

Henning Holle: The Comprehension of Co-Speech Iconic Gestures:  
Behavioral, Electrophysiological and Neuroimaging Studies. Leipzig:  
Max Planck Institute for Human Cognitive and Brain Sciences, 2007  
(MPI Series in Human Cognitive and Brain Sciences; 95)

---

# **The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies**

Der Fakultät für  
Biowissenschaften, Pharmazie und Psychologie  
der Universität Leipzig  
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades  
doctor rerum naturalium  
Dr. rer. nat.

vorgelegt  
von Henning Holle  
geboren am 12. Juli 1974 in Thuine / Emsland

Eingereicht am 15.6.2007

Verteidigt am 29.11.2007

Gutachter:

Prof. Dr. Angela D. Friederici (Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig)

Prof. Dr. Erich Schröger (Universität Leipzig)

Prof. Dr. Marco Iacoboni (Ahmanson-Lovelace Brain Mapping Center, Los Angeles)

## Acknowledgements

The work presented here could not have been accomplished without the personal and professional support of many people. In particular, I would like to thank

Prof. Dr. Angela D. Friederici for providing the means to conduct the experiments presented here, as well as for agreeing to read and evaluate this dissertation.

Dr. Thomas C. Gunter for his excellent supervision and continuing support.

Prof. Dr. Marco Iacoboni for his knowledgeable contributions to the fMRI Experiment, as well for agreeing to read and evaluate this dissertation.

Prof. Dr. Erich Schröger for taking the time to read and evaluate the work.

Dr. Shirley-Ann Rüschemeyer and Dr. Andreas Hennenlotter, for introducing me to the fMRI methodology and helpful discussions.

Christian Obermeier for his contribution to the gating experiment.

Dr. Korinna Eckstein, Dr. Jörg Bahlmann, Dr. Jutta Müller, Anna Hasting and Dr. Dirk Köster for helpful comments and discussions.

Meiner Freundin, Sandra Mroske, meinen Eltern, Martin und Maria Holle und meiner Schwester, Sigrid Beerboom, mit ihrem Mann Andreas und den Kindern Pia und Sina für ihre unbedingte Liebe und Unterstützung.



# Table of Contents

<b>Chapter 1: Characteristics of Iconic Gestures .....</b>	<b>1</b>
1.1 Key Characteristics of Iconic Gestures .....	2
1.1.1 Pointing .....	2
1.1.2 Emblems .....	3
1.1.3 Beats .....	4
1.1.4 Metaphoric Gestures .....	4
1.1.5 Summary of Properties of Iconic gestures .....	5
1.2 The Time Phases of Iconic Gestures .....	5
1.2.1 The Stroke Phase (obligatory) .....	5
1.2.2 Optional Gesture Phases .....	5
1.3 The Temporal Synchrony of Gesture and Speech .....	8
1.4 Iconic Gestures and Speech Convey Meaning Differently .....	9
1.4.1 Relation of the Parts to the Whole .....	9
1.4.2 Combinatoric vs. Noncombinatoric .....	10
1.4.3 Arbitrariness vs. Iconicity .....	10
1.5 The Function of Iconic Gestures in Comprehension .....	11
<b>Chapter 2: The Interaction of Iconic Gestures and Speech in Comprehension .....</b>	<b>13</b>
2.1 Theoretical Views on the Impact of Iconic Gestures in Comprehension .....	13
2.2 Method of Investigation: The Electroencephalogram (EEG) .....	14
2.2.1 Nature of the EEG Signal .....	14
2.2.2 The N400 .....	16
2.3 Behavioral and EEG Studies on Iconic Gesture Comprehension .....	17
2.3.1 Data Supporting the “Weak-Impact View” .....	17
2.3.2 Data Supporting the “Strong-Impact View” .....	19
2.4 Summary and Open Issues .....	20

2.5	Lexical Ambiguity as a Testing Ground for the Impact of Co-Speech Gestures in Comprehension.....	21
2.6	Experiment 1 .....	24
2.6.1	Introduction .....	24
2.6.2	Methods.....	24
2.6.3	Results .....	30
2.6.4	Discussion .....	33
2.7	Experiment 2 .....	34
2.7.1	Introduction .....	34
2.7.2	Methods.....	34
2.7.3	Results .....	35
2.7.4	Discussion .....	36
2.8	Experiment 3 .....	37
2.8.1	Introduction .....	37
2.8.2	Methods.....	40
2.8.3	Results .....	42
2.9	General Discussion of Experiments 1 - 3.....	44
2.9.1	Conclusion.....	48
<b>Chapter 3:</b>	<b>Temporal Aspects of Iconic Gesture Comprehension .....</b>	<b>51</b>
3.1	Theoretical Views on Temporal Aspects of Iconic Gesture Comprehension .....	52
3.2	Method of Investigation: Gating .....	54
3.3	Experiment 4 .....	55
3.3.1	Introduction .....	55
3.3.2	Methods.....	56
3.3.3	Results and Discussion.....	59
3.4	Experiment 5 .....	61
3.4.1	Introduction .....	61
3.4.2	Methods.....	61
3.4.3	Results .....	63

3.5	General Discussion of Experiments 4 and 5 .....	64
<b>Chapter 4: Neural Correlates of Iconic Gesture Comprehension.....</b>		<b>67</b>
4.1	Theoretical Views on the Neural Correlates of Iconic Gesture Comprehension .....	68
4.2	Method of Investigation: Functional Magnetic Resonance Imaging (fMRI).....	70
4.2.1	Nature of the Signal.....	70
4.2.2	The BOLD Effect .....	76
4.2.3	Analysis of fMRI Data .....	79
4.3	fMRI Studies Relevant to Iconic Gesture Comprehension .....	81
4.4	Experiment 6 .....	86
4.4.1	Methods.....	88
4.4.2	Results .....	94
4.4.3	Discussion .....	100
4.4.4	Conclusion.....	107
<b>Chapter 5: General Discussion .....</b>		<b>109</b>
5.1	Requirements of a to-be-developed Theory of Iconic Gesture Comprehension ....	110
5.1.1	How Does Gesture Help the Listener?: Main Effect of Gesture or Gesture-Speech Interaction? .....	111
5.1.2	Temporal Aspects.....	113
5.1.3	The Potential Automaticity .....	116
5.2	Functional Neuroanatomical Correlates of Iconic Gesture Comprehension.....	118
5.3	How Iconic Gestures Contribute to Language Comprehension: A Tentative Model .. .....	122
5.3.1	Scope of the Model .....	122
5.3.2	Stimulus Example .....	123
5.3.3	Model Architecture and Involved Processes.....	124
5.3.4	How the Model Accounts for the Existing Data .....	129
5.3.5	Some Testable Predictions .....	131
5.3.6	Potential Neural Correlates .....	132
5.4	Concluding Remarks .....	133

<b>List of Figures .....</b>	<b>135</b>
<b>List of Tables.....</b>	<b>137</b>
<b>References .....</b>	<b>139</b>
<b>Appendix A: Sentence Materials .....</b>	<b>151</b>
<b>Curriculum Vitae .....</b>	<b>157</b>
<b>Bibliographic Details.....</b>	<b>159</b>

## Chapter 1: Characteristics of Iconic Gestures

When engaged in a face-to-face conversation, speakers almost inevitably produce hand movements that show a meaningful relation to the contents of speech, i.e., they gesture<sup>1</sup>. Imagine two students having a chat, where one student (the speaker) is complaining to the other one (the listener) about the essay she has to write for class. In this situation, the speaker says “I was up all night writing” accompanied by typing hand movements. In this case, the gesture provides an illustration of the related speech unit (i.e., the word *writing*) by re-enacting the typing movement. Furthermore, the gesture is synchronized with the related speech unit in that the beginning of the typing movement (the so-called stroke, see below) coincides with the onset of the related speech unit. From the perspective of experimental cognitive neuroscience, the psychological reality of these co-speech gestures provokes some interesting research questions. These questions can be divided into issues related to the production and those issues related to the comprehension of gesture.

On the production side, one obvious question concerns the functional significance of gesture. Why do speakers frequently produce these kind of gestures? It has been shown that speakers gesture even when the listener clearly cannot benefit from these hand movements. For instance, speakers gesture when talking on the phone (Krauss, Dushay, Chen, & Rauscher, 1995) and even when talking to a blind person (Iverson & Goldin-Meadow, 1998). This suggests that gesturing is somehow related to speech production and various suggestions have been made to specify this relationship (de Ruiter, 1998; Kita & Özyürek, 2003; Krauss, Chen, & Chawla, 1996).

Co-speech gestures do also provoke questions related to their comprehension, some of which are addressed in the present dissertation. Given that many co-speech gestures are in

---

<sup>1</sup> Most gestures (including the type under study in this dissertation) do not represent a codified system. Instead, the meaning of the hand movement has to be inferred on the basis of the gesture form and/or contextual information. Therefore, a pragmatic (and admittedly quite circular) definition of gesture is employed: A gesture is a communicatively intended meaningful hand movement (for similar definitions, see Goldin-Meadow, 2003; Kendon, 1997)

themselves only improvised, elusive hand movements, one question concerns the extent to which listeners pick up the additional information provided by gesture. This issue will be addressed in a series of three experiments (Experiments 1 – 3). Provided that listeners do take the gestural information into account, one follow-up question is at what point in time the meaning is available, i.e., how much of a gesture a listener needs to see before it becomes meaningful (Experiments 4 and 5). Finally, as will be argued below, comprehending a gesture of the type exemplified above requires a listener to integrate auditory (i.e., speech) and visual (i.e., gesture) information. One exciting question from a neurocognitive perspective is what brain areas are involved in these audiovisual integration processes. Experiment 6 will deal with this issue.

The remainder of this chapter will describe the properties of gesture in more detail. Following this, each of the three research questions will be addressed in separate chapters. In the final chapter, I will summarize the experimental findings and attempt to give an integrative discussion about the role of co-speech gestures in comprehension.

## **1.1 Key Characteristics of Iconic Gestures**

The gesture example given above is an instance of an iconic gesture (McNeill, 1992). This category of gesture is characterized by a formal relationship between the hand movement and the co-expressive speech unit. In the given example, the repeated up-and-down movement of the fingers is formally related to the speech unit "writing". Whereas in this case, the gesture illustrates a bodily action, iconic gestures are also often used to depict spatial relations (for a example an arcing-downward motion while saying "You go through the undercrossing) and features of objects (e.g., producing a circular hand movement when talking about a roundabout). Because iconic gestures are in the focus of this dissertation, I will highlight their key characteristics in the following by showing similarities and differences between iconic gestures and other types of meaningful hand movements that can accompany speech.

### **1.1.1 Pointing**

A co-speech pointing gesture establishes a relationship between a point in space and a referent. For example, a host might address his guest saying "Please sit down" while pointing to an empty chair. While in this example the referent (i.e., chair) is physically present, pointing can also refer to objects described only in speech ("There was this ... picture on the

wall”) or to abstract concepts (e.g., “at this point in time...”). Note that the meaning of a pointing movement relies absolutely on context. Without contextual information, a pointing movement is meaningless. In contrast, decontextualized iconic gestures retain some meaning. Observing the “typing” gesture described above without its co-speech context still gives an impression that an object with many keys is manipulated. However, it can not be decided on the basis of the gesture form what kind of object is being manipulated (e.g., a piano or a typewriter)<sup>2</sup>.

### 1.1.2 Emblems

Emblems are very conventionalized gestures, such as the victory sign. Emblems typically have a label or paraphrase, they have to be learned in order to be understood and can be used as if they were spoken words (cf. McNeill, 1992, p. 56). It has been speculated that every society has a specific set of emblems. For example, Morris (1979) has conjectured that the hand purse (thumb and index make contact while the remaining finger point upwards) has quite distinct meanings in different European countries, e.g., “query” in Italy, “good” in Portugal, Greece and Turkey, “fear” in Belgium and France and “emphasis” in Holland and Germany. This points to an important difference between iconic gestures and emblems. Whereas emblems have to be learned in order to be understood, the comprehension of iconic gestures is not dependent on cultural knowledge, because their meaning is generated in an ad-hoc fashion on the basis of the gesture form and the co-speech context in which the gesture is observed (Goldin-Meadow, 2003).

That being said, it has to be noted that some degree of iconicity is also inherent in emblems, with some being more iconic than others. One example for a quite iconic emblem is the Eyelid Pull (pulling down the lower eyelid with the index finger) meaning “Watch out!”, “I am alert” (according to Morris, 1979) or “You can’t fool me” (according the mental lexicon of the author). Pulling down the eyelid enlarges the eye and displays increased alertness,

---

<sup>2</sup> Note that the examples represent idealizations for the sake of illustration. There is often a smooth transition between the different categories of gesture. Deictic movements often contain iconic elements (e.g. index finger pointings for nearby objects vs. flat-palm pointings for distant objects) as well as iconic gestures often contain deictic elements. In fact, in his most recent book, McNeill (2005) abandoned his traditional categorical distinction of gesture types and proposed a new dimensional view. In this scheme, a given gesture can have loadings on several factors (e.g. iconicity, deixis) simultaneously.

which is directly related to all three meanings of this emblem. On the basis of such observations, it has been speculated that many emblems initially were iconic gestures which subsequently underwent social conventionalization (cf. Kendon, 1981).

Another important difference between emblems and iconic gestures is the degree to which they depend upon a co-speech context in order to be understood. Due to the high degree of conventionalization, emblems can be effortlessly understood in the absence of speech (Gunter & Bach, 2004). In fact, emblems are often used to replace speech, for example when talking is difficult because of social (e.g., emblems with a sexual content) or environmental restrictions (e.g., Tic-Tac-Toe, the emblems used by bookies to accept bets at the horse race track, see MacSweeney et al., 2004). Iconic gestures are much more dependent on a co-speech context to convey meaning. As has been argued above, iconic gestures deprived of their co-speech context retain some meaning of their own, however, this meaning is rather vague. As this point is central for the current dissertation, it will be developed in more detail and supported by empirical findings in section 2.3.

### **1.1.3 Beats**

Beats are short quick movements that can put stress on their co-expressive speech units. The name derives from the fact that beats look like beating musical time. It has been suggested that beat gestures index the accompanying speech unit as being important on a discourse level (McNeill, 1992). For example, they may mark the introduction of new characters, new themes, etc. In contrast to iconic gestures, the form of a beat movement remains mostly unchanged regardless of content.

### **1.1.4 Metaphoric Gestures**

Metaphoric and iconic gestures have in common that they both provide an imagistic description of the speech content. However, whereas the iconic gestures describe concrete objects or events, metaphoric gestures illustrate abstract ideas. For example, a speaker might make an argument and simultaneously turn the hand so that the palm faces upward, as if an object is presented with the hand. The gesture in this case indicates that the speaker considers the argument as a spatially localizable, bounded object that can be offered to the listener. Thus, the gesture provides a pictorial description of the metaphor the speaker has in mind (AN ARGUMENT IS A PHYSICAL OBJECT). In this way, observing the metaphorical

gestures a person produces while performing certain cognitive tasks can potentially reveal how concepts are internally represented. Susan Goldin-Meadow (2003) has in this context coined the phrase of gesture as “a window on the mind”.

### **1.1.5 Summary of Properties of Iconic gestures**

Iconic gestures are meaningful hand movements that co-occur with speech. These hand movements are formally related with the co-expressive speech unit. Iconic gestures provide an illustration of concrete objects or events, not of abstract concepts. The meaning of an iconic gesture is determined both by its form as well as the co-speech context in which it is performed.

## **1.2 The Time Phases of Iconic Gestures**

Iconic gestures constitute a complex signal that can be divided into smaller segments. In the terminology established by Kendon (1972; 1980), a gesture phrase corresponds to what we intuitively call a ‘gesture’. Such a gesture phrase consists of several consecutive phases which are described below in more detail.

### **1.2.1 The Stroke Phase (obligatory)**

The most important phase of a gesture is the stroke which is defined as the phase during which the peak effort of movement occurs. For example, a speaker might say “I knocked at the door” while the closed fist makes two rapid tilting movements around the wrist. In this case, the stroke phase would start at the beginning of the first tilting movement and end after second tilting movement has been completed. The stroke phase is considered as an obligatory element of a gesture phrase, i.e., in the absence of a stroke, a gesture is not said to occur (cf. McNeill, 2005).

### **1.2.2 Optional Gesture Phases**

To illustrate the other gesture phases that can precede or follow a stroke, I will refer to an example described by David McNeill in his most recent book (2005). Figure 1.1 depicts a speaker who is retelling from memory a story from a previously read comic book. In the illustrated sequence, the main character grabs a tree and bends it backwards until it is arcuated. Immediately after this, the tree catapults the main character to a nearby building.

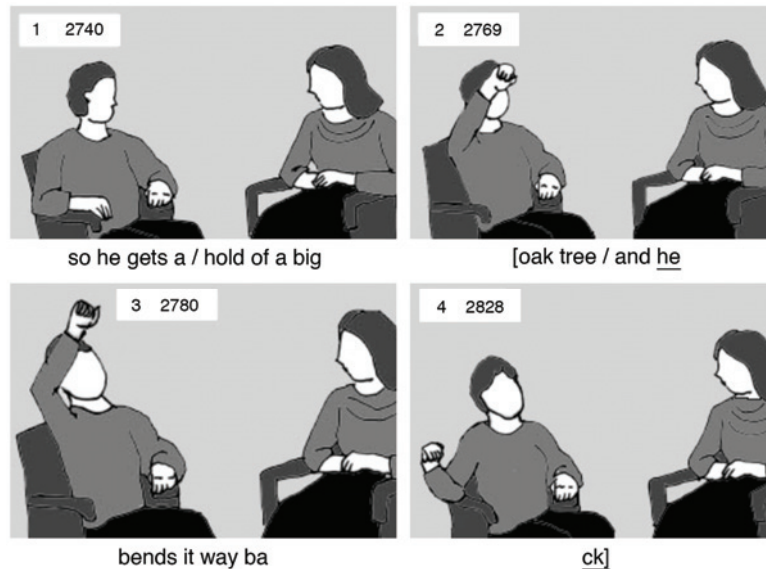


Figure 1.1: Gesture phases of the „and he bends it way back“ gesture. (slightly adapted from McNeill, 2005). The insert is a frame counter (1 frame = 1/30 sec.) The total elapsed time is about 1.5 sec.

*Panel 1. Pre-preparation position.* The hand is shown just before it leaves the armrest.

*Panel 2.* An example of a **prestroke hold**. The hand has travelled to gesture space where the stroke is to be performed. The hand is ready to perform the stroke (pull back and the rear). However, the stroke is not immediately executed, but delayed until the word *bends*.

*Panel 3.* The hand is shown half-way through the **stroke phase** (while saying *way*). The hand has closed around the ‘oak tree’ and is moving downward and to the rear.

*Panel 4.* End of stroke and beginning of the **poststroke hold** (while saying *back*). Hand is at its farthest point to the rear. After the poststroke hold, the hand immediately launched into a new stroke, showing how the character used the tree to catapult himself onto a nearby building (cf. McNeill, 2005, p. 30).

### Preparation

The preparation phase is the time period from the beginning of the gesture movement up to the stroke. Typically, the hands start to rise from a resting position and travel to the gesture space where the stroke is to be performed. Very often, some ‘phonological’ features of the stroke are already present during the preparation phase. At the beginning of the preparation phase (see Panel 1 of Figure 1.1), the right hand is in a relaxed position at the right armrest. However, at the end of the preparation phase, the right hand has already assumed a gripping

handshape (see Panel 2). It seems almost as if the hand prepares itself for the upcoming stroke during the preparation phase. This is why it has been claimed that a key characteristic of the preparation phase is anticipation (McNeill, 1992).

#### Prestroke Hold

During some gestures, the movement temporarily stops after the preparation phase, a phenomenon that has been termed prestroke hold. An example of this can be seen in Panel 2 of Figure 1.1, where the right hand is already in a position to initiate the bending-back movement, however, this movement is not initiated yet but postponed until the next word has been uttered (i.e., *bends*). It has been hypothesized that during a prestroke hold, the gesture is ready and ‘cocked’ but waits for a specific linguistic segment (McNeill, 2005). In this view, the function of prestroke holds would be to maintain the synchrony between gesture and speech (see below).

#### Poststroke Hold

Following the stroke phase, the hand sometimes freezes in midair before starting a retraction or the next gesture phrase. In Panel 4 of Figure 1.1, it can be seen that the right hand of the speaker maintains its position at the farthest point to the rear, before the next gesture phrase is initiated (describing how the character uses the tree to catapult to the next building). Poststroke holds have been suggested to occur when the speech co-expressive with the stroke continues to roll out, while the stroke itself has completed its motion (McNeill, 2005). Thus, both prestroke and poststroke holds have been theorized to be involved in the synchronization of the stroke with its co-expressive speech.

#### Retraction

The retraction phase begins where the hands start to return to the resting position<sup>3</sup> and ends when they have reached this position. This final gesture phase mirrors the preparation phase in that the effort is focused on reaching an end position.

---

<sup>3</sup> The resting position does not necessarily have to correspond to the starting position.

To sum up, iconic gestures do always contain a stroke phase. The stroke phase can be preceded by a preparation and/or a prestroke hold and followed by poststroke hold and/or a retraction.

### **1.3 The Temporal Synchrony of Gesture and Speech**

Having described the different time phases of iconic gestures, I will now look at how the different gesture phases synchronize with the co-expressive speech unit. Several studies have addressed this question.

Morrel-Samuels and Krauss (1992) investigated the temporal relationship between preparation onset and the related speech unit. In the first step of their experiment, subjects described the content of stimulus photographs to a female confederate. During these descriptions, the subjects were videotaped. From these materials, those descriptions were selected where the amount of hand movement surpassed a defined threshold. Next, a new group of subjects were presented both with transcripts of the descriptions as well as an audiovisual recording. Their task was to underline places in their transcript where the meaning of a word or a phrase was related to the meaning of a hand movement. Words or phrases that were consistently marked by at least 80% of the judges were considered as a related speech unit. When analyzing the temporal relationship between the onset of the gesture preparation phase and the related speech unit, the authors found that for all of the 60 gestures in the stimulus set, the preparation onset occurred before the onset of the related speech unit. On average, the preparation of gesture preceded the related speech unit by approx. 1 second.

Focusing on pointing gestures, Levelt and co-workers (1985) analyzed the temporal relationship between the onset of the preparation, the onset of the stroke and the onset of the related speech unit. In this experiment, subjects were seated in front of an array of four light-emitting diodes (LEDs). Two of the LEDs were placed far away from the centerline (far right, far left) and two were placed closer to centerline (near right, near left). The task of the participants was to indicate which LED was momentarily illuminated. This was done by pointing to the light and/or by using a deictic expression (e.g., *this light* for the near LEDs, *that light* for far LEDs). In all four experiments, it was found that the onset of preparation preceded the onset of the deictic expression by approx. 300 ms. With respect to the temporal

relationship between the stroke onset and the related speech unit, the authors observed that the participants had a strong tendency to synchronize these two events.

McNeill (1992, p. 92) found in a large sample of iconic gestures that the stroke phase overlaps with the related speech unit about 90 percent of the time. In the remaining cases, asynchrony was only observed in the direction that the stroke slightly preceded the related speech unit, usually because of brief hesitations (McNeill, 2005, p. 32).

