss and Rappelsberger 1998) observed a less wide-spread alpha band desynchronisation in reaction to auditory presented words as compared to visually presented words. Similarly, Krause and colleagues observed less strong desynchronisation in the alpha frequency band to auditory presented words as compared to visually presented words (Krause et al. 2006);(see also Krause et al. 1997). In the light of the present findings, it is important to note that these studies suggest that hearing language leads to different effects in the alpha frequency band as compared to reading language. Auditory language leads to less alpha band power as compared to visually presented language. This may seem at odds with our present findings of a decrease in alpha band power for a mismatching word. However, it is possible that the detection of a mismatch in acoustic form interferes with the standard processing of an auditory presented word and leads to a power decrease.

Our explanation implies that the upcoming word is at least to some extent predicted. That is, detection of acoustic form which is not in accordance with the previous context, necessitates that another form was expected. Prediction has been shown to play a role in language comprehension and to influence the N400 effect (DeLong et al. 2005; Van Berkum et al. 2005). There is some previous literature which suggests that decrease in alpha band power is sensitive to effects of context on the processing of linguistic stimuli. For instance, Krause et al. (Krause et al. 1999) observed a larger decrease to repetition of the same word as compared to the presentation of two different words (see also Karrasch et al. 1998). Klimesch et al. observed strong effects of

expectancy of words and numbers on the amount of alpha band desynchronisation (Klimesch et al. 1990). That is, power decreases in alpha power were stronger when only words were presented as compared to when words and numbers were randomly intermixed (Klimesch et al. 1990). Although this result indicates that decreases in alpha power are related to expectancy of an item, it should be noted that all stimuli in Klimesch et al. were presented visually. That is, the relation to expectancy was not found to be specific for auditory stimuli, as was the case in our study. Finally, it has been found that the processing of a novel, unexpected stimulus leads to less synchronization in the alpha frequency band between cortical areas in the cat as compared to an expected stimulus (von Stein et al. 2000). Given these previous findings of an influence of expectancy / prediction on alpha power, we considered the effect of prediction in our data. In the present study, the cloze probability of the correct critical words (the amount of participants that filled in the critical word in a pretest, see Experimental Procedure) was low (16% on average), so that it is unlikely that expectation was high for the critical words. Moreover, there was no correlation between the effect size of Language mismatch minus Correct condition or of Double mismatch minus Correct condition in the alpha cluster with the order of items. Put differently, the size of the alpha decrease did not increase or decrease over the time course of the experiment. Such an effect might have been expected if participants are able to predict the upcoming critical word when they have become familiar with the set of critical words. However, an alternative explanation is that the language system does predict upcoming words, but that the expectation is not stable over individuals (as reflected in low cloze probabilities). Unfortunately, there is no way in which we can test this assertion in the present study.

An early decrease in the gamma frequency band was observed for the Picture mismatch condition, both in comparison to the Correct condition as compared to the Language mismatch condition. Therefore,

this seems to be a specific effect to increased semantic load for information conveyed visually, as through a picture. Analogously to the alpha decrease, we interpret this finding as reflecting an early detection of a mismatch between visual information from the picture and the preceding context. That this effect is manifested in the gamma frequency band is in line with a large body of literature showing successful object recognition to be associated with gamma power increases (see e.g. Rodriguez et al. 1999; Tallon-Baudry 2003). Our data extend the role of gamma band oscillations in visual object processing to be also related to early detection of a mismatch as compared to a preceding sentence context. That is, these oscillations are not only sensitive to the presentation of an object in isolation, but also to the semantic fit of that object with regard to a preceding (sentence) context. One may argue that the effect occurs too early (around 125 ms) to be a viable candidate of a context based visual form analysis. Especially the onset of the effect at 0 milliseconds (as determined in the cluster randomization analysis) is very unlikely. This is however a consequence of the necessary smearing over time as a result of the moving time-window used in conversion of the signal into a time-frequency representation. Moreover, early indices (<150 ms) of a rough semantic analysis based on visual form properties have been reported before in the ERP literature (Thorpe et al. 1996). That is, Thorpe and colleagues had subjects classify rich visual scenes (photographs) on the basis whether an animal was present in the scene or not. It was found that within 150 ms after picture onset, the ERP started to diverge based upon the presence or absence of an animal (Thorpe et al. 1996; VanRullen and Thorpe 2001). Related to this, it has been shown that the visual context in which a face is presented modulates the ERP response 170 ms after stimulus onset (Righart and de Gelder 2006). Summarized, there is evidence from previous ERP literature that recognition of objects in complex scenes as well as effects of context can

have an effect upon the ERP response within or around 150 ms after stimulus presentation.

Interestingly, a recent study reports a similarly early gamma decrease to an incongruous *word* (no pictures were presented) (Hald et al. 2006). Stimuli were written sentences, presented word-by-word, as opposed to the spoken Materials in the present study. In analogy to our explanation of the present gamma decrease, it is possible that an early, context-based visual analysis also occurs in the case of a visual word form. That is, in the case of written language, an early context-based detection of visual word form mismatch may have caused the early decrease in gamma band power in the study by Hald and colleagues. Such an explanation is indirectly supported by the fact that an early decrease in alpha band power that we linked to acoustic mismatch (above), was not observed in the Hald et al. study.

One reason to be cautious about the interpretation of the gamma decrease is that such decrease is not observed in the Double mismatch condition. If the early gamma decrease in the Picture mismatch condition reflects the context-based detection of an early mismatch of sentence context and visual form features of the picture, it is unclear why this effect is absent in the Double mismatch condition. After all, also in the Double mismatch condition the picture is not in accordance with the preceding sentence context. A different explanation is that instead of being a picture-specific effect, the gamma decrease reflects an early detection of conflicting information coming from co-occurring critical word and picture regardless of the previous context. Such an explanation is supported by the recent finding that incongruent sound and picture of an animal leads to reduced gamma frequency band activation as compared to matching sounds and pictures (Yuval-Greenberg and Deouell 2007). If this were the case, we would also expect a gamma decrease in the Language mismatch condition, in which co-occurring picture and word also convey different information. This was not observed however.

A recent hypothesis links gamma power increases to a successful match between stored representations in memory and incoming stimulus information (Herrmann et al. 2004b). For instance, Herrmann and colleagues observed a gamma (30-40 Hz) increase in response to pictures of existing objects as compared to nonsense objects, 90 ms after stimulus presentation (Herrmann et al. 2004a). From this it was concluded that the successful matching of the stimulus picture with the presence of a representation of the object in long-term memory is reflected in increased gamma power over occipital electrodes. Along these lines, it may be the case that in the present study the sentence context primed expectation of a certain picture (cf. Engel et al. 2001). In the case of the Picture mismatch condition, this expectation is violated which leads to a lack of increase in gamma power. This is in line with our explanation for the decrease in gamma power in the Picture mismatch condition. The fact that we did not observe differences in early gamma power to a mismatching word in the Language mismatch condition however seems to contradict the hypothesis that early gamma band increase is a reflection of matching of incoming stimulus information with a long-term representation *independent of input format* (Herrmann et al. 2004b; Lenz et al. 2007). Given our findings, this model could be extended in the sense that it seems that the format of the incoming information makes a crucial difference. After all, no effects in the gamma band were observed in the Language mismatch condition.

In conclusion, we hypothesize that the late increases in theta band power reflect general working memory processes related to the integration of semantic information into a representation of the sentence context. An early decrease in alpha band power most likely reflects early analysis of the acoustic input which was not in line with the acoustic input as expected from the preceding sentence context. Finally, an early decrease in gamma band power is tentatively explained as reflecting detection of a mismatch of visual form features

on the basis of the preceding context. With regard to our main question we conclude that semantic integration during sentence comprehension is neurally implemented in a similar way regardless of the format the input is in. Early differences in specific frequency bands however code for differences in the way the incoming information is conveyed. Crucially, these early differences are context-dependent in that they only occur when the item is semantically incongruous with the preceding (sentence) context. As such they are hypothesized to reflect an early, coarse semantic analysis based upon the acoustic or visual form of the input.

As a final note we want to point out that this study is one in a series in the neurocognition of language that shows the value of doing complementary analyses in the time-frequency domain next to more traditional ERP analysis. Qualitative differences in neural processing between visual and linguistic information that could not be detected in the grand average ERPs, were observed in induced changes in specific frequency bands. However, we acknowledge that firm interpretation of the results is hindered by the lack of a considerable amount of studies in this field of research. It will be a challenge for future EEG research on language to employ frequency analysis to find a tighter explanation of the role of oscillations during language comprehension.

### **Acknowledgements**

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO ‘Cognition’ program, grant no 051.02.040). Petra van Alphen is acknowledged for expertly voicing the sentences. We thank Marcel Bastiaansen for valuable comments on an earlier version of our manuscript, Heidi Koppenhagen and Niels Schiller for providing the naming consistency information of the line drawings and two anonymous reviewers for valuable comments on our manuscript.

## **Chapter 6** The neural integration of language and action information: Co-speech gestures versus pantomimes\*

### **Abstract**

Language and action are tightly related domains of cognition. How meaningful information from language and action is integrated in the brain is however ill-understood. Language and action information can be related to each other in radically different ways. For instance, the relationship between speech and co-speech gestures is very tight in the sense that they are produced together and that gestures are not unambiguously understood without speech. This is not the case for pantomimes (enactments of an action), which are not necessarily produced together with speech and whose meaning can be understood without speech. Here we looked at how this difference in relationship between language and action information results in differential effects on cortical areas involved in multimodal integration. We found that left posterior superior temporal sulcus was involved in semantic integration of speech and pantomimes, but not of speech and co-speech gestures. Inferior frontal gyrus on the other hand was involved in integration of information of both speech and co-speech gestures as well as for speech and pantomimes. Effective connectivity analyses showed that besides being involved in integrating multimodal information, LIFG can also play a modulatory role, influencing activation levels in pSTS. Again, this was dependent upon the relationship between action and language. Our data show that the way language and action are related to each other crucially influences the neural networks involved in their multimodal integration.

---

\*This Chapter is based upon Willems, R. M., Özyürek, A., & Hagoort, P. (under review). The neural integration of language and action: Co-speech gestures versus pantomimes

## Introduction

In recent years it has become increasingly clear that language and action are tightly related domains of cognition. At the neural level it has for instance been shown that listening to speech sounds (syllables) activates part of the cortical motor system (e.g. Wilson et al. 2004; Meister et al. 2007). Similarly, the understanding of action verbs activates those parts of premotor cortex that are also involved in execution of these actions (Hauk et al. 2004; e.g. Aziz-Zadeh et al. 2006; see Willems and Hagoort 2007 for review). An important but relatively understudied language-action relationship is that of spoken language and concurrently produced hand actions called *co-speech gestures*. Speech and co-speech gestures are tightly coupled during language production and both influence the understanding of a speaker's message (e.g. McNeill 1992; Goldin Meadow et al. 1999; Goldin Meadow and Momeni Sandhofer 1999; McNeill 2000; Goldin Meadow 2003; Kita and Özyürek 2003). As such they are hypothesised to be part of one underlying system for communication (e.g. McNeill 1992; Bernardis and Gentilucci 2006). Interestingly, the tight interrelatedness of speech and co-speech gestures is reflected in the fact that gestures are hard to interpret when presented without the speech they are spontaneously produced with. This is not true for all hand actions: for instance pantomimes (i.e. enactions or demonstrations of an action without using the object involved in performing the action) can be easily understood without accompanying spoken language.

In this study we investigated the neural integration of information expressed in language and in action. Specifically we wanted to assess whether the strong dependence between speech and co-speech gestures means that neural integration is different than for actions that can be reliably interpreted when presented without language. Previous literature suggests that this may indeed be the case. Before discussing

this, we will first briefly review behavioural evidence for a tight link between speech and co-speech gestures.

The general picture that emerges from behavioural studies on Speech-Gesture comprehension is that i) listeners do pick up information that is expressed in gesture and use it in their understanding of a speaker's message, and ii) that gestures do not reliably convey information when presented without speech. The first claim, that information from gestures is picked up and used by listeners has been repeatedly found (e.g. Singer and Goldin Meadow 2005). For instance, Graham and Argyle (1975) had speakers describe abstract line drawings with and without gestures, and required listeners to make drawings on the basis of the speakers' input. Listeners were more accurate in their drawings in the speech-and-gesture condition than in the speech-alone condition. Similarly, Beattie and Shovelton (1999) showed that listeners answer questions about the size and relative position of objects in a speaker's message more accurately when gestures are part of the description than when gestures are absent. Finally, one study has found that gestures that convey additional and even contradictory information influence later retelling of a story (McNeill et al. 1994).

Second, the tight relationship between speech and gestures is reflected in the fact that gestures when presented without speech do not reliably convey information. For instance, Feyereisen and colleagues (1988) had participants watch video recordings of lectures with or without audio. They found that participants most of the time did not ascribe a meaning which coincided with the speech that originally accompanied the gestures. In a similar vein, Krauss and colleagues (1991) presented gestures without speech to participants who were tested on their understanding of the meaning of a gesture by several measures. Participants had to choose words that had originally accompanied gestures, simply write down what they thought a gesture was about, assign gestures to semantic categories or indicate whether

they had seen a gesture before or not. Although performance was above chance in all measures that were used, it was far from perfect, even on the seemingly simple task of choosing between two words as to which one matches the observed gesture best (see also Beattie and Shovelton 2002).

In summary, behavioural evidence shows that co-speech gestures - simply 'gestures' from now on - do play a role in communication, but that they 'need' language to be understood unambiguously. In this sense they are different from more pantomimic actions that are produced with the intention to be understood without speech.

One reason to suspect that neural integration of speech and gestures may be different as compared to neural integration of speech and pantomimes is that in an earlier study we found that left inferior frontal gyrus (LIFG) was implicated in integration of information from co-speech gestures into a previous sentence context (Willems et al. 2007; see also Willems et al. 2008b). In contrast, several other studies point to (the posterior part of) superior temporal sulcus (pSTS) and middle temporal gyrus (MTG) as a crucial site for multimodal integration at the semantic level. For instance Beauchamp and colleagues (2004b) showed that integration of a picture of an object and its related sound occurs in pSTS / MTG (e.g. picture of a hammer with the sound of a hammer). Specifically, activation in pSTS / MTG was modulated by whether the picture and the sound were congruent to each other or not (e.g. the picture of a hammer with the sound of a telephone) (see e.g. Calvert 2001; Callan et al. 2003; Calvert and Campbell 2003; Beauchamp et al. 2004a; Callan et al. 2004; Calvert and Thesen 2004; van Atteveldt et al. 2004; Amedi et al. 2005; van Atteveldt et al. 2007 for other examples of pSTS / MTG involvement in multimodal integration). From these and other studies it has been hypothesised that pSTS integrates two streams of information by mapping them onto a common representation in long term memory (Amedi et al. 2005). An intriguing possibility for the difference between

these studies and our previous study could be the tight relationship between gestures and spoken language. That is, for co-speech gestures, no stable memory representation exist, whereas for the action depicted by a pantomime it does.

Here we set out to investigate whether integration of speech and co-speech gestures taxes different neural systems than the integration of speech with a type of action that is not tightly coupled to language and whose meaning is easily recognisable when presented in isolation. We did this by presenting participants with Speech-Gesture and Speech-Pantomime combinations in which speech and action content could either be in accordance or in discordance with each other. A semantic congruence manipulation is repeatedly employed in neurocognitive studies of language (e.g. Kutas and Hillyard 1980; van Berkum et al. 1999; Brown et al. 2000; Kuperberg et al. 2000; Baumgaertner et al. 2002; Friederici et al. 2003; Kuperberg et al. 2003; Hagoort et al. 2004; Kelly et al. 2004; Ruschemeyer et al. 2005; Kelly et al. 2006; Willems et al. 2007; Willems et al. 2008b) as well as in neurocognitive studies of multimodal integration (Beauchamp et al. 2004b; Hein et al. 2007; van Atteveldt et al. 2007; Fuhrmann Alpert et al. 2008). The rationale for having this congruence manipulation is that we will be able to assess whether a multimodal area is involved in integrating the two streams of information at the semantic level. An alternative approach is to look at responses of an area to unimodal presentation of two streams of information and contrast these with the response to bimodal presentation. However, if it is found that an area responds more strongly to bimodal presentation as compared to unimodal presentation it cannot be concluded that it integrates the two streams of information *at the level of meaning*. It is very well possible that the area is sensitive to input from two streams of information as compared to one stream of information. If however an area is involved in integration at the semantic level, it should be sensitive to the semantic congruence in bimodal presentations of the stimuli.

In short, we investigated the neural correlates of integration of concurrently presented language and action information. The relationship between language and action could be of two types: 1) speech and co-speech gestures and 2) speech and pantomimes. The crucial difference between the two is that speech and co-speech gestures are more strongly interrelated to each other in the sense that they are produced together and that the meaning of co-speech gestures cannot be unambiguously understood outside of a speech context. On the contrary, speech and pantomimes are not necessarily produced together and the meaning of a pantomime is easily understandable when observed without language. The congruence between speech and gesture and between speech and pantomime was manipulated so that we could investigate multimodal integration at the level of semantics. If integration of speech and co-speech gestures is indeed different from integration of speech and pantomimes we expect different neural correlates for both integration processes. Alternatively it may be the case that these information types are integrated with speech in the same brain areas. Two candidate regions to show an effect of semantic multimodal integration are LIFG and pSTS / MTG. In an effective connectivity analysis we investigated how these areas may influence each other during multimodal processing.

### **Materials and Methods**

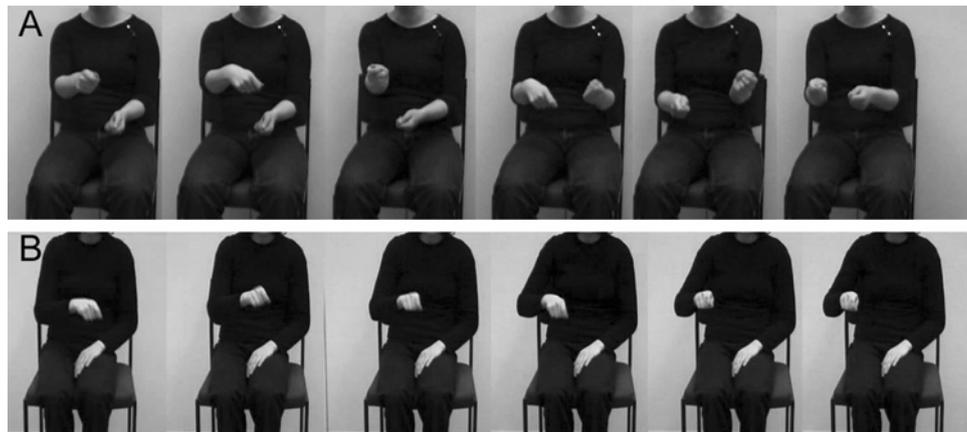
*Participants* Twenty healthy right-handed (Oldfield, 1971) participants without hearing complaints and with normal or corrected-to-normal vision took part in the experiment. None of the participants had any known neurological history. Data of four participants were not analysed because they did not perform significantly above chance level (see below). This means that data from 16 participants (11 female; mean age=22.3 years, range=19.3 - 27.4 years) were entered into the analysis. All participants were paid for participation and gave informed

consent prior to the experiment in accordance with the Declaration of Helsinki. The study was approved by the local ethics committee.

*Stimuli* Iconic gestures were taken from a natural retelling of cartoon movies by a female native speaker of Dutch (Fig. 6.1A). For the pantomimes we asked another female native speaker of Dutch to pantomime common actions (Fig. 6.1B). All videos were recorded in a sound-shielded room with a Sony TCR-TRV950 PAL camera. The actor's head was kept out of view to eliminate influences of lip or head movements. Short segments of Speech-Gesture combinations were cut from the overall retelling using Adobe Premier Pro software (version 7.0; <http://www.adobe.com>). All gesture segments contained one or more gestures with iconic content, such as referring to motion events (see Appendix at end of chapter for a literal transcription of the materials). The audio content of the pantomimes (i.e. spoken verbs) was re-recorded after recording of the video and were spoken by the same actor as in the videos. All audio content was band-pass filtered from 80 to 10500 Hz and equalised in sound level to 80 dB using 'Praat' software (version 4.3.16; <http://www.praat.org>).

There were two pretests of the materials, involving different participants than that participated in the fMRI experiment (for a summary of results see Table 6.1). In pretest 1, 25 gesture segments and 24 pantomimes were presented without speech. Participants (n=20) had to indicate what they thought was being depicted in the gesture / pantomime videos. Participants were encouraged to guess the meaning of the actions, but were allowed to fill in a question mark. On the basis of this pretest (results follow below), 14 co-speech gesture segments and 16 pantomimes were selected and presented together with speech in a second pretest. For this pretest, mismatching combinations of gestures and speech and pantomimes and speech were created by combining the audio of one gesture segment with video of

another gesture. This was also done for the pantomimes. Participants (n=16) indicated how well action and speech matched on a 1-5 scale.



**Fig. 6.1.** Examples of video content of the stimulus materials. **A)** Six stills of one of the gestures. This gesture is taken from a segment in which the speaker describes a character writing and drawing on a paper on a table. For exact speech see Appendix **B)** Six stills of one the pantomimes ('to write').

On the basis of the results of this second pretest the final set of stimuli was selected. This set contained 12 matching Speech-Gesture combinations and 12 matching Speech-Pantomime combinations as well as an equal amount of Speech-Gesture and Speech-Pantomime mismatches. The total amount of stimuli thus was 48 (4x12). The results from the two pretests for the final set of stimuli are described below and are summarised in Table 6.1. The meaning of the 12 co-speech gestures was not easily recognised without speech (results first pretest, mean percentage of participants (n=20) that indicated the correct meaning to a gesture: mean=8.8%, standard deviation (s.d.)=13.73%), and the original combinations of gesture and speech were scored as matching whereas the mismatching pairs were scored as mismatching (results second pretest: matching: mean=3.90, s.d.=0.64; mismatching: mean=1.74, s.d.=0.49, on a 1-5 scale). The meaning of the

12 pantomimes on the contrary was easily recognised without speech (first pretest, mean percentage of participants (n=20) that assigned the correct meaning to a pantomime: mean=88.4%, s.d.=14.7%). The matching combinations were consistently recognised as matching, whereas the mismatching combinations were not (matching: mean=4.95, s.d.=0.07; mismatching: mean=1.09, s.d.=0.13, on a 1-5 scale). The scores for matching and mismatching Speech-Gesture and Speech-Pantomime stimuli (second pretest) were not significantly different from each other ( $t(23)=-1.01$ ,  $p=0.33$ ).

Stimulus type	Pretest 1		Pretest 2	
	Mean (% participants)	s.d.	Mean (score)	s.d.
Pantomimes	88.4	14.7		
Gestures	8.8	13.7		
Pant-Speech match			4.95	0.07
Pant-Speech mismatch			1.09	0.13
Gest-Speech match			3.90	0.64
Gest-Speech mismatch			1.74	0.49

**Table 6.1.** Characteristics of stimuli. Displayed are the results of two pretests. Pretest 1 involved presenting pantomimes and gestures without speech and asking participants to indicate what they thought was depicted in the actions. Displayed is the mean percentage of participants that indicated the meaning that matched the meaning in the original speech fragment (left column). In pretest 2 matching and mismatching combinations of Speech-Pantomime and Speech-Gesture were presented. Participants had to indicate how well they thought audio and video fit together on a 1-5 point scale.

Mean duration of the stimuli was 2028 ms (s.d.=506; range=1166-3481) for the Pantomimes and 2209 ms (s.d.=400; range=1366-3182) for the Gestures. Note that in the analysis we did not directly compare

Pantomimes and Gesture given that these stimulus sets were not matched on basic characteristics such as duration.

From the materials that were not selected on the basis of the pretests, a set of filler items was created. These filler items were included to assess behaviourally whether participants were attending the stimuli (see below). There were 16 filler items (4 per condition: Pant-Match, Pant-Mismatch, Gest-Match, Gest-Mismatch).

*Experimental Procedure* Stimuli were presented using 'Presentation' software (version 10.2; <http://www.nbs.com>). The visual content was displayed from outside of the scanner room onto a mirror above the participant's eyes, mounted onto the head coil. The auditory content was presented through sound reducing MR-compatible head phones. The sound level was adjusted to the preference of each participant during a practice run in which ten items which were not used in the remainder of the experiment were presented while the scanner was switched on. All participants indicated they could hear the auditory stimuli well and none of the participants asked for the sound level to be increased to more than its half-maximum. Participants were instructed to attentively watch and listen to the items. After each filler item, a screen was presented with 'yes' and 'no' on either the left or the right side of the screen. Participants had to indicate whether they had observed that specific stimulus item before or not, by pressing a button with either the left or the right index finger. Response side was balanced over trials such that sometimes 'yes' could be indicated with the left index finger and sometimes with the right index finger. Participants had 2.5 seconds to respond and were instructed to respond as accurately as possible. Feedback was given after each response by appearance of the word 'correct', 'incorrect' or 'too late' on the screen.

There were three experimental runs: Audio only (AUDIO), video only (VIDEO) and audio and video (AV). In the AUDIO run, participants heard the short utterances or verbs from the gesture and

pantomime recordings without visual content on the screen. There were 12 pantomime and 12 gesture audio stimuli, which were all replicated three times, leading to 36 items for each condition (Gest-Audio, Pant-Audio). In the VIDEO run, participants saw the gestures and pantomimes presented without speech. Again, there were 12 gesture and 12 pantomime stimuli which were replicated three times leading to 36 items per conditions (Gest-Video, Pant-Video). In both the AUDIO and the VIDEO runs, eight filler stimuli were presented, four gestures and four pantomimes. Fillers were replicated two times, leading to a total of 16 filler items (8 gesture, 8 pantomime).

In the AV run participants saw the Speech-Gesture and Speech-Pantomime combinations, in matching and in mismatching versions. The conditions are labelled as Gest-Match, Gest-Mism, Pant-Match and Pant-Mism. The label 'Match' / 'Mism' refers to the match of gesture / pantomime with speech. The 12 matching and 12 mismatching combinations were replicated three times each, leading to 36 items per condition (Gest-Match, Gest-Mism, Pant-Match, Pant-Mism). The 4 matching and 4 mismatching filler items for both gestures and pantomimes were replicated two times, leading to a total of 32 filler trials (16 gesture, 16 pantomime).

Stimuli were presented in an event-related fashion, with an average intertrial interval (ITI) of 3.5 seconds. Onset of the stimuli was effectively jittered with respect to volume acquisition by varying the ITI between 2.5 and 4.5 seconds in steps of 250 ms (Dale 1999). The order of conditions was pseudo-randomised with the constraint that a condition did never occur three times in a row. Four different versions of stimulus lists were created which were evenly distributed over participants. The order of runs was varied across participants. The unimodal runs were included to test whether integration areas would also be activated during unimodal presentation of the stimuli (see Results section below).

*Image Acquisition* Data acquisition was performed using a Siemens ‘Trio’ MR-scanner with 3 Tesla magnetic field strength. Whole-brain echo-planar images (EPIs) were acquired using a standard bird-cage head coil with single pulse excitation with ascending slice order (TR=2130 ms, TE=30 ms, flip angle=80 degrees, 32 slices, slice thickness=3mm, 0.5 mm gap between slices, voxel size 3.5x3.5x3.5 mm). A high resolution T1 weighted scan was acquired for each subject after the functional runs using an MPRAGE sequence (192 slices, TR=2300 ms; TE=3.93 ms; FoV=256 mm; slice thickness=1 mm).

*Data analysis* Data were analysed using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>). Preprocessing involved discarding the first four volumes, correction of slice acquisition time to time of acquisition of the first slice, motion correction by means of rigid body registration along 3 rotations and 3 translations, normalisation to the standard MNI template, high-pass filtering (time constant of 128 s) and spatial smoothing with an 8 mm FWHM gaussian kernel. Statistical analysis was performed in the context of the General Linear Model (GLM) with regressors ‘Gestures’ and ‘Pantomimes’ in the AUDIO and VIDEO runs and regressors ‘Gest-Match’, ‘Gest-Mism’, ‘Pant-Match’, ‘Pant-Mism’ in the AV run. Responses (i.e. button presses), filler items and the realignment parameters from the motion correction were modelled as regressors of no interest. All regressors except for the motion parameters were convolved with a canonical two-gamma hemodynamic response function. Visualisation of statistical maps was done using MRICroN software (<http://www.sph.sc.edu/comd/rorden/mricron/>).

As explained in the introduction we had an a priori hypothesis that LIFG and pSTS / MTG would be involved in integration of action and language information. Therefore we created regions of interest (ROIs) in these areas. First, for LIFG we took the mean of the maxima from inferior frontal cortex from a recent meta-analysis of

neuroimaging studies of semantic language processing (Vigneau et al. 2006) (centre coordinate: MNI [-42 19 14]). Second, the ROI in pSTS was based upon the local maximum reported in the study of Beauchamp and colleagues (2004b), who found this area to be involved in multimodal integration of objects and their sounds (centre coordinate: MNI [-50 -55 7]). Regions of interest were spheres with an 8 mm radius. The activation levels of all voxels in a ROI were averaged for every subject separately and differences between conditions were assessed by means of dependent samples t-tests with  $df=15$ . We subsequently tested whether there was a relationship between the *degree* of congruence between speech and gesture or speech and pantomime and activation levels in these two ROIs. The scores from pretest 2 (in which participants had to indicate how well they thought action and speech were in accordance with each other, see Table 6.1) show that all Speech-Pantomime combinations were judged as clearly matching (mean on 1-5 point scale=4.95 , s.d.=0.07) or mismatching (mean=1.09, s.d.=0.13). However, in the Speech-Gesture pairs there was considerably more spread in these scores, both in the matching combinations (mean=3.90, s.d.=0.64) as well as in the mismatching combinations (mean=1.74, s.d.=0.49). Therefore we reasoned that by using a parametrically varying regressor based upon these scores, we would be able to pick up effects of Speech-Gesture congruence in a more sensitive way than in the analysis described above. For each stimulus item, the mean score (ranging from 1 to 5) from the pretest was taken and a parametric linearly varying regressor was constructed (Buchel et al. 1998).

To test for areas other than the ROIs that may be sensitive to integration of language and action information, a whole brain analysis was performed, by taking single subject contrast maps to the group level with factor ‘subjects’ as a random factor (random effects analysis). Statistical maps were controlled for multiple comparisons using a two-step procedure. First, statistical maps were thresholded at  $p=0.001$ ,

uncorrected at the voxel level. Second, cluster sizes were taken into account by to determine the chance of occurrence of a certain cluster size in the data by chance (Forman et al., 1995). Each map was corrected for this cluster size such that all clusters are reported at an alpha level of  $p < 0.05$  corrected. Anatomical localisation was done with reference to the atlas by Duvernoy (Duvernoy 1999).

Finally we investigated effective connectivity of LIFG and pSTS onto other cortical areas by means of whole-brain Psycho-Physiological Interactions (PPIs) (Friston et al. 1997; Friston 2002). A PPI reflects a change in the influence of one area onto other areas depending upon the experimental context. The time course of LIFG or pSTS was taken from the apriori defined ROIs (described above, LIFG: MNI [-42 19 14] and pSTS: MNI [-50 -55 7]). Consequently, we performed two analyses: one looking for effective connectivity of pSTS or LIFG with other areas modulated by Speech-Pantomime match / mismatch and one looking for modulations in connectivity between each of these two areas and other areas during Speech-Gesture match / mismatch. Time courses were deconvolved to allow for inferences at the neural level, as described by Gitelman and colleagues (Gitelman et al. 2003). Statistical maps were thresholded at  $p < 0.001$  at the voxel level and corrected for multiple comparisons by taking the cluster extent into account (Forman et al. 1995).

## **Results**

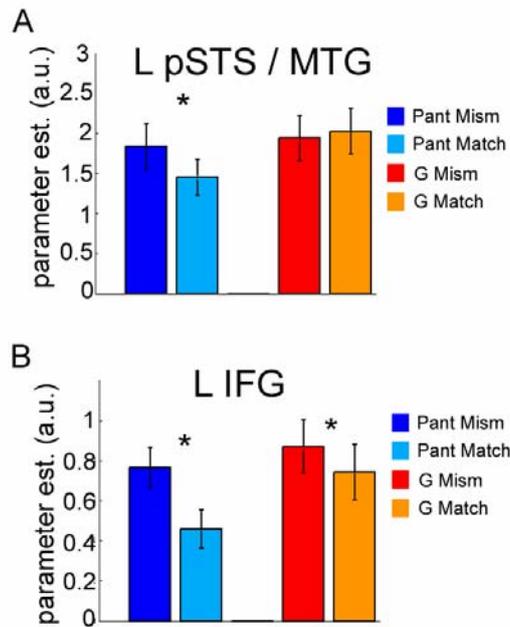
*Behavioural results* Four participants did not score above chance level to the filler items in at least one of the runs and were discarded from the analysis. Performance of the remaining 16 participants was well above chance level indicating that participants did pay attention to the stimuli (AUDIO: mean percentage correct=83.75, range=64.3 - 93.8, s.d.=9.26; VIDEO: mean percentage correct=77.21, range=62.5 - 92.3, s.d.=10.27; AV: mean percentage correct=75.42, range=62.1 - 87.5, s.d.=7.84).

*Region of interest analysis* Note that a *direct* comparison between pantomimes and gestures was not feasible given that the stimulus materials differed on important dimensions such as stimulus length and amount of movement. Therefore, we assessed differences between mismatching and matching combinations *within* each stimulus set (i.e. pantomimes and gestures), because the matching and mismatching items within each stimulus set were crossed and therefore perfectly matched in terms of low-level stimulus characteristics.

Region	Pant-Mism vs. Pant-Match		Gest-Mism vs. Gest-Match	
	t(15)	p	t(15)	p
Left pSTS / MTG	3.79	<0.001	<1	n.s.
LIFG	6.01	<0.001	1.75	0.050

**Table 6.2.** Results in a priori defined Regions of Interest. Left pSTS / MTG was only sensitive to congruence in Speech-Pantomime combinations, but not in Speech-Gesture combinations. However, LIFG was sensitive to congruence both in Speech-Pantomime combinations as well as in Speech-Gesture combinations.

In the ROI in pSTS / MTG, activation levels were significantly higher in the Pant-Mism as compared to Pant-Match condition ( $t(15)=3.79$ ,  $p<0.001$ ) (Fig. 6.2A, Table 6.2). No such effect was observed for Speech-Gesture combinations (Gest-Mism vs. Gest-Match:  $t(15)<1$ ). However, in LIFG, activation levels were higher both for Pant-Mismatch as compared to Pant-Match conditions ( $t(15)=6.01$ ,  $p<0.001$ ) as well as for Gest-Mism as compared to Gest-Match conditions ( $t(15)=1.75$ ,  $p=0.050$ ) (Fig. 6.2B, Table 6.2). Testing the degree of congruence (based upon the results of pretest 2 in which participants had to indicate how well Speech-Pantomime or Speech-Gesture combinations matched), showed a similar pattern of results. There was an effect of degree of congruence



**Fig. 6.2.** (For colour version see Appendix, p. 272). Results in Regions of Interest. Mean parameter estimates of all bimodal conditions in left pSTS / MTG (**A**) and LIFG (**B**), averaged over all voxels in the ROI. **A**) In left pSTS / MTG there was a difference between mismatching and matching Pantomime-Speech combinations (mismatch: dark blue, match: light blue), but not between mismatching and matching Gesture-Speech combinations (mismatch: red; match: orange). On the contrary, in LIFG, there was an influence of congruence both in the Speech-Pantomime combinations as well as in the Speech-Gesture combinations (**B**). A.u.: arbitrary units.

for both Speech-Gesture as well as Speech-Pantomime combinations in LIFG, but only for the Speech-Pantomime combinations in L pSTS / MTG (see Supplementary Table S6.1). This confirms the sensitivity of these areas to the congruence between speech and pantomime and / or speech and gesture. It also rules out the possibility that the fact that we did not find an effect of G-Mism versus G-Match in left pSTS is due to the larger spread of congruence scores in the Speech-Gesture combinations. That is, also when taking the spread in these scores into account, activation in left pSTS was not correlated to the degree of congruence between speech and gesture ( $t < 1$ , see Supplementary Table S6.1).

Region	T(max)	Coordinates (MNI)		
		x	y	z
<b>Pant-Mism versus Pant-Match</b>				
L Posterior STS / MTG	4.72	-56	-46	6
	4.59	-56	-64	2
R Posterior STS	5.94	62	-32	4
L Inferior Frontal Gyrus	11.33	-40	10	22
L Intraparietal Sulcus	7.88	-34	-54	46
L Insula	5.30	-42	24	-2
R Insula	4.55	40	24	4
L Cingulate Sulcus	10.27	-8	10	58
R Cingulate Sulcus	5.93	8	20	48
<b>Gest-Mism vs. Gest-Match</b>				
-	-	-	-	-

**Table 6.3.** Results of whole brain analysis comparing Pant-Mism versus Pant-Match and Gest-Mism versus Gest-Match. Displayed are an anatomical description of the region, the T-value of the maximally activated voxel in the region and the centre coordinates of the region in MNI space.

Both ROIs were activated above baseline in all unimodal conditions (Supplementary Table S6.2).

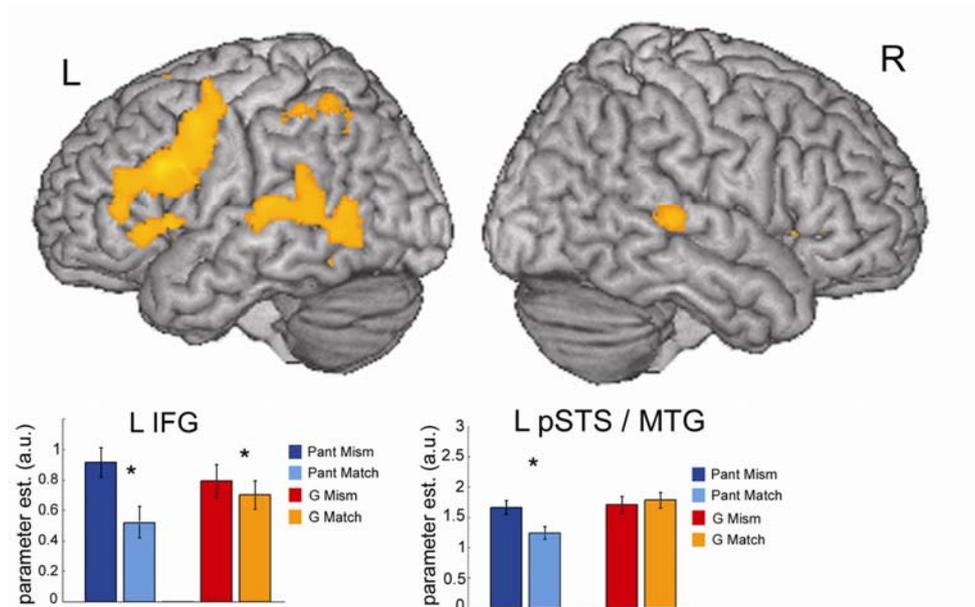
*Whole brain analysis* To test for areas other than the ROIs that may be sensitive to integration of language and action information, a whole brain analysis was performed. Contrasting Pant-Mism with Pant-Match led to a network of activations encompassing left and right pSTS / MTG, LIFG, left intraparietal sulcus, bilateral insula and bilateral cingulate sulcus (Fig. 6.3 and Table 6.3).

There were no areas which survived the statistical threshold to the Gest-Mism versus Gest-Match comparison. However, informal inspection at a lower, uncorrected threshold ( $p < 0.01$  uncorrected) showed increased activation in LIFG, but not in pSTS, as would be expected from the ROI analysis. To test whether the areas activated in the whole brain contrast comparing Pant-Mism versus Pant-Match were specific to the pantomime-speech combinations, we compared Gest-Mism versus Gest-Match as well as Pant-Mism with Pant-Match in two activation clusters, namely in LIFG (MNI [-40 10 22]) and in L pSTS (MNI [-56 -46 6]). The results (Fig. 6.3) confirm the analysis with a priori defined ROIs. In LIFG, Pant-Mism differed from Pant-Match as well as Gest-Mism from Gest-Match ( $t(15)=6.83$ ,  $p < 0.001$  and  $t(15)=1.66$ ,  $p=0.053$ , respectively). In pSTS however, only Pant-Mism and Pant-Match differed from each other, but not Gest-Mism versus Gest-Match ( $t(15)=7.13$ ,  $p=0.001$  and  $t(15)=-1.16$ ,  $p=0.26$ ).

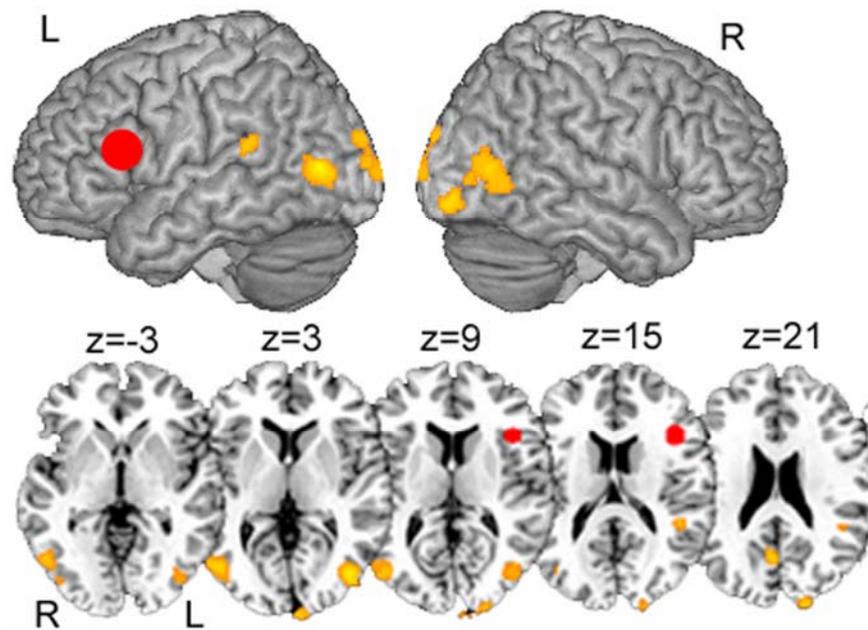
*Effective connectivity analysis* The PPI analysis with the time course from LIFG showed that effective connectivity from this region is increased in the pantomime-speech mismatch condition as compared to the pantomime-speech condition in left pSTS and bilateral lateral occipital sulcus (Fig. 6.4 and Table 6.4). The area in left pSTS overlaps with the cluster in this area that was found to be activated in the main contrast, reported above (see Supplementary Fig. S6.1). On the contrary, no areas showed effective connectivity with LIFG as a

function of the gesture-speech mismatch condition as compared to the speech-gesture match condition<sup>1</sup>. The PPI analysis in which the time course from left pSTS / MTG was taken showed that connectivity from this region is increased in the Pant-Mism condition as compared to Pant-Match condition in left middle occipital gyrus and right superior frontal sulcus (Fig. 6.5 and Table 6.5). No areas showed effective connectivity with left pSTS / MTG as a function of the Gest-Mism condition as compared to the Gest-Match condition.

In summary, in the ROI analysis we found that pSTS / MTG was sensitive to congruence of pantomimes and speech, but not of gestures and speech, whereas LIFG was sensitive to congruence in both Speech-Pantomime and Speech-Gesture combinations. Testing the parametrically varying degree of perceived congruence between Speech-Pantomime and Speech-Gesture combinations confirmed that these areas were also sensitive to the degree of congruence. This rules out the alternative explanation that we did not observe an effect of Speech-Gesture combinations in pSTS / MTG due to the greater spread of congruence in these stimuli as compared to Speech-Pantomime combinations (as was observed in the pretest). In the whole brain analysis these findings were replicated. We found a network of areas more strongly activated to incongruent Pantomime-Speech combinations as compared to congruent Speech-Pantomime combinations including left pSTS / MTG and LIFG. Of these areas, left pSTS / MTG was exclusively sensitive to Pantomime-Speech mismatch and not to Gesture-Speech mismatch (Fig. 6.3). However, LIFG was sensitive to both Pantomime-Speech mismatch as well as to Gesture-Speech mismatch (Fig. 6.3). Finally, we found that LIFG has stronger effective connectivity with pSTS during Pant-Mism condition as compared to Pant-Match condition. Such an influence of IFG onto pSTS was not observed for the Gest-Mism condition as compared to the Gest-Match condition.



**Fig. 6.3.** (For colour version see Appendix, p. 272). Areas activated in whole brain analysis to the Pant-Mism versus Pant-Match contrast. Map is thresholded at  $p < 0.05$ , corrected for multiple comparisons, and overlain on a rendered brain. Activation levels (parameter estimates in arbitrary units (a.u.)) of the clusters of activation in LIFG and left pSTS are displayed. Analysis in these clusters confirms the results from the analysis with a priori defined regions of interest: in pSTS there only is a difference between Pant-Mism and Pant-Match, but in LIFG there is a significant difference between Pant-Mism and Pant-Match as well as between Gest-Mism and Gest-Match.



**Fig. 6.4.** (For colour version see Appendix, p. 273). Results of effective connectivity analysis taking the a priori defined region of interest in LIFG as seed region, indicated in red. Statistical maps are thresholded at  $p < 0.05$ , corrected for multiple comparisons and overlain on a rendered brain. Areas that are more strongly modulated by LIFG in the Pant-Mism condition as compared to the Pant-Match condition. The rendered image is possibly misleading since it displays activations at the surface of the cortex that are actually 'hidden' in sulci. Therefore, we also display the result on multiple coronal slices. In the latter view, localisation of the activation in pSTS is more straightforward. No areas were found to be more strongly modulated by LIFG in the Gest-Mism as compared to Gest-Match condition.

Contrast	Region	T(max)	Coordinates		
			x	y	z
Pant-Mism vs. Pant-Match	L Posterior STS	3.73	-46	-42	14
	L Lateral Occ. Sulcus	8.63	-44	-73	9
	R Lateral Occ. Sulcus	3.84	42	-71	14
Gest-Mism vs. Gest-match	-	-	-	-	-

**Table 6.4.** Results of effective connectivity analysis with time course from the a priori defined ROI in LIFG as seed region. The table displays regions that were influenced by LIFG, depending upon whether the condition was Pant-Mism versus Pant-Match. An area in left pSTS overlapping with the area found in the main contrast in the whole brain analysis was found to be modulated by LIFG. No areas were influenced by LIFG depending upon whether the condition was Gest-Mism versus Gest-Match.

Contrast	Region	T(max)	Coordinates		
			x	y	z
Pant-Mism vs. Pant-Match	L Middle Occ. Gyrus	7.20	-20	-100	16
	R Superior Frontal Sulcus	5.19	18	20	46
Gest-Mism vs. Gest-match	L Inferior Frontal Gyrus	4.93	-54	32	20

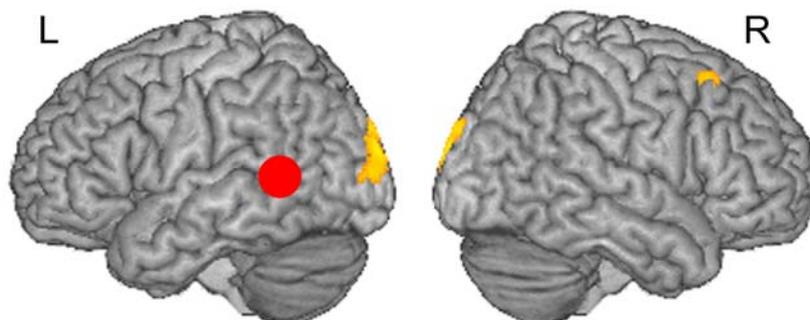
**Table 6.5.** Results of effective connectivity analysis with time course from the a priori defined ROI in left pSTS / MTG as seed region. The table displays regions that were influenced by left pSTS / MTG, depending upon whether the condition was Pant-Mism versus Pant-Match. No areas were influenced by left pSTS / MTG depending upon whether the condition was Gest-Mism versus Gest-Match.

## **Discussion**

In this study we set out to investigate the neural correlates of integration of action and language information. Two types of action-language combinations were investigated: speech combined with co-speech gestures and speech combined with pantomimes. Spoken language and co-speech gestures are strongly and intrinsically related to each other in the sense that they are produced together and that gestures cannot be unambiguously recognised or understood when they are presented without speech (e.g. Riseborough 1981; Feyereisen et al. 1988; Krauss et al. 1991; McNeill 1992; Beattie and Shovelton 2002; Goldin Meadow 2003; Kita and Özyürek 2003; Kendon 2004). This is not the case for pantomimes, which are not necessarily produced together with speech and are produced to be understood without speech. We found that this difference in relationship between language and action is reflected in different neural correlates involved in integration of the two information types.

Specifically, we found that the posterior part of left posterior superior temporal sulcus (pSTS) and middle temporal gyrus (MTG) is only sensitive to congruence of simultaneously presented speech and pantomimes, but not to simultaneously presented speech and co-speech gestures. On the contrary, left inferior frontal gyrus (LIFG) was modulated by congruence of both Speech-Gesture as well as Speech-Pantomime combinations. Below we will speculate as to what this pattern of results reveals about the respective roles of pSTS and LIFG in multi-modal integration.

Posterior STS / MTG has been implicated in multimodal integration in a multitude of studies, for instance investigating integration of phonemes and lip movements (e.g. Calvert et al. 2000; Calvert 2001; Callan et al. 2003; Callan et al. 2004), phonemes and written letters (van Atteveldt et al. 2004; van Atteveldt et al. 2007), pictures of objects and their related sounds (Beauchamp et al. 2004b; Taylor et al. 2006) and pictures of animals and their sounds



**Fig. 6.5.** (For colour version see Appendix, p. 274). Results of effective connectivity analysis taking the a priori defined region of interest in left pSTS / MTG as seed region. The ROI is displayed in red. Statistical maps are thresholded at  $p < 0.05$ , corrected for multiple comparisons and overlain on a rendered brain. Left middle occipital gyrus and right superior frontal gyrus were more strongly modulated by left pSTS in the Pant-Mism condition as compared to the Pant-Match condition. No areas were found to be more strongly modulated by left pSTS / MTG in the Gest-Mism as compared to Gest-Match condition.

(Hein et al. 2007). Here we show that this area is also involved in integration of information from actions (pantomimes) and verbs that describe the pantomime. Together with these previous findings this suggests that pSTS / MTG is also involved in integration at the semantic level, besides integration at the form level as indicated by for instance studies investigating speech sounds and lip movements.

A recent neuroimaging study indicates that also LIFG is involved in semantic multimodal integration. LIFG was found to be sensitive to the amount of semantic congruity of simultaneously presented picture of an animal and the sound of the animal (Hein et al. 2007). Moreover, in a large amount of language studies LIFG has been repeatedly found to be involved in semantic processing. This is true both when integrating linguistic information into a wider context (e.g. Friederici et al. 2003; Kuperberg et al. 2003; Hagoort et al. 2004; Rodd et al. 2005;

Ruschemeyer et al. 2005; Davis et al. 2007) as well as when in integrating non-linguistic information with language information (Hagoort et al. 2004; Willems et al. 2007; Willems et al. 2008b).

Here we replicate our earlier finding of LIFG as being involved in integration of speech and co-speech gestures (Willems et al. 2007). We extend this previous finding by showing that LIFG is not only involved in integration of information with respect to a relatively rich semantic (sentence) context. In our previous study, the semantic ‘fit’ of words or co-speech gestures was constrained by a preceding sentence context. This was not the case in the present study in which speech consisted of single words or of short segments of speech, providing little contextual information. Still, LIFG was modulated by the semantic congruence between the two streams of information.

What partially distinct roles do pSTS and LIFG play in multimodal integration? Neuroimaging literature suggests that pSTS plays its role in multimodal integration by means of matching the content of two information streams onto a representation of for example the object in long-term memory (Amedi et al. 2005). This explains why we find modulation of pSTS only to (in)congruence of speech and pantomimes and not for speech and co-speech gestures. That is, the content of both the words and the pantomimes can be mapped onto a relatively stable representation of that specific action (word) in memory. This is crucially not the case for co-speech gestures. This is a neural reflection of the strong links between speech and co-speech gestures. The dependency of gestures on accompanying language necessitates that semantic integration happens only at a higher level of processing than for input streams that can be mapped onto a representation lower in the cortical hierarchy. Our findings show that LIFG and not pSTS is involved in such higher level integration. Converging evidence for this comes from a study in which it was found that LIFG (but not pSTS) was involved in integration of novel associations of non-existing objects and sounds (Hein et al. 2007). On

the contrary, both LIFG and pSTS were found to be sensitive to semantic congruence of animal pictures and their sounds (Hein et al. 2007)

The effective connectivity results illuminate the interplay between LIFG and pSTS during multimodal integration. That is, it seems that in reaction to a mismatching speech-pantomime combination LIFG modulates activation levels in areas lower in the cortical hierarchy, most notably pSTS and an area in the vicinity of previously reported Extrastriate Body Area (EBA) (Peelen et al. 2006). This modulatory function of IFG has been hypothesised by others (Gazzaley and D'Esposito 2007) and is in line with the proposed function of this area in regulatory functions such as semantic selection / control (Thompson-Schill et al. 1997; Badre et al. 2005; Thompson-Schill et al. 2005). Put differently, in this scenario, LIFG and pSTS work together to integrate multimodal information, with a modulatory role of LIFG and a more integrative role for pSTS. On the contrary, when integration is impossible in pSTS, as was the case in the Speech-Gesture combinations, there is no such modulatory signal from LIFG to pSTS. So it seems that in this case LIFG is involved in integration of information from action and language. This scenario fits nicely with a recent report in which it was shown that during integration of the picture of an animal and its sound, activation in pSTS precedes activation in LIFG in time (Fuhrmann Alpert et al. 2008). Our findings are suggestive of the possibility that LIFG can subsequently modulate pSTS.

It is perhaps misleading to draw a sharp distinction between modulation on the one hand and integration on the other hand. Hagoort (2005b; 2005a) has characterised IFG's function as *unification*, which crucially implies both modulation of areas lower in the cortical hierarchy (e.g. through semantic selection) as well as integration of information into e.g. a sentence context (see Hagoort 2005a for

discussion). Our present findings seem to be in line with such an account.

An interesting difference between this and some other multimodal studies is that activation levels increase in response to mismatching stimulus combinations (see also Hein et al. 2007). On the contrary, some multimodal integration studies report activation increases to *matching* stimulus combinations. For instance, Van Atteveldt and colleagues (2007) observed higher activation level in left pSTS in response to a matching phoneme and letter combination (e.g. letter 'p' with phoneme [p]) as compared to a mismatching combination (e.g. letter 'k' with phoneme [p]) (see also Calvert et al. 2000 for the integration of lip movements and speech sounds). The same is true in the study by Beauchamp et al. (2004b) who found higher activation in left pSTS to the matching combination of a picture of an object and its sound versus a mismatching combination. Note however that our finding of the opposite is very common in studies of for instance sentence comprehension that modulate the (semantic) integration load of a word into a preceding sentence context (e.g. Bookheimer 2002; Friederici et al. 2003; Kuperberg et al. 2003; Hagoort et al. 2004; Rodd et al. 2005; Ruschemeyer et al. 2005; Davis et al. 2007; Willems et al. 2007; Willems et al. 2008b; Menenti et al. under review). An intriguing but speculative explanation is that the presence of language stimuli at and beyond the word level creates this difference. Moreover, it seems that task factors could be of crucial influence. Future research should investigate these possibilities in a more systematic way.

A recent study showed that gesture-speech combinations evoke increased activation in pSTS as compared to combinations of speech and so called self-adaptors such as scratching and touching the body (Holle et al 2008). It should be noted that in the present study left pSTS / MTG was also activated above baseline in the two Speech-Gesture conditions (Fig. 6.3B). However, activation in this area was not modulated by the semantic congruence of speech and gesture. In the

control condition in the Holle et al. study, speech was accompanied by meaningless touches of the body. The hand movements in such a scenario are very low in informational content. A critical difference between our stimuli and the stimuli used in Holle et al., is that in our case there always were two streams of meaningful information, whereas in their control stimuli the hand actions were essentially meaningless with respect to task of understanding the message of the speaker. Therefore, the results of the two studies seem to be in line with each other in the sense that gesture plus speech leads to increased activation in left pSTS / MTG compared to baseline (present study) or compared to speech plus self-adaptors that are very low on semantic information (Holle et al. 2008). We however conclude that pSTS is not involved in semantic integration of speech and gesture, because it is not sensitive to the match in content between speech and gestures (see also Willems et al. 2007).

A possible criticism to our study is the use of a mismatch paradigm. It may be argued that our data show involvement of several areas in the detection of a mismatch between speech and action instead of involvement of these areas in the integration of action information and speech. The mismatch paradigm is widely used in the neurocognition of language and is believed to successfully increase integration load of an item into a previous context (Kutas and Van Petten 1994; Brown et al. 2000; Özyürek et al. 2007). Also studies of multimodal integration have repeatedly and successfully employed a mismatch paradigm (Beauchamp et al. 2004b; Ojanen et al. 2005; Pekkola et al. 2006; Hein et al. 2007; van Atteveldt et al. 2007). More importantly, there are fMRI studies which show that similar neural networks show increased activation levels in paradigms which manipulate semantic integration load without using a mismatch paradigm (Rodd et al. 2005; Davis et al. 2007). Moreover, the two ROIs were also activated above baseline during presentation of the matching Speech-Gesture and Speech-Pantomime combinations (Fig. 6.3B).

In summary, we have shown that areas known to be involved in multimodal integration are also involved in integration of language and action information. Importantly, the relationship between language and action information crucially changes the neural networks involved in integration of the two information types. Our data shed light upon the roles played by pSTS and LIFG during multimodal integration. It seems that whereas pSTS integrates information from input for which a relatively stable representation in long-term memory is available, LIFG integrates information from action and language at a higher level as well as can modulate activation in pSTS when integration is more difficult.