In sum, the literature suggests that the preparation of a gesture precedes the related speech unit, whereas the gesture stroke coincides with this speech segment. This speech-stroke synchrony has been found to be a remarkably robust phenomenon. Levelt and co-workers (1985, Experiment 4) applied during some trials a load of 1600 grams to the gesturing hand. Importantly, participants could not predict at what point in time the load was applied. Of course, the stroke occurred later during the load trials as compared to the non-load trials. More interesting is that participants accounted for the stroke delay and correspondingly delayed the utterance of the deictic expression in order to maintain stroke-speech synchrony. Only when the load was applied very late during the preparation phase of gesture, the synchrony was disrupted in that the speech unit preceded the stroke. Further evidence for a strong speech-stroke synchrony was obtained in a number of delayed-auditory-feedback (DAF) experiments conducted by McNeill and colleagues (1992, p. 273).

### **1.4 Iconic Gestures and Speech Convey Meaning Differently**

In this section, I will compare the way meaning is represented in iconic gestures with the way meaning is conveyed in speech. In a next step, I will attempt to illustrate how these two quite different information streams may work together in communication, e.g., as a composite signal.

#### **1.4.1 Relation of the Parts to the Whole**

One important difference between both information streams concerns the way how greater meanings are assembled from smaller pieces of information. Consider for example a downward motion of both hands, with the palm facing down, while saying “He pushed up from the table”. The gesture describes the event as a whole, depicting manner (“push”) and trajectory (“up”) simultaneously. However, in speech the message unfolds over time, broken up into smaller meaningful segments (i.e., the words *push* and *up*). McNeill has in this context

characterized language as having the effect of “segmenting and linearizing meaning” (1992, p. 19). The act of speaking requires to transform an initially multidimensional communicative intention (e.g., inform an addressee about the own need for food) into a hierarchically organized string of words (e.g., “I am hungry!”). In the resulting utterance, each of the single parts (the words) is meaningful and the single parts are combined to create a greater whole (the sentence). Thus, in language, meaning projects from part to whole.

In contrast, the meaning of iconic gestures shows no segmentation or linearization. Although the push-up gesture can be divided into smaller meaningful parts (e.g., the pushing of the left and the pushing of the right hand), the meaning of the parts depend on the meaning of the whole. The parts are not independently meaningful morphemes or words in a language. For example, the very same downward motion can mean “calm down” in a different speech context. It has been suggested that iconic gestures are *global-synthetic* in that the meaning projects from whole to part (McNeill, 1992). Additionally, the “parts” of a gesture can be created simultaneously (e.g., manner and trajectory), whereas the parts of speech are necessarily uttered one after another.

#### 1.4.2 Combinatoric vs. Noncombinatoric

Iconic gestures are noncombinatoric. Two iconic gestures cannot be combined to form a greater meaning, let alone form a hierarchically structured sequence, such as a sentence. This contrasts with the combinatoric property of language, which allows a speaker to form arbitrarily long meaningful sequences.

#### 1.4.3 Arbitrariness vs. Iconicity

Arbitrariness means that the signal does not resemble the thing that it denotes. For example, the sound of the word “door” does not resemble a door. Arbitrariness has been suggested as one of the design features of human language (Coleman & Keith, 2006; Hockett, 1960), and the vast majority of spoken words show this property<sup>4</sup>. In contrast, the form of an iconic

---

<sup>4</sup> Note that a small number of spoken words are not arbitrary but iconic, e.g. *hiccup*, *whoosh* or the German word *Tuff-Tuff* (meaning train). The pronunciation of words is also sometimes changed to establish an iconic relationship between form and meaning: “We waited for a loooooong time!”.

gesture *does* resemble the denoted thing (hence the name). Explaining the meaning of an iconic gesture always involves explaining the form.

To sum up, iconic gestures convey meaning in a very different manner than speech does. Whereas during speaking, the meaning of the parts (i.e., words) determines the meaning of the whole (e.g., a sentence), the meaning of iconic gestures projects from the whole of the gesture to the parts of it. Moreover, in contrast to speech, iconic gestures are noncombinatoric and the relation between signal and the denoted thing is not arbitrary. Finally, iconic gestures lack the combinatoric ability of language.

### **1.5 The Function of Iconic Gestures in Comprehension**

In the beginning of this chapter, one of the questions raised was whether listeners are able to pick up the additional information provided by gesture. Given the unlike modes in which iconic gestures and speech represent meaning, the answer to this question is far from self-obvious. In the case of speech, the word form allows the access of the internally stored word meanings. In contrast, there is no one-to-one mapping between of form onto meaning for iconic gestures. It rather seems to be the case that one iconic gesture activates an array of possible meanings (Feyereisen, Van de Wiele, & Dubois, 1988). The accompanying speech may serve to select the appropriate gesture meaning. Thus, understanding speech does not require understanding gesture, but comprehending gesture necessitates the comprehension of speech *and* determining the relation of gesture form to the co-expressive speech.

Up this point in the current dissertation, the notion that iconic gestures convey additional meaning not found in speech has been made from an *interpretative* or *hermeneutic* perspective. This means that the claim was based on detailed observation of a gesture and the accompanying speech. Recall the previously given example of a student making typing hand movements while saying “I was up all night writing” (see also page 1). In an interpretative approach (see, for instance, Kendon, 1997), the interpretation would be that the gesture does convey additional information, because the information that the writing was performed on a keyboard and not by paper and pen is only indicated by gesture. It is, however, a completely independent *empirical* question whether listeners in a natural face-to-face conversation do also take this additional gestural information into account.

Chapter 2 will give an overview of the literature relevant to this question and describe a series of three experiments investigating the degree to which listeners make use of gestural information in speech disambiguation.

## **Chapter 2: The Interaction of Iconic Gestures and Speech in Comprehension**

In the following, I will outline the two opposing theoretical views on the degree to which iconic gestures convey additional information not found in speech. Next, the method used to address this question, namely the Event Related Potential (ERP) of the human Electroencephalogram (EEG) will be introduced. Following this, I will review the existing behavioral and ERP studies on the topic. Finally, a series of three ERP experiments will be described which investigated the degree to which listeners make use of gestural information in speech comprehension.<sup>5</sup>

### ***2.1 Theoretical Views on the Impact of Iconic Gestures in Comprehension***

With respect to the question whether listeners are able to benefit from the additional information provided by iconic gestures, two opposing views have been put forward. Whereas McNeill and co-workers (Cassell, McNeill, & McCullough, 1999; McNeill, Cassell, & McCullough, 1994) have suggested a strong impact of gesture in comprehension (henceforth called the “strong-impact view”), Krauss and colleagues (1995) maintain that gestures are generated as an epiphenomenon of speech production processes, but have little semantic value for the listener (henceforth the “weak-impact view”).

According to the “weak-impact view”, gestures primarily facilitate the speaker’s lexical access but are only subject to minimal semantic analysis in comprehension. Krauss et al. stress that the meaning that iconic gestures convey is determined largely by the speech which accompanies them: “...it may be that much of the gesture’s meaning is illusory. In the absence of speech, the very same gesture’s meaning can be quite opaque, communicating little, if anything” (Krauss, Morrel-Samuels, & Colasante, 1991, p. 744).

---

<sup>5</sup> The work described in Chapter 2 has also been published: Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19(7), 1175-92.

An opposing view put forward by McNeill (1992) holds that iconic gestures do convey additional information to the listener. In McNeill's model, it is assumed that gesture and speech are part of a tightly integrated system. Gesture and speech each convey some unique and some redundant information and the comprehension system routinely combines the bimodal information into an enriched unified representation. Thus, the model assumes an obligatory interaction between gesture and speech in the comprehension process. It predicts that iconic gestures are easily decoded and have a high impact on speech comprehension.

## **2.2 Method of Investigation: The Electroencephalogram (EEG)**

Both gesture and speech are communication streams with a high information density, where potentially distinguishing information bits (e.g., phonemes) are uttered in the millisecond range. Accordingly, in order to investigate the interaction between the two information streams, one needs a method with a suitable temporal resolution. One method with an excellent temporal precision is the Event-Related Potential (ERP) of the electroencephalogram (EEG). After a synopsis about the nature of the EEG signal, I will outline the characteristics of the ERP component that was used as dependent variable in Experiments 1 - 3, i.e., the N400.

### **2.2.1 Nature of the EEG Signal**

Neural signal transmission occurs via flow of charged particles across neural membranes, resulting in an electric potential in the conductive media inside and outside the cell (cf. Kutas & Federmeier, 2000). These synaptic currents can be monitored by measuring the electric potential difference between at least one scalp electrode and a reference electrode (placed at sites that are somewhat more insulated from brain activity, such as the nose or the mastoid bone behind the ear). Typically, larger arrays of electrodes are placed on the scalp in a standard montage. A popular standard montage is the 10-20 system ("American Electroencephalographic Society: Guidelines for standard electrode position nomenclature," 1991), which was also employed in EEG experiments of the present dissertation (see Figure 2.1). The recorded signal is amplified, digitized at a sufficiently high frequency (i.e., at 500 Hz in the present ERP studies) and stored on hard disk for subsequent analysis.

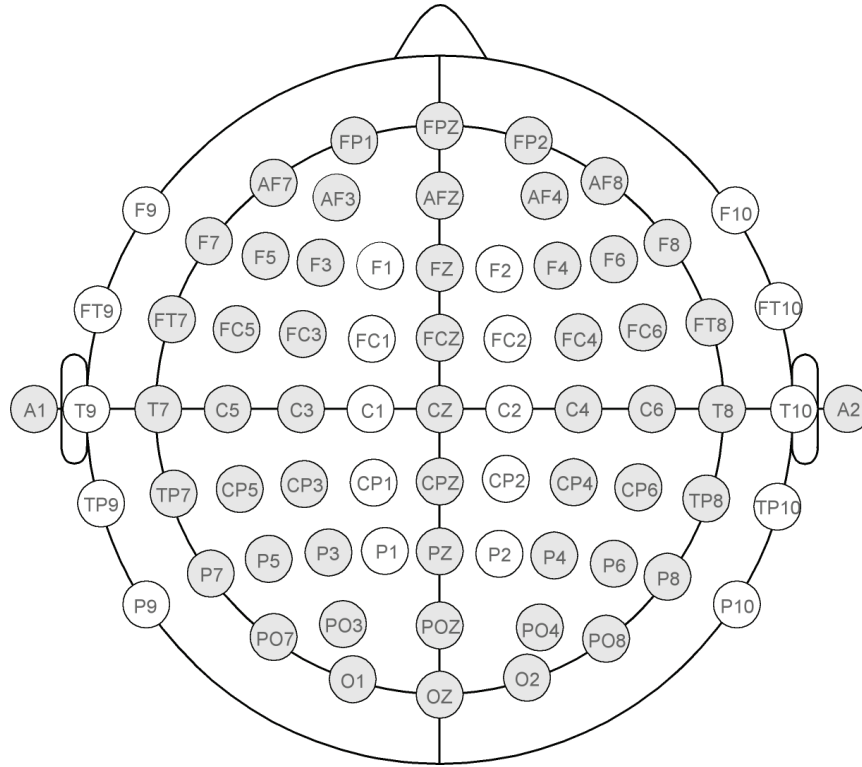


Figure 2.1: Schematic drawing of an extended 10-20 system montage showing the positions and labels of the electrodes. Electrodes not measured in the present series of experiments are not filled. Mastoid electrodes are labelled A1 and A2.

Some of the electrochemical activity of the brain can be picked up by the EEG, because the structural organization of the human cortex allows the generation of large extracellular field potentials. Especially activations of the pyramidal cells in cortical layers II, IV, and V that are vertically aligned to the scalp signal have been suggested to dominate the EEG signal (Kutas, van Petten, & Kluender, 2006). However, activity of a sufficiently large number of pyramidal cells (one estimation suggests at least 1000) has to summate in order to produce a measurable effect on the scalp surface (Rugg & Coles, 1995).

Brain activity as well as artefacts both contribute to the EEG signal. The amplitude of potential changes caused by cognitive factors is very small (around  $\pm 5 \mu\text{V}$ ) compared to the amplitude of most artefacts ( $\pm 50 - 100 \mu\text{V}$ ). Because of this poor signal-to-noise ratio, it is almost impossible to detect the effect of cognitive manipulations in spontaneous EEG. One

way to increase the signal-to-noise ratio of cognitive components is to average the EEG time-locked to the onset of repeated events. The outcome of this process is called Event Related Potential (ERP) and reflects the characteristic pattern of electrical activity to an event.

The first 200 ms of an ERP have a characteristic pattern that varies mainly as a function of the stimulation modality, whereas later segments of ERPs are more sensitive to cognitive factors. In cognitive ERP research, the continuously varying waveform of voltage across time is conventionally dissected into “components” defined by their polarity (positive or negative-going), latency, the spatial distribution across the scalp, or their sensitivity to certain experimental manipulations (cf. Van Petten & Luka, 2006). ERP components have proven to be useful dependent variables, as their presence, amplitude, timing and topography in a given experimental manipulation can give important insights about the underlying cognitive and neural processes of the phenomenon under study (Kutas & Federmeier, 2000).

Due to its excellent temporal resolution (in the range of few milliseconds), the ERP method is particularly suited to investigate cognitive processes that occur very fast, such as language comprehension. As a consequence, the past two decades have seen a boom of language-related ERP studies. One of the major findings of these research endeavors is that ERP components related to the basic subcomponents of language have been described in the literature, including quite distinct ERP components for semantic, syntactic and prosodic processes (for overviews, see Friederici, 2004; Kutas et al., 2006). The properties of a component particularly associated with semantic processing, the N400, will be described in more detail in the following.

### **2.2.2 The N400**

The N400 is a negative-going waveform that peaks about 400 ms after stimulus onset. It varies systematically with the processing of semantic information. On the level of sentence processing<sup>6</sup>, it has been found that the easier a word can be integrated into a sentence context, the smaller the amplitude of the N400. For example, Kutas and Hillyard (1984) observed that

---

<sup>6</sup> Note that an N400 has also been observed in response to single words presented in isolation (see, for instance, Barber, Vergara, & Carreiras, 2004). However, because the present study focuses on the integration of words into a sentence, the single-word literature on the N400 is beyond scope.

highly expected sentence-final words (as determined by a cloze procedure) elicited a small N400, moderately expected words a moderate N400 and unexpected words a large N400. Thus, the amplitude of the N400 varied as an inverse function of that word's rated cloze probability. Other studies have supported the notion of the N400 as an index of the degree to which a word is expected in the semantic context of a sentence. For instance, the amplitude of the N400 is also inversely related to the position of a word in a sentence (Kutas & Federmeier, 2000). One explanation of this effect is that during the first few words of a sentence, there is little context available to build up a semantic expectancy, therefore the N400 is large. As more and more words of a sentence are being processed, it becomes easier to anticipate the upcoming word, which is reflected in increasingly smaller N400 amplitudes to words at later positions in a sentence.

N400 like potentials were also observed in response to non-verbal stimuli, including line drawings, photos and environmental sounds (Orgs, Lange, Dombrowski, & Heil, 2006; West & Holcomb, 2002). The N400 in response to such non-verbal stimuli parallels the “verbal” N400 in that the amplitude varies as a function of context congruency. For example, West and Holcomb (2002) have found that the N400 in response to a picture is smaller, when it was preceded by a conceptually related sequence of pictures and larger, when preceded by a conceptually unrelated sequence.

In sum, the N400 can be seen as an established indicator of how well a meaningful (verbal or non-verbal) item semantically fits into its preceding context. In the following, I will review the existing studies on iconic gesture comprehension, some of which have employed the N400 as a dependent variable.

## **2.3 Behavioral and EEG Studies on Iconic Gesture Comprehension**

Two opposing views on the degree to which listeners take gestural information into account have been outlined above; the “weak-impact view” and the “strong-impact view”. Several studies have been conducted that speak to this issue.

### **2.3.1 Data Supporting the “Weak-Impact View”**

Several findings support the assumption that iconic gestures are only subject to minimal semantic analysis in comprehension. In an experiment by Krauss, Morrell-Samuels, and Colasante (1991, Exp. 1), video clips of gestures without sound were presented and

participants were asked subsequently which one of two phrases was the corresponding lexical affiliate. Subjects chose the correct lexical affiliate in 76 % of all cases which was significantly above chance level but far from perfect. A more recent study by Hadar and Pinchas-Zamir (2004) also investigated the semantic specificity of gestures. The participants had to decide which of five words best described the meaning of a previously seen gesture. The lexical affiliate of the iconic gestures was chosen in 40 % of cases (with chance level being at 20 %), interpreted as a rather vague semantic specificity. Feyereisen and co-workers (1988) presented iconic gestures without sound and asked their participants which of three verbal descriptions best described the meaning of the gesture. The three response alternatives were (1) the correct description, (2) a plausible, but incorrect description, (3) an implausible incorrect description. It was found that participants more often selected the plausible but incorrect description than the correct description. This was interpreted as reflecting that iconic gestures do not have one precise meaning but can encompass a wide range of possible meanings. In a study by Krauss and colleagues (1995), participants described abstract stimuli to a conversation partner. In the first part of the Experiment (the encoding phase), the descriptions were videotaped in two different conditions: In one condition, the conversation partner was facing the participant. In the other condition, the conversation partner was in a different room and only accessible via an intercom system. The gesture rate was significantly higher in the face-to-face condition. In the second part of the Experiment (the decoding phase), new participants were presented with either an audiovisual or an audio-only recording of the descriptions resulting in a total of four experimental conditions: 1) Encoding face-to-face, Decoding audio-visual; 2) Encoding face-to-face, Decoding audio-only; 3) Encoding intercom, Decoding audio-visual; 4) Encoding intercom, Decoding audio-only. The task was to select the stimulus described in the recording from a list. The identification accuracy was higher for stimuli encoded over the intercom system. There was no interaction between the encoding and decoding phase. Most importantly, the manipulation in the decoding phase had no effect, i.e., identification accuracy was not higher when the decoders could see the describer. This led the authors to conclude that “there is no compelling evidence that these gestures enhance, modify, or affect in any material way the semantic content of the message the speaker conveys” (Krauss et al., 1995, p. 550).

### 2.3.2 Data Supporting the “Strong-Impact View”

Evidence suggesting a strong impact of iconic gestures in comprehension has mainly been obtained in two different experimental paradigms. In one approach, the effect of bimodal presentation (speech with accompanying gestures) is contrasted with the effect of unimodal presentations (speech only; gesture only). The dependent variable in these experiments is typically some measure of comprehension, for instance, the amount of recalled details. Participants can recall more events after bimodal presentation (Beattie & Shovelton, 1999a, 1999b, 2001, 2002b), especially if gestures give some information about the relative size or position of objects (Beattie & Shovelton, 1999b). The bimodal-vs-unimodal paradigm allows, however, only limited inferences about the processes underlying gesture-speech integration. For instance, one cannot rule out the possibility that the advantage of the bimodal presentation is partly due to more attentive speech processing. Another drawback of the paradigm is the poor temporal resolution, which makes it impossible to determine whether gesture-speech integration is achieved by fast online or a slower offline processes.

Further data in support of a strong impact of gesture comes from another approach, in which gesture and speech provide clearly conflicting information. In these so-called mismatch paradigms, the dependent variable is the amount of interference caused by the incompatible gesture-speech combination. An interference effect is taken as evidence that the comprehension system attempted to integrate gesture and speech. A great advantage of the mismatch paradigm is that it allows for the investigation of the time-course of gesture-speech integration, provided that a method with a sufficiently high temporal resolution is employed, such as ERPs. All of the existing ERP studies on gesture-speech integration focused on the previously introduced N400 component (Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2005). In the study by Kelly et al. (2004), participants saw video clips of a person that gestured to one of two objects in front of him, namely a short, wide dish and a tall, thin glass. Directly after the offset of the gesture, one of four speech tokens was auditorily presented, namely *tall*, *thin*, *short* or *wide*. ERPs were time-locked to the onset of the speech tokens. The N400 was found to be smaller when gesture and speech referred to the same object and larger when they referred to different objects (e.g., gesturing *tall* and saying *wide*). This result suggests that it is difficult to integrate an incongruent target word into a gesture context. In an experiment by Wu and Coulson (2005) participants judged the relatedness between probe words and preceding cartoon-gesture pairs. In the ERPs time-

locked to the gestures, they reported an enhanced negativity in the N400 time range for incongruent gestures suggesting that it is more difficult to integrate an incongruent gesture into a cartoon context. The study by Özyürek et al. (2007) directly compared the time-course of semantic integration for gesture and speech. In this experiment, participants were presented with an initial sentence context that was subsequently matched or mismatched either in the gesture channel, in the speech channel, or in both. The synchrony between gesture and speech was manipulated so that the stroke onset of gesture always coincided with the onset of the target word. All mismatch conditions showed similar N400 effects. The authors concluded that the time window of semantic integration is similar for gestures and speech. In sum, all three ERP studies have provided important evidence that iconic gestures have an impact on online brain measures that are associated with semantic processing. However, studies employing a mismatch paradigm are somewhat limited in their external validity, because speakers do not produce such clear-cut gesture-speech mismatches in spontaneous conversations. Another disadvantage of the mismatch paradigm is that it focuses on potential conflict between gesture and speech in comprehension but does not address the issue of how gesture may aid language comprehension.

## **2.4 Summary and Open Issues**

In sum, there are both studies suggesting a strong impact of gesture in comprehension as well as studies which come to the opposite conclusion. As Wu has pointed out (2006), some of this contradiction may stem from the fact that many researchers interpret their data to confirm pre-existing biases without considering alternatives. For example, on the one hand Krauss and co-workers (1991) interpreted their finding that participants' accuracy in selecting the appropriate co-expressive speech was a (significant) 26 % above chance level as reflecting that gestures are only subject to minimal semantic analysis in comprehension. On the other hand, Beattie and Shovelton (1999b) observed that participants were 11 % more accurate in recalling information after bimodal stimulation (gesture + speech) than after a audio-only stimulation and interpreted this as evidence for a strong impact of gesture. Because both studies employed different measures of the semantic impact of gesture (forced-choice vs. a questionnaire-based approach), it is difficult to compare the results. In light of such problems, it seems desirable to establish a testing ground for gesture where the testing arena features some kind of criterion defining what constitutes a "strong" and what a "weak" impact of gesture in language comprehension (for an elaboration of this idea, see below).

Another open question are the mechanisms through which gesture may facilitate speech comprehension. All of the discussed ERP studies on iconic gesture comprehension have employed a mismatch paradigm, where the degree of interference caused by semantically incompatible gesture-speech combinations is taken as an indicator for the impact of gesture in comprehension. However, because of this logic, these studies can only explain how gesture interferes with speech processing. So far, there is no study that has investigated through which mechanisms gesture may actually facilitate online speech processing. One possible mechanism, namely the disambiguation of lexically ambiguous words, will be introduced in the next section.

## **2.5 Lexical Ambiguity as a Testing Ground for the Impact of Co-Speech Gestures in Comprehension**

Indeterminacy is one of the most significant challenges in language comprehension. We are constantly required to disambiguate stimuli with uncertain identities using whatever environmental and experiential context is available (see also Twilley & Dixon, 2000). Despite this massive ambiguity, selecting the contextually appropriate interpretation is an effortless process for a listener suggesting that our comprehension system is very efficient in disambiguation. One frequent kind of ambiguity is lexical ambiguity. For example, a sentence such as *The woman observed the ball* is lexically ambiguous, because the contained homonym allows two plausible interpretations. A study by Holler & Beattie (2003) investigated the role of gesture in disambiguation. In this experiment, participants were asked to read sentences containing an underlined homonym. After each sentence, the experimenter asked which of the two word meanings the sentence referred to. In almost half of all explanations, participants produced co-speech gestures to illustrate the relevant meaning. Thus, gestures produced in the context of a homonym are a phenomenon that actually occurs in face-to-face conversations. The question of the current study is whether listeners make use of this gestural information in comprehension.