#### **Notes**

1) Neither were any areas found to be modulated at an uncorrected statistical threshold of  $p < 0.01$ . Please note that a direct statistical comparison between effective connectivity from IFG in the pant-speech mismatch versus pant-speech match contrast as compared to the speech-gesture mismatch versus speech-gesture match contrast showed a similar result. This shows that connectivity of LIFG with pSTS and the other areas reported in the text were found depending upon the match or mismatch in speech-pantomime combinations, but not for match or mismatch in speech-gesture combinations.

#### **Acknowledgments**

Supported by a grant from the Netherlands Organization for Scientific Research (NWO), 051.02.040 and by the European Union Joint-Action Science and Technology Project (IST-FP6-003747). We thank Cathelijne Tesink and Nina Davids for help in creation of the stimuli and Caroline Ott for help at various stages of the project. Paul Gaalman is acknowledged for his expert assistance during the scanning sessions.

## Appendix Chapter 6

Transcription of speech segments used in the experiment. Below is a transcription of the verbs (used in the Pantomime-Speech combinations) and the short speech segments (used in the Gesture-Speech combinations) that were used in the experiment. Next to each transcription in Dutch is a translation in English.

### Pantomimes

Dutch	English
typen	to type
schudden	to shake
schrijven	to write
scheuren	to tear
roeren	to stir
kloppen	to knock
iets opendraaien	unscrew?
iets intoetsen	to type in something
iets inschenken	to pour
grijpen	to grasp
gewichtheffen	to lift weight
breken	to break

### Gestures

Dutch	English
en ... dan komt 'ie aanlopen	and then he walks in
dan loopt 'ie snel weg	then he quickly walks away
en .. valt 'ie weer terug naar beneden	and .. he falls down again

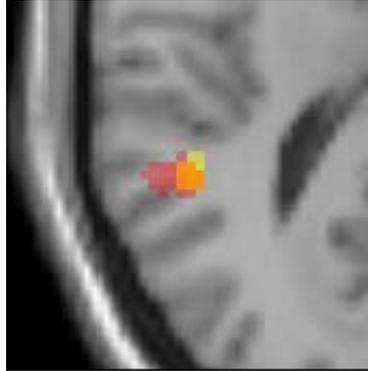
hij zwaait tegen de muur aan	he swings into the wall
eh die komt eh binnenlopen	uh he uh comes and walks in
is hij eh heel druk aan het schrijven en aan het rekenen	he is uh very busy writing and calculating
dan gaan ze elkaar achterna zitten	then they go and chase each other
loopt onder aan de regenpijp op en neer	walks from one side to the other
en die gaat naar beneden en die rolt er zo naar binnen	and he goes down and he rolls in
en de ene die smijt 'ie weg	and the one he throws away
hij staat er vrolijk aan te draaien	he is happily turning it

Region	Speech-Pant congruence		Speech-Gest congruence	
	t(15)	p	t(15)	p
Left pSTS / MTG	3.89	<0.001	<1	n.s.
LIFG	5.58	<0.001	2.39	<b>0.015</b>

**Supplementary Table S6.1.** Response of ROIs to the analysis taking the *degree* of congruence between speech and pantomime or speech and gesture into account. The degree of congruence was assessed in a separate pretest and was taken as a parametrically varying regressor into the model.

Region	AUDIO-only				VIDEO-only			
	Pant		Gest		Pant		Gest	
	t(15)	p	t(15)	p	t(15)	p	t(15)	p
Left pSTS / MTG	5.05	<0.001	4.16	<0.001	2.70	<b>0.006</b>	3.24	<b>0.001</b>
LIFG	3.03	<b>0.002</b>	3.85	<0.001	3.31	<b>0.001</b>	4.14	<0.001

**Supplementary Table S6.2.** Response of ROIs during unimodal presentation of the stimuli. Both left pSTS / MTG and LIFG are activated above baseline during unimodal presentation of the stimuli.



**Supplementary Fig. S6.1.** (For colour version see Appendix, p. 274). Visualisation of the overlap in pSTS of activation in the main contrast Pant-Mism versus Pant-Match (Red) and influence of LIFG onto pSTS as revealed in the effective connectivity analysis (Yellow).



## **Chapter 7 Embodied action understanding in the motor system: Evidence from left- and right-handers\***

### **Abstract**

What is the role of our own motor system in understanding the actions of others? Neural simulation theory states that an observed action is implicitly simulated, using the observer's own motor system. Indeed, empirical findings show that the specific make-up of the observer's motor system influences neural correlates of action understanding. However, in the understanding of common activities, too strong a coupling between motor production and action understanding might be detrimental. Rather, action meaning may be abstracted away from an individual's motor practice. Here we used handedness to critically test the nature of cortical motor activation in understanding common actions. Our results show that although the motor cortex is involved in gleaning meaning from actions, its activation is not influenced by the observer's hand preference. This shows that neural simulation to common actions occurs in terms of the meaning or goal of the action. Embodied cognition - of which neural simulation is an instantiation - allows for a flexible relationship between an individual's motor system and the neural processing of action meaning. We conclude that action understanding involves our own motor system, but not necessarily in the form of a one-to-one mapping between motor production repertoire and neural correlates of action observation.

---

\*This Chapter is a slightly modified version of Willems, R. M., Özyürek A., de Lange F. P., & Hagoort, P. (under review). Embodied action understanding in the motor system: Evidence from left- and right-handers

## **Introduction**

What is the role of our own motor system in understanding the actions of others? Neural simulation theory asserts that we use our own motor system to understand the meaning of an observed action. That is, to understand another person's action we implicitly simulate that action using motor structures of the brain. Evidence from neuroimaging largely supports the notion of neural motor simulation (see Rizzolatti et al. 2001; Gallese et al. 2004; Rizzolatti and Craighero 2004; Iacoboni and Dapretto 2006 for review). Many studies report activation increases in areas of the cortical motor system (most notably ventral premotor and inferior parietal cortex) both during observation and production of actions (e.g. Fadiga et al. 1995; Grafton et al. 1996; Hari et al. 1998; Nishitani and Hari 2000; Buccino et al. 2001; Rizzolatti et al. 2001; Rizzolatti and Craighero 2004; Caetano et al. 2007). However, neural simulation implies that the coupling between action production and action observation may be stronger than 'just' common motoric activation during production and observation. Neural simulation theory predicts that the neural correlates of action observation are influenced by the motor production system of the observer.

Previous neuroimaging studies have provided evidence for an influence of motor expertise on action observation, mainly by comparing neural activation in motor experts versus non-experts. For instance, it was found that dancers show stronger activation in the cortical motor system when they observe a type of dance they are familiar with (Calvo-Merino et al. 2005; Calvo-Merino et al. 2006). Interestingly, this was both the case in comparison with non-experts (controls that were not dancers) as well as compared to expert dancers who were trained in another type of dance (specifically, classical ballet versus capoeira) (Calvo-Merino et al. 2005; see also Calvo-Merino et al. 2006; Cross et al. 2006); (see also Reithler et al. 2007).

These studies support a strong influence of action production experience on action observation. However, it seems that for the

understanding of many common actions too strong a coupling between the observer's motor repertoire and action observation could be detrimental. For example, understanding that someone is lacing her shoe should not depend upon the exact technique with which the observer usually laces his shoes. Rather it seems that the neural coding of the meaning or goal of such common activities could be more flexible, in the sense that it is not strictly tied to the observer's motor production preference.

In this study we set out to answer at which level the meaning of common actions is represented in the brain. On the one hand it may be the case that simulation occurs at a level which is strictly tied to the observer's motor production preference. As a result neural simulation will be different for left- and right-handers. On the other hand it may be that simulation takes place at the level of the goal or meaning of an observed action, in a way that is relatively independent of the motor production preference of the observer. If this is the case, left- and right-handers may use their motor system in a similar way to understand the meaning of an action, despite their difference in hand preference. We aimed to distinguish between these possibilities by measuring neural activation in left- and right-handers who watched depictions of common activities. Actions were observed as performed with either the left or the right hand. Participants observed the actions during passive viewing as well as with a task in which they had to indicate the meaning of the observed action. By means of this design we were able to separate effects of hand preference (between-subjects factor), the hand that was observed performing the action (within-subjects factor) and the influence of the goal or intention with which the actions were observed (within-subjects factor).

We had three main questions. First, we tested whether parts of the neural motor system are modulated by action understanding. If so, one would expect increased neural activation in areas of the motor system when observing meaningful actions as compared to meaningless

actions. Although there is some neuroimaging evidence which suggests that premotor and inferior parietal cortex are activated in response to the understanding of action *words* (Hauk et al. 2004; Noppeney et al. 2005; Aziz-Zadeh et al. 2006), Decety and colleagues did not find increased activation in premotor or parietal cortex, when they compared meaningful actions (pantomimes) to meaningless actions (sign language signs) (Decety et al. 1997). To resolve these conflicting findings, we directly tested whether premotor and parietal cortex are modulated by whether an action has a meaning or not. If so, this would support the idea of these areas as not passively reacting to the observation of an action, but are actively involved in coding the meaning of the action, in analogy to what has been suggested for action words (e.g. Gallese and Lakoff 2005; Pulvermuller 2005).

Second, we tested the *nature* of motor cortex activation during action observation, by exploring whether hand preference of the observer influences neural correlates of action understanding. If action understanding is characterized by a strict neural coupling between motor repertoire and neural correlates of action understanding, we expect lateralization of neural activity in the parts of the motor system that are involved in understanding the meaning of an action. Alternatively, if action understanding is less tightly coupled to the observer's motor preference, there should be no different lateralization of motor cortex activation between the groups. Note that several studies report that producing hand actions such as writing leads to strongest activation in the premotor and primary motor areas contralateral to the hand that performs the action, both in left- and in right-handers (Siebner et al. 2002; Longcamp et al. 2003, 2005; Longcamp et al. 2006; Harrington et al. 2007). Moreover, a neuroimaging study with healthy right-handed participants indicates that execution of a pantomime with the left or the right hand leads to strongest increases in premotor and primary motor cortex contralateral to the hand that performs the pantomime (Johnson-Frey et al. 2005).

Therefore, our rationale that simulating an observed action onto the preferred hand should lead to strongest activation in the contralateral hemisphere seems valid, at least for premotor and primary motor cortex.

Third, we tested whether activation of the cortical motor system in coding the meaning of an action is automatic or not. It has been suggested that motor cortex activation in reaction to an action concept is automatic, and occurs even when an action word takes on a metaphorical meaning (Gallese and Lakoff 2005). A recent study however questions this assertion (Aziz-Zadeh et al. 2006). It was found that reading sentences describing actions performed with different effectors (such as foot and hand) led to activation of the premotor cortex in a somatotopic manner. However, phrases in which action verbs were used in a metaphorical manner (such as ‘chewing over the details’) did not lead to such an effect. We have argued before that this may mean that an action concept does not automatically activate the cortical motor system, but that this is dependent upon the context the action word occurs in (Willems and Hagoort 2007). In the present study, participants watched the depictions of common activities under two task conditions: passive viewing and a task in which they had to indicate whether they thought the action was meaningful or not. If motor cortex activation during action understanding is automatic, we expect no effect of this task manipulation on areas of the action observation system. However, if involvement of the action system is dependent upon the goal or intention of the observer, we expect to see an effect of meaning when participants have to actively process the meaning of the actions, but not when they are passively viewing the actions.

In summary, in this study we set out to investigate the role of the observer’s motor system in understanding meaning from observed actions. First, we assessed whether parts of the cortical motor system are involved in coding the meaning of an action. Second, we tested

whether left- and right-handed individuals map an observed action onto their own preferred hand or whether there is a more flexible relationship between an action's concept and the observer's production preference. Finally, we assessed the influence of the intention or goal with which the actions are observed by varying the task from passive viewing to actively encoding the meaning of the observed action.

### **Materials and Methods**

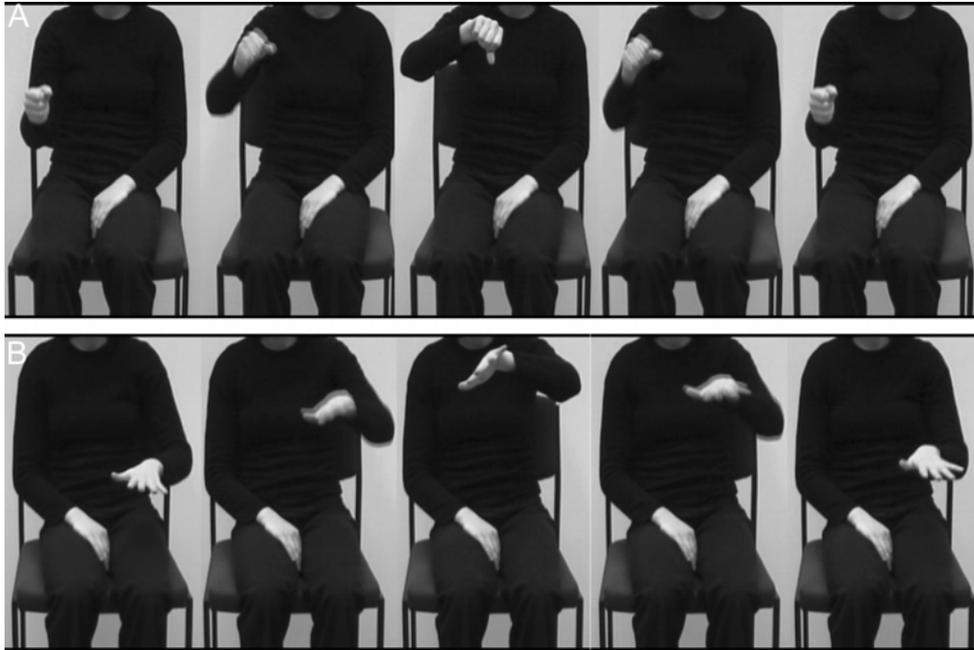
*Participants* Thirty-two healthy participants with Dutch as their mother tongue took part in the study. Handedness was assessed by means of a translated and adapted version of the Edinburgh handedness inventory (Oldfield 1971). Half of the participants were left-handed (N=16, 9 female, mean age 25.5 years, range 19-42; handedness score: mean=-88, median=-100, modus=-100, range -33 to -100) and half of the participants were right-handed (N=16, 9 female, mean age 22.2 years, range 18-29; handedness score: mean=82, median=85, modus=100, range 33 to 100). A between groups comparison of the absolute values of the handedness score indicated that the groups did not differ in terms of degree of handedness ( $t(30)=0.99$ ,  $p=0.33$ ). Besides the standard questions of the Edinburgh inventory, we also included questions about hand preference for the actions that were depicted in the stimuli. All participants indicated for all these actions that they used their dominant hand to perform the action. None of the participants had a known neurological history and all had normal or corrected-to-normal vision. Participants gave written informed consent in accordance with the declaration of Helsinki. The participants were paid for participation. The study was approved by the local ethics committee.

After completion of data collection, we were worried that the degree to which our participants had a preference to use their dominant hand to perform the actions would differ between actions. It seems conceivable that whereas for instance most people have a very

strong preference to write with their dominant hand, the degree of hand preference may be more variable for some of the other actions that we used as stimuli. We therefore post-tested the degree of hand preference for all actions in our participant group by asking participants to again indicate their preferred hand for each action and to additionally score how often they used the *non-preferred* hand on a 1-5 scale ranging from 'almost as often with the other (non-preferred) hand' to 'almost never with the other (non-preferred) hand'. We got these ratings for 27 out of 32 participants, 15 left-handers and 12 right-handers. There was some spread in the degree of hand preference between the actions in both groups: left-handers: mean=3.72, s.d.=0.61, range 2.73 - 4.93; right-handers: mean=4.00, s.d.=0.70, range 2.33 - 5.00. The groups did not differ in their degree of hand preference, however ( $t(25)=-1.03$ ,  $p=0.31$ ). These scores were used as parametrically varying linear regressor in the fMRI analysis (see below).

*Materials* Stimuli were short movie clips of an actress acting out (pantomiming) simple actions or their meaningless counterparts (Fig. 7.1). Meaningless actions were created by changing the hand shape of the pantomime (Fig. 7.1). Meaningless actions were matched to the pantomimes in terms of duration and direction and overall amount of movement. Video clips were selected out of a large sample of pantomimes and meaningless actions on the basis of a separate pretest. In this pretest, a different set of participants than the ones that participated in the fMRI study ( $N=16$ ) indicated how meaningful they thought an action was on a 1-5 scale and what they thought the actor was depicting in the action. On the basis of the results of the pretest, fourteen different hand actions and their meaningless counterparts were selected (see Appendix with this chapter). The scores on the 1-5 scale were significantly higher in the meaningful as compared to the meaningless actions (mean meaningful=3.94, mean meaningless=2.01,  $t(13)=11.30$ ,  $p<0.0001$ ). The correct meaning was given to the meaningful actions in on average 89.4 % of the responses (range 75%-

100%, modus=100%).



**Fig. 7.1.** Example of the stimuli. Five stills are shown from the right-hand version of **A)** the meaningful action ‘pour’ and **B)** the left-hand version of its meaningless counterpart. The meaningless action was rendered meaningless by using a different hand shape. In terms of duration and overall direction of movement the two video clips were equal. Note that the meaningless action is the ‘mirrored’ version of the original, i.e., it appears as if the action is performed with the left hand, whereas in reality it was performed with the right hand. In the experiment, of every action there was a version as if presented with the left hand and a version as if presented with the right hand.

Only in on average 2.2% of the responses (range 0-12.5% modus=0%) the meaning of the meaningful action on which the meaningless action was based, was given to a meaningless action. Mean duration of the video clips was 1761 ms (range 633 – 3248 ms). The head was kept out of view to avoid influences of face and / or lip movements. Two actions were performed bimanually, the other actions were performed unimanually. The bimanual actions were included to have a higher

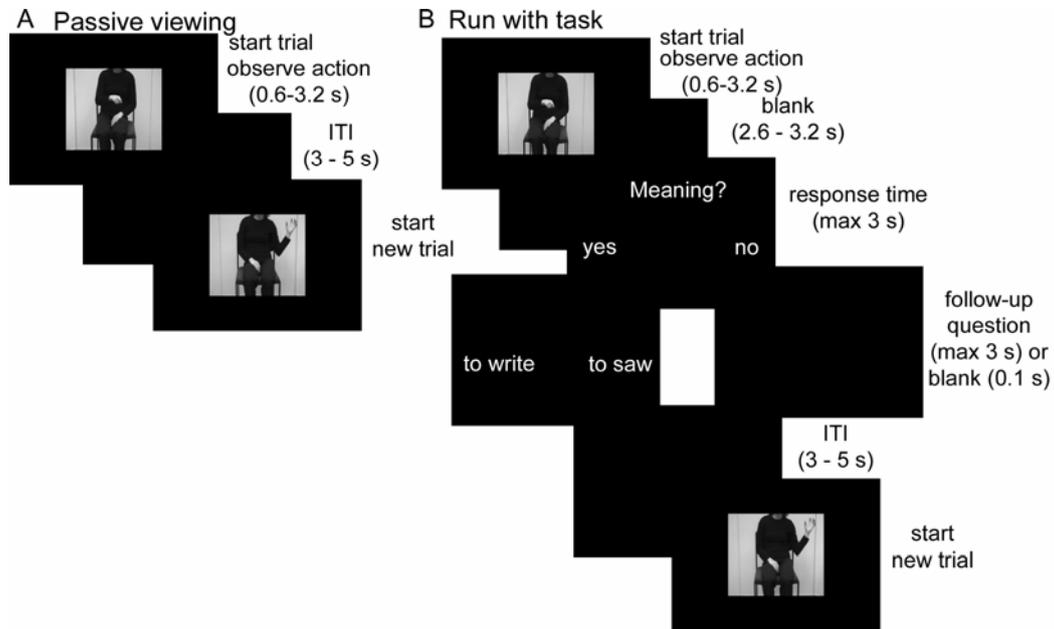
degree of variation in the materials and to increase power for the Meaningful versus Meaningless comparison. The bimanual actions were however modelled separately in the data analysis, since there was no dominant hand in these movie clips as opposed to the unimanual actions. The original video clips were mirrored along the vertical axis to create a video clip of the same action, but as if performed with the other hand. That is, the actions were performed with the right (dominant) hand in the original recording, but after mirroring it looked as if the actress used her left hand to perform the action (Fig. 7.1). There were 56 stimuli (14 movies x 2 (meaningful / meaningless) x 2 (original / mirrored)=56).

*Experimental design and Procedure* Movie clips were projected onto a screen from outside of the scanner room and were visible to the participant through a mirror mounted onto the head coil. Stimuli subtended 13 x 10 cm at a viewing distance of 80 cm (9.3° x 7.2° visual angle).

The experiment consisted of two separate runs. A total of 168 trials (three repetitions of all stimuli) were presented in each run. In the first run, which lasted 16 minutes, participants were instructed to attentively watch someone performing actions. They were told they would receive a recognition test at the end of the run. Intertrial interval was effectively jittered across trials and ranged between 3 to 5 seconds in steps of 250 ms in random order (Fig. 7.2A). In the second run, which lasted 28 minutes, participants observed the same actions but now had to indicate whether they thought the action was meaningful by pressing a button with the left or the right index finger after every action (Fig. 7.2B). On average 2900 ms (see below) after every action movie participants saw a screen with 'yes' or 'no' on the left or right side of the screen, with 'yes' indicating meaningful and 'no' indicating meaningless. The mapping between response hand (left or right) and 'yes' or 'no' response was randomized so that participants

could not predict which button they had to press before they saw the response alternatives on the screen. Before the start of the experiment participants were instructed that some of the actions had a clear meaning, whereas others did not. If participants indicated the action to be meaningful, they got a forced-choice alternative between two meanings which they had to choose from by again pressing a button with the left or the right index finger (Fig. 7.2B). Intertrial interval was effectively jittered across trials and ranged between 3 to 5 seconds in steps of 250 ms (randomly distributed) plus the additional time between end of the stimulus and start of the response period, which was 2600 to 3200 ms, in steps of 200 ms (randomly distributed). Participants had 3 seconds to respond. They were told that no speeded response was needed, but that response time was limited. Stimulus presentation and response periods were separated in order to prevent confounding influences of response planning or execution to the neural activations evoked by observation of the actions (Fig. 7.2B). Before the start of the second run, participants saw six practice items (which were different from the actions used in the main experiment) to become familiar with the response procedure.

*Data acquisition* Data were acquired on a Siemens ‘Trio’ MR-scanner with 3 Tesla magnetic field strength. Whole-brain echo-planar imaging (EPI) was performed using a standard bird-cage head coil with single pulse excitation (TR=2130 ms, TE=30 ms, flip angle=80 degrees, 32 slices, slice thickness=3mm, 0.5 mm gap between slices, voxel size 3.5x3.5x3.5 mm). During the scanning session, eye movements were measured using an infrared IviewX eyetracker (<http://www.smi.de>) with custom-built shielding which avoided that the eye-tracking equipment would interfere with EPI data collection. Eye movements were recorded to account for influences of the amount of eye movements that may correlate with differences in the experimental conditions. Moreover, eye-tracking was used to assess vigilance (i.e.



**Fig. 7.2.** Time course of a trial in the passive viewing run (A) and in the run with a task (B). **A)** In the passive viewing run, the end of a stimulus was followed by a blank screen for a random intertrial interval (ITI) of on average 4 seconds (range from 3-5 seconds in steps of 250 ms). **B)** In the run with a task, stimulus offset was followed by a blank screen for on average 2.9 seconds (range from 2.6 to 3.2 seconds in steps of 200 ms) after which participants had to indicate whether they thought the observed action was meaningful or not by pressing a button with the left or the right index finger. Response (yes / no) to finger (left / right) mapping was varied randomly and response time cut-off was set at 3 seconds. If participants responded 'no', a blank screen appeared for 100 ms, followed by the ITI. If participants responded 'yes' they got a follow-up question which was a forced choice between two possible meanings for the action. After responding, there was a variable ITI of on average 4 seconds (range 3-5 seconds in steps of 250 ms) before the next trial started.

wakefulness). All participants remained awake during the scanning session. Eye movements of one participant were not recorded due to technical failure.

*Data analysis* SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>) was used for fMRI data analysis. Preprocessing was done by discarding the first three volumes, motion correction by means of rigid body transformation (rotation and translation along all three dimensions), slice timing correction of all slices to the onset of the first slice, normalization of images to the Montreal Neurological Institute (MNI) EPI template, interpolation of voxel sizes to 2x2x2 mm, and spatial smoothing with a kernel of 8 mm FWHM. First level single subject statistics were computed in the context of the general linear model with six regressors of interest (Meaningful left hand, Meaningful right hand, Meaningful both hands, Meaningless left hand, Meaningless right hand, Meaningless both hands) for each run separately. Trials in which an incorrect response was given were included in the model as regressor of no interest. The estimates derived from the motion correction algorithm, the amount of eye movements in the x and y plane and responses (button presses) were included in the model as regressors of no interest. The eye movement regressors were obtained by computing the total amount of eye movements in the x and y plane separately for every volume (TR). All regressors except for the motion parameters were convolved with a canonical two gamma hemodynamic response function.

A separate analysis was performed on the data of the participants for whom we got scores on the degree of hand preference for the actions that were used as stimuli in this experiment (see *Participants* section). These were 27 out of 32 participants (15 / 16 left-handed and 12 / 16 right-handed). On top of the main model as described above, the mean-corrected subject-specific scores for each stimulus were added as an extra parametrically varying linear regressor which was orthogonal to

the MFr (meaningful right hand) and MFl (meaningful left hand) regressors (Buchel et al. 1998).

Single subject contrast maps were taken to a second level random effects group analysis of variance (ANOVA) with factors Group (left-handed / right-handed), Meaning (meaningful / meaningless) and (observed) Hand (left / right / both). Correction for violation of the sphericity assumption was applied when appropriate. Whole brain analysis results are corrected for multiple comparisons at a family-wise error rate of  $p < 0.05$  by using the theory of Gaussian random fields (Friston et al. 1996). Anatomical localisation of results was done with reference to the Duvernoy atlas (Duvernoy 1999).

## Results

*Behavioural results* Overall, participants gave correct responses in 88.8% of the trials. In 7.3% of the responses an incorrect response was given to a meaningless action, compared to 3.6% of the responses in which an incorrect response was given to a meaningful action. This difference was statistically significant ( $t(31)=2.90$ ,  $p=0.007$ ). There was no difference in the amount of incorrect responses between left- and right-handed participants ( $t(30) < 1$ ).

Analysis of the eye-movement data indicated that participants made more eye movements during meaningless actions than during meaningful actions, although the difference was not statistically significant in the overall comparison ( $t(30)=-1.87$ ,  $p=0.071$ ). This effect did reach statistical significance in the passive viewing run ( $t(30)=-2.07$ ,  $p=0.047$ ), but not in the run with a task ( $t(30)=-1.65$ ,  $p=0.079$ ).

*fMRI results* In the fMRI data analysis we explored all main effects and interactions between conditions. First we explored which areas were modulated by the main effect of Meaning (meaningful / meaningless). In the passive viewing run, there were no areas activated to this contrast<sup>1</sup>. In the run in which participants had to indicate whether they

thought an action was meaningful or not, the main effect of Meaning evoked activations in a wide-spread network of areas, encompassing left inferior frontal sulcus, left and right precentral sulcus, left supplementary motor area, left inferior parietal sulcus, left middle temporal sulcus, left inferior temporal sulcus, bilateral inferior occipital gyrus and the thalamus bilaterally. In all these areas activation levels were higher for the meaningful as compared to the meaningless actions (Fig. 7.3 and Table 7.1). Second, we explored which areas were modulated by the hand that acted out the pantomime in the movie clips (main effect of Hand). In the passive viewing run, activations were found in bilateral inferior occipital gyrus, left inferior temporal sulcus and in bilateral postcentral sulcus (Fig. 7.4 and Table 7.2). In left-hemispheric areas, observation of the action performed with the left hand led to stronger activation, whereas the opposite pattern was observed in the areas in the right hemisphere. This is probably due to the dominant (i.e. ‘acting’) hand being in the contralateral visual hemifield in these cases. That is, when the observed action was performed with the left hand, this was on the right side of the screen, hence - probably - in the contralateral hemifield for the left-hemisphere. Note that participants could freely view the actions and were not required to fixate to the middle of the screen. A similar pattern was observed for the run in which participants performed a task: the main effect of Hand evoked activations in bilateral inferior

**Table 7.1.** (*opposite page*) Areas differentially activated to meaningful and meaningless actions (main effect of Meaning) in the whole brain analysis. Displayed are a description of the region, its size in voxels (2x2x2 mm), the F value and MNI coordinates of the maximally activated voxel. Results are corrected for multiple comparisons at family-wise error rate of  $p < 0.05$ . Local maxima are reported which are more than 8 mm apart

Main effect of Meaning	Region	Size	F(max)	x	y	z
<b>Passive viewing run</b>	-	-	-	-	-	-
<b>Run with task</b>						
Meaningful>Meaningless	Left inferior frontal sulcus / precentral sulcus	833	49.38	-40	4	30
			45.91	-40	0	38
			35.55	-42	26	16
	Right precentral sulcus	355	41.09	32	-4	56
			31.01	36	-18	56
	Left inferior parietal sulcus	666	38.76	-32	-78	30
			33.18	-26	-56	44
	Supplementary motor area	197	39.29	-4	2	66
	Right inferior occipital gyrus	654	83.28	24	-98	-6
	Left inferior occipital gyrus	556	51.39	-20	-100	-8
			49.03	-32	-92	-10
	Left middle temporal sulcus	57	28.09	-58	-42	4
	Left inferior temporal sulcus	224	41.15	-32	-50	-18
			30.65	-50	-56	-14
	Thalamus	179	37.30	-4	-30	-6
	-	-	-	-	-	-
Meaningless>Meaningful						

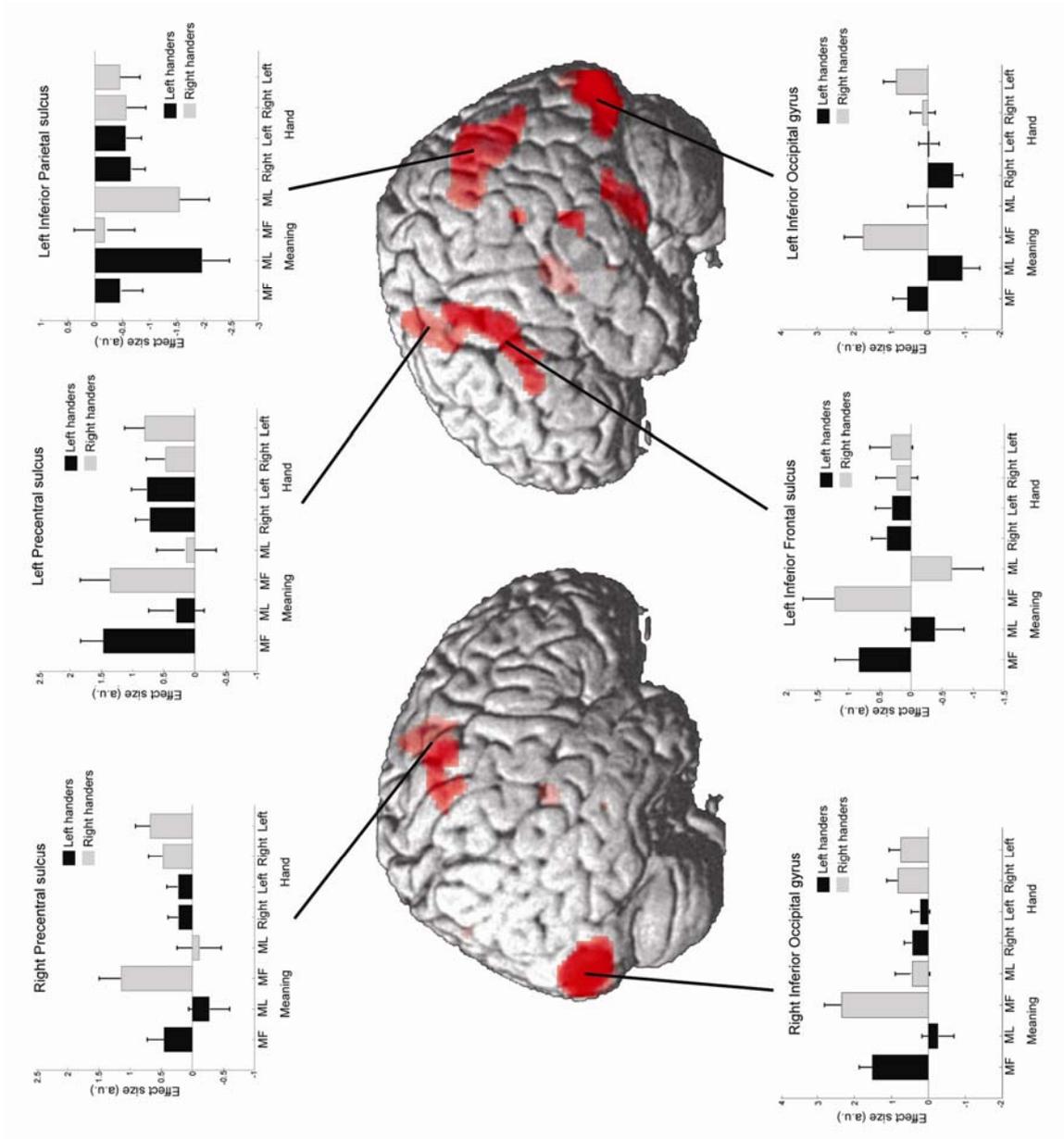
<b>Main effect of Hand</b>						
	<b>Region</b>	<b>Size</b>	<b>F</b>	<b>x</b>	<b>y</b>	<b>z</b>
<b>Passive viewing run</b>						
Left hand>Right hand	Left inferior occipital gyrus	483	55.76	-18	-98	-10
			27.70	-30	-92	-16
	Left postcentral sulcus	30	19.99	-32	-40	60
Right hand>Left hand	Left inferior temporal sulcus	20	17.42	-40	-74	2
	Right inferior occipital gyrus	329	36.21	20	-96	-8
	Right postcentral sulcus	14	17.67	14	-62	66
<b>Run with task</b>						
Left hand>Right hand	Left postcentral sulcus	275	26.81	-32	-48	66
			25.62	-36	-40	58
	Left superior occipital gyrus	91	22.61	-20	-90	32
	Left inferior occipital gyrus	467	37.53	-14	-98	-10
			17.22	-26	-92	-14
Right hand>Left hand	Right inferior occipital gyrus	119	22.10	14	-100	16
			17.19	12	-102	8
	-	-	-	-	-	-

**Table 7.2.** Areas differentially activated to the hand that acted out the pantomime (main effect of Hand) in the whole brain analysis. Displayed are a description of the region, its size in voxels (2x2x2 mm), the F value and MNI coordinates of the maximally activated voxel. Results are corrected for multiple comparisons at family-wise error rate of p<0.05. Local maxima are reported which are more than 8 mm apart.

occipital gyrus, left superior occipital gyrus and in left postcentral sulcus. The observation of actions performed with the left hand evoked stronger activation than actions performed with the right hand in left-hemispheric regions (Fig. 7.5 and Table 7.2). However, the opposite pattern was not observed in right inferior occipital gyrus, as was the case in the passive viewing run.

Third, the main effect of Group revealed areas in which activation levels generally differed between left- and right-handers. In the passive viewing run this led to increased activation in bilateral inferior occipital gyrus, as well as in bilateral inferior temporal sulcus (supplementary Fig. S7.1 and Table 7.3). In the run with a task there again was a main effect of Group in bilateral inferior temporal sulcus, as well as in left superior occipital gyrus (supplementary Fig. S7.2 and Table 7.3). In all these areas activation levels were higher for left-handed participants than for right-handed participants. The inferior temporal regions overlap with or are in the vicinity of earlier findings of extrastriate body area (EBA) and human motion area MT, which have been found to be sensitive to the observation of the human body (Downing et al. 2001; Astafiev et al. 2004; Peelen et al. 2006).

**Fig. 7.3.** (next page; For colour version see Appendix, p. 275). Neural differences between meaningful and meaningless actions in the run in which participants had to indicate whether the action was meaningful or not (main effect of Meaning). Activation levels are strongest for the meaningful actions in all areas. Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI coordinates) in right precentral sulcus (32 -4 56), left precentral sulcus (-42 -2 50), left inferior parietal sulcus (-26 -68 46), right inferior occipital gyrus (24 -98 -6), left inferior frontal / ventral premotor cortex (-40 4 3) and left inferior occipital gyrus (-20 -100 -8). Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



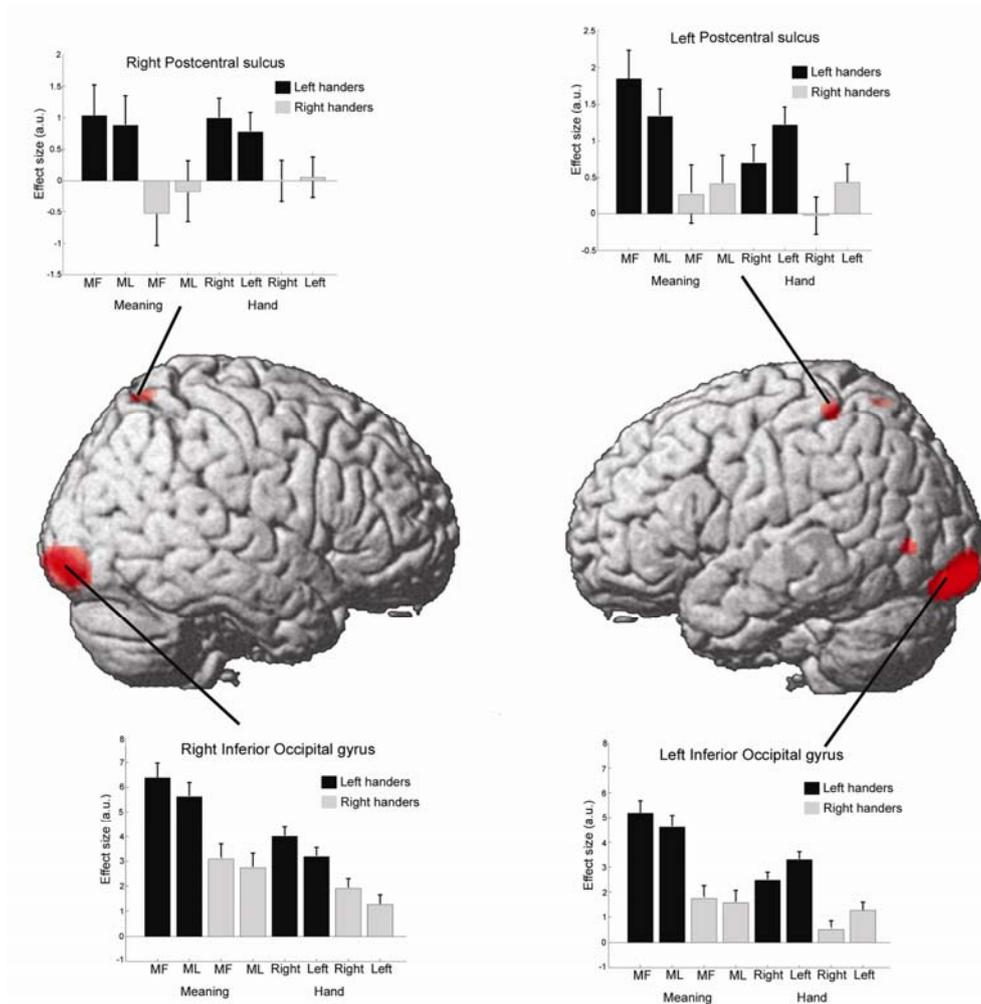
<b>Main effect of Group</b>						
	<b>Region</b>	<b>Size</b>	<b>F</b>	<b>x</b>	<b>y</b>	<b>z</b>
<b>Passive viewing run</b>						
Left-handers>Right-handers	L inferior occipital gyrus	591	50.83	-16	-90	-12
			42.98	-6	-96	-2
			39.88	-12	-102	2
	R inferior occipital gyrus	12	29.45	16	-104	8
	R inferior temporal sulcus	427	46.15	44	-64	-2
			42.42	40	-76	-2
		35.74	46	-70	-12	
	L inferior temporal sulcus	67	42.47	-38	-72	2
		24	33.33	-32	-88	-12
Right-handers>Left-handers	-	-	-	-	-	-
<b>Run with task</b>						
Left-handers>Right-handers	L superior occipital gyrus	23	33.04	-10	-104	8
	R inferior temporal sulcus	367	72.38	46	-70	-2
	L inferior temporal sulcus	80	41.37	-38	-74	2
Right-handers>Left-handers	-	-	-	-	-	-

**Table 7.3.** Areas differentially activated in left- and right-handed participants (main effect of Hand) in the whole brain analysis. Displayed are a description of the region, its size in voxels (2x2x2 mm), the F value and MNI coordinates of the maximally activated voxel. Results are corrected for multiple comparisons at family-wise error rate of  $p < 0.05$ . Local maxima are reported which are more than 8 mm apart.

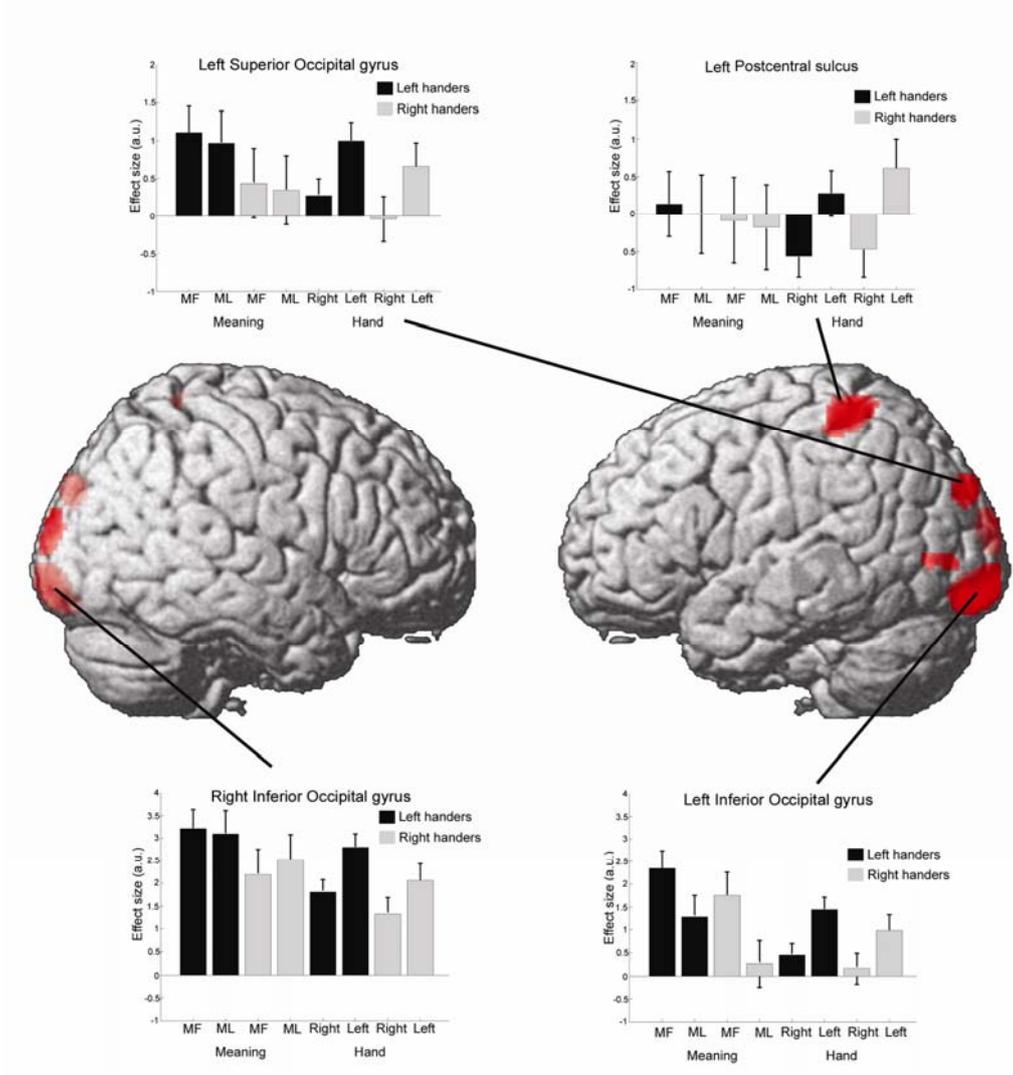
Importantly, no areas were activated to the Group x Meaning interaction or any of the interaction terms in any of the two task settings<sup>1</sup>.

Finally, no areas were significantly correlated to the linear parametrically varying regressors based upon the degree of handedness scores in any of the two task settings. The main findings reported above were also present in this subset of the data (N=27), but the degree of hand preference did not correlate with activation levels in any of the areas. This was also not the case when the search space was restricted to areas activated in any of the main contrasts described above by means of small volume correction.

In summary, we found strong effects of Meaning in a network of areas including bilateral dorsal premotor cortex and left inferior frontal gyrus when participants were required to actively process the meaning of the actions. However, this effect was not modulated by the hand preference of the observer. That is, no areas were sensitive to the Group x Meaning interaction, nor did activity in any area correlate with the subject- and item-specific degree of hand preference.



**Fig. 7.4.** (For colour version see Appendix, p. 277). Neural differences between observed hand in the run in which participant passively viewed the actions (main effect of Hand). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI) in right postcentral sulcus (14 -62 66), left postcentral sulcus (-32 -40 60) and right (22 -94 -10) and left (-18 -98 -10) inferior occipital gyrus. Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



**Fig. 7.5.** (For colour version see Appendix, p. 278). Neural differences between observed hand in the run in which participants had to indicate whether the action was meaningful or not (main effect of Hand). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left- (black bars) and right-handers (grey bars). Effect sizes are taken from local maxima (MNI) in left superior occipital gyrus (-20 -90 32), left postcentral sulcus (-32 -48 66) and right (14 -100 16) and left (-14 -98 -10) inferior occipital gyrus, expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .

## **Discussion**

In this study we investigated the role of the observer's motor system in understanding the meaning of common hand actions. First, we assessed whether cortical motor areas play a role in coding the meaning of an observed action or not. Neural simulation theory predicts involvement of parts of the cortical motor system in understanding the meaning of an action, as compared to a traditional 'cognitivist' stance in which meaning is thought to be represented only in a purely symbolic way, outside of the visuo-motor parts of the brain (e.g. Fodor 1975). Indeed, we found that neural activation in the cortical motor system was higher when participants judged an action to be meaningful compared to when they judged the action to be meaningless. In contrast to earlier findings (Decety et al. 1997), we show that besides inferior frontal cortex, also premotor and inferior parietal cortex are activated more strongly in response to a meaningful action as compared to a meaningless action. An extensive body of literature suggests that these areas are part of the human 'mirror neuron system' which is activated both during action production and during action observation (Rizzolatti et al. 2001; Rizzolatti and Craighero 2004; Iacoboni and Dapretto 2006). Here we extend these findings by showing that premotor and inferior parietal cortex are sensitive to whether the action conveys a meaning or not. This is in line with the assertion that an action's meaning is at least partially represented in motor structures of the brain. In this framework, parts of the cortical motor system do not passively react to the observation of an action, but are actively involved in coding the meaning of an action (Gallese and Lakoff 2005). Given its known role in processes of semantic selection / unification (Thompson-Schill et al. 1997; Hagoort 2005b, a; Thompson-Schill et al. 2005), it is likely that inferior frontal cortex is involved in top-down modulation of areas of the action observation network, such as inferior parietal and premotor cortex. The fact that activation in inferior frontal cortex was left-lateralised

suggests that such modulation may have occurred through the language system.

An interesting finding is that parts of the visual system were also modulated by the meaning of the actions. Since our stimuli were matched on motion and visual characteristics, this difference cannot be evoked by low-level differences in the input material. Therefore, it is best explained as a top-down effect of assigning a meaning to an action. This conclusion is strengthened by the difference in response in the visual cortex between the two runs. In the run without a task, visual areas show an effect of the visual hemifield in which the hand was presented (Fig. 7.4). However, in the run with a task, this effect is not present (Fig. 7.5). This suggests that during passive viewing, visual cortex is most influenced by visual characteristics of the stimulus, whereas during the other run, this region is mostly influenced by top-down task effects. The fact that we find this effect in the visual cortex is in line with embodied cognition, in which meaning is thought to be grounded in *sensori*-motor representations.

Interestingly, and in contrast to the premotor cortex activations, the parietal cortex activation was left-lateralized. This is in line with neurological literature which shows that a deficit in pantomiming tool use is mostly associated with left-hemispheric parietal regions in right-handers (e.g. Goldenberg et al. 2003) as well as in left-handers (Frey et al. 2005). Neuroimaging literature also indicates that specifically the left parietal cortex is activated during the execution of pantomimes (Moll et al. 2000).

Second, our study extends previous research with respect to the level of representation of motor simulation during action understanding. Although all participants had a strong preference to produce each of the actions that they observed with either the left or the right hand, they did not map the observed actions onto their cortical motor system according to their hand preference. That is, there was no difference in activation patterns in motor areas between left-

and right-handers. We acknowledge that this conclusion is based on a null-result, however, for several reasons we argue that our conclusion of no subject-specific mapping of observed actions onto the motor system is justified. First, the parameter estimates from the cortical motor system show no trend towards a difference between left and right hemisphere and left- and right-handed participants (Fig. 7.3). Note that simulation according to hand preference would imply strong and clear-cut differences in the pattern of activation in the motor system of both groups. That is, left-handers should consistently map the observed action onto their preferred (left) hand, whereas right-handers should consistently map the observed actions onto their preferred (right) hand. Second, a lack of statistical power is unlikely to be the cause of the null-finding, since we did observe differences between the two groups in several regions outside of the motor system (Fig. S7.1 and S2). It is tempting to speculate about these group differences in EBA as reflecting a difference in the way left- and right-handers observe body parts or biological motion. That is, it could be argued that left-handers mostly observe (right-handed) people who perform actions differently than they do which may increase attention to observed body parts / biological motion. This is a highly speculative suggestion which does seem to deserve further investigation nonetheless. Third, a potentially more sensitive analysis, taking the subject-specific degree of hand preference for each action into account, did also not reveal any area to be sensitive to the degree of hand preference of the observer. Finally, this result is not due to an overly strict thresholding of the statistical maps. Also at a much more liberal statistical threshold ( $p < 0.25$  corrected), no areas in the motor system were activated to the main effect of Group or to the Group x Meaning or Group x Hand interactions.

How is it possible that the motor system is responsive to the meaning of actions, but not in a way that is specific to the observer's motor preference? The embodied cognition framework - of which neural

simulation is a particular instantiation (e.g. Gallese and Lakoff 2005) - allows for such a less tight coupling between an observer's motor repertoire and neural correlates of action understanding, while at the same time maintaining the importance of sensori-motor representations for action understanding. That is, in this framework an action's concept can be 'abstracted away' from the motor specifics of the observer. Anderson (2003) describes how this is the case for an extreme example: "... the concept of 'walking', in so far as it is logically and semantically related to various concepts of movement, [...] ought to be easily acquirable by an individual who cannot, and who perhaps never could, walk. The concept can be placed in a logical and semantic network *which is on the whole grounded, even given that there is no specific experience of walking which directly grounds the concept.*" (p. 113, our emphasis). In other words, although parts of the cortical visuo-motor system are involved in representing the meaning of an action, this representation is not strictly coupled to the observer's motor production system. Empirical support for this position comes from a recent study which found that two aplasic individuals born without hands activate parts of the motor system involved in controlling other effectors (foot and mouth) when observing hand actions (Gazzola et al. 2007). In our study, clearly all participants were capable of producing the actions that they observed. Still however, the neural correlates of understanding the actions are flexible in the sense that they are not strictly coupled to the observer's action production preference.

Our results might seem at odds with earlier neuroimaging studies that did show an influence of motor repertoire of expert dancers on the neural correlates of action observation (Calvo-Merino et al. 2005; Calvo-Merino et al. 2006; Cross et al. 2006). The present study differs in important ways from earlier studies comparing experts and non-experts. That is, all actions we showed were within the motor repertoire of all participants. In the 'expert versus non-experts' studies, the non-expert group had not performed the actions they observed

(Calvo-Merino et al. 2006), and / or were not capable of doing so since the actions were complicated dance movements (Calvo-Merino et al. 2005). It is conceivable that the level of ‘understanding’ and / or neural simulation is less in these individuals compared to experts who have performed the actions before<sup>2</sup>. This however is a question for future research. Our data suggest that in the case of actions we have all performed and observed many times before, the system has generalized the action’s meaning to a conceptual level that is less strictly tied to the motor specifics of the observer.

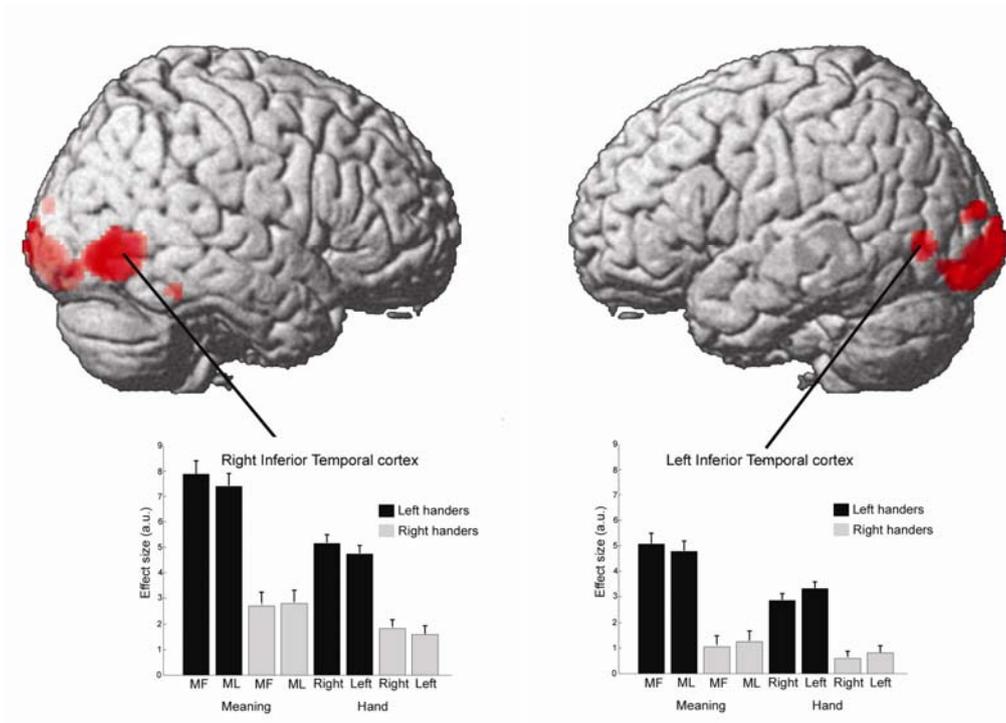
Finally, we looked into the automaticity of motor cortex activation to action understanding. We found that the motor system was only modulated by the meaning of an action when participants were required to actively observe the action. When participants passively viewed the actions, no effect of meaning was observed in premotor and / or inferior parietal cortex. It seems therefore that the action observation system is under top-down influence, perhaps by inferior frontal cortex (see also Jonas et al. 2007; but see de Lange et al. 2008). This nicely relates to an earlier study in which we showed that premotor cortex (BA 6) was more strongly activated when a hand gesture was incongruent with the previous sentence context as compared to when the hand gesture was congruent with a sentence context (Willems et al. 2007). These data showed that premotor cortex is not only responsive to the observed action, but is also sensitive to the (language) context the action occurs in. Note that it may well be the case that conceptualization of the actions during the active run occurred through linguistic mediation. This does however not invalidate our conclusion that the cortical motor system plays a role in coding the meaning of an action.

It may be objected that perhaps participants were not paying attention to the stimuli during the passive viewing condition. The pattern of eye movements was however the same in the passive viewing condition as in the active condition, with more eye movements during

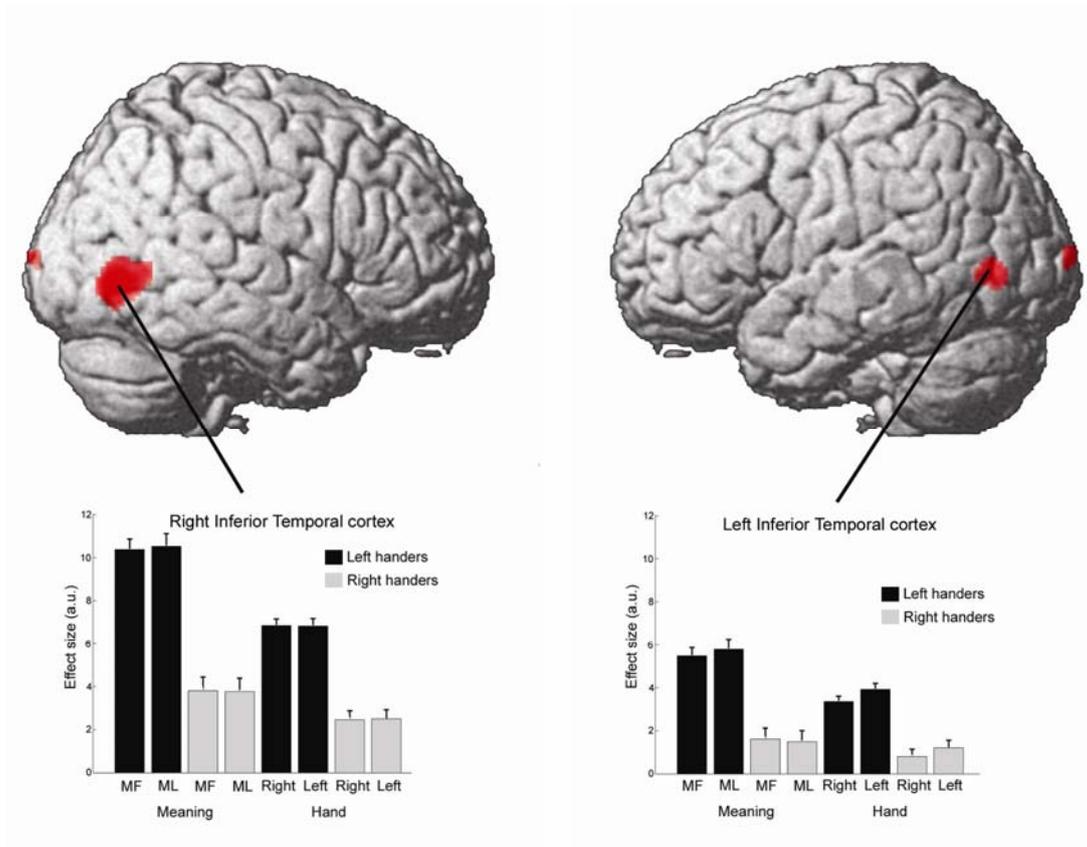
observation of meaningless as compared to meaningful actions. This indicates that the lack of an effect of meaning in the passive viewing data is unlikely to be due to participants not looking at the stimuli.