Whereas some homonyms have equally frequent meanings, most homonyms are unbalanced (e.g., *ball*) in that they have a more frequent dominant meaning (e.g., *game*) and a lesser frequent subordinate meaning (e.g., *dance*). During the processing of unbalanced homonyms, the comprehension system can use two sources of information to activate the appropriate word meaning: (1) the context in which the homonym is encountered and (2) word meaning

frequency. The vast literature on this topic (for a review, see Twilley & Dixon, 2000) allows some predictions about how these two sources of information interact in activating the word meanings of a homonym.<sup>7</sup>

In the absence of context or in a neutral context, word meaning frequency determines the activation pattern. In such a situation, the dominant meaning is activated to a stronger degree than the subordinate meaning (Simpson, 1981; Simpson & Burgess, 1985; Simpson & Krueger, 1991; Vu, Kellas, & Paul, 1998). For example, in the experiment by Simpson & Burgess (1985) homonyms were presented without prior context. Participants made lexical decisions to target words that were associates of the dominant or the subordinate meaning of a homonym prime. The pattern of reaction times for the different SOAs indicates that the dominant meaning was activated more quickly and maintained longer than the subordinate meaning. Even when participants were explicitly asked to think of all possible meanings of a homonym, the dominant meaning was found to be more active than the subordinate meaning suggesting that the activation process is not under strategic control (Simpson & Burgess, 1985, Experiment 3). Thus, in the absence of a contextual cue, word meaning frequency determines how the different meanings of a homonym are activated.

Once the homonym is preceded by a biasing sentence context, the activation of the subordinate word meaning always seems to be affected. It has consistently been reported that the subordinate meaning is more active after a congruent subordinate context and less active after an incongruent dominant context (Onifer & Swinney, 1981; Paul, Kellas, Martin, & Clark, 1992; Simpson & Krueger, 1991; Vu et al., 1998). Thus, the activation of the subordinate word meaning of a homonym varies reliably as a function of context congruency.

In contrast, the activation of the dominant word meaning does not always seem to be affected by context congruency. Whereas some studies reported that the dominant meaning was more active after a dominant context and less active after a subordinate context

---

<sup>7</sup> In this overview on the literature on homonym processing, I will refrain from absolute statements about the activation of the two word meanings, such as *the dominant meaning is active, the subordinate meaning is not active*. Instead, the focus is on whether the activation of a word meaning varies as a function of the preceding context (e.g. *The dominant meaning is more active after a dominant context than after a subordinate context*). The rationale of this focus is to facilitate the comparability between the literature and the present experiments.

(Onifer & Swinney, 1981; Paul et al., 1992; Simpson & Krueger, 1991; Vu et al., 1998), others have found that the dominant meaning is always active, after both a dominant context as well as after a subordinate context (Tabossi, 1988; Tabossi, Colombo, & Job, 1987). These seemingly contradictory findings can be nicely explained by data from Simpson (1981) and Martin and colleagues (1999). These two studies systematically varied the degree to which a preceding sentence context biased either the dominant or the subordinate meaning of a homonym. Their results show that only a strongly biasing context was able to modulate the activation of the dominant meaning, i.e., the dominant meaning was more active after a strong dominant context and less active after strong subordinate context. Both of the weakly biasing contexts activated the dominant meaning to a similar degree, i.e., the weak subordinate context was as effective in activating the dominant meaning as the weak dominant context. Thus, it seems to be the case that once the contextual constraints become weak, the comprehension system makes increased use of other sources of information, in this case word meaning frequency. As a result, the dominant word meaning is always activated after weak contexts, even if the context biased the subordinate meaning.

Taken together, the literature on homonym processing suggests that a biasing context generally modulates the activation of the subordinate word meaning. Modulatory context effects of the dominant word meaning are restricted to strongly biasing contexts. It is a characteristic feature of weakly biasing contexts that they are unable to modulate the activation of the dominant word meaning. Instead of using simple sentences, the present dissertation investigates the extent to which co-speech iconic gestures can constitute a contextual cue for homonym disambiguation. The use of unbalanced homonyms has in this case the particular advantage that one can infer from the observed pattern of results whether gesture had a strong or a weak impact on disambiguation. If gestures have the status of a strong contextual cue for a listener, the activation of both the dominant and the subordinate word meaning should vary as a function of the congruency of the preceding gesture context (*strong context pattern*). If, however, iconic gestures have only a weak impact on disambiguation, only the activation of the subordinate word meaning should be modulated by context congruency (*weak context pattern*).

## 2.6 Experiment 1

### 2.6.1 Introduction

Experiment 1 examines whether co-speech iconic gestures influence the disambiguation of homonyms in auditory sentence processing. To this end EEG was recorded as participants watched videos of a person simultaneously gesturing and speaking. The experimental sentences contained an unbalanced homonym in the initial part of the sentence (e.g., *Sie beherrschte den Ball ... / She controlled the ball ...*) and were disambiguated at a target word in the subsequent clause (*was sich im Spiel ... / which during the game ...* vs. *was sich im Tanz ... / which during the dance ...*). Coincident with the initial part of the sentence the speaker produced an iconic gesture, which supported either the dominant or the subordinate meaning. ERPs were time-locked to the onset of the target word.

The literature cited above suggests a systematic relation between the observed activation pattern and context strength. If iconic gestures have the status of a strong contextual cue for a listener, the N400 time-locked to both dominant and subordinate target words should vary reliably as a function of context congruency. More precisely, the N400 time-locked to the subordinate target words should be smaller after a congruent subordinate gesture context and larger after an incongruent dominant gesture context. Conversely, the N400 time-locked to the dominant target words should be smaller after a congruent dominant gesture context and larger after an incongruent subordinate context. However, if iconic gestures constitute a weak contextual cue, only the N400 time-locked to the subordinate target words should vary as a function of context congruency. Based on the literature, it is hypothesized that iconic gestures have a strong impact on speech disambiguation.

### 2.6.2 Methods

#### Participants

Twenty-seven native German-speaking students were paid 7 Euro per hour for their efforts and signed a written informed consent. Three participants had to be excluded based on rejection criteria. The remaining 24 participants (14 female, mean 25 years of age, range 21-30) were right-handed (mean laterality coefficient 93, Oldfield, 1971). All participants had normal or corrected-to-normal vision and none reported any known hearing deficit.


### Stimuli

#### *Homonyms:*

The present dissertation is based on a set of 91 unbalanced German homonyms (for a description how the set was obtained, see Gunter, Wagner, & Friederici, 2003). Each of the homonyms had a more frequent dominant and a lesser frequent subordinate meaning, which shared identical phonological and orthographical surface features (e.g., *ball* – dominant meaning: *game*; subordinate meaning: *dance*). Target words representing the dominant meaning as well as target words representing the subordinate meaning were assigned to each of the homonyms. The relatedness of the target words to the homonyms had been previously tested using a lexical decision task in the visual modality (see also Wagner, 2003). For all target words, the lexical decision time was significantly shorter as compared to an unrelated item. In cases where either the dominant or the subordinate meaning was very abstract, the homonym was excluded from the set, resulting in a reduced set of 55 homonyms. For each of these 55 homonyms, two two-sentence utterances were constructed including either the dominant or the subordinate target word. The utterances consisted of a short introductory sentence introducing a character followed by a longer complex sentence describing an action of that character. The complex sentence was composed of a main clause containing the homonym and a successive sub-clause containing the target word. Previous to the target word, the sentences for the dominant and subordinate versions were identical (see Table 2.1).

Table 2.1: Experiment 1 Stimulus Examples. Introductory sentence was identical for all four conditions. The first two columns indicate the conveyed meaning of gesture and the subsequent target word: Dominant (D) or Subordinate (S). Target word in bold. Literal translation in italics. Cross-splicing was performed at the end of the main clause (i.e., in this case after the word "Ball").

Introduction: Alle waren von Sandra beeindruckt.  
Everybody was impressed by Sandra.

gesture	target word	gesture / homonym	target word
D	D	Sie kontrollierte den Ball <sub>amb</sub> , was sich im <i>She controlled the ball<sub>amb</sub>, which during the</i> 	<b>Spiel</b> beim Aufschlag deutlich zeigte. <i>game at the serve clearly showed.</i>
D	S	Sie kontrollierte den Ball <sub>amb</sub> , was sich im <i>She controlled the ball<sub>amb</sub>, which during the</i> 	<b>Tanz</b> mit dem Bräutigam deutlich zeigte. <i>dance with the bridegroom clearly showed.</i>
S	S	Sie kontrollierte den Ball <sub>amb</sub> , was sich im <i>She controlled the ball<sub>amb</sub>, which during the</i> 	<b>Tanz</b> mit dem Bräutigam deutlich zeigte. <i>dance with the bridegroom clearly showed.</i>
S	D	Sie kontrollierte den Ball <sub>amb</sub> , was sich im <i>She controlled the ball<sub>amb</sub>, which during the</i> 	<b>Spiel</b> beim Aufschlag deutlich zeigte. <i>game at the serve clearly showed.</i>

#### *Gesture Recording:*

A professional actress was videotaped while uttering the sentences. The exact recording scenario was as follows. The actress stood in front of a video camera with her hands hanging comfortably in a resting position. In a first step, she memorized one two-sentence-utterance until she could utter it fluently. Then she was asked to utter the sentence and simultaneously perform a gesture that supported the meaning of the sentence. The gestures were created by the actress and not choreographed in advance by the experimenter. She was instructed to perform the gesture to coincide with the initial part of the complex sentence (e.g., *Sie kontrollierte den Ball / She controlled the ball*) and to return her hands to the resting position

afterwards. About two thirds of all gestures re-enacted the actions in the sentence from a first-person perspective (typing on a keyboard, swatting a fly, peeling an apple) while the remainder of gestures typically depicted salient features of objects (the shape of a skirt, the height of a stack of letters). To minimize influences of mimic, the face of the actress was covered with a nylon stocking. All gestures resembling emblems or gestures directly related to the target words were excluded.

#### *Pre-Test:*

The selected video material was edited using commercial editing software (Final Cut Pro 5). A pre-test was conducted to assess how effective the gestures were in disambiguating the homonyms. In this modified cloze procedure, the videos were displayed to twenty German native speakers with sound muted one word before the onset of the target word. The participants had to select the most probable sentence continuation. The two alternatives on the response sheet were the dominant and the subordinate sentence continuations (e.g., *was sich im Spiel* / *which during the game* vs. *was sich im Tanz* / *which during the dance*). Overall, the gestures elicited a cloze probability of 93.7 %, which was significantly above chance level ( $p < .01$ ). Only homonyms which could be disambiguated by gesture in at least 80% of all participants were kept, resulting in the final set of 48 homonyms. In this final set, dominant and subordinate gestures did not differ significantly in their cloze probability (paired  $t(1,47) = 0.69, p > 0.4$ ).

#### *Splicing:*

The speech of the sentences was re-recorded in a separate session to improve the sound quality. Because listeners may also use prosodic cues to resolve lexical ambiguities, half of the sentences were realized via a cross-splicing procedure, i.e., sentence parts from different recordings were combined using audio editing software. The aim of this procedure was to keep the speech in the dominant and the subordinate version of an item physically identical for as long as possible. For example, the sentence depicted in Table 2.1 was realized in the following way. The dominant sentence was used as it was recorded. The subordinate sentence was created by substituting the final part of the dominant sentence with the final part of a recording of the subordinate sentence (see Table 2.1 for more details). Thus, in this case only the subordinate sentence was realized via cross-splicing. However, across the complete stimulus set, dominant and subordinate sentences were equally often cross-spliced.

The speech material was combined with the gesture videos resulting in a 2 x 2 design with Gesture (**D**ominant vs. **S**ubordinate) and Target Word (**D**ominant vs. **S**ubordinate) as within-subject factors. The nylon stocking masked the mouth movements of the actress so naïve participants could not identify the speech-lip mismatch in the two incongruent conditions. The final set consisted of 48 item quartets resulting in a total of 192 sentences (see Table 2.1).

#### *Rating of Gesture Phases:*

The onset of the gesture preparation as well as the on- and offset of the gesture stroke were independently assessed by two persons (inter-rater reliability > .90). These values did not differ significantly across gesture conditions (all  $F(3,94) < 1$ ). For more information about the temporal relationship between gesture and speech in the experimental set, see Table 2.2.

Table 2.2: Experiment 1 Stimulus Properties. Mean on- and offset values are in seconds relative to the onset of the introductory sentence (SD in parenthesis).

gesture	target word	gesture onset	stroke	gesture offset	stroke	homonym onset	target onset	word	target offset	word
D	D	2.07 (0.46)		2.91 (0.48)		2.84 (0.40)	3.78 (0.38)		4.16 (0.38)	
D	S	2.07 (0.46)		2.91 (0.48)		2.84 (0.40)	3.80 (0.38)		4.17 (0.38)	
S	S	2.17 (0.52)		3.01 (0.51)		2.84 (0.40)	3.80 (0.38)		4.17 (0.38)	
S	D	2.17 (0.53)		3.01 (0.51)		2.84 (0.40)	3.78 (0.38)		4.16 (0.38)	
Mean		2.12 (0.49)		2.96 (0.50)		2.84 (0.40)	3.79 (0.38)		4.17 (0.38)	

#### Procedure

The participants were seated in a dimly lit, sound-attenuated chamber facing a computer screen. They were instructed to watch and listen carefully. Their task was to judge after each trial whether gesture and speech had been compatible. Note that in order to perform this task, participants had to compare the meaning indicated by the homonym/gesture combination in the initial part of the sentence with the meaning expressed by the target word in the following sub-clause. The videos were centered on a black background and extended for 10° visual angle horizontally and 8° vertically. A trial started with a fixation cross on the screen, which was presented for 2000 ms, followed by the video presentation. Immediately after the offset of the video, a question mark prompted the participants to respond. Feedback was only given if participants failed to respond within 2000 ms after the response cue. Response reaction times (RTs) were measured starting with the presentation of the question mark.

An experimental session (excluding time for electrode application) lasted approximately 90 minutes. The experiment had four blocks each consisting of 48 items. One block lasted approximately eight minutes. A different, completely unrelated experiment was sandwiched between blocks 1 and 2 (part one) and 3 and 4 (part two) to reduce memory strategies. The presentation order of the videos was varied in a pseudo-randomized fashion, separately for each of the two parts. The order of the parts was reversed for half of the participants. In addition, the key assignment for correct (left or right) was counter-balanced across participants, resulting in a total of four experimental lists. One of the four lists was randomly assigned to each participant. That is, each list was seen by six participants.

#### ERP Recording

The EEG was recorded from 56 Ag/AgCl electrodes (Electrocap International). It was amplified using a PORTI-32/MREFA amplifier (DC to 135 Hz) and digitized online at 500 Hz. Electrode impedance was kept below 5 k $\Omega$ , and the left mastoid served as a reference. Vertical and horizontal electrooculograms (EOG) were also measured.

#### Data Analysis

Single-subject ERPs were calculated for each of the four conditions. The epochs were time-locked to the onset of the target word and lasted from 200 ms pre-stimulus onset to 1000 ms post-stimulus onset. A 200 ms pre-stimulus baseline was used. Four Regions of Interest (ROIs) were defined: anterior-left (AL): AF7, AF3, F7, F5, F3, FT7, FC5, FC3; anterior-right (AR): AF4, AF8, F4, F6, F8, FC4, FC6, FT8; posterior-left (PL): TP7, CP5, CP3, P7, P5, P3, PO7, PO3; posterior-right (PR): CP4, CP6, TP8, P4, P6, P8, PO4, PO8. An automatic artifact rejection using a 200 ms sliding window was performed on the EOG channels ( $\pm 30 \mu\text{V}$ ) and on the EEG channels ( $\pm 40 \mu\text{V}$ ). Overall, approximately 30% of the trials did not enter statistical analysis due to artifacts or incorrect responses. Based on visual inspection of the data, a time window from 300 to 500 ms was used to analyze the N400 effects. The N400 is of crucial importance in the current paradigm to examine the impact of gesture on the integration of the target words. The potential effects after 500 ms were beyond the scope of the empirical question and were therefore not statistically analyzed. Recall that the dominant and subordinate version of an item were matched up to the target word, which had a mean length of 380 ms (see Table 2.2). After the target word, the dominant and subordinate sentences continued differently. Thus, the current design does not allow for a clear

interpretation of effects occurring after the target word offset, because such an effect could reflect an impact of the preceding gesture, the specific sentence continuation, or the interaction of both. For the ERP data, a repeated-measure ANOVA using the within-subject factors Gesture (D, S), Target Word (D, S), Part (one, two), Region (anterior, posterior) and Hemisphere (left, right) was calculated. Only effects that involve the critical factors Gesture Target Relation or Target Word Meaning are reported. Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied where necessary. In such cases, the uncorrected degrees of freedom ( $df$ ), the corrected  $p$ -values and the correction factor  $\epsilon$  are reported. Before entering statistical analysis, the data were filtered offline with a high-pass filter of 0.2 Hz. For presentation purposes only, an additional 10-Hz low pass filter was used.

### 2.6.3 Results

#### Behavioral Data

Performance was accurate for all four conditions (DD: 92.0 %; SS 89.8 %; SD 83.6 %; DS 83.9 %) and increased during the experimental run (part 1: 84.3 %; part 2: 91.4 %). An ANOVA with the factors Gesture (2), Target Word (2) and Part (2) revealed a significant three-way interaction between Gesture, Target Word and Part ( $F(1,23) = 10.9$ ;  $p < .0001$ ) as well as a two-way interaction between Gesture and Target Word ( $F(1,23) = 35.66$ ;  $p < .0001$ ). Based on the three-way interaction, separate ANOVAs within each of the four conditions DD, DS, SD and SS were carried out to analyze the simple main effects of Part. The step-down analysis indicated that the increase in performance during the experimental run was only significant for conditions DS ( $F(1,23) = 36.63$ ;  $p < .0001$ ), SD ( $F(1,23) = 10.68$ ;  $p < .0001$ ), and SS ( $F(1,23) = 4.56$ ;  $p < .05$ ) but not for condition DD ( $F(1,23) = 1.15$ ;  $p > .29$ ). To investigate the two-way interaction of Gesture by Target Word, the four conditions were compared via a series of Bonferoni-corrected post-hoc tests. These tests indicated that the responses for condition DD were significantly more accurate than the responses for condition SD ( $F(1,23) = 42.44$ ;  $p_{Bon} < .001$ ) and DS ( $F(1,23) = 43.21$ ;  $p_{Bon} < .001$ ). Similarly, the accuracy for condition SS was significantly greater than the accuracy for condition SD ( $F(1,23) = 21.36$ ;  $p_{Bon} < .001$ ) and DS ( $F(1,23) = 21.24$ ;  $p_{Bon} < .001$ ). No significant differences were observed between condition DD and SS ( $F(1,23) = 4.81$ ;  $p_{Bon} > .23$ ) as well as between condition SD and DS ( $F(1,23) < 1$ ).

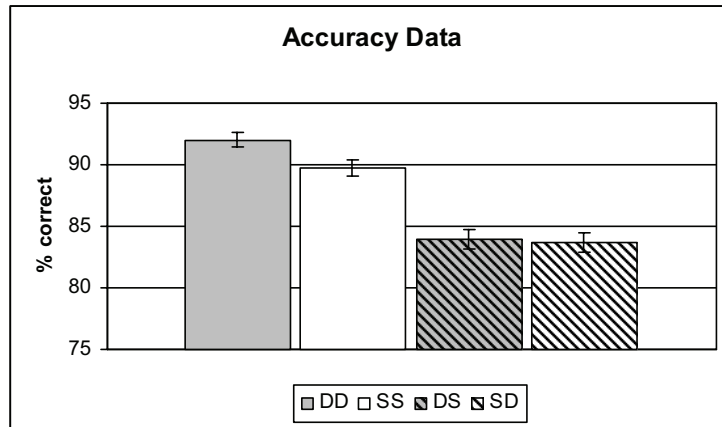


Figure 2.2: Accuracy Data for Experiment 1. The error bars indicate the standard error of the mean (*SEM*).

The RT showed a similar pattern in the four conditions (DD: 601 ms; SS 621 ms; SD 655 ms; DS 660 ms). The corresponding ANOVA showed only a two-way interaction between Gesture and Target Word ( $F(1,23) = 13.57$ ;  $p < .01$ ), indicating that the reaction time was longer for the incompatible conditions DS and SD as compared to the compatible conditions DD and SS.

#### ERP Data

As can be seen in Figure 2.3, the ERPs show an increased negativity for the incongruent conditions SD and DS starting around 300 ms. On the basis of its latency and scalp distribution, the negativity was identified as an N400. After the N400, a sustained negativity for in the incompatible conditions is visible at anterior sites. In addition, there is a positivity for condition SS at posterior sites.

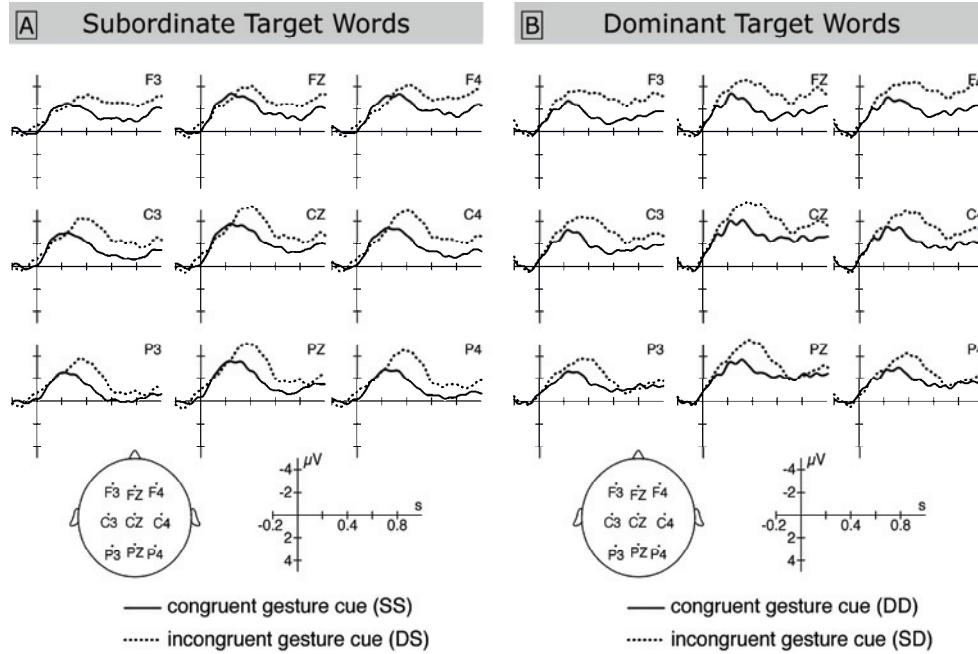


Figure 2.3: ERP Data for Experiment 1. Pairwise presentation of the ERPs time-locked to the target word for all four conditions ( $N = 24$  for each). Negativity is plotted up. In both the left and right panel, the solid line represents the instances in which the preceding nominal gesture cue and the target word were compatible. The dotted line represents the cases in which nominal gesture cue and target word were incompatible.