This study is one in a series of few that draw attention to the importance of the kind of actions under study in action understanding. For instance, Newman-Norlund and colleagues (2007) found that the motor mirror system is differentially reactive to complementary and imitative actions. Recent data from our lab underscore the importance of the type of action under study (Willems and Hagoort under review). It was found that observation of simple hand movements (repeated contractions and extensions of all fingers) *does* differentially activate ventral premotor and parietal cortex in left- and right-handed participants. We hypothesize that this difference is due to the fact that the understanding of the common actions that were the stimuli in the present paper is ‘abstracted away’ from the motor specifics of the observer. It seems that the action observation system in such cases codes the *meaning* or *goal* of the observed action rather than the exact way in which the action is performed (see also Gazzola et al. 2007; Rijntjes et al. 1999). This is arguably not the case in the observation of simple repeated contractions and extensions of the hand, which were the stimuli in our other study (Willems and Hagoort under review). It will be a challenge for future research to describe what crucially drives the neural action observation system and how its activation is exactly influenced by the motor production preference of the observer.

[Conclusion section on p. 214]



**Fig. S7.1** (For colour version see Appendix, p. 279). Neural differences between left- and right-handers in the run in which participants passively viewed the actions (main effect of Group). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI coordinates) in right (44 -64 -2) and left (-38 -72 2) inferior temporal sulcus, overlapping with previously reported location of extrastriate body area and human motion area MT (Peelen et al. 2006). Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



**Fig. S7.2** (For colour version see Appendix, p. 280). Neural differences between left- and right-handers in the run in which participants had to indicate whether the action was meaningful or not (main effect of Group). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI coordinates) in right (46 -70 -2) and left (-38 -74 2) inferior temporal sulcus, overlapping with previously reported location of extrastriate body area and human motion area MT (Peelen et al. 2006). Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .

## **Conclusion**

We provide evidence for the involvement of the cortical motor system in action understanding. Crucially, our findings shed light on the nature of motor activation during action understanding. That is, action understanding does not necessarily involve a one-to-one mapping of the observed action onto the observer's motor system. Rather, when a strict coupling between motor repertoire / preference and neural simulation might be detrimental for action understanding – as was the case in our study -, the action observation system is capable of generalizing beyond the motor specifics of the observer. This is in line with the embodied cognition framework, which asserts that action understanding is grounded in sensori-motor processes, but not necessarily in a way that is strictly tied to the observer's motor preference. This is a neural reflection of the flexibility of action understanding which allows us to use our own motor system to understand the actions of others, but in a flexible manner.

## **Notes**

- 1) Also at a more liberal statistical threshold (i.e.  $p < 0.25$  corrected) no areas were found to be activated.
- 2) For discussion about the fundamental difference in observing actions by experts as compared to non-experts see Merleau-Ponty ([1945] 1962) and Rietveld (in press).

## **Acknowledgements**

Supported by a grant from the Netherlands Organization for Scientific Research (NWO), 051.02.040 and by the European Union Joint-Action Science and Technology Project (IST-FP6-003747). We thank Cathelijne Tesink and Nina Davids for help in creation of the stimuli and Paul Gaalman for assistance during the scanning sessions.

## Chapter 8 Summary and Discussion

### Summary

The topic of this thesis is the neural basis of understanding and integrating meaning conveyed through hand actions and through spoken language. We mostly studied iconic co-speech gestures, which are hand actions that are naturally produced together with speech and that express information in an iconic way. Another type of hand actions under study were pantomimes, which are depictions of common actions, acted out as if the actor is performing the action, but without using or acting upon real objects. Here I will briefly summarise the results of each chapter.

In Chapters 2 and 3 we compared the neural integration of semantic information conveyed through gestures or through spoken words into a preceding language context. In Chapter 2 we used ERPs to investigate the neural time course of this process. It was found that increased semantic integration load from information conveyed through a co-speech gesture as well as through a spoken word elicited similar effects in the N400 component, which is known to be sensitive to the semantic ‘fit’ of an item in relation to the preceding context. We concluded that information from gesture and speech is integrated in a similar way at the semantic level in the brain and that no temporal precedence is given to spoken language. In Chapter 3 we employed the same experimental design but now the target question regarded the neural loci of semantic integration of gestures and speech. It was found that the left inferior frontal cortex is involved in semantic integration of information conveyed through spoken words as well as through co-speech gestures. This means that the function of this classical language area is not restricted to integrating information from language, but that it is also activated when action-related information is harder to integrate within the prior sentence context. In this study, we however

also found effects that were specific to integration of either word or gesture. Increased integration load for words led to specific activation of left superior temporal cortex, whereas increased integration load of co-speech gestures led to increased activation in premotor cortex (BA 6). These areas are believed to be part of the neural language and action networks, respectively. Their activation is explained as evoked by top-down influence, probably asserted by inferior frontal cortex, onto areas lower in the cortical hierarchy. This top-down influence is stronger when information integration is harder which is reflected in an activation increase.

Chapters 4 and 5 constitute a detour in this language- and action-oriented thesis. A highly similar experimental design was employed as compared to the studies in Chapters 2 and 3, but now comparing neural integration processes of semantic information conveyed through a word or through a picture (line drawing) of an object. The rationale was that the integration of visual information which has a clear meaning without speech might be different than the integration of co-speech gestures whose meaning is not easily recognized outside of a speech context. The ERP and fMRI results were highly similar in these studies as compared to the integration of co-speech gestures and spoken language (Chapters 2 and 3). That is, again, we observed a similar neural time course as well as overlap in inferior frontal cortex both for integration of semantic information from spoken words as well as from pictures into a preceding sentence context. It was argued that this is convincing evidence for the claim that different types of information are processed in a similar way by the language comprehension system, at least at the level of semantics. In Chapter 5, time-frequency analysis was performed on the EEG data from the study reported in Chapter 4. Specific effects for integration of a spoken word and of a picture were observed in decreases in the alpha and gamma frequency bands respectively. These effects occurred rather early (around 100 ms after

onset of the critical word / picture) and were argued to reflect an early, context-based detection of incongruent acoustic or visual form.

In Chapter 6 we directly compared neural areas involved in integration of information from speech and co-speech gestures or from speech and pantomimes. A crucial difference between co-speech gestures and pantomimes is that gestures cannot be unambiguously recognised without the speech they are normally co-expressed with. This is not the case for pantomimes: the action that they express is easily recognisable without accompanying speech. We reasoned that this may lead to areas being differentially involved in the integration of gestures and speech on the one hand and pantomimes and speech on the other hand. We replicated the finding from Chapter 3 in the sense that mismatching speech-gesture combinations lead to increased activation levels in inferior frontal cortex as compared to matching speech-gesture combinations. This was also true for mismatching pantomime-speech combinations as compared to matching speech-pantomime combinations. However, in left superior temporal sulcus there was an effect of semantic integration for speech-pantomime combinations, but not for speech-gesture combinations. It was argued that only when there is a relatively stable memory representation of an observed action / word, pSTS plays a role in multimodal integration. If this is not the case (as with co-speech gestures) integration only happens at a higher level in the cortical hierarchy, reflected in activation in inferior frontal cortex.

Finally, in Chapter 7, it was investigated how meaning from actions that can convey meaning without language, is represented in the brain. Left- and right-handed participants observed pantomimes of common actions and actions that were comparable in terms of overall movement parameters, but that were rendered meaningless by means of changing the hand shape of the action. The handedness manipulation was used to test a hypothesis from neural simulation theory which assumes that action understanding occurs through

mapping the observed action onto the observer's own motor system. If this is indeed the case we would expect differences in lateralisation in the neural motor system between the hand preference groups. Alternatively, it may be the case that although the cortical motor system is involved in action understanding, its activation is less tightly coupled to the motor preference of the observer. It was found that parts of the cortical motor system are more strongly activated when participants judged an action to be meaningful as compared to when they judged an action to be meaningless. This argues for part of the cortical motor system to be involved in coding the meaning of an action. Importantly, this was not influenced by the hand preference of the observer. It was argued that these data provide support for a role of the cortical motor system in coding the meaning of an observed action at the level of the meaning or goal of the action.

## **Discussion**

### *Neural correlates of co-speech gestures during language comprehension*

The results on the neural basis of understanding information from co-speech gestures during language comprehension are best characterized by the similarity that we observed between gestures and words. As described above, overlapping neural correlates were observed for increased semantic integration load of both spoken words as well as from gestures. The first thing to conclude from this is that co-speech gestures do elicit semantic processing. As described in the introductory chapter (Chapter 1), this has been a debated topic in gesture research. From the findings in Chapters 2 and 3, as well as from an increasing body of comparable literature, it can be concluded that semantic information from co-speech gestures can impact understanding and that they are not mere 'hand waving'. The present studies add to this that the nature of integration of co-speech gestures is very similar to that of words. It was argued that the neural correlates of semantic integration of gestures reflect 'unification' or integration of incoming

information with the previous content of the message. The insight from the studies described here is that such unification occurs in similar ways for information from words, co-speech gestures and pictures. We show that in neural terms, no precedence is given to language information over non-language information.

What do these findings mean for our understanding of the role of co-speech gestures during language comprehension? Above I have argued that speech is not given temporal precedence in the neural integration process, but one could similarly argue that gestures are not treated differently by the brain, despite the interestingly different format they are in. Such a conclusion is however not justified for at least three reasons.

The first concerns the cortical areas that were found to be activated specifically for gestures, words or pictures. In the summary I hinted at these to reflect differences in the format the semantic information is in, and that these areas may be modulated by top-down processing by higher-order areas after detection of the semantic anomaly. So a different way to look at our findings is that there is overlap between the networks involved in integration of words / gestures / pictures, with some areas performing similar functions, and others specific functions. This means that the common mechanism of semantic integration is similar, regardless of the format of the incoming information. However, the overall network of areas underlying the specific process is different, depending upon the format of the incoming information.

Second, differences in oscillatory dynamics may be indicators of differences in multi-modal integration. This is suggested by the differences we observed for integration of information from words and pictures in specific frequency band power (Chapter 5). It should be stressed that very firm interpretation of these findings is not easy. The first reason is that the findings are not fully in line with the conclusion that oscillatory dynamics underlie coding of differences in format. The

second reason is that there is little previous literature to compare our results with. Still, it seems that early differences in power in specific frequency bands distinguish integration of words and pictures. It would be interesting to see whether this is similarly true for integration of co-speech gestures and words. Unfortunately, the low quality in higher frequency bands of the EEG data of the gesture study (Chapter 2) did not allow for reliable estimation and statistical testing in the time-frequency domain.

Third, we found that posterior STS - an area which is involved in multimodal integration of a large variety of audio-visual stimuli -, is not sensitive to semantic congruency of speech and gesture. We found that STS was modulated by congruency of pantomimes and words, but not by congruency of gestures and speech. On the contrary, the anterior part of left inferior frontal cortex was modulated by congruency of both speech-pantomime and speech-gesture pairs. This indicates not so much a difference in integration of linguistic versus non-linguistic information, but rather shows that different action types are integrated with language in a qualitatively different way. This is neural evidence for the fact that gestures and speech are tightly integrated in natural language use and that without speech, gestures seem to lose their informational value. Another characteristic of co-speech gestures which is highlighted by these findings is what I call the 'flexibility' of gestures. When understanding and observing a gesture one cannot easily map the gesture onto an existing memory trace, as one can do in the case of for instance a pantomime, or the picture of an object. Rather, the inherent ambiguity of gestures necessitates that integration occurs at a higher level in the system, most probably involving parts of the language network.

It should be noted that for reasons of experimental design we have treated 'co-speech gestures' as a rather homogeneous group in our studies. This is an over-simplification of the real-life situation in which it is known that types of co-speech gestures differ in subtle ways. An

example concerns the commonly made distinction between observer- and character-viewpoint gestures. Moreover, we have sometimes (Chapter 2 and 3) - but not always (Chapter 6) - used modelled co-speech gestures that were artificially combined with speech. Although there were good experimental reasons to do this, such as control over the exact materials and relative timing of speech and gesture, it is possible that subtle effects of speech and gesture combinations were lacking in our stimuli. However, ERP studies of co-speech gestures using more natural materials (e.g. Wu and Coulson 2005) report similar findings as were obtained here.

In conclusion, our data show that semantic information in different formats is integrated in similar ways with language information in the brain. This is strong evidence for the assertion that the language system takes in information from a variety of sources in a similar way when understanding language.

Future research should investigate different types of gestures as well as different gesture-speech combinations to get a more complete picture of the role of co-speech gestures in language comprehension. Differences are suggested in the shifting of cortical networks depending upon the format the semantic information is in, and / or specific changes in certain frequency bands of the EEG signal. Moreover, it seems that integration of action and language information is qualitatively different depending upon the relationship between action and language information.

#### *Action meaning in the brain*

Traditionally, meaning representations have been regarded as fully symbolic and amodal. The position that on the contrary meaning is at least partially 'grounded' in bodily activity is mostly associated with the embodied cognition framework. A remaining question is *how* meaning is embodied in the sense how motor experience or preference influences action meaning representation in the brain. In the research

described in Chapter 7, we have used neuroimaging to investigate the level of representation of action meaning in the brain.

We found that parts of the cortical motor system are more strongly activated when participants had to extract the meaning of an action. However, this activation was not influenced by the hand preference of the observer. As was argued above (Chapter 7), we have taken this to mean that action understanding does not involve a one-to-one mapping of the observed action onto the observer's motor system. Rather, we argued for the meaning representation of common actions to be partially abstracted away from the exact way in which an observer normally performs these actions. That is, the concept is overall grounded, but in a way that is not specifically tied to the exact motor practice of the observer.

This raises an important question: Can an action's meaning be grounded without the observer having any motor practice with the action? Anderson (2003) has argued that this is possible. He describes the case of an individual who is born without the capacity to walk. Still, this person will be able to understand the concept 'to walk' and he / she can understand what other people are doing when they walk. This is possible, Anderson describes, because the concept 'walk' will be part of a bigger semantic network of 'movement' or 'going from one place to the other'. Clearly, these concepts are embodied, also for a person who is in a wheel-chair. Since the concept 'walk' is embedded within this network, its meaning has become grounded. This explanation is not fully satisfactory. It seems that the very nature of embodied cognition necessitates that differences in bodily / action experience have some effect on meaning representation. In the case of the action that we studied in chapter 7, it is conceivable that understanding is not too strictly bound to the exact way in which the observer performs the action. However, Anderson's claim is more extreme; it is suggested that *no* action experience is needed to come to understanding of for instance the action 'to walk'. This however seems to be only true in the limited

sense that a person who is incapable of walking understands that walking means ‘displacement in space’ or something related. However, it seems that this is not a similar type of understanding of what it means to walk to a person who is capable of walking. How understanding in experts and non-experts may be different is an issue that I will not go into any deeper. However, this discussion on types of actions and types of motor expertise and their consequence for action understanding, serves to illustrate that the nature of motor cortex activation during action observation / understanding may crucially depend upon the type of actions that are observed, the motor expertise of the observer, and / or the goal / motivation with which the action are observed. Along these lines, my conclusion is not that action observation does not lead to subject-specific neural simulation *per se*. Actually, in a related study we observed strong evidence for neural simulation in parts of the motor brain when left- and right-handed participants observed very simple, essentially meaningless finger movements (Willems and Hagoort under review).

#### *Role of inferior frontal cortex*

Finally, I will consider the role of left inferior frontal cortex. This brain region was found implicated in several of the cognitive processes under study in this thesis. Since the formulation of a role for inferior frontal cortex (‘Broca’s area’) as a neural correlate of language production, it has become one of the classical language areas of the brain. During the history of neurology, the area has been claimed to be language-specific. However, with the advent of cognitive neuroimaging it has become increasingly clear that inferior frontal cortex is also involved in rather different cognitive functions. Examples include action observation, sequencing, semantic selection and unification (Thompson-Schill et al. 1997; Hagoort 2005b, a; Molnar-Szakacs et al. 2005; Koechlin and Jubault 2006). Such findings leave little room for a role of inferior frontal cortex as uniquely tied to language functioning. The multitude

of cognitive tasks in which this area seems to be involved, raises another, more fundamental question. That is, how is it possible that one and the same cortical area is capable of doing such seemingly distinct cognitive tasks? There are two answers to explain the multitude of conditions that make inferior frontal cortex 'light up'. First, one might try to classify all seemingly distinct tasks that activate this region under one common denominator. Thompson-Schill and colleagues, for instance, suggest that the inferior frontal cortex is involved in 'the regulation of mental activity' (Thompson-Schill et al. 2005). Hierarchical processing is another suggested overarching function of the inferior frontal cortex (Koechlin and Jubault 2006; Tettamanti and Weniger 2006; Koechlin and Summerfield 2007). Underlying these proposals is the notion that a given cortical area performs one function, the so-called 'one area one function' rule. However, another way of conceiving the broad range of tasks activating inferior frontal cortex is to think of an area as a node in multiple different networks, in which the network and not the area instantiates a function (Mesulam 1990, 1998; Fuster 2003). That is, higher order cortex will be implicated in different functional networks. Note that this does not imply equipotentiality of cortical areas. It is clear that there is specialization in the brain, the degree of which might be different between areas, however (see Mesulam 1990, 1998; Fuster 2003). Neither of these views are mutually exclusive nor does strong evidence exist in favour of the one or the other. It does however seem that the latter view might prove to be more fruitful than searching for ever more abstract 'superfunctions' to be able to comply to the 'one area one function' rule.

How does this relate to the findings of inferior frontal cortex activation in the research described here? We found increased activation in left inferior frontal cortex to increased integration load of semantic information from gestures and pictures (Chapters 3, 4 and 6) and in response to meaningful as compared to meaningless actions

(Chapter 7). In all these cases it seems that its role can be characterised by semantic selection or unification. Irrespective of the exact labelling one wants to assign to this function, it seems that inferior frontal cortex is implicated in top-down modulation of areas lower in the cortical hierarchy (see Gazzaley and D'Esposito 2007). In the research reported here this modulation is of semantic nature and seems to involve modulation of areas coding the meaning and / or format of the word / gesture / action to which semantic processing is applied. The results from the connectivity analysis reported in Chapter 6 are evidence for this. However, the top-down modulatory role the area plays can take many forms depending upon the input to the system and the task at hand. In that sense it is perhaps misleading to label inferior frontal cortex with one function. Rather, it is the task setting and the network in which it is activated which describes the role the area plays.



## References

- Aboitiz, F., Garcia, R., 1997. The evolutionary origin of the language areas in the human brain: A neuroanatomical perspective. *Brain Research Reviews* 25(3), 381-396.
- Aboitiz, F., Garcia, R. R., Bosman, C., Brunetti, E., 2006. Cortical memory mechanisms and language origins. *Brain and Language* 98(1), 40-56.
- Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., Naumer, M. J., 2005. Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research* 166(3-4), 559-571.
- Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. *Journal Comparative Neurology* 412(2), 319-341.
- Anderson, M. L., 2003. Embodied cognition: A field guide. *Artificial Intelligence* 149, 91-130.
- Arbib, M. A., 2005. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28(2), 105-124.
- Astafiev, S. V., Stanley, C. M., Shulman, G. L., Corbetta, M., 2004. Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience* 7(5), 542-548.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., Iacoboni, M., 2006. Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology* 16(18), 1818-1823.
- Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., Wagner, A. D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47(6), 907-918.
- Barrett, S. E., Rugg, M. D., 1990. Event-related potentials and the semantic matching of pictures. *Brain and Cognition* 14(2), 201-212.
- Bastiaansen, M. C., Hagoort, P., 2006. Oscillatory neuronal dynamics during language comprehension. *Prog Brain Res* 159, 179-196.
- Bastiaansen, M. C., van Berkum, J. J., Hagoort, P., 2002. Syntactic processing modulates the theta rhythm of the human EEG. *Neuroimage* 17(3), 1479-1492.

- Bastiaansen, M. C., van der Linden, M., Ter Keurs, M., Dijkstra, T., Hagoort, P., 2005. Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience* 17(3), 530-541.
- Baumgaertner, A., Weiller, C., Buchel, C., 2002. Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage* 16(3 Pt 1), 736-745.
- Beattie, G., Shovelton, H., 1999. Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology* 18(4), 438-462.
- Beattie, G., Shovelton, H., 2002. An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology* 93(2), 179-192.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., Martin, A., 2004a. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience* 7(11), 1190-1192.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., Martin, A., 2004b. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41(5), 809-823.
- Bernardis, P., Gentilucci, M., 2006. Speech and gesture share the same communication system. *Neuropsychologia* 44(2), 178-190.
- Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Reviews Neuroscience* 25, 151-188.
- Brown, C. M., Hagoort, P., Kutas, M. (2000). Postlexical integration processes in language comprehension: Evidence from brain-imaging research. *The cognitive neurosciences*. M. S. Gazzaniga. Cambridge, Mass., MIT Press: 881-895.
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., Freund, H. J., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience* 13(2), 400-404.
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C. A., Rizzolatti, G., 2004. Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *Journal of Cognitive Neuroscience* 16(1), 114-126.

- Buchel, C., Holmes, A. P., Rees, G., Friston, K. J., 1998. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage* 8(2), 140-148.
- Butterworth, G., Shovelton, H. (1978). Gesture and silence as indicators of planning in speech. *Recent advances in the psychology of language*. N. Campbell and P. Smith. New York, Plenum: 347-360.
- Caetano, G., Jousmaki, V., Hari, R., 2007. Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. *Proceedings of the National Academy of Sciences USA* 104(21), 9058-9062.
- Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., Vatikiotis-Bateson, E., 2003. Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14(17), 2213-2218.
- Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., Vatikiotis-Bateson, E., 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience* 16(5), 805-816.
- Calvert, G. A., 2001. Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex* 11, 1110-1123.
- Calvert, G. A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience* 15(1), 57-70.
- Calvert, G. A., Campbell, R., Brammer, M. J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology* 10(11), 649-657.
- Calvert, G. A., Thesen, T., 2004. Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris* 98(1-3), 191-205.
- Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., Haggard, P., 2005. Action observation and acquired motor skills: an fMRI study with expert dancers. *Cerebral Cortex* 15(8), 1243-1249.
- Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., Haggard, P., 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology* 16(19), 1905-1910.

- Caramazza, A., Hillis, A. E., Rapp, B. C., Romani, C., 1990. The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology* 7(3), 161-189.
- Church, R. B., Goldin-Meadow, S., 1986. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23(1), 43-71.
- Clark, H. H., 1996. *Using language*. New York, NY, US: Cambridge University Press.
- Corbetta, M., Shulman, G. L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 3(3), 201-215.
- Corina, D. P., Knapp, H., 2006. Sign language processing and the mirror neuron system. *Cortex* 42(4), 529-539.
- Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaro, A., Romani, G. L., Aglioti, S. M., 2005. Neural systems underlying observation of humanly impossible movements: An fMRI study. *Cerebral Cortex* 15(11), 1761-1767.
- Cross, E. S., Hamilton, A. F., Grafton, S. T., 2006. Building a motor simulation de novo: observation of dance by dancers. *Neuroimage* 31(3), 1257-1267.
- Culicover, P. W., Jackendoff, R., 2006. The simpler syntax hypothesis. *Trends in Cognitive Sciences* 10(9), 413-418.
- Cutler, A., Clifton, C. E. (1999). *Comprehending spoken language: A blueprint of the listener. The neurocognition of language*. C. M. Brown and P. Hagoort. Oxford, Oxford University Press.
- Dale, A. M., 1999. Optimal experimental design for event-related fMRI. *Human Brain Mapping* 8(2-3), 109-114.
- Davidson, D. J., Indefrey, P., 2007. An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Research* 1158, 81-92.
- Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M., Menon, D. K., 2007. Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences USA* 104(41), 16032-16037.
- de Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C., Cayeux, I., 2005. Cognitive modulation of olfactory processing. *Neuron* 46(4), 671-679.

- de Lange, F. P., Hagoort, P., Toni, I., 2005. Neural topography and content of movement representations. *Journal of Cognitive Neuroscience* 17(1), 97-112.
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., Bekkering, H., 2008. Complementary systems for understanding action intentions. *Current Biology* 18, 454-457.
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., Fazio, F., 1997. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120(Pt 10), 1763-1777.
- DeLong, K. A., Urbach, T. P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8), 1117-1121.
- Downing, P. E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. *Science* 293(5539), 2470-2473.
- Duvernoy, H. M., 1999. *The human brain: Surface, three-dimensional sectional anatomy with MRI, and blood supply*. Springer, Vienna.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25(4), 1325-1335.
- Emmorey, K. (2006). *The Role of Broca's Area in Sign Language. Broca's region*. Y. Grodzinsky and K. Amunts. New York, Oxford University Press:169-184.
- Engel, A. K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2(10), 704-716.
- Fadiga, L., Fogassi, L., Pavesi, G., Rizzolatti, G., 1995. Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology* 73(6), 2608-2611.
- Federmeier, K. D., Kutas, M., 1999. Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. *Brain research. Cognitive Brain Research* 8(3), 373-392.
- Federmeier, K. D., Kutas, M., 2001. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(1), 202-224.

- Federmeier, K. D., Kutas, M., 2002. Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia* 40(7), 730-747.
- Feyereisen, P., Van de Wiele, M., Dubois, F., 1988. The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* 8(1), 3-25.
- Fodor, J. A., 1975. *The language of thought*. Harvard University Press, Cambridge, MA.
- Fodor, J. A., 1983. *The modularity of mind*. MIT press, Cambridge, MA.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., Noll, D. C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance Medicine* 33(5), 636-647.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. *Sentence processing: Psycholinguistic essays presented to Merrill Garrett*. W. E. Cooper and C. T. Walker. Hillsdale, NJ, Erlbaum: 27-85.
- Frazier, L. (1987). *Sentence processing: A tutorial review*. Attention and performance 12: The psychology of reading. M. Coltheart. Hillsdale, NJ, England, Lawrence Erlbaum Associates: 559-586.
- Frey, S. H., Funnell, M. G., Gerry, V. E., Gazzaniga, M. S., 2005. A dissociation between the representation of tool-use skills and hand dominance: insights from left- and right-handed callosotomy patients. *Journal of Cognitive Neuroscience* 17(2), 262-272.
- Friederici, A. D., Ruschemeyer, S. A., Hahne, A., Fiebach, C. J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex* 13(2), 170-177.
- Friston, K., 2002. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual Review Neuroscience* 25, 221-250.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., Dolan, R. J., 1997. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6(3), 218-229.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. *Neuroimage* 7(1), 30-40.

- Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., Frith, C. D., 1996. Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* 4(3 Pt 1), 223-235.
- Fuhrmann Alpert, G., Hein, G., Tsai, N., Naumer, M. J., Knight, R. T., 2008. Temporal characteristics of audiovisual information processing. *Journal of Neuroscience* 28(20), 5344-5349.
- Fuster, J. M., 2003. *Cortex and mind*. Oxford University Press, New York.
- Gallagher, H. L., Frith, C. D., 2004. Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia* 42(13), 1725-1736.
- Gallese, V., Keysers, C., Rizzolatti, G., 2004. A unifying view of the basis of social cognition. *Trends in Cognitive Sciences* 8(9), 396-403.
- Gallese, V., Lakoff, G., 2005. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* 22(3-4), 455-479.
- Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. *Cognitive Brain Research* 16(2), 123-144.
- Ganis, G., Kutas, M., Sereno, M. I., 1996. The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience* 8(2), 89-106.
- Gazzaley, A., D'Esposito, M. (2007). *Unifying prefrontal cortex function: executive control, neural networks and top-down modulation. The human frontal lobes*. J. Cummings and B. Miller. New York, Guildford.
- Gazzola, V., van der Worp, H., Mulder, T., Wicker, B., Rizzolatti, G., Keysers, C., 2007. Aphasics born without hands mirror the goal of hand actions with their feet. *Current Biology* 17(14), 1235-1240.
- Gitelman, D. R., Penny, W. D., Ashburner, J., Friston, K. J., 2003. Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* 19(1), 200-207.
- Glenberg, A. M., Kaschak, M. P., 2002. Grounding language in action. *Psychonomic Bulletin and Review* 9(3), 558-565.
- Goldenberg, G., Hartmann, K., Schlott, I., 2003. Defective pantomime of object use in left brain damage: apraxia or asymbolia? *Neuropsychologia* 41(12), 1565-1573.