To test the N400 effect, the mean amplitude from 300 to 500 ms time-locked to the onset of the target word was computed for all conditions. An ANOVA with the factors Gesture (2), Target Word (2), Part (2), Region (2) and Hemisphere (2) yielded significant two-way interactions between Gesture and Target Word ( $F(1,23) = 30.64$ ;  $p < .0001$ ) and between Gesture and Region ( $F(1,23) = 10.78$ ;  $p < .01$ ). The Gesture by Region interaction indicated that the N400 at target words following subordinate gestures was in general slightly more negative at anterior sites ( $F(1,23) = 3.64$ ;  $p < .07$ ) but not at posterior sites ( $F(1,23) = 1.03$ ;  $p > .32$ ). Licensed by the Gesture by Target Word interaction, the simple main effects of Gesture at both types of target words were analyzed. At dominant target words, the N400 was larger after a subordinate gesture ( $F(1,23) = 26.23$ ;  $p < .0001$ ). Conversely, the N400 at subordinate target words was larger if preceded by a dominant gesture ( $F(1,23) = 18.44$ ;  $p < .001$ ). Thus, the activation of both the dominant and the subordinate word meaning varied reliably as a function of the congruency of the gesture context. The observed N400 effects

were stable across the experimental run, i.e., no significant interactions with the factor Part were observed.

#### 2.6.4 Discussion

The question addressed in Experiment 1 was whether dynamic co-speech gestures can be used as disambiguation cues. The behavioral data shows that identifying an incongruent gesture-target word relation was associated with more errors and a longer RT. The ERP data shows that the N400 at a dominant target word was larger after a subordinate gesture. Similarly, the N400 at a subordinate target word was larger following a dominant gesture.

The lower accuracy and higher reaction times for conditions SD and DS suggest that identifying an incongruent relationship between the initial gesture-homonym context and the subsequent target word was a bit more difficult for participants. It can be taken as initial evidence that participants used the gestural information for meaning selection, which caused some degree of interference at incongruent target words. However, because the congruency became already evident at the target word, but response was delayed until the offset of the video, the behavioral data should be interpreted with caution.

The ERP results indicate that the iconic gestures influenced the activation of the word meanings. The N400 at the subordinate target words was smaller after a subordinate gesture and larger after a dominant gesture. Thus, the subordinate meaning was more active in working memory after a subordinate gesture context and less active after a dominant gesture context (see Figure 2.3). Conversely, the N400 at dominant target words was smaller after dominant gesture and larger after a subordinate gesture. Thus, the dominant word meaning was more active in working memory after a dominant gesture and less active after a subordinate gesture. Taken together, the activation of both word meanings varied reliably as a function of the preceding gesture context. Because such a pattern of results is characteristic for strongly biasing context, it is concluded that the iconic gestures constituted a strong contextual cue.

In summary, Experiment 1 demonstrated listeners use gestural information to disambiguate speech. The pattern of results suggests that the gestures strongly biased the activation of the word meanings.

## **2.7 Experiment 2**

### **2.7.1 Introduction**

One limitation of Experiment 1 might be the task that was employed. The participants had to compare the information from both gesture and speech and explicitly judge their compatibility. Thus, the task forced participants to combine gesture and speech. Experiment 2 investigates whether iconic gestures are still used as disambiguating cues once the task is less explicit and no longer requires an integration between gesture and speech.

### **2.7.2 Methods**

#### Participants

In Experiment 2, 25 native German speakers participated, none of whom had participated in Experiment 1. One subject had to be excluded due to excessive artifacts. The remaining 24 participants (12 female) had a mean age of 25 (range 21 – 29) and were right-handed (laterality coefficient 95, Oldfield, 1971). All participants had normal or corrected-to-normal vision and none reported any known hearing deficit.

#### Stimuli

The same stimuli as in Experiment 1 were used.

#### Procedure

Presentation of stimuli was identical to Experiment 1, however, participants were instructed to perform a different task. The aim of the task was to ensure that participants attended to both the visual and the auditory stream of the video. However, the task should not require participants to combine both streams of information and give no cue as to how the arm movements might be related to the contents of speech. The participants received the following instructions: “In this experiment, you will be seeing a number of short videos with sound. During these videos the speaker moves her arms. After some videos, you will be asked whether you have seen a certain movement or heard a certain word in the previous video”.

A visual prompt cue was presented after the offset of each video. After 87.5 % of all videos, the prompt cue indicated the upcoming trial, i.e., no response was required in these trials.

After 6.25 % of all videos, the prompt cue asked participants to prepare for the movement task. A short silent video clip was presented as a probe. The probes were taken from the experimental stimuli and only contained that portion during which the arm movement was executed. After the offset of the probe video, a question mark prompted the participants to respond. Feedback was given if participants answered incorrectly or if they failed to respond within 2000 ms after the response cue. RTs were measured starting with the presentation of the response cue.

After the remaining 6.25 % of the videos, the prompt cue informed the participants that the word task had to be performed. Participants had to indicate whether a visually presented probe word had been included in the previous sentence. The probe words were selected from sentence-initial, -middle and -final positions of the complex sentence. Response and feedback were identical to the movement task trials.

#### ERP Recording and Data Analysis

The parameters for the recording and the analysis of the data were the same as in Experiment 1. Because the movement task was performed after only 6.25 % of all trials, very few gesture related behavioral data (i.e., only 4 responses per condition) were obtained. Therefore, a statistical analysis of the behavioral data is not reported. Behavioral responses were also not a rejection criterion for the ERP trials. Based on the artifact rejection, approximately 11 % of the trials were excluded from statistical analysis. As in Experiment 1, a time window ranging from 300 to 500 ms was selected based upon visual inspection of the data for the statistical analysis of the N400 effects.

### **2.7.3 Results**

#### ERP Data

An enhanced N400 for the incompatible conditions SD and DS is visible starting around 300 ms (see Figure 2.4). In addition, the ERPs for subordinate target words appear more negative as compared to dominant target words.

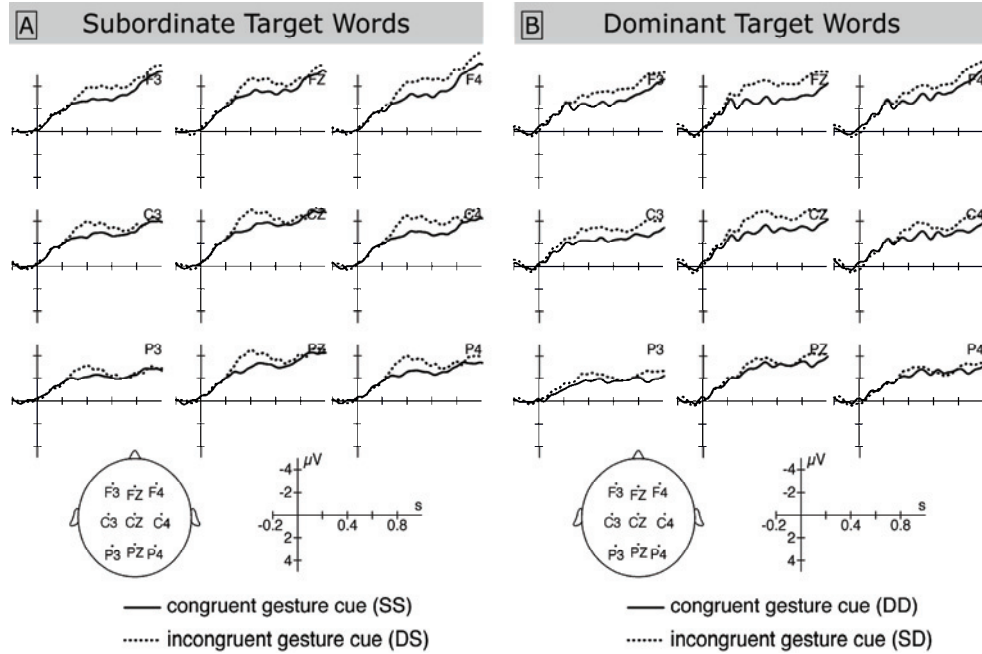


Figure 2.4: ERP Data for Experiment 2. Pairwise presentation of the ERPs time-locked to the target word for all four conditions ( $N = 24$  for each). The solid line represents the instances in which the preceding nominal gesture cue and the target word were compatible, the dotted line the incompatible instances.

The ANOVA for the time window from 300 to 500 ms revealed a significant two-way interaction between Gesture and Target Word ( $F(1,23) = 15.46$ ;  $p < .001$ ) and a main effect of Target Word ( $F(1,23) = 7.84$ ;  $p < .05$ ). The main effect of Target Word indicated that the N400 was slightly more negative at subordinate target words. On the basis of the two-way interaction, the simple main effects of Gesture were tested separately. At dominant target words, the N400 was larger after subordinate gesture ( $F(1,23) = 4.72$ ;  $p < .05$ ). Similarly, at the subordinate target words, the N400 was larger after a dominant gesture ( $F(1,23) = 10.32$ ;  $p < .01$ ). Thus, both word meanings varied reliably as a function of the preceding gesture context. The N400 effects did not interact with the factor Part and thus were stable across the experimental run.

#### 2.7.4 Discussion

The aim of Experiment 2 was to clarify the extent to which the results obtained in Experiment 1 were task-dependent. As in Experiment 1, broadly distributed N400 effects with similar

latencies were observed for the incongruent conditions DS and SD as compared to the congruent conditions SS and DD. This is in analogy to the discussion of Experiment 1 interpreted to reflect a strong impact of the gesture context on disambiguation. Thus, iconic gestures are used as disambiguation cues, even if disambiguation is not explicitly required by the task.

In contrast to the results of Experiment 1, there appears to be a general negative-going trend for all conditions. To investigate this negative shift, longer ERPs were extracted, which included the prompt cue that was presented about 1400 ms after target word onset. These prolonged ERPs show a slowly rising and frontally distributed negativity peaking at 1600 ms. After the peak, the ERPs for all conditions clearly return to baseline level. Based on the scalp distribution and the task manipulation, this negative shift is interpreted as a contingent negative variation (CNV, see, for instance, Rugg & Coles, 1995). In Experiment 2, the participants did not know whether a task was coming up or not until the offset of the video. It is therefore not surprising that the participants built up expectations about the upcoming prompt cue during the final part of the gesture clip. This general expectation is suggested to be reflected in a CNV.

In sum, the results from Experiments 1 and 2 are in line with previous ERP studies (Kelly et al., 2004; Wu & Coulson, 2005) in showing that an initial gesture context can modulate the processing of a subsequent target word. Experiment 1 and 2 extend the previous findings in showing that contextual effects of gesture are not restricted to the processing of isolated target words but can be generalized to auditory sentence processing.

## **2.8 Experiment 3**

### **2.8.1 Introduction**

One important open question in co-speech gesture comprehension is the degree to which the integration between both modalities is an obligatory or “automatic” process. Does the listener always take gesture into account as has been suggested by McNeill and co-workers (1994)? Or are there situations in which the information from the gesture channel has no detectable influence on comprehension? In terms of experimental manipulations, there are different ways of addressing this issue. One way is to test whether the effect of gesture is task-independent. For example, both the studies by Özyürek et al. (2007) as well as Kelly et al. (2004) found an

interference effect, although the task in these experiments did not force the participants to take gesture into account. These findings suggest that at least in situations where gesture and speech provide clearly conflicting information, the interaction between both domains is obligatory.

A recent study by Kelly et al. (2007) approached the question of automaticity in a different way. The experimental paradigm was similar to the previously described study by the same first author (Kelly et al., 2004), but contained an additional manipulation of the intentional relationship between gesture and speech. Gesture and subsequent target word were either produced by the same speaker or by two different speakers. When participants knew that the same speaker produced gesture and speech, the processing of mismatching vs. matching words elicited a bilateral N400 effect. In contrast, when participants knew that gesture and target word were produced by two different speakers, the N400 effect had a markedly different topography and was significant only at right frontal electrode sites. This can be seen as initial evidence that the processing of gesture and speech may not be an entirely automatic process.

A third possibility of testing the relative amount of automaticity is to manipulate the proportion of related vs. unrelated prime-target combinations in an experiment. Two-process theories of information processing (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) state that automatic processes are fast-acting, occur without intention or awareness, and do not use limited-capacity resources. In contrast, controlled processes are slower, are under a person's strategic control, and use limited capacity resources. On the basis of these assumptions, several studies that sought after the automaticity of semantic priming have used a relatedness proportion manipulation (Chwilla, Brown, & Hagoort, 1995; Holcomb, 1988; Koyama, Nageishi, & Shimokochi, 1992). In these experiments, it was tested whether a given indicator of priming (e.g., N400 effect for unrelated vs. related targets) varied as a function of the proportion of related word pairs within an experiment. If the N400 effect was unaffected by the relatedness manipulation, it was concluded that semantic priming is a primarily automatic process. If, however, the N400 effect was smaller in the context of many unrelated word pairs, it was suggested that semantic priming involves a considerable degree of controlled processes. Hahne and Friederici (1999) adopted a similar strategy in determining the automaticity of two different syntactic ERP components.

One possibility to apply a relatedness proportion manipulation to the field of gesture comprehension is the use of hand movements that are unrelated to the contents of speech. One obvious candidate for these meaningless hand movements are the self-touching movements speakers frequently produce, e.g., a speaker might scratch his/her chin, rub his/her temple, squeeze his/her nose and so on. These grooming movements (also called adaptors or manipulators) are typically very repetitive and the speaker is hardly aware of them (Goldin-Meadow, 2003). People probably differ not only in their individual set of grooming movements, but also in the frequency with which they exhibit these behaviors (Ekman, 1999). The impact of grooming on comprehension is not well understood, although there is some evidence that excessive grooming movements can cause a speaker to appear less trustworthy (DePaulo et al., 2003). An important difference between gesture and grooming is that grooming is not systematically tied to a segment of speech. Thus, a grooming movement in the current disambiguation paradigm gives the listener no cue for meaning selection.

One advantage of grooming is therefore that it constitutes a neutral context, which allows to investigate whether the effects of gesture are inhibitory or facilitatory in nature. In Experiments 1 and 2 it was observed that the N400 at subordinate target words was larger after a dominant gesture and smaller after a subordinate gesture. Without an unrelated condition, it is impossible to tell whether this effect occurred because the dominant gesture inhibited the subordinate meaning or because the subordinate gesture actually facilitated the processing of the subordinate meaning.

Another advantage of adding meaningless hand movements to the paradigm is that it allows to test whether gesture-speech integration is a primarily automatic process. Experiment 1 and Experiment 2 have shown that speakers use gestural information to disambiguate speech. The results from Experiment 2 suggest that this disambiguating effect is somewhat task-independent. This can be taken as some initial evidence that a listener performs an automatic integration of gesture and speech as discussed by McNeill and coworkers (1994). However, as has been outlined above, the addition of meaningless grooming movements constitutes another important test for the potential automaticity of gesture-speech integration. Adding the grooming movements makes the manual domain less informative, because under such circumstances only a portion of all observed hand movements provide the listener with a helpful cue for disambiguation. If the integration of gesture and speech is a primarily automatic process, the addition of meaningless hand movements should not weaken the disambiguating effects of gesture, i.e., the N400 of both the dominant and the subordinate

target words should vary as a function of the congruency of the preceding gesture context, as it was the case in Experiments 1 and 2. If, however, the integration of gesture and speech is also substantially influenced by controlled factors, a different pattern of results should emerge. A listener may start to consider *all* manual cues (including the gestures) as less informative once meaningless hand movements are added. Such a devaluation of gesture could result in a pattern typical for weakly biasing contexts, i.e., only the N400 at subordinate target words would vary as a function of the preceding gesture. Finally, it is possible that listeners completely disregard the semantic content of gesture once grooming is added. In this case the N400 at the target words should vary only as a function of word meaning frequency. Based on the results of Experiment 1 and 2 as well as the data from Özyürek et al. (2007) and Kelly et al. (2004), it is hypothesized that gesture-speech integration is an automatic process, i.e., the addition of meaningless grooming movements should not weaken the impact of gesture on homonym disambiguation.

### 2.8.2 Methods

#### Participants

29 participants participated in Experiment 3, none of whom had participated in Experiments 1 or 2. Five participants did not enter statistical analysis because of excessive artifacts. The remaining 24 participants (12 female) had a mean age of 24 (range 19 – 28) and were right-handed (laterality coefficient 93, Oldfield, 1971). All participants had normal or corrected-to-normal vision, and none reported any known hearing deficit.

#### Stimuli

In addition to the stimuli used in Experiment 1 and 2, a third gesture condition was added. Again, our professional actress was videotaped while uttering the sentence stimuli. Instead of the disambiguating gestures, she performed meaningless grooming hand movements (scratching, rubbing, etc.). The sentence material described in Experiment 1 was combined with the newly recorded material resulting in a 3 x 2 design with Gesture (**D**ominant, **S**ubordinate, **G**rooming) and Target Word (**D**ominant, **S**ubordinate) as within-subject factors (see Table 2.3). The three types of hand movements did not differ significantly in the on- and offset of the movement stroke (both  $F(2,141) < 1$ ).

Table 2.3: Examples of additional stimuli used in Experiment 3



Table 2.4: Properties of additional stimuli in Experiment 3. Mean on- and offset values are in seconds relative to the onset of the introductory sentence (SD in parenthesis).

Gesture	Speech	gesture onset	stroke	gesture offset	stroke	homonym onset	target onset	word	target offset	word
G	D	2.16 (0.49)		2.96 (0.50)		2.84 (0.40)	3.78 (0.38)		4.16 (0.38)	
G	S	2.16 (0.49)		2.96 (0.50)		2.84 (0.40)	3.80 (0.38)		4.17 (0.38)	

### Procedure

The less explicit task from Experiment 2 and the same instructions were employed. The experiment consisted of six blocks consisting of 48 items, in which each block lasted approximately seven minutes. In contrast to Experiment 1 and 2, there was no unrelated second experiment embedded at half time, thus the total time of the experimental session (excluding time for electrode application) was reduced to 50 minutes. For the statistical analysis, the data of each participant was divided into three parts (part 1: blocks 1 & 2; part 2: blocks 3 & 4; part 3: blocks 5 & 6). Two pseudo-randomized lists were created. The key assignment for correct (left or right) was also balanced across participants, resulting in a total of four experimental lists.

### ERP Recording and Data Analysis

The data were amplified using a BrainAmp MR plus amplifier (DC to 250 Hz). For the ERP data, a repeated-measure ANOVA using the within-subject factors Gesture (D, S, G), Target Word (D, S), Region (anterior, posterior), Hemisphere (left, right) and Part (1, 2, 3) was calculated. Based on the artifact rejection, approximately 14 % of trials were excluded from statistical analysis. The time window for the statistical analysis of the N400 effect was set from 300 to 500 ms based on visual inspection of the data. All other recording and analysis details were as described in Experiment 1.

### 2.8.3 Results

#### ERP Data

As can be seen in Figure 2.5, the processing of subordinate target words is associated with a larger N400 if preceded by a incongruent dominant gesture or an unrelated grooming movement. The ERPs for these conditions (DS & GS) remain more negative than condition SS even after the N400 time window. At the dominant target words, there may be anteriorly an increased negativity for dominant targets following a subordinate gesture. Finally, there is a general negative-going trend for all conditions, especially at anterior sites.

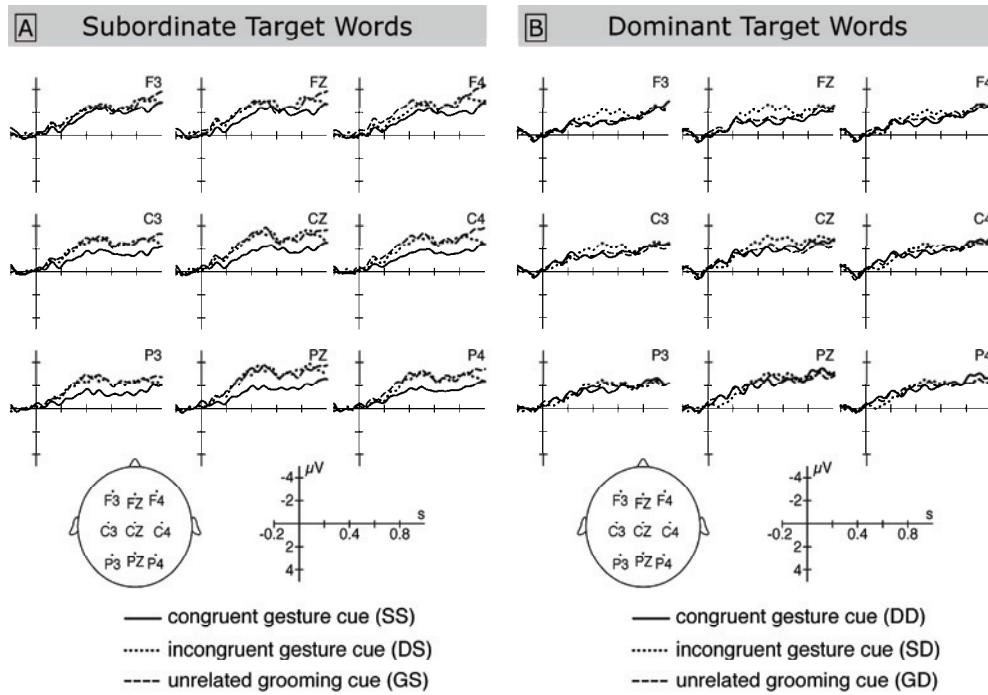


Figure 2.5: ERP Data for Experiment 3. Presentation of the ERPs time-locked to the target word in two sets of three conditions ( $N = 24$  for each). The solid line represents the instances in which the preceding nominal gesture cue and the target word were compatible, the dotted line the incompatible instances. The dashed line represents those cases in which a target word was preceded by an unrelated grooming movement.

An ANOVA for the time window from 300 to 500 ms yielded a significant three-way interaction between Target Word, Part and Region ( $F(2,46) = 4.1$ ;  $p < .05$ ;  $\epsilon = .95$ ) a

marginally significant two-way interaction between Gesture and Target Word ( $F(2,46) = 3.28$ ;  $p < .06$ ;  $\epsilon = .81$ ) as well as a main effect of Target Word ( $F(1,23) = 7.16$ ;  $p < .05$ ).

A number of step-down analyses were performed to clarify the origin of the three-way interaction. In a first step, separate ANOVAs with the factors Target Word and Region were calculated within each level of Part. There were no significant effects or interactions in Part 1 (all  $F(1,23) < 2.03$ ; all  $p > .17$ ) or in Part 3 (all  $F(1,23) < 1$ ). However, it was found that in Part 2 of the Experiment, the N400 was significantly larger at subordinate target words ( $F(1,23) = 7.02$ ;  $p < .05$ ). This was especially the case at anterior sites, as indicated by a significant Target Word by Region interaction ( $F(1,23) = 4.91$ ;  $p < .05$ ). Thus, there was a significant deviation in the way the subordinate target words were processed during the second third of the Experiment. Note, however, that the interaction did not involve the factor Gesture, which is of crucial interest in the current design.

The main effect of Target Word was found to indicate a larger N400 at subordinate words. Based on the two-way interaction of Gesture by Target Word, the simple main effects of Gesture at both types of target words were analyzed. The simple main effect of Gesture was significant at subordinate target words ( $F(2,46) = 4.56$ ;  $p < .01$ ;  $\epsilon = .88$ ) but not at dominant targets ( $F(2,46) < 1$ ). Finally, the three levels of gesture at subordinate targets were contrasted via three Bonferroni-corrected post-hoc tests. It was found that the N400 at subordinate target words was larger after grooming than after a subordinate gesture ( $F(1,23) = 6.99$ ;  $p_{Bon} < .05$ ) and larger after a dominant gesture than after a subordinate gesture ( $F(1,23) = 6.93$ ;  $p_{Bon} < .05$ ). The difference between grooming and dominant gestures at subordinate targets was not significant ( $F(1,23) < 1$ ).

Based the two-way interaction of Gesture by Target Word, the simple main effect of speech was analyzed for grooming. It was found that the N400 at subordinate targets following grooming was more negative than the N400 at dominant targets following grooming ( $F(1,23) = 9.96$ ,  $p < .05$ ). Thus, in the absence of a contextual cue, word meaning frequency determined the activation pattern.

Because it was found that the N400 at dominant target words did not vary as a function of context congruency, one immediate follow-up question concerned the extent to which the dominant meaning was activated at all in Experiment 3. To address this issue, the N400 amplitude of the conditions with a dominant target word (DD, GD, SD) was compared to a

condition where the gesture context had clearly caused a higher activation of the contextually appropriate word meaning (condition SS). The corresponding ANOVA was not significant ( $F(3,69) < 1$ ) suggesting that the dominant word meaning was activated in Experiment 3.

As in Experiment 2, long epochs ranging from -200 ms to 2000 ms relative to target word onset were extracted to clarify the nature of the negative-going trend. Again, an anteriorly distributed and slowly rising negativity with its peak at 1600 ms was revealed. After the peak, the negativity quickly returned to the baseline level. Thus, the negative-going trend is suggested to reflect a CNV as in Experiment 2.

## **2.9 General Discussion of Experiments 1 - 3**

Experiments 1 - 3 investigated whether listeners use the information from iconic gestures to disambiguate unbalanced homonyms. In Experiment 1, participants were explicitly asked to judge the compatibility between an initial homonym-gesture-combination and a subsequent target word. ERPs time-locked to the target word revealed that the N400 was smaller after a congruent gesture context and larger after an incongruent gesture context suggesting that listeners can use gestural information to disambiguate speech. Experiment 2 replicated the results using a less explicit task indicating that the disambiguating effect of gesture is somewhat task-independent. Unrelated grooming movements were added to the paradigm in Experiment 3. This manipulation changed the pattern of results. Only the N400 at the subordinate target words varied as a function of the preceding gesture context whereas the N400 at dominant target words did not.

In Experiment 3, the activation of the subordinate meaning varied as a function of context congruency just as it was observed in Experiments 1 and 2. Grooming as an unrelated condition allows for a clearer interpretation of the N400 effect. Both grooming as well as a dominant gesture context make the integration of a subsequent subordinate target word more difficult as reflected by the increased N400. Only a subordinate gesture context leads to an attenuated N400 at subordinate target words. This result suggests that the underlying mechanism is facilitatory and not inhibitory, because the N400 is smaller after a congruent subordinate gesture than after a neutral grooming context. Thus, a gesture that supports the lesser frequent meaning of a homonym actually facilitates the processing of a related target word in online sentence comprehension. This is an important extension of the existing ERP literature on iconic gesture comprehension (Kelly et al., 2004; Özyürek et al., 2007; Wu &

Coulson, 2005), because previously conducted ERP studies on iconic gesture comprehension have only demonstrated how gesture can impair speech processing<sup>8</sup>. Experiment 3 shows that disambiguation is one mechanism through which gesture information can also *facilitate* speech comprehension. A listener can save resources by using the gestural information to activate the lesser frequent meaning of a homonym. Maybe it is also possible to divert the saved resources to other sources of information (e.g., prosody, body posture), which would have been missed without the facilitatory effect of gesture. However, this is subject to further research.

Because it has been frequently observed that the N400 is sensitive to the degree to which a target word fits into a given context (Kutas & Federmeier, 2000), one could have expected that the N400 at subordinate target words would be larger after an incongruent dominant gesture context than after a neutral grooming context. However, the size of the N400 effects that grooming and dominant gestures elicited at subordinate target words did not differ. One possible explanation is that this is due to a floor effect. Given that the target word is processed (on average) some 900 ms after the homonym/hand-movement combination, it may very well be that in the case of grooming, the subordinate meaning is no longer active at all at the position of the target word. Simpson & Burgess (1985) reported data suggesting that in a neutral context, the subordinate meaning is maintained only for 500 ms following homonym presentation. In light of such data, the similar N400 effects at subordinate target words may reflect that after the processing of both types of hand movements (i.e., grooming and dominant gesture), the subordinate meaning is completely de-activated by the time the subordinate target word is encountered.

There were no significant N400 differences at the dominant target word in Experiment 3. The N400 for the three context types at dominant target words (conditions DD, GD & SD) did not differ significantly from the N400 for condition SS suggesting that the dominant word meaning was activated after all three types of hand movements. Taken together, it was found in Experiment 3 that the activation of the subordinate meaning varied reliably as a function of context congruency whereas the dominant meaning was equally active after a congruent dominant gesture as well as after an incongruent subordinate gesture. As has been mentioned

---

<sup>8</sup> It should be noted that Wu & Coulson (2005) discussed their findings as a facilitatory effect of iconic gestures. Their paradigm, however, did not focus on co-speech gestures.

in the introduction (see page 21), such a pattern of results is typical for weakly biasing context cues. It seems that once a listener is confronted with a mixture of meaningless grooming movements and meaningful gestures, the impact of gesture on speech disambiguation is weakened. Thus, the integration of gesture and speech in comprehension is not a purely automatic process but is modulated by situational factors, in this case the proportion of meaningful and meaningless hand movements. In the following, some possible mechanisms through which the addition of grooming may have weakened the impact of gesture are discussed.

It is in principle possible that the cause for the weaker impact of gesture is simply the increased number of homonym repetitions in Experiment 3. For example, participants may initially have taken gesture into account but as they became more familiar with the stimulus set during the experiment, they may have ceased to pay attention to gestures, as they realized that the gestures actually provided no helpful cue. In terms of statistical factors, such a scenario would imply an interaction between the factors Gesture and Part. However, such an interaction was not found in the statistical analysis. The only significant interaction involving the factor Part was a three-way interaction between Target Word, Region and Part (see results section), which cannot explain the different pattern of results in Experiment 3.

Another possible explanation is that the processing of grooming movements interfered with the processing of the gestures. It is conceivable that the participants misinterpreted some grooming movements as gestures on the one hand and on the other hand mistook some gestures as grooming. The outcome of such misinterpretations would be a weaker impact of gesture. However, such an “active interference” explanation is considered as not very likely because grooming caused the expected neutral context pattern, i. e. the N400 was smaller at dominant targets and larger at subordinate targets after a grooming movement. This result suggests that grooming was an acceptable neutral context for the participants with respect to meaning selection and not a strange distracting movement that interfered with speech processing.

An alternative explanation for the weaker impact of gesture in Experiment 3 is that the addition of grooming decreased the degree to which listeners took gestural information into account. In this experiment, there was only a 66 % chance that an observed hand movement conveyed meaning. This reduced probability may have caused listeners to put less weight on

gestural information and more weight on other sources of information (in this case word meaning frequency) during the meaning selection process. If this explanation is valid, an interesting question for future research will be how much meaningless hand movements are tolerable for a listener before the impact of gesture is weakened. Experiment 3 used a 2:1 ratio of meaningful vs. meaningless hand movements. Although the ratio in natural face-to-face conversation is unknown, it may very well be the case that listeners are used to seeing a higher proportion of meaningful movements (e.g., 3:1).

Whatever the underlying mechanism, it is clear that the addition of grooming weakened the impact of gesture. This result is incompatible with the automaticity notion of gesture-speech integration from McNeill and colleagues: “the point we wish to emphasize is the involuntary, automatic character of forming an idea unit out of information from the two channels” (1994, p. 236). Experiment 3 suggests that gesture-speech integration does not operate in such a modular fashion. Instead, external factors like the proportion of meaningful to meaningless hand movements can also influence the degree to which listeners take gesture into account. This finding may have implications for research on gesture-speech production. For example, it might be the case that speakers who produce few to no grooming movements are more effective communicators than speakers with a high individual grooming frequency. The finding that gesture-speech integration is not an entirely automatic process is also in line with recent data by Kelly et al. (2007) who suggested that the intentional relationship between gesture and speech also influences the degree to which both channels of information interact in comprehension.

Another factor that seems to moderate the impact of gesture is the amount of semantic overlap between gesture and speech. Özyürek et al. (2007) found that in a clear-cut mismatch with no semantic overlap between gesture and speech (e.g., gesturing *rolling down* while saying *knock*) the processing of speech is negatively affected as indexed by an enlarged N400. Kelly et al. (2004) realized different degrees of semantic overlap in their experiment. Gesture and speech were either highly overlapping (match), partially overlapping (complementary) or not overlapping (mismatch). In this study, the processing of target words that mismatched the preceding gesture was associated with an enlarged N400 as compared to the match or complementary condition. In neither of the two studies did the task require taking gesture into account, suggesting that the interference caused by semantically non-overlapping gesture-speech combinations was inevitable in these experiments. The present series of experiments contained stimuli with a moderate semantic overlap. Gesture and speech were semantically

overlapping in that they both referred to the same homonym. However, both domains were also non-overlapping in that gesture always conveyed additional information about the appropriate word meaning of the homonym. In Experiment 3, it was found that such a moderate amount of semantic overlap can facilitate the processing of the lesser frequent word meaning. It is suggested that the amount of semantic overlap determines whether an iconic gesture has an interfering or a facilitating effect on comprehension. If there is almost no overlap between gesture and speech as in the case of clear-cut mismatches, gesture has an interfering effect on comprehension. If there is a moderate amount of semantic overlap as in the case of disambiguation, gesture can facilitate speech comprehension<sup>9</sup>. The idea that a listener benefits from a moderate semantic overlap between gesture and speech is also in line with behavioral data from Goldin-Meadow and colleagues (Alibali, Flevares, & Goldin-Meadow, 1997; Goldin-Meadow & Momeni-Sandhofer, 1999).

Experiment 3 demonstrated that the integration of gesture and speech in comprehension can be modulated by situational factors such as the amount of meaningful hand movements in an experiment. It would be interesting to see whether the addition of meaningless hand movements also weakens the impact of gesture in a mismatch paradigm. Such an experiment would allow some conclusions about which of the two factors – semantic overlap or proportion of meaningful hand movements – plays a more prominent role in determining the degree to which gesture is taken into account.

### **2.9.1 Conclusion**

Experiments 1 - 3 sought after the extent to which listeners make use of the additional information provided by iconic gestures in speech comprehension. Two opposing theoretical views on this issue have been put forward in the literature, i.e., the “strong-impact view” and

---

<sup>9</sup> The data from Kelly et al. (2004) may appear incompatible with this suggestion because in that study the N400 in the complementary condition (with a moderate gesture-speech overlap) was not more attenuated as compared to the no-gesture condition. Note, however, the possibility that a facilitatory effect of gesture was not observed in this study because of a potential ceiling effect. Since each of the four target words was repeated 48 times during that experiment, the N400 for the target words may already have been attenuated to a large degree during the no-gesture condition, leaving “no room” for an additional facilitatory effect of gesture in the complementary condition.

the “weak-impact view” of gesture. In sum, the data are in line with the “strong-impact view”, because it has been demonstrated that listeners use gestural information to disambiguate speech. Particularly the processing of a lesser frequent meaning of a homonym can be facilitated by iconic gestures. Another important finding is that the integration of gesture and speech is not an obligatory process, but is modulated by situational factors. Once the listener is confronted with a mixture of meaningful and meaningless hand movements, the impact of gesture is weakened.



## Chapter 3: Temporal Aspects of Iconic Gesture Comprehension

The three conducted ERP experiments suggest that there is a substantial amount of interaction between gesture and speech in comprehension. This finding is in line with behavioral studies which have shown that a listener can extract the additional information provided by iconic gestures (e.g. Alibali et al., 1997; Beattie & Shovelton, 1999a, 2002a). However, little is known so far about *how much* of a gesture an addressee needs to see before it becomes meaningful.

Determining such a “gesture recognition point” is interesting for both theoretical as well as methodological reasons. On the basis of an interpretative approach to gesture, McNeill (McNeill, 1992, p. 24, pp. 375/376) has claimed that most of the meaning of gesture is conveyed during the stroke phase (for a definition of the different phases of gesture, see section 1.2). However, to date, there are no empirical studies available that have systematically investigated this issue. Determining a gesture recognition point is also important for methodological reasons. Event related methods in cognitive neuroscience, such as ERPs or event related functional magnetic resonance imaging, require the use of similar repeated events (e.g. a set of gestures). Event related methods assume that for each event, certain cognitive operations (e.g., semantic integration) occur at approximately the same point in time relative to event onset. If this assumption holds true, aggregating data over a sufficient number of events will result in a characteristic pattern (e.g., an ERP) reflecting the neural correlate of the cognitive operation. However, if the assumption is violated, because the time from event onset to the onset of the cognitive operation is variable across the different events, the method will not yield meaningful results. The time phases of iconic gestures show a tremendous temporal variation, especially the time from preparation onset to stroke onset varies considerably across gestures. Thus, on the one hand, the preparation onset of a gesture is probably not a good choice for time-locking the events, because the time from event onset to the suggested meaningful part of gesture (the stroke) is too variable. On the other hand, the stroke onset represents also a suboptimal choice for time-locking the events, because one cannot rule out the possibility that the observed differences in the dependent variable are partly due to semantic differences in the preceding preparation phase. In this dilemma, a

gesture recognition point may represent a suitable alternative, because it represents the earliest point in time at which the meaning of gesture is available.

### **3.1 Theoretical Views on Temporal Aspects of Iconic Gesture Comprehension**

In contrast to static visual stimuli (e.g., pictures), iconic gestures are a dynamic signal which can be separated into formally distinct time phases. These phases have been described in detail earlier (see section 1.2). Here, I will only recapitulate the key characteristics of the prototypical gesture phases, i.e., preparation, stroke and retraction.

The preparation phase begins when the hands have left the resting position. The hands rise to the location where the stroke is to be performed. During the preparation phase, the hands often already assume the hand shape and orientation needed to execute the stroke (e.g., the gripping hand shape in the bending back example, see p. 6). This anticipatory nature of the preparation phase is also reflected in the temporal gesture-speech synchrony in that the onset of preparation tends to precede the onset of the co-expressive speech unit (McNeill, 1992, 2005; Morrel-Samuels & Krauss, 1992). The preparation phase ends with the onset of the stroke.

The stroke phase is the time period during which the peak effort of movement occurs. A stroke is the defining feature of a gesture in that without a stroke, a gesture is not said to occur (McNeill, 2005, p. 32). In contrast, the preparation and/or the retraction phase can sometimes be omitted, for example when immediately following a stroke, the stroke of a new gesture is executed. The stroke of an iconic gesture frequently coincides with co-expressive speech unit, a phenomenon known as stroke-speech synchrony (see also section 1.3).

The retraction phase begins, when following a stroke, the hands start to return to a resting position. It ends when they have reached this resting position.

Whereas these time phases can be defined on a formal level<sup>10</sup>, it has been suggested that meaning is conveyed to a different degree throughout the three gesture phases. Particularly the stroke phase has been suggested as that segment of iconic gesture in which the meaning is expressed (McNeill, 1992, 2005).

Concerning the degree to which the preparation phase of gesture may also convey meaning to the listener, the literature is not very explicit. Throughout his book on gestures (1992), McNeill makes various suggestions about the functional significance of the preparation phase. For instance, one passage conjectures that the preparation phase is predominantly involved in gesture-speech synchrony: “The stroke is the phase that carries the gesture content. The preparation phase is crucial for the question of gesture timing.” (1992, p. 84). Other passages stress the fact that the form of the hand movement in the preparation phase often anticipates the upcoming stroke and its accompanying speech unit. For example, he notes the “early signalling of the image in the preparation phase” (p. 30) and characterizes this initial gesture phase as “anticipation of speech by gesture” (p. 26).

In sum, the literature suggests that the meaning of gesture is conveyed primarily in the stroke phase of gesture. However, because the preparation phase already contains certain ‘phonological’ features of the stroke (e.g., hand shape or location), it may very well be the case that listeners can extract a substantial amount of gestural information from preparatory movements. Importantly, to date, all claims that have been made with respect to the impact of the different gesture phases on communicating meaning are based solely on an interpretative approach to gesture (Krauss et al., 1996). So far there are no empirical studies available that have investigated this issue. Thus, it is an important open empirical question at what point in time a gesture becomes meaningful for a listener.

---

<sup>10</sup> In fact, McNeill *includes* meaningfulness in his definition of the stroke-phase: „Semantically it [i.e. the stroke phase] is the content-bearing part of the gesture. Kinesically, it is the phase carried out with the quality of ,effort‘“ (1992, p. 375-376). Note, however, that this bipartite definition of the stroke phase is circular in semantic terms and precludes a systematic empirical investigation of the question at what exact point in time a gesture becomes meaningful. Therefore, the stroke phase is defined solely on formal properties throughout the present dissertation.

In the following, I will introduce the method that was employed to address this issue, i.e., the gating paradigm. Next, I will present two experiments designed to determine and validate the point in time at which a gesture becomes meaningful.

### **3.2 Method of Investigation: Gating**

Gating is a very popular paradigm in research on spoken word recognition (Grosjean, 1996). During the gating procedure, linguistic stimuli are presented in segments of increasing duration. After each segment, participants are requested to propose the word being presented. Additionally, they are often required to give a confidence rating for their proposal.

The rationale of the gating procedure is based on the assumption that word recognition is a discriminative process. For example, according to the cohort model of spoken word recognition (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987), processing the first phoneme of a word activates all lexical entries that are positionally consistent with that information (i.e., the word's cohort). As subsequently encoded phonemic information becomes available, activation is withdrawn from lexical entries with which the information is inconsistent, and the cohort is resolved when only a single candidate remains.

Although the gating paradigm was first used in spoken word recognition, it can be virtually used with any kind of sequential stimuli, including the processing of native vs. non-native phonemes, the recognition of signs in the American Sign Language (ASL, Emmorey & Corina, 1990) and the processing of musical sequences (Jansen & Povel, 2004).

The gating method can be adapted to suit the needs of the research question. Among others, the task can differ with respect to the increment size of the gates (20 – 100 ms), the presentation format (successive or duration-blocked), context (with or without context) and type of response (written vs. oral). Some of these variables affect the outcome of gating. For example, it has been found that the successive presentation format is associated with a certain degree of response perseveration and thus may yield a more conservative picture of comprehension than the blocked format (Walley, Michela, & Wood, 1995).

Besides the various possible designs there are two general types of independent variables that can be manipulated. On the one hand, there are stimulus characteristics like word frequency, length and morphology, on the other hand there are diverse types of context, for example

phonetic or visual cues for words and prosodic variables for sentences. Both types of independent variables can affect the three most often employed dependent variables of the gating paradigm, which are isolation point, confidence ratings and candidates proposed.

The isolation point is the size of the segment needed to identify a stimulus without change in response thereafter (Grosjean, 1996). In contrast to the isolation point, which is the basic, obligatory measure of the gating paradigm, the two other measures (confidence rating as well as number of candidates proposed) do not necessarily have to be obtained. Furthermore, one can calculate the recognition point, which is the size of the segment needed to reach a certain confidence level (usually above 80%) as well as a number of further measures (missing values, total acceptance point).

Over the past twenty-five years, many different well known effects of language processing have been shown using the gating paradigm, e.g., the effect of context on word recognition (e.g., Craig, Kim, Rhyner, & Chirillo, 1993; McAllister, 1988), the effect of word frequency on word recognition (e.g., Tyler & Wessels, 1985; Walley et al., 1995) as well as an effect of word length (e.g., Craig & Kim, 1990; Grosjean, 1980). It has also been found that words can be recognized earlier when a listener is additionally presented with the visual display of the speaker (Hardison, 2005).

In summary, the gating paradigm is very useful for investigating the recognition of any kind of sequential stimulus material. Its biggest disadvantage is that it might not only reflect online processes but possibly also slower offline operations in stimulus recognition. Note that there is some disagreement about this issue (Grosjean, 1996).

### **3.3 *Experiment 4***

#### **3.3.1 Introduction**

So far, no empirical data is available with regard to the question at what point in time an iconic gesture becomes meaningful for a listener. The available literature, which is based on an interpretative approach to gesture, suggests that the stroke phase of gesture is the content-bearing part of these hand movements (McNeill, 1992, 2005). On the basis of these conjectures, it is hypothesized that listeners should be able to determine the meaning of gesture at some point during the stroke phase. Experiment 4 tests this hypothesis by means of a gating procedure.

As has been mentioned in the general introduction (see section 1.4), iconic gestures convey meaning in a different way than speech does. One iconic gesture typically activates an array of possible meanings, and these meanings are often difficult to verbalize. This makes a gating procedure of iconic gestures with a free proposal of the meaning impossible. Instead, the gestures have to be placed in some context to reduce the number of possible interpretations.

Unbalanced homonyms, as they have been used in Experiments 1 – 3, may be especially suited to provide such a constraining context for gating. According to exhaustive access models of homonym processing (Onifer & Swinney, 1981; Swinney, 1979), a homonym initially activates both word meanings. This initial multiple-access phase (suggested to last a few hundred ms, Swinney, 1991) is followed by a meaning-selection phase, where the contextually appropriate word meaning is selected. As has been shown in Experiment 1 – 3, listeners use iconic gestures as contextual cues during meaning selection. Thus, in a gating experiment, there are only two possible interpretations for a segment of an iconic gesture observed in the context of an unbalanced homonym: a more frequent dominant meaning and a lesser frequent subordinate meaning. Such a forced-choice version of the gating paradigm has been employed previously, for example in word onset detection (Arciuli & Cupples, 2004) or the processing of native vs. non-native phonemes (Schulpen, Dijkstra, Schriefers, & Hasper, 2003).

#### The Present Experiment

Experiments 1 - 3 have shown that listeners use the additional information provide by gesture for disambiguation. In a next step, Experiment 4 will explore at what point in time gesture starts to exert its disambiguating influence. To this end, the point at which participants can reliably select the correct meaning of a homonym on the basis of a gesture will be determined in a gating paradigm. Based on the literature, it is hypothesized that disambiguation should occur at some point during the stroke phase.

### **3.3.2 Methods**

#### Participants

Forty native German-speaking students (21 female, age range 19-29 years, mean 24.7 years) participated in Experiment 1. All were right-handed (mean laterality coefficient 93.7, Oldfield, 1971). The participants in this and the following experiments were paid 7 Euros per

hour for their efforts and signed a written informed consent. All had normal or corrected-to-normal vision. None of the participants had taken part in any previous experiments using the identical stimulus material.

### Stimuli

The same set of gesture stimuli, which was described in detail in Experiment 1, was also used in this experiment (see Table 2.1, p. 26). The experimental set consisted of 96 gestures which were presented without sound, starting at the preparation phase (for more presentation details, see below). The average length of the preparation phase across the 96 gestures was 0.421 sec (*SD* 0.188 sec), whereas the average length of the stroke phase was 0.839 sec (*SD* 0.332 sec).