- Goldin Meadow, S., 2003. *Hearing gesture: How our hands help us think*. Cambridge, MA, US: Belknap Press of Harvard University Press.
- Goldin Meadow, S., Kim, S., Singer, M., 1999. What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology* 91(4), 720-730.
- Goldin Meadow, S., Momeni Sandhofer, C., 1999. Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science* 2(1), 67-74.
- Goldin-Meadow, S., Alibali, M. W., Church, R. B., 1993. Transitions in concept acquisition: using the hand to read the mind. *Psychological Review* 100(2), 279-297.
- Grafton, S. T., Arbib, M. A., Fadiga, L., Rizzolatti, G., 1996. Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Experimental Brain Research* 112(1), 103-111.
- Graham, J. A., Argyle, M., 1975. A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology* 10(1), 57-67.
- Grezes, J., Armony, J. L., Rowe, J., Passingham, R. E., 2003. Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage* 18(4), 928-937.
- Grezes, J., Costes, N., Decety, J., 1999. The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain* 122(Pt 10), 1875-1887.
- Hagoort, P., 2003a. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *NeuroImage* 20 Suppl 1, S18-29.
- Hagoort, P., 2003b. Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience* 15(6), 883-899.
- Hagoort, P. (2005a). Broca's Complex as the Unification Space for Language. *Twenty first century psycholinguistics: Four cornerstones*. A. Cutler. Mahwah, NJ, Lawrence Erlbaum Associates Publishers: 157-172.
- Hagoort, P., 2005b. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* 9(9), 416-423.

- Hagoort, P., Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. *Perspectives in sentence processing*. L. Frazier, J. Clifton Charles and K. Rayner. Hillsdale, NJ, England, Lawrence Erlbaum Associates: 45-80.
- Hagoort, P., Brown, C. M., 2000. ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia* 38(11), 1518-1530.
- Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K. M., 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304(5669), 438-441.
- Hagoort, P., van Berkum, J., 2007. Beyond the sentence given. *Philosophical Transactions Royal Society London B* 362(1481), 801-811.
- Hald, L. A., Bastiaansen, M. C., Hagoort, P., 2006. EEG theta and gamma responses to semantic violations in online sentence processing. *Brain Language* 96(1), 90-105.
- Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C., Buchel, C., 2003. The human action recognition system and its relationship to Broca's area: an fMRI study. *Neuroimage* 19(3), 637-644.
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., Rizzolatti, G., 1998. Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proceedings of the National Academy of Sciences USA* 95(25), 15061-15065.
- Harrington, G. S., Farias, D., Davis, C. H., Buonocore, M. H., 2007. Comparison of the neural basis for imagined writing and drawing. *Human Brain Mapping* 28(5), 450-459.
- Hasson, U., Nusbaum, H. C., Small, S. L., 2007. Brain networks subserving the extraction of sentence information and its encoding to memory. *Cerebral Cortex* 17(12), 2899-2913.
- Hauk, O., Johnsrude, I., Pulvermuller, F., 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41(2), 301-307.
- Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., Naumer, M. J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience* 27(30), 7881-7887.
- Herrmann, C. S., Lenz, D., Junge, S., Busch, N. A., Maess, B., 2004a. Memory-matches evoke human gamma-responses. *BMC Neuroscience* 5, 13.

- Herrmann, C. S., Munk, M. H., Engel, A. K., 2004b. Cognitive functions of gamma-band activity: memory match and utilization. *Trends in Cognitive Sciences* 8(8), 347-355.
- Holcomb, P. J., McPherson, W. B., 1994. Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition* 24(2), 259-276.
- Holle, H., Gunter, T. C., 2007. The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience* 19(7), 1175-1192.
- Holle, H., Gunter, T. C., Ruschemeyer, S. A., Hennenlotter, A., Iacoboni, M., 2008. Neural correlates of the processing of co-speech gestures. *Neuroimage* 39(4), 2010-2024.
- Huettel, S. A., Song, A. W., McCarthy, G., 2004. *Functional Magnetic Resonance Imaging*. Sinauer Associates, Sunderland, MA.
- Huynh, H., Feldt, L., 1976. Estimation of the Box correction for degrees of freedom from sample data in randomized block and splitplot designs. *Journal of Educational Statistics* 1, 69-82.
- Iacoboni, M., Dapretto, M., 2006. The mirror neuron system and the consequences of its dysfunction. *Nat Rev Neurosci* 7(12), 942-951.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. *Science* 286(5449), 2526-2528.
- Jeannerod, M., 2001. Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* 14(1 Pt 2), S103-109.
- Jensen, O., Gelfand, J., Kounios, J., Lisman, J. E., 2002. Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex* 12(8), 877-882.
- Jensen, O., Kaiser, J., Lachaux, J. P., 2007. Human gamma-frequency oscillations associated with attention and memory. *Trends in Neuroscience* 30(7), 317-324.
- Johnson-Frey, S. H., Newman-Norlund, R., Grafton, S. T., 2005. A distributed left hemisphere network active during planning of everyday tool use skills. *Cerebral Cortex* 15(6), 681-695.
- Jokisch, D., Jensen, O., 2007. Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *Journal of Neuroscience* 27(12), 3244-3251.
- Jonas, M., Siebner, H. R., Biermann-Ruben, K., Kessler, K., Baumer, T., Buchel, C., Schnitzler, A., Munchau, A., 2007. Do simple

- intransitive finger movements consistently activate frontoparietal mirror neuron areas in humans? *Neuroimage* 36 Suppl 2, T44-53.
- Josephs, O., Turner, R., Friston, K., 1997. Event-Related fMRI. *Human Brain Mapping* 5, 243-248.
- Karrasch, M., Krause, C. M., Laine, M., Lang, A. H., Lehto, M., 1998. Event-related desynchronization and synchronization during an auditory lexical matching task. *Electroencephalography and Clinical Neurophysiology* 107(2), 112-121.
- Kelly, S. D., Barr, D. J., Church, R. B., Lynch, K., 1999. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language* 40(4), 577-592.
- Kelly, S. D., Church, R. B., 1998. A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development* 69(1), 85-93.
- Kelly, S. D., Kravitz, C., Hopkins, M., 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain and Language* 89(1), 253-260.
- Kelly, S. D., Ward, S., Creigh, P., Bartolotti, J., 2006. An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*.
- Kendon, A., 2004. *Gesture: Visible action as utterance*. Cambridge University Press, Cambridge.
- Kita, S., Özyürek, A., 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48(1), 16-32.
- Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Brain Research Reviews* 29(2-3), 169-195.
- Klimesch, W., Pfurtscheller, G., Mohl, W., Schimke, H., 1990. Event-related desynchronization, ERD-mapping and hemispheric differences for words and numbers. *International Journal of Psychophysiology* 8(3), 297-308.
- Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. *Neuron* 50(6), 963-974.
- Koechlin, E., Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. *Trends Cognitive Sciences* 11(6), 229-235.

- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A. D., 2004. Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience* 7(3), 302-307.
- Koski, L., Wohlschlagel, A., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C., Iacoboni, M., 2002. Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex* 12(8), 847-855.
- Krause, C. M., Astrom, T., Karrasch, M., Laine, M., Sillanmaki, L., 1999. Cortical activation related to auditory semantic matching of concrete versus abstract words. *Clinical Neurophysiology* 110(8), 1371-1377.
- Krause, C. M., Gronholm, P., Leinonen, A., Laine, M., Sakkinen, A. L., Soderholm, C., 2006. Modality matters: the effects of stimulus modality on the 4- to 30-Hz brain electric oscillations during a lexical decision task. *Brain Research* 1110(1), 182-192.
- Krause, C. M., Porn, B., Lang, A. H., Laine, M., 1997. Relative alpha desynchronization and synchronization during speech perception. *Brain Res Cogn Brain Res* 5(4), 295-299.
- Krauss, R. M., Morrel Samuels, P., Colasante, C., 1991. Do conversational hand gestures communicate? *Journal of Personality and Social Psychology* 61(5), 743-754.
- Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., Caplan, D., 2003. Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience* 15(2), 272-293.
- Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., Lythgoe, D. J., Williams, S. C., David, A. S., 2000. Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: an fMRI study. *Journal of Cognitive Neuroscience* 12(2), 321-341.
- Kutas, M., Hillyard, S. A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207(4427), 203-205.
- Kutas, M., Hillyard, S. A., 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307(5947), 161-163.
- Kutas, M., Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. *Handbook of*

- psycholinguistics. M. A. Gernsbacher. San Diego, CA, Academic Press: 83-143.
- Lattner, S., Friederici, A. D., 2003. Talker's voice and gender stereotype in human auditory sentence processing--evidence from event-related brain potentials. *Neuroscience Letters* 339(3), 191-194.
- Lebreton, K., Desgranges, B., Landeau, B., Baron, J. C., Eustache, F., 2001. Visual priming within and across symbolic format using a tachistoscopic picture identification task: a PET study. *Journal of Cognitive Neuroscience* 13(5), 670-686.
- Lenz, D., Schadow, J., Thaerig, S., Busch, N. A., Herrmann, C. S., 2007. What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *International Journal of Psychophysiology* 64(1), 31-38.
- Longcamp, M., Anton, J. L., Roth, M., Velay, J. L., 2003. Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage* 19(4), 1492-1500.
- Longcamp, M., Anton, J. L., Roth, M., Velay, J. L., 2005. Premotor activations in response to visually presented single letters depend on the hand used to write: a study on left-handers. *Neuropsychologia* 43(12), 1801-1809.
- Longcamp, M., Tanskanen, T., Hari, R., 2006. The imprint of action: motor cortex involvement in visual perception of handwritten letters. *Neuroimage* 33(2), 681-688.
- Luck, S. J., 2005. An introduction to the event-related potential technique. MIT Press, Cambridge, MA.
- MacSweeney, M., Campbell, R., Woll, B., Brammer, M. J., Giampietro, V., David, A. S., Calvert, G. A., McGuire, P. K., 2006. Lexical and sentential processing in British Sign Language. *Human Brain Mapping* 27(1), 63-76.
- MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., Calvert, G. A., Brammer, M. J., 2004. Dissociating linguistic and nonlinguistic gestural communication in the brain. *Neuroimage* 22(4), 1605-1618.
- MacSweeney, M., Woll, B., Campbell, R., Calvert, G. A., McGuire, P. K., David, A. S., Simmons, A., Brammer, M. J., 2002a. Neural correlates of British sign language comprehension: spatial processing demands of topographic language. *Journal of Cognitive Neuroscience* 14(7), 1064-1075.

- MacSweeney, M., Woll, B., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C., Suckling, J., Calvert, G. A., Brammer, M. J., 2002b. Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain* 125(7), 1583-1593.
- Maris, E., 2004. Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology* 41(1), 142-151.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164(1), 177-190.
- Martin, A., Chao, L. L., 2001. Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology* 11(2), 194-201.
- McCarthy, R. A., Warrington, E. K., 1988. Evidence for modality-specific meaning systems in the brain. *Nature* 334(6181), 428-430.
- McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. *Nature* 264(5588), 746-748.
- McNeill, D., 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL, US: University of Chicago Press.
- McNeill, D., 2000. *Language and gesture*. Cambridge University Press, Cambridge.
- McNeill, D., Cassell, J., McCullough, K. E., 1994. Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction* 27(3), 223-237.
- McPherson, W., Holcomb, P. J., 1999. An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* 36(1), 53-65.
- Medendorp, W. P., Kramer, G. F., Jensen, O., Oostenveld, R., Schoffelen, J. M., Fries, P., 2007. Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cerebral Cortex* 17(10), 2364-2374.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. *Current Biology* 17(19), 1692-1696.
- Melinger, A., Levelt, W. J. M., 2004. Gesture and the communicative intention of the speaker. *Gesture* 4(2), 119-141.
- Menenti, L., Petersson, K. M., Scheeringa, R., Hagoort, P., under review. When elephants fly: Differential sensitivity of left and right inferior frontal gyri to discourse and world knowledge.

- Merleau-Ponty, M., [1945] 1962. *Phenomenology of Perception*. Routledge & Kegan, London.
- Mesulam, M. M., 1990. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology* 28(5), 597-613.
- Mesulam, M. M., 1998. From sensation to cognition. *Brain* 121(6), 1013-1052.
- Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., Buckner, R. L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* 11(6 Pt 1), 735-759.
- Miller, J., Patterson, T., Ulrich, R., 1998. Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35(1), 99-115.
- Mitra, P. P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. *Biophysical Journal* 76, 691-708.
- Moll, J., de Oliveira-Souza, R., Passman, L. J., Cunha, F. C., Souza-Lima, F., Andreiuolo, P. A., 2000. Functional MRI correlates of real and imagined tool-use pantomimes. *Neurology* 54(6), 1331-1336.
- Molnar-Szakacs, I., Iacoboni, M., Koski, L., Mazziotta, J. C., 2005. Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cerebral Cortex* 15(7), 986-994.
- Morrel Samuels, P., Krauss, R. M., 1992. Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(3), 615-622.
- Mottronen, R., Krause, C. M., Tiippana, K., Sams, M., 2002. Processing of changes in visual speech in the human auditory cortex. *Brain Research Cognitive Brain Research* 13(3), 417-425.
- Mottronen, R., Schurmann, M., Sams, M., 2004. Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience Letters* 363(2), 112-115.
- Nakamura, A., Maess, B., Knosche, T. R., Gunter, T. C., Bach, P., Friederici, A. D., 2004. Cooperation of different neuronal systems during hand sign recognition. *Neuroimage* 23(1), 25-34.

- Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., Braun, A., Clark, V., Jezzard, P., Turner, R., 1998. Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proceedings of the National Academy of Sciences U S A* 95(3), 922-929.
- Newman, A. J., Bavelier, D., Corina, D., Jezzard, P., Neville, H. J., 2002. A critical period for right hemisphere recruitment in American Sign Language processing. *Nature Neuroscience* 5(1), 76-80.
- Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M., Bekkering, H., 2007. The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience* 10(7), 817-818.
- Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., Shaywitz, B. A., Gore, J. C., Shankweiler, D., 2000. An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience* 12(1), 120-133.
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J. B., 2005. Valid conjunction inference with the minimum statistic. *Neuroimage* 25(3), 653-660.
- Nigam, A., Hoffman, J. E., Simons, R. F., 1992. N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience* 4(1), 15-22.
- Nishitani, N., Hari, R., 2000. Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences of the United States of America* 97(2), 913-918.
- Nishitani, N., Schurmann, M., Amunts, K., Hari, R., 2005. Broca's region: from action to language. *Physiology* 20, 60-69.
- Noppeney, U., Josephs, O., Kiebel, S., Friston, K. J., Price, C. J., 2005. Action selectivity in parietal and temporal cortex. *Brain Research Cognitive Brain Research* 25(3), 641-649.
- Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I. P., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. *Neuroimage* 25(2), 333-338.
- Oldfield, R. C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1), 97-113.
- Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., Jensen, O., 2006. Theta and gamma oscillations predict encoding

- and retrieval of declarative memory. *Journal of Neuroscience* 26(28), 7523-7531.
- Özyürek, A., 2002. Do speakers design their cospeech gestures for their addressees?: The effects of addressee location on representational gestures. *Journal of Memory and Language* 46(4), 688-704.
- Özyürek, A., Willems, R. M., Kita, S., Hagoort, P., 2007. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19(4), 605-616.
- Pecher, D., Zwaan, R. A., Eds. 2005. *Grounding cognition: The role of perception and action in memory, language, and thinking.* Cambridge University Press, Cambridge, UK.
- Peelen, M. V., Wiggett, A. J., Downing, P. E., 2006. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron* 49(6), 815-822.
- Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jaaskelainen, I. P., Kujala, T., Sams, M., 2006. Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T. *Neuroimage* 29(3), 797-807.
- Petersson, K. M., Forkstam, C., Ingvar, M., 2004. Artificial syntactic violations activate Broca's region. *Cogn Sci* 28, 383-407.
- Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., Gabrieli, J. D., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10(1), 15-35.
- Pulvermuller, F., 2005. Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 6(7), 576-582.
- Reithler, J., van Mier, H. I., Peters, J. C., Goebel, R., 2007. Nonvisual motor learning influences abstract action observation. *Current Biology* 17(14), 1201-1207.
- Rietveld, E. D. W., in press. The skillful body as a concerned system of possible actions: Phenomena and neurodynamics. *Theory and Psychology*.
- Righart, R., de Gelder, B., 2006. Context influences early perceptual analysis of faces--an electrophysiological study. *Cerebral Cortex* 16(9), 1249-1257.
- Riseborough, M. G., 1981. Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior* 5(3), 172-183.

- Rizzolatti, G., Arbib, M. A., 1998. Language within our grasp. *Trends in Neuroscience* 21(5), 188-194.
- Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. *Annual Review of Neuroscience* 27, 169-192.
- Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience* 2(9), 661-670.
- Rodd, J. M., Davis, M. H., Johnsrude, I. S., 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex* 15(8), 1261-1269.
- Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., Varela, F. J., 1999. Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397(6718), 430-433.
- Rohm, D., Klimesch, W., Haider, H., Doppelmayr, M., 2001. The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. *Neuroscience Letters* 310(2-3), 137-140.
- Ruschemeyer, S. A., Fiebach, C. J., Kempe, V., Friederici, A. D., 2005. Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping* 25(2), 266-286.
- Ruschemeyer, S. A., Zysset, S., Friederici, A. D., 2006. Native and non-native reading of sentences: an fMRI experiment. *Neuroimage* 31(1), 354-365.
- Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., Simola, J., 1991. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters* 127(1), 141-145.
- Saygin, A. P., Wilson, S. M., Dronkers, N. F., Bates, E., 2004. Action comprehension in aphasia: linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia* 42(13), 1788-1804.
- Schacter, D. L., Buckner, R. L., 1998. Priming and the brain. *Neuron* 20, 185-195.
- Schubotz, R. I., von Cramon, D. Y., 2004. Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *Journal of Neuroscience* 24(24), 5467-5474.

- Shallice, T., 1988. Specialisation within the semantic system. *Cognitive Neuropsychology* 5(1), 133-142.
- Siebner, H. R., Limmer, C., Peinemann, A., Drzezga, A., Bloem, B. R., Schwaiger, M., Conrad, B., 2002. Long-term consequences of switching handedness: a positron emission tomography study on handwriting in "converted" left-handers. *Journal of Neuroscience* 22(7), 2816-2825.
- Singer, M. A., Goldin Meadow, S., 2005. Children learn when their teacher's gestures and speech differ. *Psychological Science* 16(2), 85-89.
- Sitnikova, T., Kuperberg, G., Holcomb, P. J., 2003. Semantic integration in videos of real-world events: an electrophysiological investigation. *Psychophysiology* 40(1), 160-164.
- Spivey Knowlton, M. J., Sedivy, J. C., 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55(3), 227-267.
- Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain. Thieme Medical, New York.
- Tallon-Baudry, C., 2003. Oscillatory synchrony and human visual cognition. *Journal of Physiology Paris* 97(2-3), 355-363.
- Tanenhaus, M. K., Spivey Knowlton, M. J., Eberhard, K. M., Sedivy, J. C., 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217), 1632-1634.
- Tanenhaus, M. K., Trueswell, J. C. (1995). Sentence comprehension. *Speech, language, and communication*. J. L. Miller and P. D. Eimas. San Diego, CA, Academic Press: 217-262.
- Taraban, R., McClelland, J. L. (1990). Parsing and comprehension: A multiple-constraint view. *Comprehension processes in reading*. D. A. Balota, G. B. Flores d'Arcais and K. Rayner. Hillsdale, NJ, England, Lawrence Erlbaum Associates: 231-263.
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., Tyler, L. K., 2006. Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences USA* 103(21), 8239-8244.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., Perani, D., 2005. Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience* 17(2), 273-281.
- Tettamanti, M., Weniger, D., 2006. Broca's area: a supramodal hierarchical processor? *Cortex* 42(4), 491-494.

- Thompson, L. A., Massaro, D. W., 1986. Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology* 42(1), 144-168.
- Thompson, L. A., Massaro, D. W., 1994. Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology* 57(3), 327-354.
- Thompson-Schill, S. L., Bedny, M., Goldberg, R. F., 2005. The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology* 15(2), 219-224.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., Farah, M. J., 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences U S A* 94(26), 14792-14797.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381(6582), 520-522.
- Trueswell, J. C., Tanenhaus, M. K. (1994). *Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. Perspectives on sentence processing.* C. Clifton, Jr. Hillsdale, NJ, England, Lawrence Erlbaum Associates: 155-179.
- Tuladhar, A. M., ter Huurne, N., Schoffelen, J. M., Maris, E., Oostenveld, R., Jensen, O., 2007. Parieto-occipital sources account for the increase in alpha activity with working memory load. *Human Brain Mapping* 28(8), 785-792.
- van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. *Neuron* 43(2), 271-282.
- van Atteveldt, N. M., Formisano, E., Blomert, L., Goebel, R., 2007. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex* 17(4), 962-974.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3), 443-467.
- van Berkum, J. J. A., Hagoort, P., Brown, C. M., 1999. Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience* 11(6), 657-671.
- van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., Brown, C. M., 2003. When and how do listeners relate a sentence to the wider

- discourse? Evidence from the N400 effect. *Cognitive Brain Research* 17(3), 701-718.
- van den Brink, D., Brown, C. M., Hagoort, P., 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience* 13(7), 967-985.
- Van Petten, C., Rheinfelder, H., 1995. Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia* 33(4), 485-508.
- VanRullen, R., Thorpe, S. J., 2001. The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience* 13(4), 454-461.
- Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., Wise, R., 2006. The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex* 16(12), 1790-1796.
- Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30(4), 1414-1432.
- von Stein, A., Chiang, C., Konig, P., 2000. Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of Sciences USA* 97(26), 14748-14753.
- Vosse, T., Kempen, G., 2000. Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and lexicalist grammar. *Cognition* 75(2), 105-143.
- Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H., Gabrieli, J. D., 1997. Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. *Journal of Cognitive Neuroscience* 9(6), 714-726.
- Wagner, A. D., Pare-Blagoev, E. J., Clark, J., Poldrack, R. A., 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31(2), 329-338.
- Weiss, S., Rappelsberger, P., 1998. Left frontal EEG coherence reflects modality independent language processes. *Brain Topography* 11(1), 33-42.
- West, W. C., Holcomb, P. J., 2002. Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research* 13(3), 363-375.

- Willems, R. M., Hagoort, P., 2007. Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language* 101(3), 278-289.
- Willems, R. M., Hagoort, P., under review. Hand preference influences neural correlates of action observation.
- Willems, R. M., Oostenveld, R., Hagoort, P., 2008a. Early decreases in alpha and gamma band power distinguish linguistic from visual information during sentence comprehension. *Brain Research* 1219, 78-90.
- Willems, R. M., Özyürek, A., de Lange, F. P., Hagoort, P., under review. The role of handedness in neural representation of action meaning.
- Willems, R. M., Özyürek, A., Hagoort, P., 2007. When language meets action: the neural integration of gesture and speech. *Cerebral Cortex* 17(10), 2322-2333.
- Willems, R. M., Özyürek, A., Hagoort, P., 2008b. Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience* 20(7), 1235-1249.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* 7(7), 701-702.
- Wu, Y. C., Coulson, S., 2005. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42(6), 654-667.
- Xiong, J., Rao, S., Jerabek, P., Zamarripa, F., Woldorff, M., Lancaster, J., Fox, P. T., 2000. Intersubject variability in cortical activations during a complex language task. *Neuroimage* 12(3), 326-339.
- Yuval-Greenberg, S., Deouell, L. Y., 2007. What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience* 27(5), 1090-1096.

## Samenvatting in het Nederlands

We bewegen onze handen als we praten. Meestal valt ons dit niet op, we doen het als vanzelf. Wanneer we met iemand spreken horen we dus niet alleen spraak, maar *zien* we ook handgebaren. Onderzoek toont aan dat we informatie uit zulke handgebaren halen, dat we ze gebruiken om de boodschap van de spreker te begrijpen. In dit proefschrift heb ik bekeken hoe het brein informatie van spraak en van zulke handgebaren met elkaar verbindt.

In hoofdstuk 2 en 3 keken we naar neurale verschillen en overeenkomsten in hoe informatie die wordt overgebracht via een gebaar of via een gesproken woord wordt gecombineerd met spraak. Om dit te bestuderen lieten we gezonde deelnemers gelijktijdig luisteren en kijken naar gesproken zinnen en handgebaren. In sommige gevallen was de betekenis van een woord of een gebaar niet passend gegeven de zin die voorafging aan het woord of gebaar. Bijvoorbeeld in de zin ‘De dingen die hij op het boodschappenlijstje sloeg, moest hij niet vergeten’, is het woord ‘sloeg’ duidelijk niet op zijn plaats. Een veel voor de hand liggende zin zou zijn geweest: ‘De dingen die hij op het boodschappenlijstje *schreef*, moest hij niet vergeten’. We weten van eerder onderzoek dat zulke ‘semantische (=betekenisgerelateerde) schendingen’ tot bepaalde hersenreacties leiden die de mate uitdrukken waarin een woord ‘acceptabel’ is gegeven het eerdere deel van de zin. In de experimenten beschreven in hoofdstuk 2 en 3 was soms het woord niet acceptabel, en soms de betekenis van een gebaar. Op deze manier konden we onderzoeken hoe verwerking van semantische informatie van een woord of van een gebaar voor het brein van elkaar verschilt of niet. In hoofdstuk 2 lieten we zien dat schendingen overgedragen via een woord of via een gebaar tot eenzelfde reactie leiden in het Electro-Encephalogram (EEG) signaal van onze deelnemers. Ongeveer 400 milliseconden nadat deelnemers het ‘onacceptabele’ woord of gebaar waarnamen, was er een

verandering in het signaal in vergelijking met 'acceptabele' woorden of gebaren. De hersenen reageren dus op de mate van semantische congruentie van zowel woord als gebaar en doen dit - en dit was belangrijk - op een vergelijkbare manier. Het is dus niet zo dat bijvoorbeeld talige informatie sneller of op een andere manier gebruikt wordt. Ook informatie die wordt overgedragen via de handen leidt tot eenzelfde effect. De EEG (of ERP) methode die we gebruikten in hoofdstuk 2 heeft als voordeel dat hersenprocessen goed in de tijd te volgen zijn, tot op de milliseconde (1/1000e seconde) nauwkeurig. *Wanneer* iets gebeurt is dus goed te bepalen met deze techniek. Echter, het is zeer lastig om te bepalen welke delen van de hersenen bij een proces betrokken zijn. Om dit te onderzoeken gebruikten we in hoofdstuk 3 functionele MRI (fMRI), waarmee beter vastgesteld kan worden *waar* in het brein iets gebeurt. Nu vonden we dat een gebied in het inferieure deel van de frontale cortex meer actief werd wanneer een woord semantisch 'onacceptabel' was, alsook wanneer een gebaar 'onacceptabel' was, vergeleken met 'acceptabele' woorden en gebaren. Dit gebied is zeer belangrijk voor taal, zowel voor het produceren (bijv. spreken) als voor het begrijpen ervan. Onze conclusie was dat dit gebied betrokken is bij het samenbrengen van zowel talige informatie met waargenomen taal (woorden) als bij informatie van acties (handgebaren). We vonden ook enkele gebieden die specifiek gevoelig waren voor informatie overgebracht via een woord of via een gebaar. De algemene conclusie van hoofdstukken 2 en 3 was dat vergelijkbare hersenprocessen betrokken zijn bij het begrijpen en samenbrengen (wij gebruikten de term 'unificeren') van gesproken taal en van handgebaren. Er zijn kort samengevat gebieden in de hersenen die informatie van taal en van handgebaren combineren en ze doen dat op dezelfde manier.

In hoofdstuk 4 vroegen we ons af of hetzelfde geldt voor de combinatie van spraak met informatie die 'op zichzelf kan staan'. We kwamen tot deze vraag omdat het bekend is dat spraakgerelateerde

handgebaren alleen begrepen worden als ze gecombineerd zijn met taal. Dat wil zeggen, als we mensen vragen te beschrijven waar iemand het over heeft en we alleen de handgebaren van de spreker laten zien, dan vinden de meeste mensen dit heel lastig. Echter, met spraak begrijpt men de boodschap wel en begrijpt men ook wat de handgebaren toevoegen aan het gesprokene. Kortom, handgebaren hebben spraak nodig om begrepen te worden (dit is uiteraard niet het geval voor alle handacties, waarover later meer). Een plaatje van een object zoals een kom kan makkelijk begrepen worden zonder spraak en is in die zin anders dan een handgebaar. Plaatjes en handgebaren hebben echter allebei een niet-talige verschijningsvorm. Onze vraag in hoofdstuk 4 was of we dezelfde effecten zouden zien voor 'onacceptabele' plaatjes als voor 'onacceptabele' woorden en of deze effecten vergelijkbaar zouden zijn met die gevonden voor 'onacceptabele' gebaren. Opnieuw maten we *wanneer* (met behulp van EEG / ERP) en *waar* (met behulp van fMRI) de betekenis van plaatjes en woorden wordt samengebracht met de eerdere zins-informatie in het brein. Kort gezegd vonden we dezelfde effecten voor plaatjes als voor handgebaren (hoofdstuk 2 en 3). Opnieuw zagen we dat na ongeveer 400 milliseconden een verschil waarneembaar was in de ERP-response van 'onacceptabele' woorden of plaatjes vergeleken met 'acceptabele' woorden of plaatjes. In de fMRI metingen zagen we wederom dat hetzelfde gebied meer actief wordt voor onacceptabele plaatjes en woorden als voor acceptabele plaatjes en woorden, namelijk het inferieure deel van de frontale cortex. Onze conclusie was dan ook dat niet alleen het samenbrengen van talige informatie en informatie overgebracht via handgebaren, maar ook van talige informatie en informatie overgebracht via plaatjes op vergelijkbare wijze in het brein verloopt.