### Gating

In a typical gating experiment, participants are presented with stimuli in segments of increasing duration. After each segment, participants must classify the segment with respect to a given parameter. As has been mentioned in the introduction to this experiment, gating with iconic gestures is only possible if the number of possible response alternatives is low because of the high variability of the meaning attributed to these gestures. In the present experiment, iconic gestures are placed in the context of a homonym which greatly reduces the number of possible interpretations, as participants only have to decide whether a segment is more compatible with the dominant or the subordinate meaning of a homonym.

The increment size in the current experiment was one frame (corresponding to 40 ms), i.e., each segment of gesture was 40 ms longer than the previous one. Gating started at the onset of the preparation phase and ended either when the offset of the stroke phase was reached or when the subject gave a correct response for 10 consecutive segments. Because very short video sequences are difficult to display and recognize, each segment also contained the 500 ms directly before the onset of the preparation. Thus, the shortest segment of each gesture had a length of 540 ms (500 + 40 ms for the first frame of the preparation phase).

### Procedure

Participants were seated in front of a computer screen and were instructed that they would be seeing a German word (the homonym) for a short time followed by a soundless gesture video. They were told that each gesture would be presented several times with increasing duration. Their task was to determine whether the homonym referred to the dominant or the subordinate

meaning based on gesture information. Three response alternatives were possible: (1) dominant meaning (as indexed by the dominant target word), (2) subordinate meaning (i.e., the subordinate target word) and (3) "*next frame*". Subjects were instructed to choose the third response alternative until they felt they had some indication of which meaning was targeted.

A trial started with the homonym presented on the screen for 500 ms, followed by a blank screen (300 ms) and the gesture video. The videos were centered on the screen and extended for 10° visual angle horizontally and 8° vertically. Light gray was used as background color in order to make the experiment less stressful for the participants' eyes.

500 ms after the video offset, the three response alternatives were displayed on the screen in blue colored font, with one target word located to the left of the center of the screen and the other one to the right of the center. "*next frame*" was located at the bottom center of the screen and was presented in a smaller black font. The response buttons on the keyboard were arranged in a similar fashion. Subjects had to press the corresponding response button to confirm their choice.

After an ISI of 500 ms, the next trial started. This procedure was repeated for every single gesture with increasing duration until either the whole gesture (until stroke offset) had been displayed or the stop criterion of ten correct successive responses had been met.

The gestures items were pseudo-randomly distributed across two experimental lists (list: A vs. B). Each of the lists contained 24 dominant and 24 subordinate gestures, resulting in a total of 48 gestures per experimental list. For each homonym, either the dominant or the subordinate gesture was within one list.

Every participant had to accomplish this procedure for 48 gestures plus 1 training item. Short pauses were possible after each gesture item. An experimental session lasted about 90 minutes.

#### Data Analysis

The gesture segment at which participants chose the correct meaning without any changes in response thereafter was determined as the disambiguation point. The disambiguation point in this experiment has some similarities with the isolation point in spoken word recognition which is defined as the size of the segment needed to identify the word without any changes

in response thereafter (Grosjean, 1996). In the present experiment, the disambiguation point is the amount of gesture information needed to identify a gesture as either being related to the dominant or the subordinate meaning of a homonym without any changes in response thereafter. An important difference, however, is that the isolation point is determined in a context-independent manner, whereas the disambiguation point of gesture is dependent on the context of a homonym.

In the rare instance that even the longest segment of gesture did not elicit a correct response from a participant, the response was excluded from further analysis (68 of 1920 responses, 3.5 %). The remaining data was used to calculate the mean disambiguation point for each gesture.

The disambiguation points were analyzed using a repeated measure ANOVA with the factors word meaning frequency (dominant vs. subordinate) and list (A vs. B). Additionally, some exploratory analyses were performed. Independent-samples t-tests were calculated for one-handed gestures vs. two-handed gestures, symmetric vs. non-symmetric two-handed gestures and slow vs. fast gestures.

### 3.3.3 Results and Discussion

The mean disambiguation points for the single items ranged from 2.22 to 19.63 frames ( $M = 9.88$ ,  $SD = 3.6$ ), calculated relative to the preparation onset. Thus, on average the participants needed to see about 400 ms of gesture to disambiguate a homonym.

The ANOVA with the factors word meaning frequency (2) and list (2) revealed that dominant gestures ( $M = 9.33$ ,  $SD = 3.6$ ) were identified faster than subordinate gestures ( $M = 10.42$ ,  $SD = 3.58$ ) as indicated by the significant main effect of word meaning frequency ( $F(1,94) = 4.2$ ,  $p < .05$ ). All other main effects or interactions were not significant (all  $F$ s  $< 2.8$ , all  $p$   $> .10$ ). This result may suggest that more gesture information is needed to select the subordinate meaning because the comprehension system has a strong bias towards selecting the more frequent dominant meaning.

None of the exploratory analyses yielded significant results (one-handed vs. two-handed gestures ( $t(1,94) = .678$ ,  $p = .49$ ); symmetric vs. non-symmetric two-hand gestures ( $t(1,49) = -1.489$ ,  $p = .14$ ); slow vs. fast gestures ( $t(1,94) = 1.106$ ,  $p = .27$ )).

However, when exploring the distribution of the disambiguation points relative to the stroke onset, a surprising result was found. Disambiguation points ranged from almost twenty frames

before the stroke onset to nine frames past the stroke onset, with the disambiguation points of 60 gestures being prior to the stroke onset. This means that almost two thirds of all gestures enabled a meaning selection before the participants had actually seen the stroke (see Figure 3.1). The difference between disambiguation point and stroke onset was found to be significantly smaller than zero across participants ( $t_1(1,39) = -4.7, p < .001$ ) and items ( $t_2(1,95) = -2.3, p < .05$ ). The corresponding  $minF'$  statistic (Clark, 1973) was significant ( $minF'(1,128) = 4.26, p < .05$ ) indicating that gestures reliably enabled a meaning selection before the onset of the stroke.

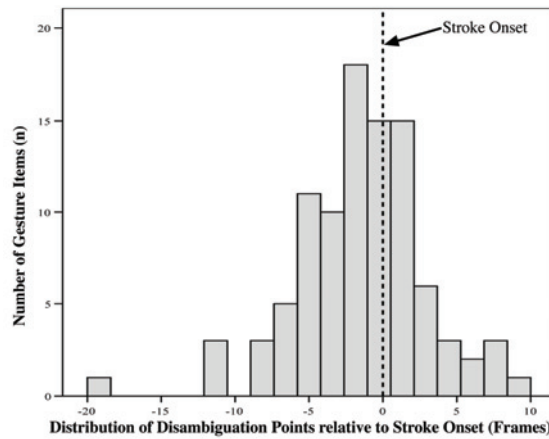


Figure 3.1: Results of Experiment 4 shown as a histogram of Disambiguation points relative to stroke onset. The x-axis shows the difference in frames between the disambiguation point and the stroke. Thus, gestures with a disambiguation point before the stroke receive a negative value, gestures with a disambiguation later than the stroke get a positive value. One frame corresponds to 40 ms. The y-axis indicates how often a specific difference occurred in the stimulus set.

Based on the literature, it was hypothesized that gestures can disambiguate homonyms at some point during the stroke phase. The findings of Experiment 4, however, suggest that in the present stimulus set very often the information in the preparation phase already suffices to select the appropriate meaning of a homonym. The way the gestures were presented in Experiment 5 differs of course dramatically from the way they are perceived in a natural environment. For example, the gestures were not accompanied by speech in the gating procedure. Instead, the visually presented homonym served as a context for gesture. Additionally, gating is characterized by massive repetition of the stimuli. Both of these factors

may have induced processing strategies very different from real-life situations. Thus, it is important to explore whether the “early” disambiguating effect of gesture is also detectable in a co-speech context.

### **3.4 Experiment 5**

#### **3.4.1 Introduction**

During the initial stimulus preparation (see Methods section of Experiment 1, p. 26), a modified version of the cloze procedure was employed to assess how effective the gestures can disambiguate the homonyms. In Experiment 5, the same task was used but this time the gesture was only presented up to the disambiguation point. If, as suggested by the results from Experiment 4, the disambiguating information of gesture is already present at this early point in time, the clipped gestures should also enable a successful disambiguation in a co-speech context, i.e., the elicited cloze probability should be significantly above chance level.

#### **3.4.2 Methods**

##### Participants

Twenty native German-speaking students (10 female, age range 20-30 years, mean 25.6 years) participated in Experiment 5. All were right-handed (mean laterality coefficient 84.6, Oldfield, 1971). None of the participants had taken part in Experiment 4 or in any previous experiments using the identical stimulus material.

##### Stimuli

The same 96 gesture items as in Experiment 4 were used as stimulus material in Experiment 5. The gesture videos were accompanied by the re-recorded speech-streams (see Methods section of Experiment 1, p. 26). Gestures were clipped at the disambiguation point as determined in Experiment 4 by inserting a recording of the empty background. The sound of the sentences was muted one word before the onset of the target word. Because all disambiguation points occurred before the target word, this manipulation created the illusion of a speaker that disappeared during the sentence while the speech went on for a bit longer.

The resulting 96 clipped gesture items were pseudo-randomly distributed to two experimental lists (A & B). Each list contained 24 dominant and 24 subordinate gesture items, resulting in a

total of 48 items per list. For a particular homonym, only the dominant or the subordinate gesture was within one list.

For both lists, response sheets were constructed containing item number, dominant continuation of the speech streams (e.g., *was sich im Spiel* / *which during the game*) and subordinate continuation of the speech streams (e.g., *was sich im Tanz* / *which during the dance*). A blank column was added for annotations. The spatial arrangement of dominant and subordinate continuations on the response sheet was pseudo-randomized.

### Procedure

For both of the two experimental lists, ten participants were tested in a lecture room in one single session. They were seated separately to prevent them from cheating. The participants were told that they would be watching gesture videos accompanied by speech both being clipped after some time and that they should pay attention to both gesture and speech. The videos were presented on a silver screen using a video projector while speech was presented via two speakers left and right of the silver screen. After each video clip, the experimenter stopped the player for approximately 5 seconds. Within in this short duration, participants had to choose the continuation which they thought fit best to the muted speech by marking it on the response sheet. If none of the given continuations seemed appropriate, participants could also write down an alternative continuation. An experimental session consisted of three training trials and 48 experimental trials and lasted about 40 min.

### Data Analysis

Items which were not answered with the given continuations or not answered at all were excluded from the analysis (30 of 960, 3.1 %). For the remaining items, the cloze probability (i.e., the percentage of subjects who chose the target word related to the gesture's meaning as sentence continuation) was computed for each dominant and subordinate gesture. A one-sample t-test was used to test whether the cloze probability elicited by the gestures was significantly above chance level. In addition, a repeated measure ANOVA with the factors word meaning frequency (dominant vs. subordinate) and list (A vs. B) was calculated. Finally, the cloze probability of the clipped gestures in this experiment was compared to the cloze probability of the identical gestures shown at full duration by calculating a paired t-test.

The probabilities for these complete gestures derive from a pre-test that was conducted during the stimulus preparation for Experiment 1 (see p. 26).

### 3.4.3 Results

The cloze probability elicited by the gestures ranged from 20 % to 100 % ( $M = 78$ ,  $SD = 19$ ) with 12 gestures having a probability of 50 % or less. The cloze probability elicited was found to be significantly above chance level both across participants ( $t_1(1,19) = 12.5$ ,  $p < .0001$ ) and items ( $t_2(1,95) = 14.4$ ,  $p < .0001$ ). The corresponding  $minF'$  statistic (Clark, 1973) was significant ( $minF'(1,52) = 89.1$ ,  $p < .0001$ ) giving rise to the assumption that even short, incomplete gestures can contain semantic information. No effect of the ANOVA was significant (all  $F < 2.2$ , all  $p > .14$ ) indicating that neither word meaning frequency nor experimental list had an influence on cloze probability.

Gestures clipped at the disambiguation point differed from the complete gestures with respect to their cloze probability. Complete gestures elicited a mean cloze probability 93.7 %, while clipped gestures elicited only a probability of 78 % (see Figure 3.2). This significant difference (paired- $t(1,95) = 8.753$ ,  $p < .01$ ) indicates that complete gestures possess more disambiguating power than the clipped gestures, probably due to their higher semantic content. Nevertheless there is substantial semantic information contained in the clipped gestures, because the elicited cloze probability was significantly above chance level.

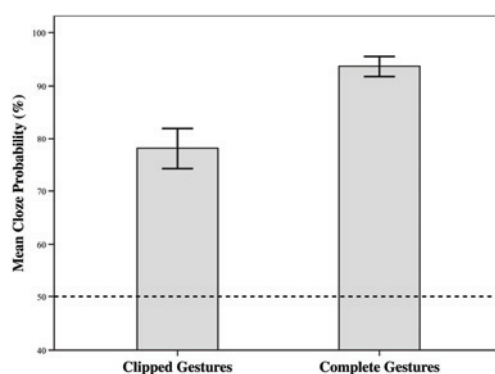


Figure 3.2: Results of Experiment 5. Mean cloze probabilities elicited by the clipped gestures and the complete gestures. The data for the complete gestures was collected during the initial stimulus preparation (see Method section of Experiment 1, p. 26). The error bars indicate a confidence interval of 95 %, i.e.,  $\pm 1.96 * SEM$

### 3.5 General Discussion of Experiments 4 and 5

Experiments 4 and 5 explored how much gesture information is needed to successfully disambiguate a homonym. In Experiment 4, the earliest point in time at which a gesture helps to disambiguate between multiple meanings of a homonym was determined. Experiment 5 aimed to validate this disambiguation point in a co-speech context. The results from Experiment 4 show that for 60 out of 96 gestures, the information contained in the preparation phase suffices for successful disambiguation. Experiment 5 extended this finding to a co-speech context in a sentence completion task.

It is surprising that participants so often made a correct meaning selection based on gesture information prior to that part of gesture thought to be most informative, i. e. the stroke. In the following some possible explanations of this unexpected finding are discussed. One reason may be that homonyms are a special testing ground for studying gesture comprehension. If taken out of its co-speech context, the meaning of an iconic gesture is very imprecise (Hadar & Pinchas-Zamir, 2004; Krauss et al., 1991). However, in the context of a homonym, the number of possible interpretations for an observed hand movement is greatly reduced: the participants only have to determine whether the gesture is compatible with either the dominant or the subordinate meaning. In some instances, this can be a quick decision. A good example is the homonym *Kamm*, meaning either *comb* or *crest*. Both meanings can be correctly selected long before the stroke occurs. In the case of the dominant gesture, the posture of the right hand at the disambiguation point indicates that an object is being held which renders the *crest* meaning somewhat implausible (see Figure 3.3a). In the case of the subordinate gesture, both hands are ascending making the *comb* meaning less probable (see Figure 3.3b). Although in both examples the gesture is not yet fully developed at the disambiguation point, participants seem to be able to *anticipate* the upcoming stroke. In this way, gesture may often allow a quick inhibition of one meaning already in the preparation phase.

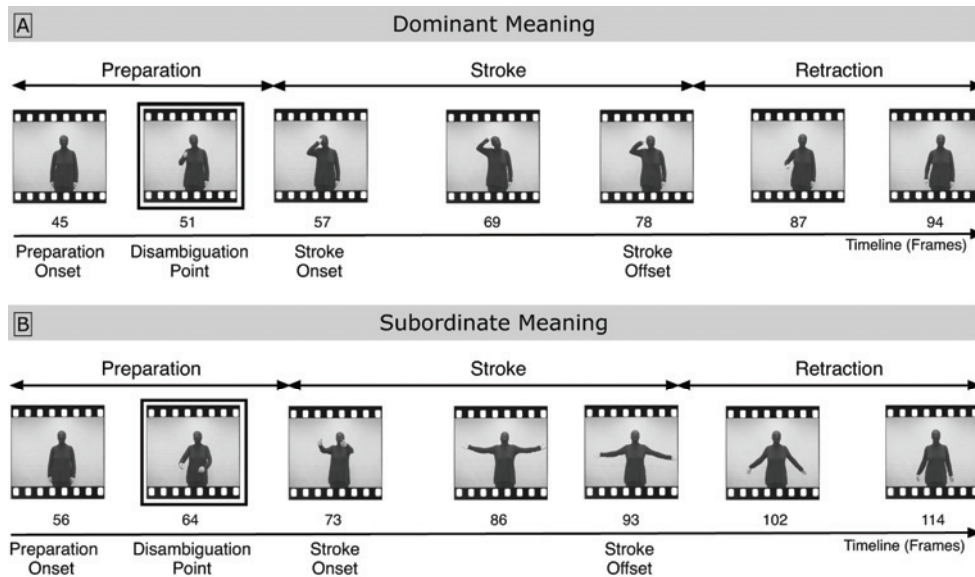


Figure 3.3: Example of “early” disambiguation. The still images exemplify the two gestures used to disambiguate the homonym *Kamm* (dominant meaning: *comb*; subordinate meaning: *crest*). The position of the stills on the x-axis corresponds roughly to the position in the video stream. Additionally, the exact frame number is given below each still. The duration of the three gesture phases is indicated by the arrows on top. a) Gesture for the dominant meaning b) Gesture for the subordinate meaning

It should be noted that the early impact of gesture may be partly due to the experimental setup. In both experiments, participants were explicitly made familiar with both meanings of a homonym and forced to select a meaning based on the gesture information. Although a large body of literature indicates that an unbalanced homonym initially activates both meanings followed by a meaning selection (e.g., Elston-Guttler & Friederici, 2005; Swaab, Brown, & Hagoort, 2003; Swinney, 1979; Van Petten & Kutas, 1987), the forced-choice paradigm may have triggered somewhat different processing strategies than in regular online language comprehension. Further research is needed to evaluate whether preparatory gesture information can also influence online measures of language comprehension.

Another important result of Experiment 5 is that complete gestures allow a more accurate meaning selection than clipped gestures. From a theoretical perspective, one might have expected that once the system has selected one meaning, further gesture information is not helpful for meaning selection. Clearly, this is not the case. The information conveyed after the

disambiguation point leads to an additional 15% increase in selection accuracy. Homonym disambiguation through gestures seems to be a process that does not stop at the disambiguation point. It is suggested that the disambiguation point represents the point in time at which gesture information allows for a first “educated guess” about the contextually appropriate meaning of a homonym. Subsequent gesture information may serve to strengthen this initial selection.

The finding that iconic gestures can convey meaning in the preparation phase has important repercussions on ERP research on gesture comprehension. Previous studies have time-locked the ERPs typically to the onset of the stroke (Özyürek et al., 2007; Wu & Coulson, 2005). On the one hand, this is a reasonable decision, because the stroke is very often the most salient feature of a gesture and can easily and reliably be identified. On the other hand, Experiments 4 and 5 suggest that a substantial amount of information is already present at the onset of the stroke. If, for example, the aim of a hypothetical study is to compare the time-course of semantic integration for gesture and speech, the stroke onset is not a good time point for time-locking the ERPs. In this case, gesture would get an unjustified head start because of the information already conveyed in the preparation phase.

In summary, the data from Experiment 4 and 5 suggest that (at least in a forced-choice situation), even the anticipatory information conveyed in the preparation of a gesture can often suffice to disambiguate a homonym. Because Experiments 4 and 5 focused on offline measures of language comprehension, future research should investigate the extent to which preparatory gesture information is also available during online language processing.

## Chapter 4: Neural Correlates of Iconic Gesture Comprehension

For a listener, the communicative value of an iconic gesture is based upon its form. As the name implies, the form of the gesture is directly related to the denoted thing. However, a listener cannot determine the *exact* meaning of a gesture solely on the basis of the form. The high variability in the meaning listeners attribute to decontextualized iconic gestures (Feyereisen et al., 1988; Hadar & Pinchas-Zamir, 2004; Krauss et al., 1991) suggests that there is a one-to-many (rather than a one-to-one) mapping of form onto meaning for these gestures. Only in combination with the co-speech context can the listener determine the precise meaning of an iconic gesture. Thus, the meaning of an iconic gesture is determined both by its form as well as by the speech context in which it is performed.<sup>11</sup>

Based on such considerations, the prevalent view in the literature has been that during comprehension, listeners actually combine the information from gesture and speech into one unified representation (Alibali et al., 1997; Cassell et al., 1999; McNeill et al., 1994). Evidence supporting such an integrative account of iconic gesture comprehension is for example that upon subsequent request, listeners cannot indicate whether a speaker has communicated a particular piece of information in gesture or speech (Goldin-Meadow & Momeni-Sandhofer, 1999).

Experiment 6 investigates which brain systems are involved in the suggested audiovisual integration processes underlying iconic gesture comprehension. This final experiment expands upon the findings of the first five Experiments of this dissertation. Experiments 1 – 3 provided evidence that listeners use the information provided by iconic gestures to disambiguate speech. That is, it was shown that gesture and speech interact during comprehension. Next, the earliest point in time at which listeners can make a meaning selection on the basis of the gesture was determined in a gating experiment. This disambiguation point can be taken as an estimation of the earliest point in time at which

---

<sup>11</sup> The work described in this chapter is currently in press for publication: Holle, H., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A., & Iacoboni, M. (in press), *NeuroImage*.

gesture and speech interact. The final Experiment of the dissertation explores what brain areas are involved when gesture and speech interact at the disambiguation point.

The question was addressed in an experiment using functional magnetic resonance imaging. This method as well as the theories and studies relevant to the brain bases of iconic gesture comprehension will be summarized in the following section. Subsequently, I will describe the conducted experiment and discuss the findings.

#### **4.1 Theoretical Views on the Neural Correlates of Iconic Gesture Comprehension**

The comprehension of co-speech iconic gestures has not attracted a lot of interest from the neuroimaging community until recently (see also Willems & Hagoort, 2007). This is rather surprising, because, as will be outlined below, a neurocognitive approach to iconic gesture comprehension holds the potential to yield insights both for a better theoretical understanding of gesture comprehension as well as for a better understanding of how the brain processes and integrates naturally occurring complex audiovisual stimuli.

The existing literature, as well as Experiment 1 – 5 of the present dissertation, strongly suggest that iconic gesture *do* communicate information to the listener. It is, however, largely unexplored how these gestures derive their capacity for signification. Methods which allow the mapping of different mental processes to different parts of the brain, such as functional magnetic resonance imaging (for an introduction of this method, see next section), may contribute to elucidate through which brain mechanisms iconic gesture are processed. One theoretical suggestion has been that iconic gestures acquire meaning mainly by providing links to objects (Feyereisen & de Lannoy, 1991), because iconic gestures often illustrate salient features of objects. For instance, in the stimulus set used for the present dissertation, the speaker outlined the shape of a skirt with both hands, when uttering the following sentence: *Er schilderte den Rock<sub>amb</sub>... / He described the skirt<sub>Dom</sub>... / rock music<sub>Sub</sub>....* Feyereisen & deLannoy (1991) have suggested that in such a case the ability to actually “see” a skirt in the speaker's gesture relies on higher-order visual processes similar to those engaged by actual objects or pictures thereof (see also Wu, 2006). If this is indeed the main mechanism through which meaning is inferred, the processing of iconic gestures should yield activation in brain areas associated with object identification. In particular certain areas of the ventral

stream of the visual system are tightly associated with object perception (Grill-Spector, 2003). Some researchers have even argued that certain areas within the ventral stream contain specialized modules for the processing of specific types of visual stimuli, such as faces (Kanwisher, McDermott, & Chun, 1997), or body parts (Downing, Jiang, Shuman, & Kanwisher, 2001). Thus, if iconic gestures acquire meaning mainly by providing links to objects, substantial activation in areas associated with object identification is to be expected.<sup>12</sup>

Alternatively, iconic gesture may derive their capacity mainly by providing links to actions, because many iconic gestures (including most of the stimuli of the present dissertation) are indeed re-enacted actions. If this is the main mechanism through which these hand movements acquire meaning, one should expect that their processing activates the brain network associated with action comprehension. In particular the putative human mirror neuron system, suggested to be located in inferior frontal and parietal cortical areas (Rizzolatti & Craighero, 2004, see also section 4.3), may play a crucial role in the comprehension of action-related iconic gestures.

Neuroimaging of iconic gesture comprehension does not only have the potential to gain further theoretical insight about how gestures are processed in the brain. In addition, comprehending a co-speech iconic gesture requires the listener to integrate auditory and visual information which may take place in specialized brain areas. Converging lines of evidence from single-cell recordings in non-human mammals as well as fMRI studies in human subjects suggest that the brain contains areas specialized for the integration of temporally synchronized information from multiple modalities (for a review, see Calvert & Thesen, 2004). It has been argued that these so-called multisensory integration sites accomplish the fusion of separate unimodal sensations (e.g., a flashing light and a tone) into a single unified percept (Stein & Meredith, 1993). As has been discussed in the introduction, iconic gestures and speech are produced in a temporally synchronized fashion in the form that the stroke of a gesture tends to coincide with the related speech unit. Most researchers agree that during comprehension, the information from iconic gestures is integrated with the information from speech (Alibali et al., 1997; McNeill et al., 1994). Thus, it is an intriguing possibility that the brain contains areas specialized for the integration of gesture and speech.

---

<sup>12</sup> The stimulus used in the present dissertation contained unfortunately too few object-related gestures to allow a test of this hypothesis.

The principles of multisensory integration, as well as the relevant studies on audiovisual integration will be reviewed in section 4.3.

However, before giving a more detailed introduction of the potential brain bases of iconic gesture comprehension, I will first introduce the method most of these studies (as well as Experiment 6) employed, namely functional magnetic resonance imaging.

## **4.2 Method of Investigation: Functional Magnetic Resonance Imaging (fMRI)**

One important aim of functional neuroimaging is the mapping of different mental processes to different parts of the brain. Along the various available neuroimaging techniques, functional magnetic resonance imaging (fMRI) has become the most popular and rapidly evolving method in recent years. This method will be introduced in the following. First, I will attempt to characterize the physical properties of the MR signal. Following this, I will outline how the MR signal can be used as an indirect indicator of brain activity by measuring the level of blood de-oxygenation on a second-by-second basis. Finally, I will describe the processing steps involved in the analysis of fMRI data.

The description of the fMRI method is mainly based on the books by Huettel et al. (2004) and Jezzard et al. (2001), as well as book chapters and review articles on the topic (Nair, 2005; Song, Huettel, & McCarthy, 2006).

### **4.2.1 Nature of the Signal**

In general, the signal of magnetic resonance imaging is generated through a series of three consecutive phases: equilibrium, absorption and reception. (1) During the equilibrium phase, the sample under investigation (e.g., a brain) is placed into a strong static magnetic field which causes the spinning hydrogen nuclei contained in the sample to move in a gyroscopic manner around the axis of the static magnetic field. This phenomenon is called precession. There are only two possible ways the precessing nuclei can align themselves to the static field, a parallel (lower-energy) state and an anti-parallel (higher-energy) state. In equilibrium, more spins assume the lower-energy state. (2) During the absorption phase, an electromagnetic pulse in the radiofrequency range (RF pulse) is applied to the sample. The RF pulse frequency matches exactly the frequency at which the protons precess around the axis of

the static magnetic field. The energy of the RF pulse is absorbed and causes some of the nuclei to become excited, as they switch from the lower-energy into the higher-energy state. (3) After the RF pulse is turned off, the reception phase begins. As the excited spins return to their initial lower-energy state, they release the absorbed energy in the form of a measurable signal that mirrors the RF pulse used for excitation (see Figure 4.1).

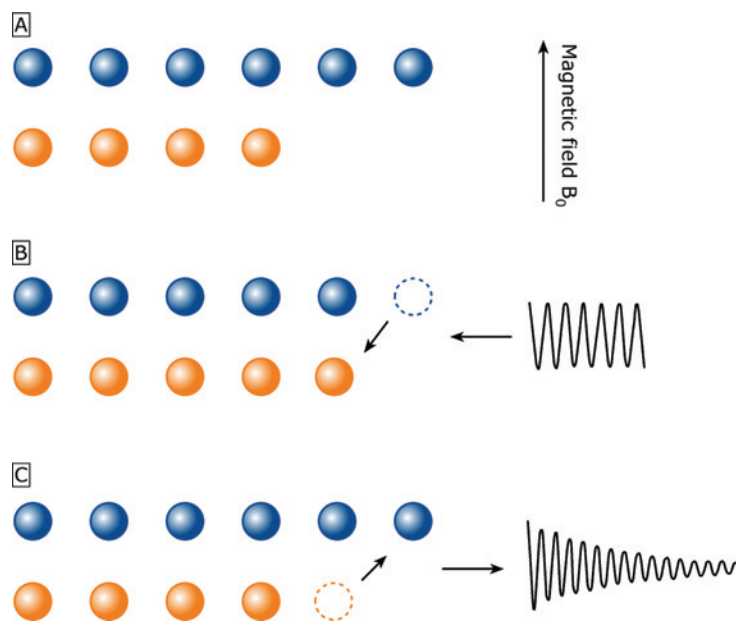


Figure 4.1: The basic principle of MR signal generation: (A) The equilibrium state, where more protons assume the lower-energy parallel state (BLUE). (B) An RF pulse matching the precession frequency is applied to the sample, causing some spins to assume the higher-energy anti-parallel state (RED). (C) As the excited spins return to their initial lower-energy state, they release the absorbed energy, which creates a measurable MR signal. Adapted from Huettel et al. (2004)

Three essential parts of an MR scanner are needed to record a signal that is useful for imaging: The static magnetic field, the RF coils and the gradient coils. The static magnetic field is needed to establish the equilibrium state, during which more spins assume the lower-energy parallel state. Another essential scanner part, the RF coils, are used to transmit and receive short-lasting electromagnetic RF impulses. The transmitter RF coils are used to excite the spins during the absorption phase. Unlike the static magnetic field, which is always present, the RF coils are only turned on for a brief period during image acquisition. After the RF pulse is turned off, the receiver coils are used to measure the energy emitted by the excited

spins as they switch back from the higher-energy to the lower-energy state. Because the ultimate goal is to acquire images of the matter under study, the gradient coils are used to vary the precession frequency in a topographically predictable manner. Gradient fields are magnetic fields that linearly vary in their strength in one of the three dimensions of the scanning space (i.e., the x-, y- and z-dimension). They are superimposed on the static magnetic field and cause a linear modulation of the frequency at which the spins precess around the longitudinal axis. The clever combination of gradient fields and RF pulses allows to excite only a selected portion of spins (e.g., a slice) which makes the MR signal spatially distinguishable.

#### Nuclear Spin as the Basis of the MR Signal

All matter consists of atoms. Atoms, in turn, contain protons, electrons and (sometimes) neutrons. Protons (and if present neutrons) form the nucleus of the atom. Atoms of different matter have different nuclear composition. The nucleus of hydrogen ( $^1\text{H}$ ), for example, consists of a single proton. Thermal causes the hydrogen proton to spin around itself. Because the proton has a positive electrical charge, the spinning movement has two consequences: First, the spin generates an electric current. When put into a strong magnetic field, this current induces a torque (i.e., a turning movement) which is called magnetic momentum or  $\mu$ .<sup>13</sup> Second, because hydrogen has an odd-numbered atomic mass (of 1), the spin causes angular momentum (called J), which is defined as the product of the mass of a spinning object multiplied by its angular velocity. Both  $\mu$  and J can be expressed as vectors that point in the same direction.<sup>14</sup>

When placed into a strong magnetic field as present in an MR scanner, the spinning nuclei align themselves either parallel or anti-parallel to the magnetic field, with more nuclei assuming the lower-energy parallel state. Because of the angular momentum, the axis of the

---

<sup>13</sup> The magnetic momentum is defined as the amount of torque exerted by the spinning nucleus on a magnet, moving electrical charge or current-carrying coil (cf. Huettel et al., 2004).

<sup>14</sup> Only nuclei that have both magnetic momentum as well as angular momentum can be studied by magnetic resonance imaging (e.g.  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ ,  $^{23}\text{Na}$ ,  $^{31}\text{P}$ ). Because hydrogen nuclei ( $^1\text{H}$ ) naturally occur in abundance in the human body and produce a strong MR signal, they are the most commonly used nucleus type in MR imaging.

spinning nuclei is not perfectly aligned to the main field. Instead, the spinning protons describe a gyroscopic motion around the axis of the static magnetic field, a phenomenon called precession.

The magnetic moment is larger than the angular momentum vector by a factor  $\gamma$ , which is known as the gyromagnetic ratio. As has been mentioned previously, spins can take one of two possible states, a parallel state (lower energy) and a state anti-parallel to the magnetic field (higher-energy, less stable). When changing from parallel to anti-parallel, a spin must absorb electromagnetic energy. Conversely, when changing from anti-parallel to parallel, a spin emits energy. It is crucial to note that the frequency of the absorbed and emitted energy depends only on the gyromagnetic ratio and the strength of the static magnetic field. Therefore, for all nuclei that possess both magnetic as well as angular momentum, the frequency at which the nuclei precess around the longitudinal axis can be calculated (the so-called Larmor frequency). For example, the Larmor frequency for hydrogen nuclei on a 3 Tesla Scanner is about 126 MHz. As a comparison, regular FM radio is transmitted in a frequency range of 87.5 to 108.0 MHz.

#### From Isolated Spins to a Spin System

In MRI, one does not measure the magnetic properties of isolated spins, but the net magnetization of many simultaneously resonating nuclei, which can also be referred to as a spin system. The net magnetization of a spin system can be indicated by a vector with a certain length (indicating the strength) and direction. In the absence of a strong magnetic field, the axes around which the single hydrogen nuclei spin are oriented in a random manner, therefore the net magnetization vector is zero. However, in a strong magnetic field, all spins assume either a state that is parallel or anti-parallel to the magnetic field. Because in equilibrium more spins are in the lower energy state, the net magnetization vector points in the direction of the main field. This is also referred to as the net longitudinal magnetization.

The length of the excitation RF pulse can be calculated to be of such a length that the net magnetization vector is flipped into the transverse plane, i.e., the plane perpendicular to the main field (spanned by the xy axes of the scanning space). For example, an RF pulse of  $90^\circ$  means that the net longitudinal magnetization (along the z-axis) was zero at the offset of the pulse and the net magnetization vector pointed  $90^\circ$  into the transverse plane (see Figure 4.2). This tipping of net magnetization into the transverse plane creates the measurable MR signal,

because the receiver coils are sensitive to the amount of net magnetization in the x- and y-dimension of the transverse plane. Since the individual spins continue precessing as the longitudinal magnetization recovers and the transverse magnetization decays, the net magnetization vector describes a wobbling motion in the coordinate system known as nutation (see Figure 4.2). The wobbling motion is also the reason why the MR signal picked up by the receiver coils shows an oscillating pattern that decays over time (see Figure 4.1, lower panel).

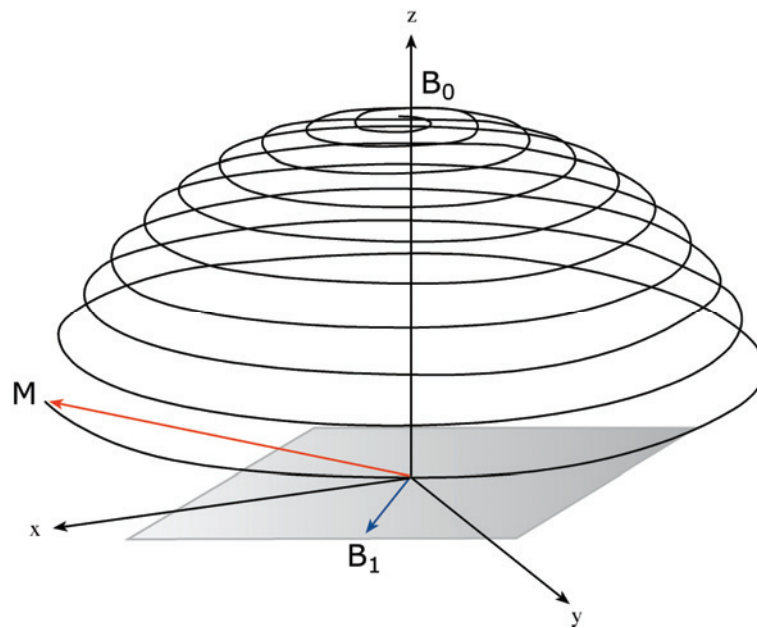


Figure 4.2: Spin nutation. After it has been tipped into the transverse plane (xy), the net magnetization vector (M) describes a “wobbling motion” as the longitudinal net magnetization recovers. Adapted from Huettel et al. (2004)

#### The Time Constants T1, T2, and T2\*

As can be seen from Figure 4.2, following an excitation pulse, two types of changes occur over time: the longitudinal magnetization recovers and the transverse magnetization decays. At the offset of an excitation pulse with a  $90^\circ$  flip angle, the net longitudinal magnetization is zero. As the excited spins return from the higher-energy anti-parallel to their initial lower-energy parallel state, the net longitudinal magnetization continues to increase until it has reached the equilibrium state. The longitudinal magnetization recovery follows a logarithmic

function and the time constant governing longitudinal recovery is known as  $T_1$ . After one  $T_1$  period, 66% of the longitudinal magnetization have recovered, whereas 95% have been regained after three  $T_1$  periods. Because the  $T_1$  constant is different for different types of brain tissue, it provides a source of image contrast. For example, cerebrospinal fluid (CSF) contains more hydrogen nuclei than gray matter; therefore the  $T_1$  of CSF is larger than the  $T_1$  of gray matter. Thus, in pulse sequences sensitive to  $T_1$  contrast (which generate so-called  $T_1$ -weighted images), brain areas containing CSF appear darker (i.e., they emit less MR signal) than areas containing primarily gray matter.

Theoretically, if one could observe the MR signal from a single isolated resonating nucleus, the decay of transverse magnetization would be *directly* related to the recovery of the longitudinal magnetization in an inverse fashion<sup>15</sup> (cf. Matthews, 2001). However, in MRI one is observing emissions from a huge number of protons simultaneously (the brain contains more than  $4 \times 10^{19}$  hydrogen protons/mm<sup>3</sup>). As will be explained below in more detail, both the exchange of energy between adjacent nuclei as well as local magnetic field inhomogeneities cause that the decay of the transverse magnetization occurs much faster than the recovery of the longitudinal magnetization.

At the offset of the excitation pulse, all spins are in phase in their precession movement around the z axis, i.e., they begin precession within the transverse plane at the same starting point. This maximal phase coherence is reflected in a strong transverse magnetization. Subsequently, phase coherence (and hence the transverse magnetization) decays. The decay is due to two factors, one intrinsic, the other extrinsic. The intrinsic factor is spin-spin interaction. Because all resonating spins are simultaneously emitting electromagnetic energy, they influence one another causing some spins to precess faster and others slower. Gradually, the spin phases fan out over time until they are completely dephased. The signal loss caused by this intrinsic mechanism is called  $T_2$  decay, and it is characterized by the time constant  $T_2$ . The complete decay of transverse magnetization due to spin-spin interactions ( $T_2$  decay) occurs always much faster (in the order tens to hundreds of ms) than the full recovery of the longitudinal relaxation, which is in the order of a few seconds (at field strengths typical for fMRI).

---

<sup>15</sup> Note that Figure 4.2 also (inadequately) suggests such a direct relationship between longitudinal recovery and transverse decay of magnetization.

An additional, extrinsic source of loss in transversal net magnetization is the external magnetic field. The external field is not perfectly homogeneous. Because each spin precesses at a frequency proportional to its local field strength, variations in field from location to location cause spins at different spatial locations to precess at different frequencies, also leading to a loss of coherence. The *combined* effects of spin-spin interaction and field inhomogeneity lead to signal loss known as  $T_2^*$  decay, characterized by the time constant  $T_2^*$ . Because  $T_2^*$  decay includes the additional factor of field inhomogeneity, for any given substance, the  $T_2^*$  constant is always smaller than the  $T_2$  time constant (cf. Song et al., 2006).

Related to these relaxation time parameters are the repetition time (TR) and the echo time (TE). The TR indicates the time period between two excitation pulses, whereas TE determines the delay from the excitation pulse to the onset of data acquisition. Depending on what combination of TE- and TR-values is used in a particular pulse sequence, influences, among other factors, to what kind of brain of tissue the sequence is sensitive. This is one of the features that makes MRI such a versatile imaging instrument, because pulse sequences can tailored to specific types of brain tissue (e.g., gray matter, white matter).

#### 4.2.2 The BOLD Effect

Like  $T_1$  and  $T_2$ ,  $T_2^*$  can also provide a source for image contrast. In particular  $T_2^*$ -weighted contrasts are important for fMRI. Ogawa et al. (1990) noticed that the  $T_2^*$  contrast is sensitive to the degree of blood deoxygenation. In this experiment,  $T_2^*$ -weighted images of rodent brains were acquired as they breathed either 100 % pure oxygen or normal air. In contrast to the pure-oxygen condition, breathing normal air made the blood vessels of the brain clearly visible. This effect, which came later to be known as the BOLD effect (Blood-Oxygen-Level-Dependent), occurs because deoxygenated hemoglobin, as opposed to oxygenated hemoglobin, possesses magnetic momentum. Because deoxygenated blood is magnetically susceptible, it creates local inhomogeneities in the magnetic field. As a result, brain areas with greater amounts of deoxygenated blood show a greater signal loss (i.e., they appear darker) in a  $T_2^*$ -weighted image than areas with a lesser amount of deoxygenated blood. This occurs because the spins in brain areas with deoxygenated blood dephase faster than the spins in brain areas with oxygenated blood.

Ogawa and other researchers quickly realized that the BOLD signal could potentially be used as indirect, non-invasive measure of human brain activity. However, this would require establishing the exact coupling between the BOLD signal and neuronal activity.

#### The BOLD Response as an Indirect Measure of Neuronal Activity

Like all body cells, neurons in the brain require energy in order to function, which is provided by adenosinetriphosphate (ATP). The most efficient way for a neuron to generate ATP is through a process known as aerob hydrolysis. In order to produce ATP through an aerob hydrolysis, the neuron needs to be provided with glucose and oxygen. These ingredients are transported to the capillary blood vessels in the vicinity of the neuron through the blood. Oxygen, which is bound to the hemoglobin of the red blood cells, is released from its bond and diffuses into the cell's soma where it is metabolized. In return, carbon dioxide, one of the waste products of aerob hydrolysis, diffuses from the soma to the capillaries, where it also binds to the hemoglobin. The now deoxygenated blood<sup>16</sup> is transported away from the capillaries, through a system of vessels that increase in diameter as they approach the right heart chamber (venules, veins, vena cava). Subsequently, the oxygen-poor blood is pumped to the lungs, where it is oxygenated again.

Given the fact that cognitive processes require signaling and integrating activity in ensembles of neurons, which in turn requires a supply of oxygen and the disposal of carbon dioxide, one would intuitively expect that the BOLD signal is negatively correlated with neural activity. However, by now a large number of studies have firmly established the fact that the BOLD signal is *positively* correlated with neuronal activity (for a review, see Nair, 2005). In other words, neural activity decreases the amount of deoxygenated blood in the surrounding brain tissue.

The main cause for this paradoxical relationship between neuronal activity and the BOLD signal seems to be that neuronal activity in a particular brain region is followed by a large increase in blood flow to that region that overcompensates the neurons' need for oxygen. For example, Fox et al. (1988) found that functional activation increases cerebral blood flow

---

<sup>16</sup> Not all oxygen is removed from the blood in the capillaries. It is estimated that the deoxygenated blood in the capillary beds, the venules and the draining veins still has an oxygen saturation of 60 – 70% at rest, as opposed to 100 % saturation in oxygenated blood (Nair, 2005).

(CBF) by about 50 %, whereas the cerebral metabolization rate of oxygen ( $CMR_{O_2}$ ) was only increased by about 5 %. The general mechanism through which neuronal activity results in this overcompensating increase in CBF seems to be dilatation of the arterioles, mediated by a complex signaling mechanism including neurons, astrocytes and vascular cells (for a review, see Iadecola, 2004).

Among the several models that have been put forward to explain the exact relationship between CBF,  $CMR_{O_2}$ , and the BOLD signal (Magistretti & Pellerin, 1999; Mandeville et al., 1999), the balloon model by Buxton and co-workers (Buxton, Uludag, Dubowitz, & Liu, 2004; Buxton, Wong, & Frank, 1998) has gained considerable significance in the research community in recent years. Using a minimal set of assumptions, it can predict all of the observed properties of the BOLD response to cognitive stimuli, i.e., the sometimes reported initial dip of the signal, the subsequent rise to a plateau level (overshoot) as well as the following post-stimulus undershoot, where the signal drops below the initial baseline level. The balloon model assumes that the venous compartments of the capillaries are expandable (a “balloon”). According to the model, the initial dip is due to an initial increase in oxygen extraction before the flow increase occurs. Subsequently, the overcompensatory increase in blood flow into the capillary bed takes place. As a first consequence, the deoxygenated blood is flushed away from the capillaries. Second, the inflow surpasses the clearance capacity of the venous system, causing an inflation of the “balloons” in the venous compartments. Both consequences contribute to the strong MR signal observed during the overshoot. Finally, the model predicts that a post-stimulus undershoot occurs when the blood volume returns to baseline more slowly than the blood inflow.

#### EPI-BOLD Sequences

The BOLD response to an isolated event is delayed by one or two seconds and reaches its maximum around 6 seconds. Thus, in order to reliably estimate the BOLD response for a whole brain volume, images covering the complete volume have to be acquired at least every 3 seconds (i.e., at the Nyquist frequency). Traditional pulse sequences did not allow such a rapid image acquisition, because they required a separate HF pulse for each column of a slice. It was not until the advent of EPI-BOLD (Mansfield & Maudsley, 1977) in combination with improved gradient hardware that such a rapid acquisition of  $T_2^*$ -weighted images became possible. An EPI-BOLD sequence allows to acquire data from a complete slice using a single

excitation pulse. Figure 4.3 illustrates how EPI-BOLD works. At first a gradient along the z axis is turned on to ensure that only the spins within a selected slice precess at the frequency of the RF pulse (slice selection). Next, the data of the slice are acquired in a zig-zagging trajectory, which is achieved by rapid switching of the x- and y gradient coils. This procedure is repeated for each slice, until a complete brain volume has been acquired.