Echter, het lijkt erg onwaarschijnlijk dat plaatjes, gebaren en woorden voor de hersenen allemaal hetzelfde zijn. Een indicatie voor hoe deze informatiedragers op verschillende wijze in het brein verwerkt

worden kregen we al van de fMRI bevindingen, waarin sommige gebieden specifiek gevoelig bleken voor informatie van woorden of handgebaren. In hoofdstuk 5 onderzochten we of verschillen in neurale activiteit in verschillende frequentiebanden van het EEG-sig-naal wellicht ook een rol spelen in het onderscheiden van verschillende wijzen waarop betekenisvolle informatie door het brein verwerkt wordt. Hiervoor analyseerden we de EEG-data van hoofdstuk 4 op een anderen manier dan gebruikelijk is. We deelden het EEG-sig-naal op in 'banden' van verschillende frequentie. Zo zijn er banden waar het sig-naal slechts zeer langzaam fluctueert (lage frequenties), maar er zijn ook banden waarin het sig-naal zeer snel fluctueert (hoge frequenties). Een analogie om dit te illustreren is geluid. We weten dat een lage toon te beschrijven valt als een sinusgolf met een lage frequentie. Een hoge toon daarentegen is uit te drukken als een golfpatroon met een veel hogere frequentie. Eerder onderzoek heeft aangetoond dat verschillende cognitieve processen in verschillende banden van het EEG sig-naal een effect kunnen hebben. Dit vonden we ook voor de zinnen met 'onacceptabele' woorden in vergelijking met zinnen met 'onacceptabele' plaatjes. Na een onacceptabel woord zagen we al na zo'n 50 milliseconden een verschil in de alpha-frequentie band (rond 8 Hz) die niet aanwezig was na een onacceptabel plaatje. Na een onacceptabel plaatje zagen we evenzo snel een verschil in de gamma-frequentie band (rond 50 Hz) die niet aanwezig was na een onacceptabel woord. Overigens zagen we voor zowel onacceptabele woorden als voor onacceptabele plaatjes een later (+- 600 ms na het onacceptabele woord of plaatje) verschil in de theta-frequentieband (rond 5 Hz). Behalve een vergelijkbaar effect (theta-frequentie) vonden we dus ook een effect dat specifiek was voor woorden (alpha-frequentie) en een specifiek effect voor plaatjes (gamma-frequentie). Het lijkt er dus op dat het brein zeer snel een 'onacceptabel' woord of plaatje detecteert en dit codeert in verschillende frequentiebanden, afhankelijk van de modaliteit waarin de informatie werd overgebracht. Vervolgens

is het mechanisme waarmee de informatie van een woord of plaatje wordt samengebracht ('ge-unificeerd') met de eerdere inhoud van de zin hetzelfde.

Het startpunt voor het onderzoek in hoofdstuk 6 was het hierboven beschreven verschijnsel dat de betekenis van spraakgelateerde gebaren niet duidelijk herkenbaar is wanneer ze worden waargenomen zonder spraak. In dit hoofdstuk onderzochten we of het samenbrengen van taal- en actie-gerelateerde informatie anders is voor gebaren vergeleken met pantomimen. Met pantomimen worden acties bedoeld waarin iemand een handeling uitbeeldt zonder het object te gebruiken dat normaal gesproken voor de handeling nodig is. Bijvoorbeeld als iemand voordoet hoe een hamer gebruikt wordt kan zij dit demonstreren door met de ene hand de denkbeeldige hamer op en neer te bewegen en met de andere hand een denkbeeldige spijker vast te houden. De betekenis van pantomimen is over het algemeen makkelijk herkenbaar zonder spraak. In hoofdstuk 6 gebruikten we fMRI om te meten of dit verschil in afhankelijkheid van taal tussen deze twee actie-soorten tot uitdrukking komt in de activiteit van verschillende hersengebieden. We lieten deelnemers luisteren en kijken naar combinaties van spraak en gebaren enerzijds en naar combinaties van spraak en pantomimen anderzijds. We manipuleerden de congruentie tussen acties en taal: soms pasten de gebaren en de spraak goed bij elkaar en soms niet. Hetzelfde gold voor de pantomimen: soms was de spraak een goede beschrijving van de pantomime, soms niet. Het idee hierachter was dat we op deze manier gebieden konden achterhalen die betrokken zijn bij integratie van informatie van spraak en van acties. We vonden dat een gebied in het inferieure deel van de frontale cortex (dat we ook meer actief zagen worden in hoofdstukken 3 en 4) meer actief werd als reactie op zowel incongruente spraak-gebaar als incongruente spraak-pantomime combinaties. Echter een gebied in het posterieure deel van de superieure temporale cortex was alleen maar gevoelig voor congruentie tussen spraak-pantomime combinaties

en niet voor spraak-gebaar combinaties. Van dit gebied is bekend dat het betrokken is bij integratie van informatie van verschillende modaliteiten zoals lipbewegingen en spraak of van letters en hun klank. Deze bevindingen laten zien dat integratie van taal- en actie-gerelateerde informatie op verschillende wijze in het brein plaatsvindt, onder andere afhankelijk van hoe sterk de acties gerelateerd zijn aan taal. Verder geven deze bevindingen een aanwijzing voor de verschillende rol die deze beide gebieden wellicht spelen in processen betrokken bij integratie van informatie van verschillende modaliteiten.

In hoofdstuk 7 onderzochten we neurale processen die een rol spelen bij het begrijpen van acties die gepresenteerd werden zonder taal. We gebruikten fMRI om hersenprocessen te meten terwijl deelnemers pantomimen (zoals hierboven beschreven) observeerden. In een deel van de pantomimen was een daadwerkelijke actie te herkennen. Bij een ander deel hadden we de handvorm van de pantomime zodanig veranderd dat de betekenis onduidelijk werd. De hoofdvraag van dit onderzoek kwam voort uit een theorie die stelt dat we acties begrijpen door een geobserveerde actie te simuleren. Dat wil zeggen, we doen de actie van een ander impliciet na in ons eigen brein om de handeling te kunnen begrijpen. Een vraag die we hadden was dan ook of gebieden in de hersenen die betrokken zijn bij motoriek ook actief zouden worden als we de betekenis van een handeling (pantomime in dit experiment) begrijpen. Voor als dit inderdaad zo blijkt te zijn, hadden we een tweede vraag, namelijk *hoe* simulatie een rol speelt bij het begrijpen van de betekenis van een handeling. Een interpretatie van simulatie kan bijvoorbeeld zijn dat de waargenomen actie door de waarnemer wordt gesimuleerd op de manier waarop hij / zij deze handeling zou uitvoeren. Als we bijvoorbeeld iemand haar veters zien strikken, begrijpen we dit dan door de manier waarop zij haar schoenen strikt te kopiëren? Of simuleren we deze handeling op de manier waarop we zelf onze schoenen strikken? Een manier die wellicht net anders kan zijn dan dat we waarnemen bij de ander. Feit

is dat het ons in het voorbeeld geen enkele moeite kost om te begrijpen wat de ander doet. De vraag die we ons stelden is hoe dit begrip gerealiseerd wordt in het brein. Om dit te onderzoeken deden we metingen bij zowel links- als rechtshandigen terwijl ze de pantomimen observeerden zoals boven beschreven. Logischerwijs konden alle deelnemers de betekenis van de pantomimen eenvoudig begrijpen, het verschil tussen de groepen was dat ze een verschillende voorkeurshand hadden waarmee ze de geobserveerde acties zelf uit zouden voeren (de deelnemers namen in ons experiment slechts waar, ze hoefden geen acties uit te voeren). Het is bekend dat er lateralisatie van motorische gebieden is. Dat wil zeggen dat acties met de linkerhand voor het overgrote deel door de rechterhemisfeer aangestuurd worden. Daarom kunnen er verschillen tussen links- en rechtshandigen verwacht worden als ze tenminste de geobserveerde acties simuleren op de manier waarop ze hem zelf uit zouden voeren. We vonden dat motorische gebieden inderdaad meer actief werden als de pantomimen een duidelijke betekenis hadden in vergelijking met de acties waarvan de betekenis niet duidelijk was. Echter dit was niet verschillend voor links- en rechtshandigen. Deze bevindingen laten zien dat hoewel we ons motorisch systeem gebruiken voor het begrijpen van de betekenis van een handeling, we dit niet doen door simulatie op exact de manier waarop wij zelf de handeling uit zouden voeren. Het lijkt er eerder op dat de manier waarop betekenis gerepresenteerd is in het brein als het ware 'losgezongen' is van de wijze waarop we de waargenomen handeling uitvoeren en dat we een waargenomen handeling simuleren op het niveau van de betekenis of het doel van de handeling in plaats van de manier waarop de handeling wordt uitgevoerd.



## **Acknowledgements**

I have greatly benefited from the help of several people in the work that is reported in this thesis. First and foremost I want to acknowledge the support of my two supervisors, Aslı Özyürek and Peter Hagoort. Aslı is especially thanked for her constructive and always very critical evaluations of my research plans and draft manuscripts. I thank Peter for giving me the freedom to pursue my ideas while at times either gently steering the wheel or firmly pushing the brakes when necessary. The supervision of both of you is characterised by ‘critical trust’, which I greatly appreciate.

The members of the manuscript committee have evaluated my thesis. I am very thankful to Harold Bekkering, Jos van Berkum and Sotaro Kita that they were willing to spend time and effort doing this.

I want to thank my colleagues at the Donders Institute, Centre for Cognitive Neuroimaging. I found a warm and intellectually rich nest at (what was formerly known as) the FCDC. I am thankful that there were many people around who wanted to discuss, answer my questions, help me out with the crucial details of doing research and above all question what I thought was obvious. Then it turns out that many of these people are clever and nice at the same time. Specifically I want to thank Giosuè Baggio, Pieter Buur, Rick Helmich, Floris de Lange, Ali Mazaheri, Karl-Magnus Petersson and Marieke Schölvink. Special Honourful thanks go out to Markus Bauer.

A special word of thanks to Jens Schwarzbach who has stimulated my enthusiasm for research by means of his own enthusiasm.

Petra van Alphen ‘voiced’ the sentences that we presented to our subjects. Nina Davids, Femke Deckers and Cathelijne Tesink ‘gestured’ their way into this thesis. Many tanks to all of you.

Over the years I happily shared an office with (in order of appearance): Sandra van Aalderen, Hubert Fonteijn, Christian Forkstam, Carinne Piekema, Liu Xiao, Kirsten Weber and Erno

Hermans. I enjoyed being roommates with you and I will refrain from complaining about the times Hubert locked me in.

Non-scientists often believe that science is only about thinking; a stereotype which is beautifully illustrated on the front cover. This is - unfortunately, I must say - not true. Successful research depends to a large degree upon what may seem 'side-issues' to an outsider: good-working hardware, computer infrastructure, and administration. The research described in this thesis is no exception and I therefore want to thank the members of the technical and administrative groups for their important contribution to my work. I am most grateful for the flexible and efficient way in which they took care of issues that I am sure would have turned to disaster if left to my own hands. Paul Gaalman is specifically acknowledged for his expert assistance during the scanning sessions.

The Rijksmuseum Amsterdam is acknowledged for kindly providing a reprint of the engraving on the front cover.

The Tuna universitaria de Maastricht has enriched my life in a way I could not have imagined when first joining a Tuesday evening rehearsal almost seven years ago. I want to thank all its members for sharing the Tuna tradition and spirit and for only occasionally mentioning my musical incapacabilities.

Doing research has given me profound joy and satisfaction - notwithstanding the also very profound annoyances that can be part of the job. It is a privilege to have had the opportunity to do the work described in this thesis. However, to be surrounded by, and to take part in the lives of your loved ones is a blessing. I want to thank my 'Sittard friends' and Renske Wassenberg and Ruud Pijnenburg for their friendship, my parents for the loving environment in which I grew up, the Willems, Heijnen and Davids families for being my family, and - of course and most dearly - Nina for her love.

## **Publications**

**Willems, R. M., Özyürek, A., & Hagoort, P.** (under review).

Differential roles of left inferior frontal and superior temporal cortex in multimodal integration of action and language.

**Willems, R. M., Özyürek, A., De Lange, F. P. & Hagoort, P.** (under review). Neural representation of action meaning: Automaticity and the role of handedness.

**Willems, R. M., & Hagoort, P.** (under review). Hand preference influences neural correlates of action observation

**Willems, R. M., Oostenveld, R., & Hagoort, P.** (2008). Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension. *Brain Research* 1219, 78-90

de Lange F. P., Spronk M., **Willems R. M.**, Toni I., Bekkering H. (2008). Complementary systems for understanding action intentions. *Current Biology*, 18(6), 454-457

**Willems, R. M., Özyürek, A., & Hagoort, P.** (2008). Similar neural correlates for the integration of words and pictures at the sentence level: Evidence from ERPs and fMRI. *Journal of Cognitive Neuroscience* 20(7), 1235-1249

**Willems, R. M., Özyürek, A., & Hagoort, P.** (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, 17(10) 2322-2333

Özyürek, A., **Willems, R. M.**, Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19(4), 605-616

**Willems, R. M.**, & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture and action: A review. *Brain and Language*, 101(3), 278-298.

**Willems, R. M.** (2007). The neural construction of a Tinkertoy. *Journal of Neuroscience* 27(7), 1509-1510 [*Journal club* review]

**Curriculum Vitae**

Roel Willems (Limbricht, 1980) finished his secondary education at the Gymnasium of College Sittard in 1998. He got his master's degree in psychology from the Universiteit Maastricht in 2003. In September 2003 he became a PhD student in the Neurocognition of Language group at the Donders Institute for Brain, Cognition and Behaviour in Nijmegen, The Netherlands. This thesis is the result of his work in that period. Currently he is a post-doctoral researcher in the Joint Action in Science and Technology project at the same institute.

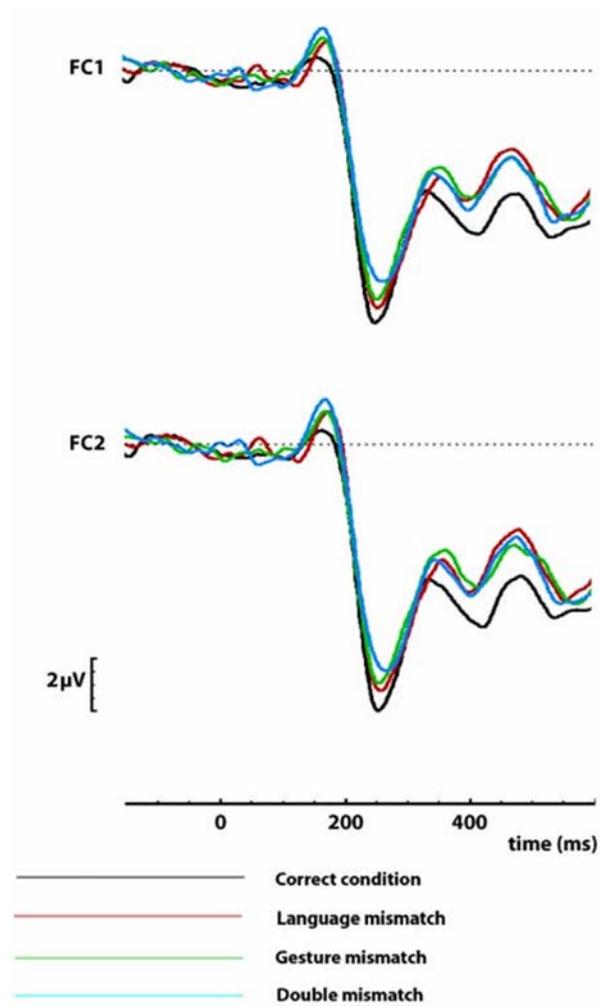


## Series Donders Institute for Brain, Cognition and Behaviour

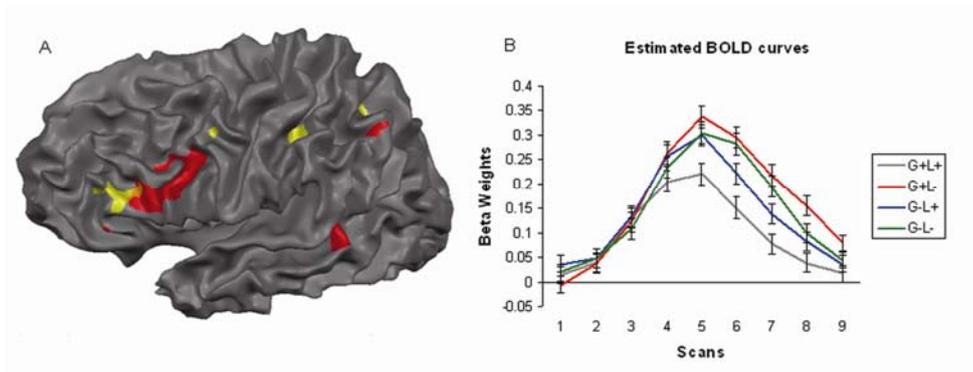
1. van Aalderen-Smeets, S.I. (2007). *Neural dynamics of visual selection*. Maastricht University, Maastricht, The Netherlands.
2. Schoffelen, J.M. (2007). *Neuronal communication through coherence in the human motor system*. Radboud University, Nijmegen, The Netherlands.
3. de Lange, F.P. (2008). *Neural mechanisms of motor imagery*. Radboud University Nijmegen, The Netherlands.
4. Grol, M.J. (2008). *Parieto-frontal circuitry in visuomotor control*. University Utrecht, Utrecht, The Netherlands.
5. Bauer, M. (2008). *Functional roles of rhythmic neuronal activity in the human visual and somatosensory system*. Radboud University Nijmegen, The Netherlands.
6. Mazaheri, A. (2008). *The influence of ongoing oscillatory brain activity on evoked responses and behaviour*. Radboud University Nijmegen, The Netherlands.
7. Hooijmans, C.R. (2008). *Impact of nutritional lipids and vascular factors in Alzheimer's Disease*. Radboud University Nijmegen, The Netherlands.
8. Gaszner, B. (2008). *Plastic responses to stress by the rodent urocortineric Edinger-Westphal nucleus*. Radboud University Nijmegen, The Netherlands.
9. Willems, R.M. (2009). *Neural reflections of meaning in gesture, language, and action*. Radboud University Nijmegen, The Netherlands.



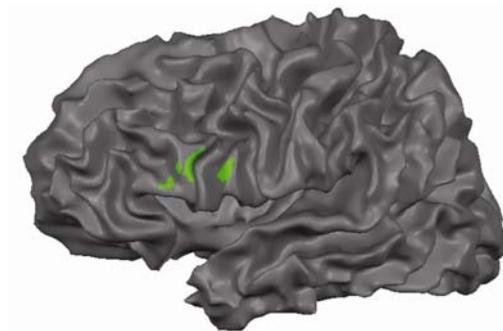
## Appendix: Colour figures



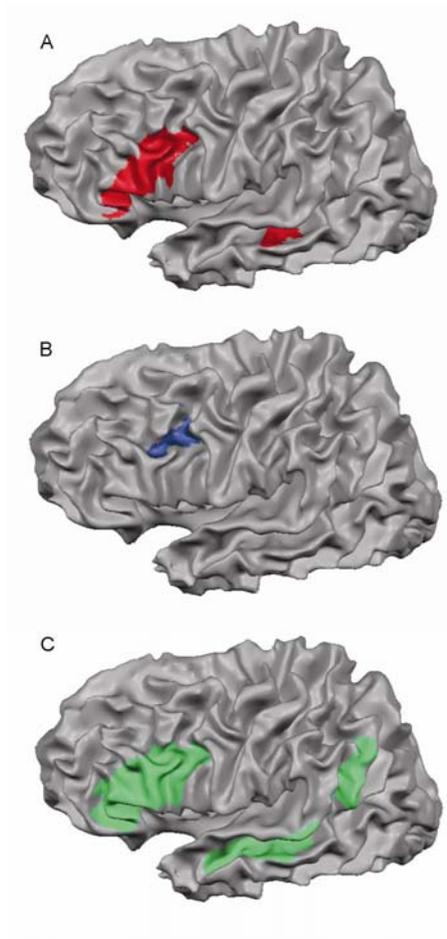
**Fig. 2.2.** Grand-average waveforms for ERPs elicited in the three mismatch conditions and the correct condition at two representative electrode sites (FC1 and FC2). Negativity is plotted upwards. Waveforms are time locked to the onset of spoken verb and gesture (0 ms).



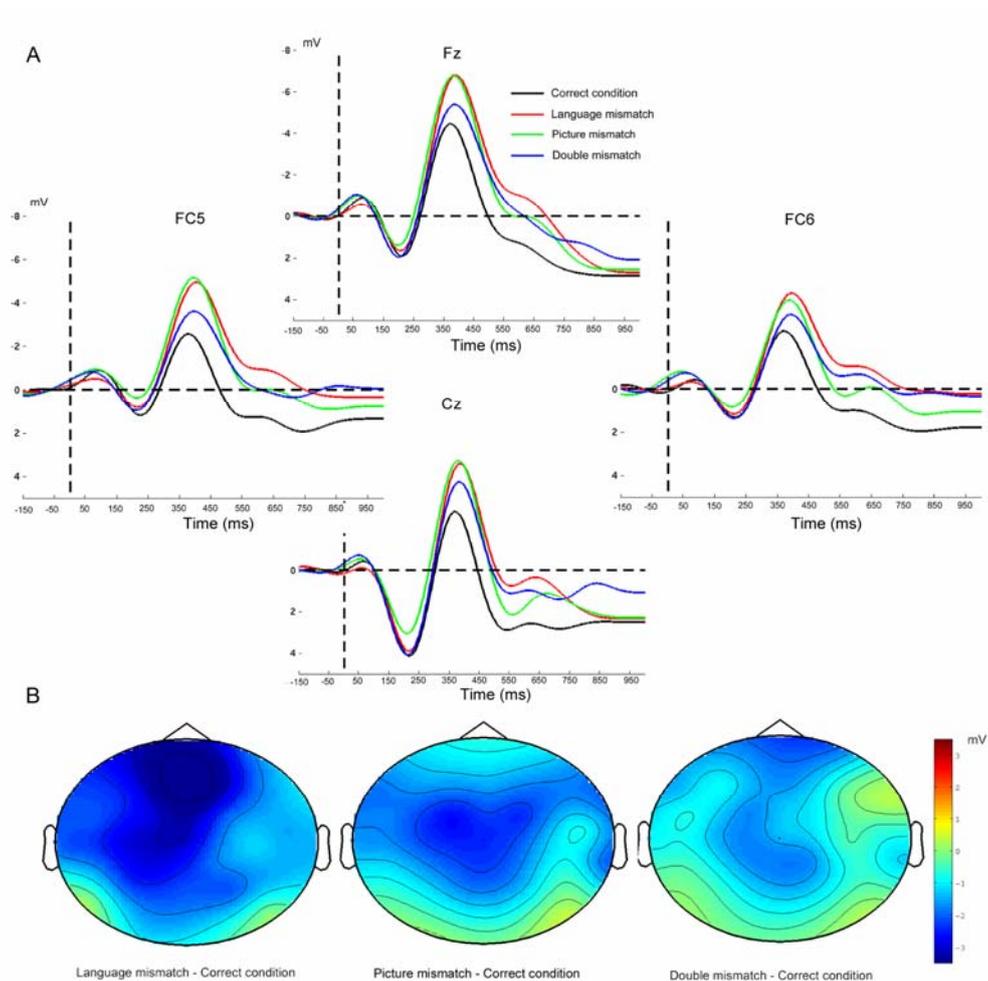
**Fig. 3.3.** Gesture and speech in a sentence context. **A)** Significant activations in the whole brain analysis for the language mismatch versus correct (red) and the gesture mismatch versus correct (yellow) comparisons. Note the overlap in inferior frontal cortex (BA 45, [x y z] [-46 23 25]). Maps are thresholded at  $t(15) > 3.5$ ,  $p < 0.05$  (corr.). No activations were found in the right hemisphere. **B)** Blood Oxygen Level Dependent (BOLD) curves from the activated regions in left inferior frontal cortex (centre coordinates [x y z] [-43 11 26]). This region is also activated in the correct condition (grey line), but more so in reaction to a semantic mismatch (red, blue and green lines). G+L+: correct condition, G+L-: language mismatch, G-L+, gesture mismatch, G-L-: double mismatch.



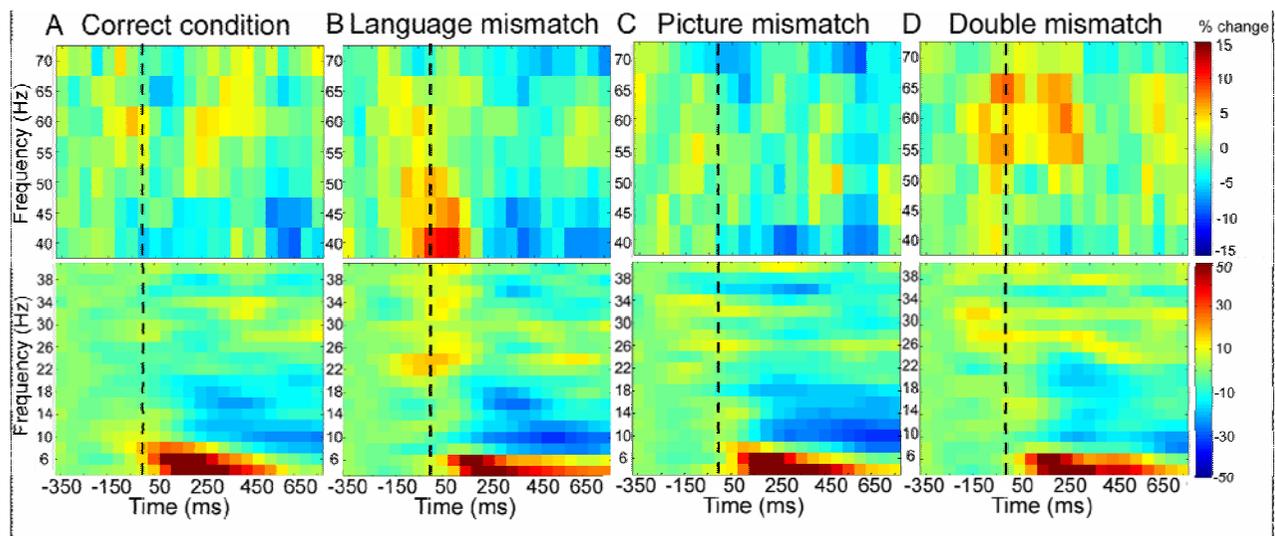
**Fig. 3.4.** Significant activations in the whole brain random effects analysis in the double mismatch versus correct condition contrast. The inferior frontal area and the precentral area in purple were more strongly activated by the double mismatch (G-L-) condition than by the correct condition (G+L+). Map is thresholded at  $t(15) > 3.5$ ,  $p < 0.05$  (corrected) and projected onto the cortical sheet of one of the participants. No activations were found in the right hemisphere.



**Fig. 4.3.** Results from the fMRI whole brain random effects group analysis ( $t(15) > 3.9$ ,  $p < 0.05$ , corrected). Areas significantly activated in the **A)** Language mismatch versus Correct condition contrast (red), **B)** Picture mismatch versus Correct condition contrast (blue), **C)** Double mismatch versus Correct condition contrast (green). Results are overlain on a cortical sheet segmented along the grey-white matter border in stereotaxic (Talairach) space.



**Fig. 5.1 A)** Averaged event-related potentials time-locked to the onset of the critical word. Presented are the waveforms from electrodes FC5 (left), Fz (upper), FC6 (right) and Cz (lower) of all four conditions. The increased negativity of the collared lines (Mismatch conditions) as compared to the black line (Correct condition) is clearly visible. Negative is plotted upwards. Waveforms are low-pass filtered for illustration purposes only. **B)** Scalp topographies of the N400 effects in the 300-600 ms range for the Language mismatch-Correct condition (left), Picture mismatch-Correct condition (middle) and Double mismatch-Correct condition (right) comparisons. Note the more anterior distribution than is normally observed for the N400 effect elicited by spoken or written words.



**Fig. 5.2.** Time-frequency representations of the four conditions. Power is normalized with respect to power in the -350 to 0 ms time window in each frequency band by computing the relative change (percent signal change) as compared to the baseline condition for each frequency band separately. That is, baseline correction involved subtracting the mean of the baseline of that specific frequency band from the measured value and dividing this number by the mean power in the baseline (value - baseline / baseline). Therefore, the values in the figure represent percentage power change as compared to baseline. It was made sure that no post-stimulus activation was included in the baseline period due to conversion into the frequency domain. Although instructive, this figure does not clearly illustrate the differences between conditions. These are displayed in Figure 5.3.

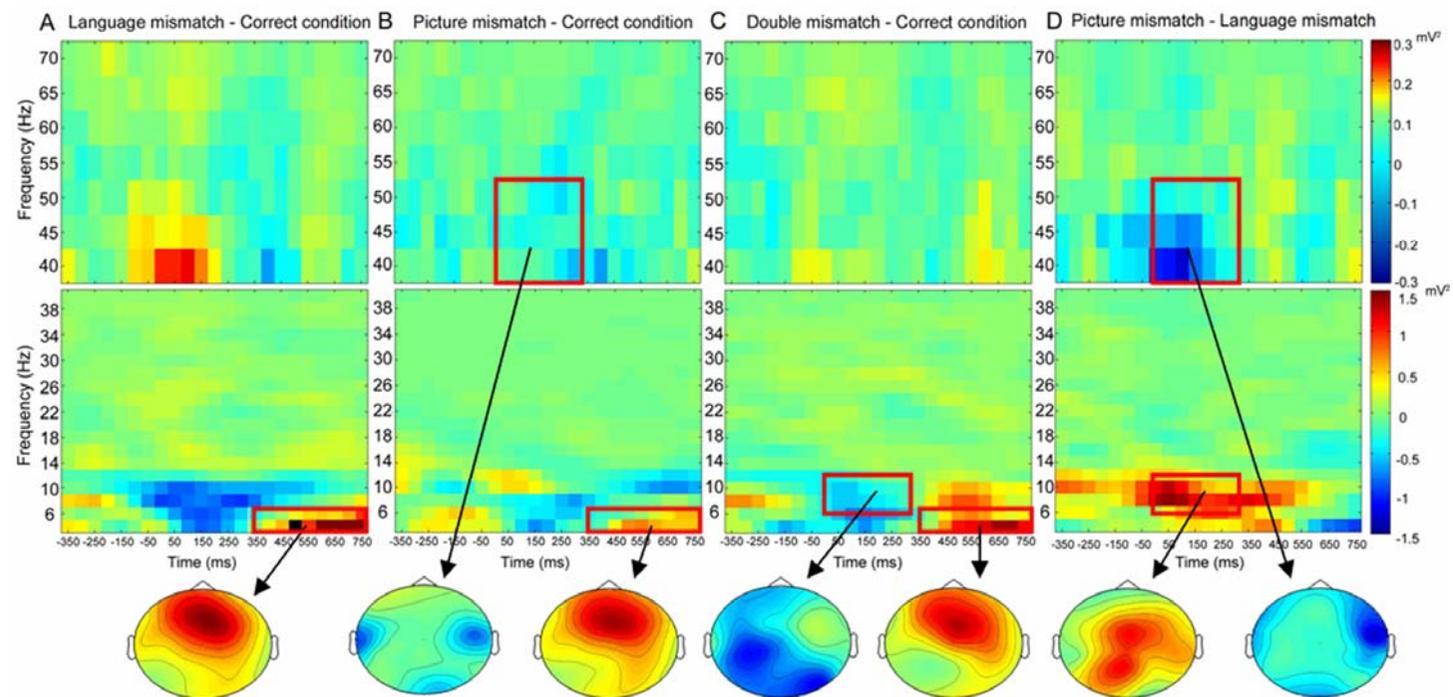
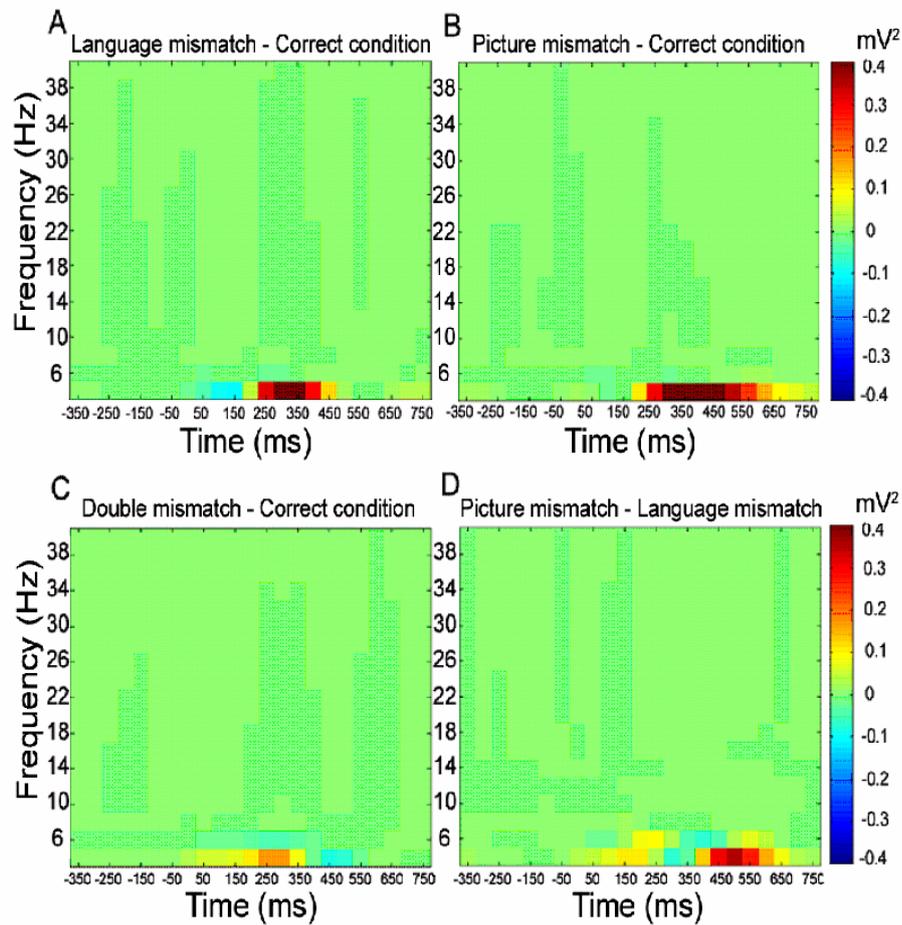
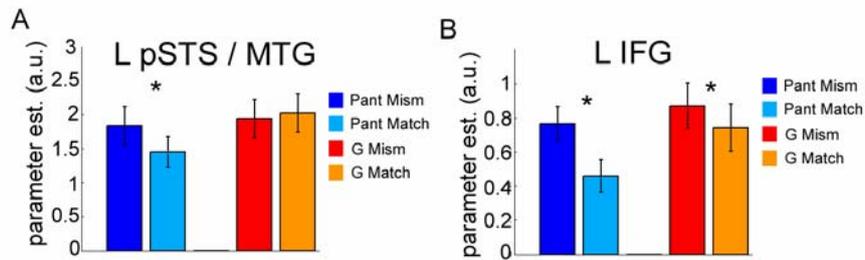


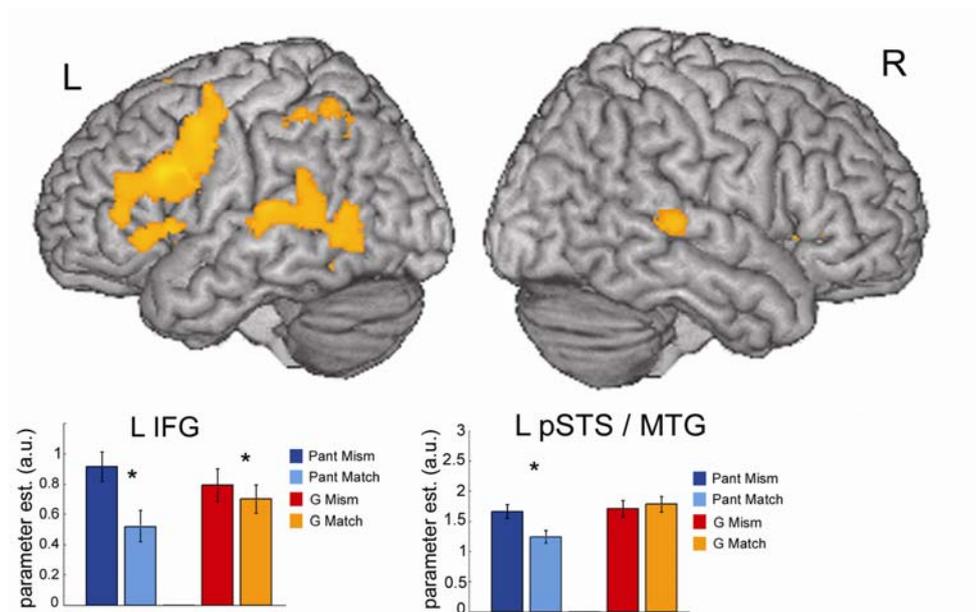
Fig. 5.3. Average time-frequency representations of **A)** Language mismatch-Correct condition, **B)** Picture mismatch-Correct condition, **C)** Double mismatch-Correct condition and **D)** Picture mismatch-Language mismatch condition. Time-frequency clusters (defined a priori based upon previous literature) in which the particular mismatch differed from the Correct condition are indicated with a red square. Displayed is the average power difference over all electrodes. Scalp topographies of significant differences between conditions are also displayed. Note the difference in scaling between lower and higher frequencies.



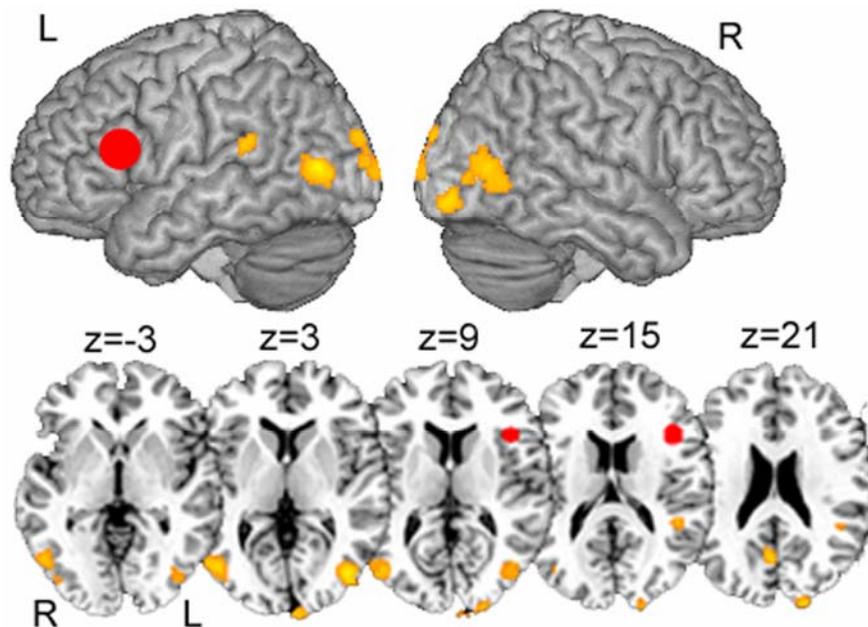
**Fig. 5.4.** Time-frequency representation of the averaged Event-Related Potentials. Displayed are the TFRs of the difference waves of the **A)** Language mismatch-Correct condition, **B)** Picture mismatch-Correct condition, **C)** Double mismatch-Correct condition and the **D)** Picture mismatch-Language mismatch condition. The manifestation of the N400 as an increase in power around 4 Hz is clearly visible. TFRs were created by applying the same analysis procedure for the averaged ERP difference waves as used in the time-frequency analysis of the single trial data.



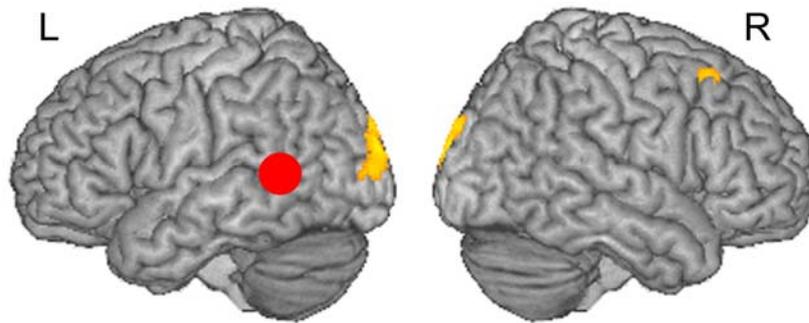
**Fig. 6.2.** Results in Regions of Interest. Mean parameter estimates of all bimodal conditions in left pSTS / MTG (A) and LIFG (B), averaged over all voxels in the ROI. A) In left pSTS / MTG there was a difference between mismatching and matching Pantomime-Speech combinations (mismatch: dark blue, match: light blue), but not between mismatching and matching Gesture-Speech combinations (mismatch: red; match: orange). On the contrary, in LIFG, there was an influence of congruence both in the Speech-Pantomime combinations as well as in the Speech-Gesture combinations (B). A.u.: arbitrary units.



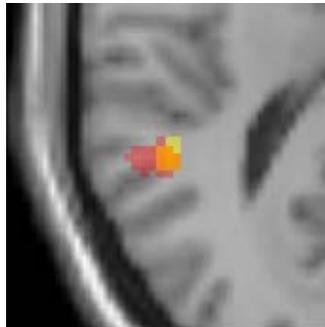
**Fig. 6.3.** Areas activated in whole brain analysis to the Pant-Mism versus Pant-Match contrast. Activation levels (parameter estimates in arbitrary units (a.u.)) of the clusters of activation in LIFG and left pSTS are displayed. Analysis in these clusters confirms the results from the analysis with a priori defined regions of interest: in pSTS there only is a difference between Pant-Mism and Pant-Match, but in LIFG there is a significant difference between Pant-Mism and Pant-Match as well as between Gest-Mism and Gest-Match.



**Fig. 6.4.** Results of effective connectivity analysis taking the a priori defined region of interest in LIFG as seed region, indicated in red. Statistical maps are thresholded at  $p < 0.05$ , corrected for multiple comparisons and overlain on a rendered brain. Areas that are more strongly modulated by LIFG in the Pant-Mism condition as compared to the Pant-Match condition. The rendered image is possibly misleading since it displays activations at the surface of the cortex that are actually 'hidden' in sulci. Therefore, we also display the result on multiple coronal slices. In the latter view, localisation of the activation in pSTS is more straightforward. No areas were found to be more strongly modulated by LIFG in the Gest-Mism as compared to Gest-Match condition.



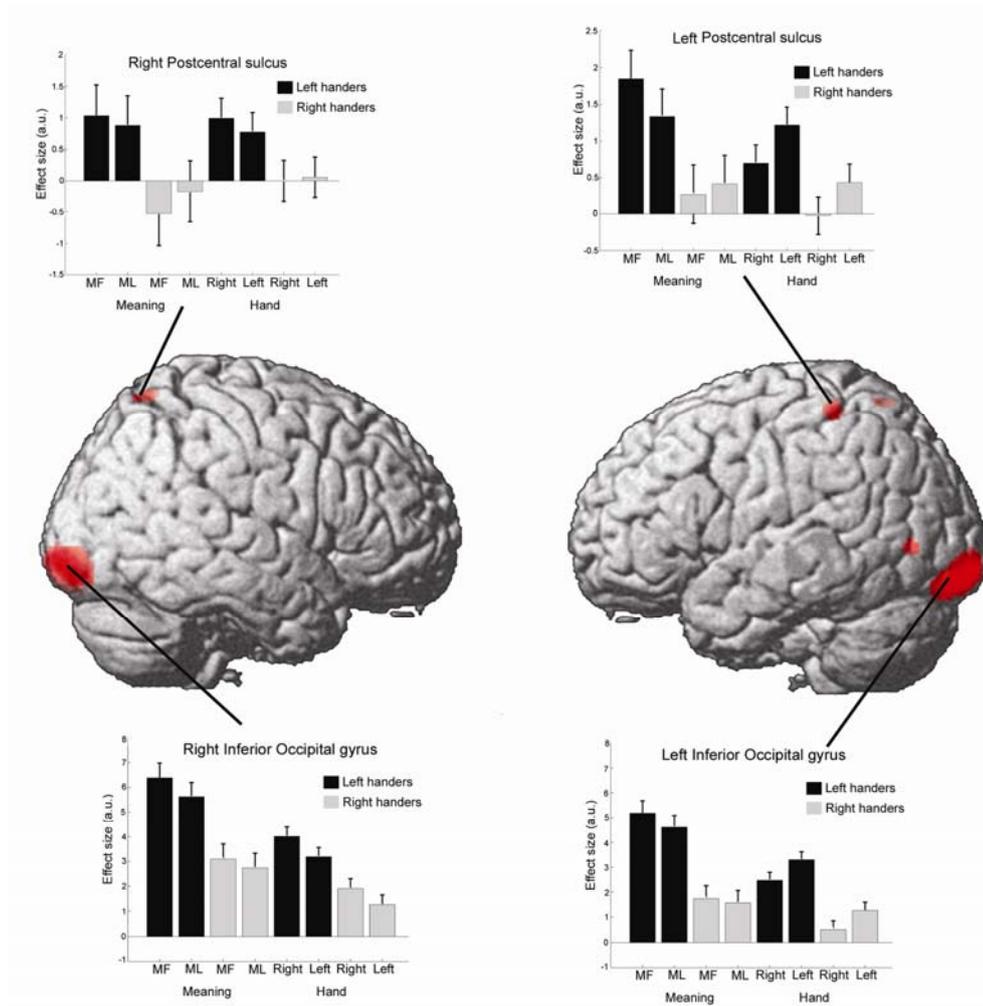
**Fig. 6.5.** Results of effective connectivity analysis taking the a priori defined region of interest in left pSTS / MTG as seed region. The ROI is displayed in red. Statistical maps are thresholded at  $p < 0.05$ , corrected for multiple comparisons and overlain on a rendered brain. Left middle occipital gyrus and right superior frontal gyrus were more strongly modulated by left pSTS in the Pant-Mism condition as compared to the Pant-Match condition. No areas were found to be more strongly modulated by left pSTS / MTG in the Gest-Mism as compared to Gest-Match condition.



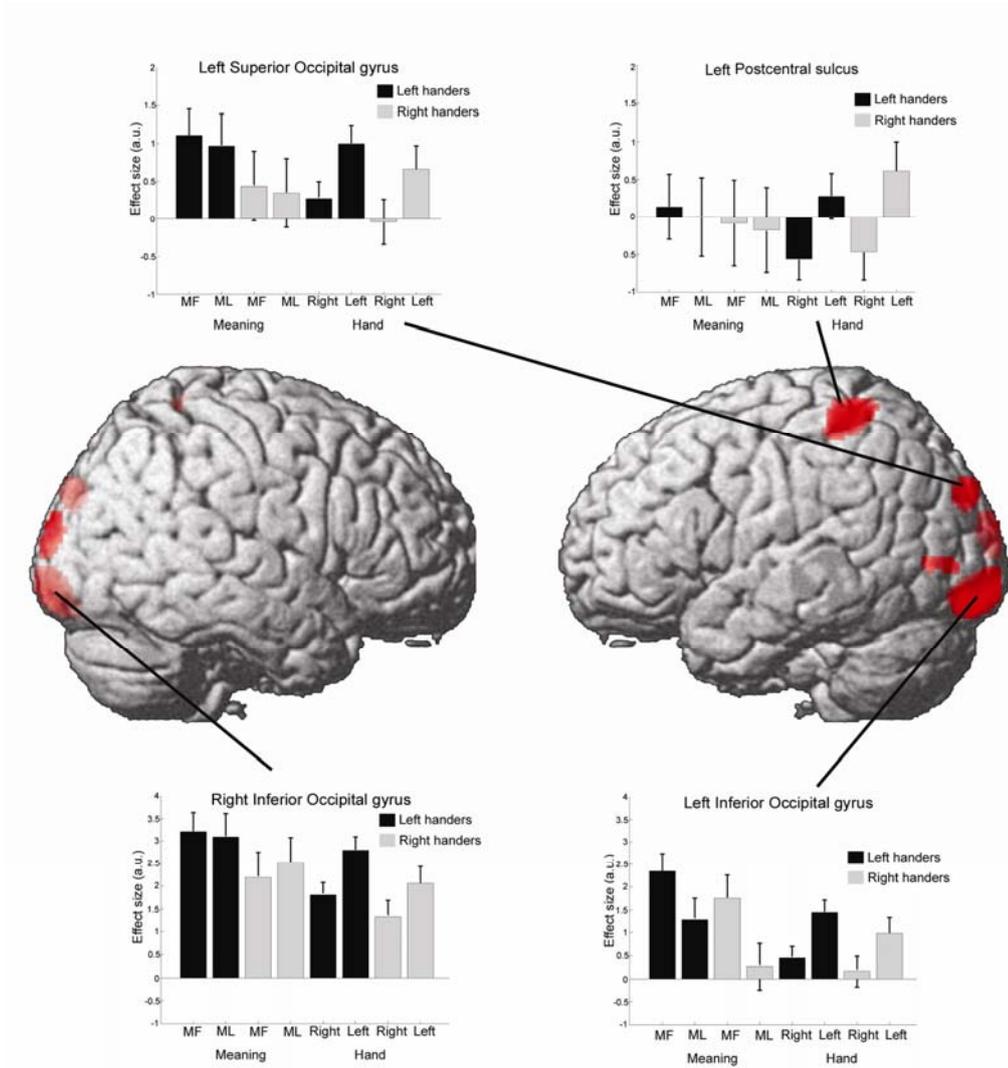
**Supplementary Fig. S6.1.** Visualisation of the overlap in pSTS of activation in the main contrast Pant-Mism versus Pant-Match (Red) and influence of LIFG onto pSTS as revealed in the effective connectivity analysis (Yellow).



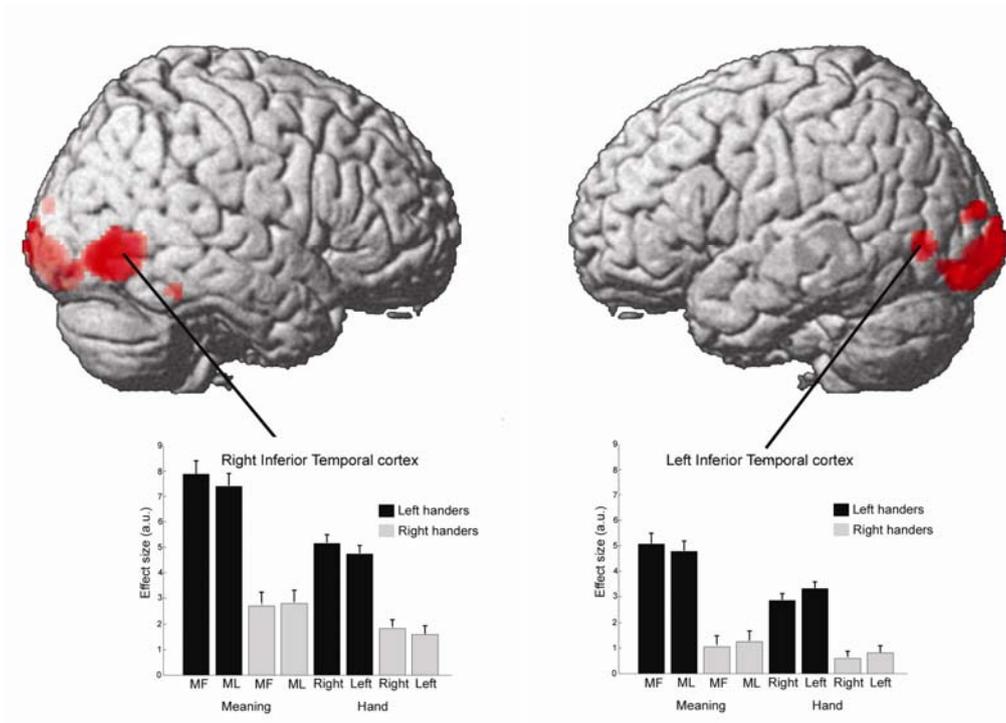
**Fig. 7.3.** (Previous page). Neural differences between meaningful and meaningless actions in the run in which participants had to indicate whether the action was meaningful or not (main effect of Meaning). Activation levels are strongest for the meaningful actions in all areas. Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI coordinates) in right precentral sulcus (32 -4 56), left precentral sulcus (-42 -2 50), left inferior parietal sulcus (-26 -68 46), right inferior occipital gyrus (24 -98 -6), left inferior frontal / ventral premotor cortex (-40 4 3) and left inferior occipital gyrus (-20 -100 -8). Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



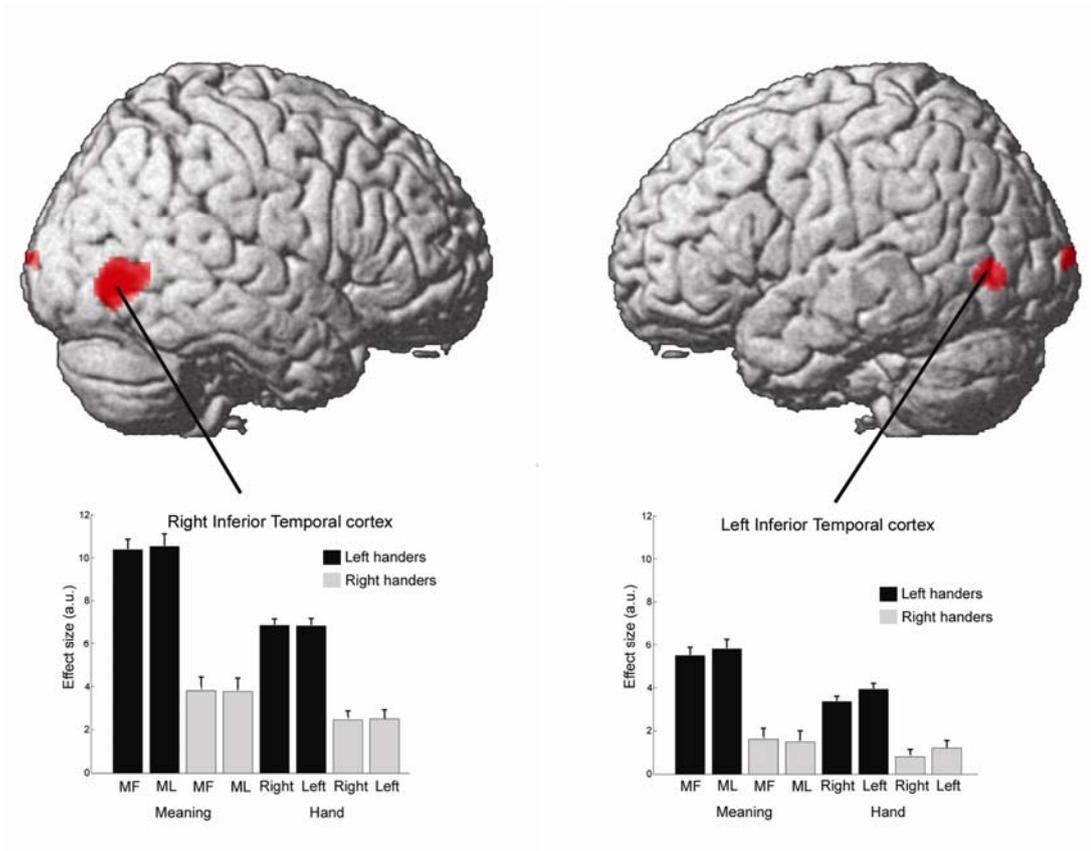
**Fig. 7.4.** Neural differences between observed hand in the run in which participant passively viewed the actions (main effect of Hand). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI) in right postcentral sulcus (14 -62 66), left postcentral sulcus (-32 -40 60) and right (22 -94 -10) and left (-18 -98 -10) inferior occipital gyrus. Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



**Fig. 7.5.** Neural differences between observed hand in the run in which participants had to indicate whether the action was meaningful or not (main effect of Hand). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left- (black bars) and right-handers (grey bars). Effect sizes are taken from local maxima (MNI) in left superior occipital gyrus (-20 -90 32), left postcentral sulcus (-32 -48 66) and right (14 -100 16) and left (-14 -98 -10) inferior occipital gyrus, expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



**Fig. S7.1** Neural differences between left- and right-handers in the run in which participants passively viewed the actions (main effect of Group). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI coordinates) in right (44 -64 -2) and left (-38 -72 2) inferior temporal sulcus, overlapping with previously reported location of extrastriate body area and human motion area MT (Peelen et al. 2006). Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .



**Fig. S7.2** Neural differences between left- and right-handers in the run in which participants had to indicate whether the action was meaningful or not (main effect of Group). Panels show effect sizes of meaningful or meaningless actions (left side) and actions performed with the right or with left hand (right side) for left-handed (black bars) and right-handed (grey bars) participants. Effect sizes are taken from local maxima (MNI coordinates) in right (46 -70 -2) and left (-38 -74 2) inferior temporal sulcus, overlapping with previously reported location of extrastriate body area and human motion area MT (Peelen et al. 2006). Effect sizes are expressed as the beta weight for a particular regressor. Error bars indicate standard error (s.e.m). Statistical map is corrected for multiple comparisons by controlling the family-wise error rate at  $p < 0.05$ .