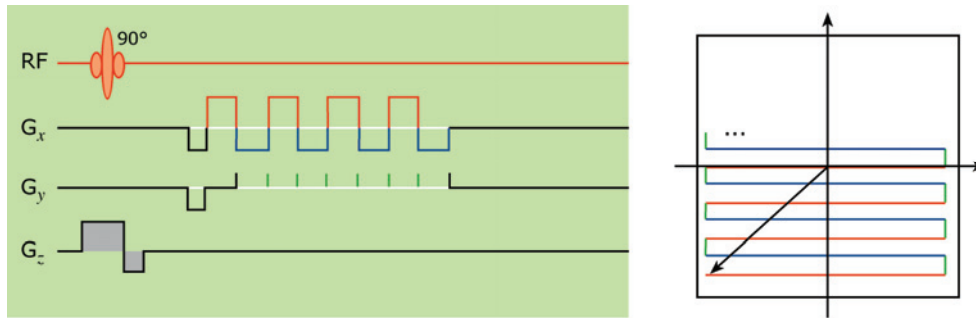


Figure 4.3: EPI-BOLD sequence. Slightly adapted from Huettel et al. (2004)

In this way, EPI-BOLD sequences allow the acquisition of the BOLD signal for a complete brain volume on a second-by-second basis. The spatial resolution, as defined by the voxel size, is usually in the range of several millimeters (e.g., 3 x 3 x 3 mm). In the next section, I will briefly describe the steps involved in the preprocessing and statistical analysis of fMRI data.

#### 4.2.3 Analysis of fMRI Data

Before fMRI data are subjected to a statistical analysis, a number of preprocessing steps are usually applied. First, the data are corrected for head movements that occurred during acquisition. Next, a slice time correction can be applied. Because the data are acquired one slice at a time, there is a considerable degree of temporal variation in the acquisition time of slices belonging to the same timestep. The correction algorithm interpolates the data of each timestep so that the data appear as if the individual slices were acquired simultaneously. To remove slow drifts from the data, a high-pass filter can be applied (baseline correction). Additionally, the data can be subjected to a certain degree of spatial smoothing during this preprocessing stage. Spatial smoothing, which is usually achieved by applying a Gaussian filter to the data, effectively spreads the intensity at each voxel in the image over nearby

voxels. Filter width is usually expressed in FWHM (full width at half maximum) and typical filter widths for fMRI range from 6 to 10 mm (Huettel et al., 2004).

At some point during the analysis (either before or after the data are subjected to statistical analysis) the data of each participant are co-registered to a high-resolution anatomical scan of that participant. Additionally, to allow an analysis of data over participants, the data are transformed into a standardized space, for example into the Talairach space (Talairach & Tournoux, 1988).

In contrast to EEG data, which are typically subjected to an event-related analysis, fMRI are most often analyzed using a deconvolution approach. Such a deconvolution analysis is typically mass-univariate, i.e., each voxel (containing a time-series of MR signal values) is treated as a separate dependent variable. The statistical analysis mirrors a multiple regression with the convolved product of stimulus onset and hemodynamic response function as predictor and the MR signal value as the to-be-predicted variable. It is performed in the context of the General Linear Model (GLM) in three steps. (1) design specification (2) GLM parameter estimation (3) the interrogation of the results using contrast vectors to produce statistical parametric maps (SPMs).

During design specification, the responses are modeled by convolving a series of delta or box functions, indicating the onset of an event or epoch, with a set of basis functions (Penny, 2006). The present study employed a synthetic hemodynamic response function as a basis function. In the resulting design matrix, each condition is represented by (at least) one column. After the design matrix has been generated, it is applied to the data. For every voxel, the observed BOLD signal is predicted on the basis of the design matrix by estimating the beta-values of the columns using a least-squares-error criterion. Having completed the GLM parameter estimation, one can construct contrast vectors to analyze main effects or interactions by contrasting the beta-values of the corresponding conditions. The resulting SPMs indicate what brain voxels show a significant activation difference in the respective contrast.

Because each voxel is analyzed by a separate statistical test and a typical fMRI study gathers data from tens of thousands of voxels, false positive activations are a continuing problem in neuroimaging research. Simply reducing the  $\alpha$  error to account for the multiple comparison problem is not a solution (e.g., via a Bonferroni-correction), because the loss of statistical

power is so dramatic that no effects at all can be detected. Several suggestions to alleviate this dilemma have been made, including the combined use of cluster size and significance level (double thresholding, Forman et al., 1995), the false discovery rate, which restricts false positive control to suprathreshold voxels only (Benjamini & Hochberg, 1995), or small volume random field corrections on the basis of a-priori hypotheses (Worsley et al., 1996). A particularly helpful tool for applying the double threshold criterion is the AlphaSim program (Ward, 2000), which is part of the AFNI toolbox. For a given experimental setup, the program determines for a given threshold criterion (e.g.,  $p < .001$ ), what cluster size is needed to reduce the probability of false positives to  $p < .05$ . This is achieved through a series of Monte-Carlo simulations. The advantage of the program is that it takes into account several factors known to influence the amount of false positives simultaneously, including the number and dimension of the voxels and the amount of spatial smoothness in the data.

Having introduced the physical properties underlying the MR signal as well the indirect relationship between the BOLD signal and brain activity, I will now give an overview of neuroimaging studies relevant to the question of what brain areas are involved in the comprehension of co-speech iconic gestures.

### **4.3 FMRI Studies Relevant to Iconic Gesture Comprehension**

In the introduction to this chapter, it was outlined how one can adopt different neurocognitive perspectives on the processing of iconic gestures. Among these different views, the action perspective is the degree to which the processing of iconic gestures recruits the brain network associated with action comprehension. Given that many iconic gestures constitute re-enacted actions, this seems intuitively plausible. Based on the findings that area F5 and PF of the macaque brain contain neurons that fire both during the observation as well as the execution of goal-directed hand movements (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996, 2002; Umiltà et al., 2001), it has been proposed that these so-called mirror neurons form the neural circuitry for action understanding (Rizzolatti, Fogassi, & Gallese, 2001). Although direct evidence (via single-cell recording) for mirror neurons in the human brain is still lacking, there is a substantial body of indirect evidence that a similar system exists in humans as well (for recent overviews, see Binkofski & Buccino, 2006; Iacoboni & Dapretto, 2006; Molnar-Szakacs, Kaplan, Greenfield, & Iacoboni, 2006). In particular, the inferior frontal gyrus (IFG) including the adjacent ventral premotor cortex and the inferior parietal lobule (IPL) have been suggested as the core components of the putative human mirror neuron system (MNS)

(Rizzolatti & Craighero, 2004). According to a recent theoretical suggestion, the human MNS is able to determine the goal of observed actions by means of an observation-execution matching process (for a more detailed description, see Iacoboni, 2005; Iacoboni & Wilson, 2006). Because many iconic gestures are re-enacted actions, it is therefore a theoretically plausible possibility that the MNS also participates in the processing of such gestures.

Second, one can adopt a multimodal perspective on iconic gesture comprehension. As has been argued in Chapter 1, iconic gestures show little conventionalization, i.e., there is no “gestionary” that can be accessed for their meaning. Instead, the meaning of iconic gestures has to be generated online on the basis of gesture form and the co-speech context in which the gesture is observed (Feyereisen et al., 1988; McNeill, 1992, 2005). Thus, comprehending a co-speech iconic gesture is a process which requires a listener to integrate auditory and visual information. Within the multimodal view on iconic gestures, a further distinction can be made between local and global gesture-speech integration (see also Willems, Özyürek, & Hagoort, 2006). Because co-speech gestures are embedded in spoken utterances that unfold over time, one can investigate the integration processes between gesture and speech both at a local level (i.e., the integration of temporally synchronized gesture and speech units) as well as on a global sentence level (i.e., how greater meaning ensembles are assembled from smaller sequentially processed meaningful units such as words and gestures).

Local integration refers to the combination of simultaneously perceived gestural and spoken information. Previous research indicates that the temporal relationship between gesture and speech in production is not arbitrary (McNeill, 1992; Morrel-Samuels & Krauss, 1992). Instead, speakers tend to produce the peak effort of a gesture, the so-called stroke, simultaneously with the relevant speech segment (Levelt et al., 1985; McNeill, 1992). This stroke-speech synchrony might be an important cue for listeners in comprehension, because it can signal to which speech unit a gesture belongs. For instance, a speaker might produce a turning hand movement while saying “*He tightened the screw*”. The gesture stroke is considered to be produced simultaneously with the verb of the sentence. In this example, local integration would refer to the interaction between the simultaneously conveyed visual information (i.e., the turning-movement gesture) and auditory information (the word *tightened*).

Although related to such local processes, the global integration of gesture and speech is a more complex phenomenon. Understanding a sentence accompanied by a gesture does not only require determining the meaning of all the individual constituents (i.e., words, gestures). In addition, the listener has to determine how the constituents are related to each other, in order to figure out who is doing what to whom (cf. Grodzinsky & Friederici, 2006). This relational process requires integrating information over time. The multimodal aspect in this integration over time is the extent which the process recruits similar or different brain areas depending on whether the to-be-integrated information is a spoken word or a gesture. Thus, local integration refers to an instantaneous integration across modalities, whereas global integration describes an integration over time, with modality as a moderating variable. Whereas interactions at the global level can be examined in an epoch-related analysis, an analysis of gesture-speech interactions at the local level can only be performed in an event-related design. More precisely, in order to investigate how gesture and speech interact at the local level, one first has to objectively identify the point in time at which gesture and speech start to interact. As will be outlined below, the gating paradigm may be used to determine such a time point.

Willems et al. (2006) investigated the neural correlates of gesture-speech interaction on a global sentence level. In this experiment, subjects watched videos in which an initial sentence part (e.g., *The items that he on the shopping list*<sup>17</sup>) was followed by one of four possible continuations: (1) a correct condition, where both gesture and speech matched the initial sentence context (e.g., saying *wrote* while producing a writing gesture), (2) a gesture mismatch (e.g., saying *wrote* while producing a hitting gesture), (3) a speech mismatch (e.g., saying *hit*, gesturing writing) and (4) a double mismatch (saying *hit*, gesturing hitting). In the statistical analysis, the complete length of the videos was modelled as an epoch. When contrasted with the correct condition, only the mid- to anterior portion of the left IFG (BA 45/47) was consistently activated in all three mismatch conditions. On the basis of this finding, Willems and co-workers suggested that the integration of semantic information into a previous sentence context (regardless whether the to-be-integrated information had been conveyed by gesture or speech) is supported by the left IFG.

---

<sup>17</sup> The example is a literal translation of the original Dutch stimuli.

Whereas the Willems study investigated the interaction of gesture and speech at a global level, it is an open issue what brain areas are involved local gesture-speech interactions. The most detailed studies of audio-visual integration processes have focused on the mammalian superior colliculus (Stein & Meredith, 1993). On the basis of single-cell recordings in response to auditory (A), visual (V) and combined cues (AV), Stein and colleagues (1993) have suggested certain properties and rules by which multisensory integration is achieved at the neuronal level. For example, the superior colliculus contains neurons that respond to both auditory and visual cues. Moreover, the response to auditory and visual cues that occur in close temporal and spatial proximity (AV) is substantially enhanced, sometimes exceeding the summed firing rate of the unimodal responses ( $A + V$ ) by a factor of 12 and above. Because the output no longer resembles a linear combination of the input, it has been suggested that the information obtained from the auditory and visual modality has been combined (or “fused”) into a single new output signal. This process has been termed multisensory integration (Stein, Meredith, & Wallace, 1993). This superadditive response pattern is most pronounced for those stimuli that produce the weakest response in unimodal presentation. For example, Stein & Meredith (1993) observed that as the amplitude of an auditory cue decreased, so did the response of most multisensory neurons to this unimodal stimulus. However, if a near-threshold auditory stimulus was accompanied by a spatially and temporally congruent visual stimulus, the relative response enhancement (i.e., the percentage that the response to AV is greater than response to  $A + V$ ) was greatest. This principle was termed inverse effectiveness and has been suggested as the mechanism underlying the perceptual enhancement of degraded auditory cues by simultaneously presented visual cues (Callan et al., 2003; Sekiyama, Kanno, Miura, & Sugita, 2003). Finally, Stein & Meredith (1993) reported that multisensory neurons show a response depression to crossmodal stimuli that are temporally or spatially disparate. This means that the response to a unimodal stimulus can be strongly attenuated by the presence of a spatially or temporally incongruent stimulus in another modality.

On the basis of these data (obtained mainly via single-cell recordings in the cat), researchers have derived different criteria for MSI areas in the human cortex, as measured by large-scale measures of neuronal activation such as fMRI (for a critical discussion of the validity of these criteria, see Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004b; Calvert & Thesen, 2004; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005).

1. *Superadditivity*: A MSI area shows a superadditive BOLD increase to bimodal presentations ( $AV > A + V$ )
2. *Inverse effectiveness*: Superadditivity is most pronounced for near-threshold stimuli. For example, a MSI area shows a greater BOLD response to degraded bimodal stimuli (e.g., AV with auditory noise) than undegraded bimodal stimuli (e.g., AV without noise)
3. *Response depression*: A MSI area shows greater levels of activation for congruent bimodal than for incongruent bimodal stimuli (e.g., speech accompanied by congruent lip movements vs. speech accompanied by incongruent lip movements).

Using these criteria, a number of brain areas have been suggested as multisensory integration sites in the human brain, including the intraparietal sulcus (Calvert, Campbell, & Brammer, 2000; Lewis, Beauchamp, & DeYoe, 2000), superior temporal sulcus (STS) (Beauchamp, 2005) and anterior cingulate cortex (Banati, Goerres, Tjoa, Aggleton, & Grasby, 2000; Calvert, Hansen, Iversen, & Brammer, 2001). Among these cortical regions, the STS seems to be particularly involved in audiovisual integration processes (for a review, see Beauchamp, 2005). For example, the STS seems to be involved in the integration of lip movements and speech sounds (Calvert et al., 2000; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). Furthermore, Skipper and colleagues (2005) observed that the activation in the posterior STS elicited by the observation of talking faces is modulated by the amount of visually distinguishable phonemes. In an experiment by Sekiyama et al. (2003) it was found that the left posterior STS is particularly involved in the McGurk effect, e.g., the fusion of an auditory /ba/ and a visual /ga/ into a perceived /da/. While in these examples, visual and auditory information can be mapped onto each other on the basis of their form, there is evidence that the STS is also involved in more complex mapping processes at a higher semantic-conceptual level. For instance, Beauchamp et al. (2004) found that the STS is associated with the integration of pictures of animals and their corresponding sounds. Saygin and co-workers (2003) have reported that patients with lesions in the posterior STS are impaired in their ability to associate a picture (e.g., a cow) with a corresponding sound (e.g., *moo*).

On the basis of these results, it is not unreasonable to assume that the STS is also involved in the multimodal interactions between gesture and speech. The integration of iconic gestures and speech during comprehension has some similarities with the integration of pictures and

their associated sounds, as it was for instance investigated by Beauchamp et al. (2004). In both cases, there is a temporal synchrony between auditory and visual information. In the audiovisual condition of the Beauchamp study, the pictures and the corresponding sounds were presented simultaneously. Likewise, as it has been introduced above, the stroke of a gesture tends to coincide with the relevant speech unit. Another similarity is that in both cases, the unimodal information first has to be processed and semantically interpreted to some extent individually, before an interaction at the semantic-conceptual level between auditory and visual information can occur. A sound (e.g., of a telephone ringing) first has to be interpreted to some extent before it can become associated with a picture (e.g., of a telephone). Likewise, a gesture also first has to be interpreted on the basis of its form, before it can become associated with a spoken word. However, the two audiovisual interaction types differ in complexity. In the Beauchamp study, the semantic relationship between auditory and visual information was fixed. The visual object was always presented with the sound that such an object typically creates. In contrast, the semantic relationship between iconic gestures and speech is not fixed. A sentence such as *During the game, he returned the ball* can be accompanied by a gesture that depicts the form of the ball, or a gesture that focuses on the returning motion. Moreover, the gesture might primarily depict the trajectory of the ball's movement, the manner (rolling, sliding, ...) or a combination of trajectory and manner. Finally, the gesture can depict the scene from a character viewpoint (i.e., the person returning the ball) or from an observer viewpoint. How the gesture is related to speech is not defined a-priori, but has to be detected by the listener on an ad-hoc basis. Thus, the comprehension of iconic gestures requires complex semantic interactions between gestural and auditory information. So far there are no studies that have investigated whether the STS also houses these complex multimodal processes underlying co-speech iconic gesture comprehension.

#### **4.4 Experiment 6**

Experiment 6 aimed to locate brain areas involved in the processing of co-speech iconic gestures. As has been described above, one intriguing view on iconic gestures is the multimodal perspective. Investigating the putative multimodal integration sites for gesture and speech would entail an experimental design with a gesture-only, speech-only as well as a bimodal gesture+speech condition, in order to demonstrate superadditivity as it was suggested by Calvert and colleagues (2004). However, the problem with such a manipulation is that it neglects the one-sided dependency between the two information channels. Whereas

understanding speech does not depend on gesture, iconic gestures are dependent upon the accompanying speech in that these gestures are only distinctly meaningful in their co-speech context. Most researchers agree upon that decontextualized iconic gestures convey only imprecise meaning to the listener (e.g., Cassell et al., 1999; Krauss et al., 1991). Thus, when presenting a gesture-only condition to participants, one runs a great risk of inducing artefactual processing strategies. As McNeill has stated: “It is profoundly an error to think of gesture as a code or 'body language', separate from spoken language. [...] It makes no more sense to treat gestures in isolation from speech than to read a book by looking at the 'g's.” (McNeill, 2005, p. 4). Another independent group of researchers around Robert Krauss have also come to the conclusion that decontextualized iconic gestures convey little meaning to the listener and that the relationship between auditory, visual and audiovisual information is not well captured by a linear model (Krauss et al., 1991, Experiment 3).

Rather than adopting a strict multisensory perspective, Experiment 6 approaches the comprehension of co-speech iconic gestures by means of a disambiguation paradigm, where lexically ambiguous sentences (e.g., *Sie berührte die Maus*, *She touched the mouse*) are accompanied either by disambiguating iconic gestures or meaningless grooming movements. The rationale is that brain regions involved in the interaction of gesture and speech should respond stronger to a gesture-supported sentence (where auditory and visual information are semantically related) than to a sentence accompanied by a meaningless grooming movement (where auditory and visual information are unrelated). Such a disambiguation paradigm has several advantages. First, it has some external validity. Holler and Beattie (2003) have shown that speakers spontaneously produce a substantial amount of iconic gestures when asked to explain the different word meanings of a homonym. Second, in a disambiguation paradigm, the iconic gestures are not removed from their co-speech context, which excludes the possibility of a gesture-only condition inducing artefactual processing strategies. Third, the influence of the speech channel, which is certainly the channel with the highest information content, is perfectly controlled for, because the sentences are physically identical in the critical experimental conditions.

Thus, all of the observed differences in a disambiguation paradigm can only be due the accompanying hand movement (i.e., the main effect) or the interaction between the hand movement and the spoken sentence. The challenge in interpreting the results is to determine which one – main effect or interaction – actually caused an observed activation difference. One can think of Experiment 6 as an exploratory study in the evolving field of co-speech

gesture comprehension. It identifies regions possibly involved in the interaction between iconic gestures and speech in a paradigm with a high external validity.

In Experiment 6, only the gestures but not the meaningless grooming movements allow a disambiguation of the homonym, i.e., only in the case of gesture there is an interaction between the visually and the auditorily conveyed information. On the basis of the literature, it is hypothesized that the processing of co-speech gestures will elicit greater levels of activation in the STS than the processing of the meaningless co-speech grooming movements.

To elucidate the role of the left IFG (i.e., BA 44, 45 & 47) in local gesture-speech interactions, an additional manipulation of word meaning frequency is included. All sentences could either be interpreted in terms of a more frequent dominant meaning (e.g., the *animal* meaning of *mouse*) or the lesser frequent subordinate meaning (e.g., the *computer device* meaning of *mouse*). Because previous studies have shown that the processing of lexically low frequent words recruits the left IFG to a stronger degree than high frequent words (Fiebach, Friederici, Muller, & von Cramon, 2002; Fiez, Balota, Raichle, & Petersen, 1999; Joubert et al., 2004), it is hypothesized that the processing of subordinate gestures will elicit greater levels of activation in the left IFG than dominant gestures. Alternatively, if the left IFG (and in particular the anterior inferior portion) is not only the site of multimodal gesture-speech interactions at the global level, as suggested by Willems et al. (2006), but also at the local level, greater levels of activation for gestures as compared to grooming should be observed in this region.

#### 4.4.1 Methods

##### Participants

Seventeen native speakers of German (10 females), age 21-30 (mean age 25.7,  $SD = 2.8$ ) participated in this experiment after giving informed written consent following the guidelines of the Ethics committee of the University of Leipzig. All participants were right-handed (mean laterality coefficient 92.7,  $SD = 11.3$ , Oldfield, 1971) and had normal or corrected-to-normal vision. None reported any known hearing deficits.

### Materials




#### *Audio Recording*

Experiment 6 used the same pool of gestures as they were recorded for Experiment 1 (see p. 26). However, Experiment 6 had a different focus than Experiments 1 – 3. In Experiments 1 – 3, the impact of gesture was measured indirectly, by analyzing the ERPs time-locked to subsequent target words that were either congruent or incongruent to the meaning of the preceding gesture. In contrast, Experiment 6 aimed to measure the disambiguating effect of gesture directly, by modelling the disambiguating point of the gestures (as determined in Experiment 4) as an event (for more details, see below). Because of this change in focus, the sentences were slightly shortened by removing the final disambiguating part of the second sentence that followed the homonym. For instance, instead of the two sentences with different endings used previously for the ambiguous word mouse (..which the cat... vs. which the computer ..., there was now only one completely ambiguous version that was ended by a homonym (She touched the mouse; see also Table 4.1) without a disambiguating sentence continuation.

The speech of these modified shortened sentences was re-recorded in a separate session. The re-recorded sentences were then combined with the original gesture and grooming movements as they were recorded for Experiment 1, yielding a total of three video clips for each ambiguous sentence (dominant gesture, subordinate gesture, grooming). In order to maintain a comparable audiovisual synchrony between the three experimental conditions, the re-recorded audio was synchronized with the video stream according to the phonological synchrony rule, which states that “the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech” (McNeill, 1992, p. 26). The exact procedure for combining the re-recorded sentences with the gesture videos was as follows. First, the recording of each sentence that seemed most compatible with both word meanings was selected. Next, the most strongly stressed syllable in the second sentence was determined. For instance, in the stimulus example this was the second syllable of the verb *berührte* (see Table 4.1). Following this, a video segment was combined with the selected re-recorded sentence so that the onset of the stroke of each hand movement (dominant gesture, subordinate gesture, grooming) coincided with the peak syllable. In the resulting audiovisual stream, the onset of the sentence always marked the onset of the video. In this way, each experimental sentence set was realized by combining one audio recording with three different types of hand

movements: (1) a gesture supporting the dominant meaning, (2) a gesture supporting the subordinate meaning, (3) a grooming movement.

Table 4.1: Stimulus example for Experiment 6

Introduction: Korinna streckte die Hand aus. <i>Korinna reached her hand out.</i>	
Type of hand movement	Ambiguous sentence
Dominant meaning	Sie berührte die Maus <sub>amb</sub> <i>She touched the mouse<sub>amb</sub></i>
	
Subordinate meaning	Sie berührte die Maus <sub>amb</sub> <i>She touched the mouse<sub>amb</sub></i>
	
Grooming	Sie berührte die Maus <sub>amb</sub> <i>She touched the mouse<sub>amb</sub></i>
	

Introductory sentence was identical for all three conditions. Literal translation in italics.

### Pre-Test

The selected video material was edited using commercial editing software (Final Cut Pro 5). A pre-test was conducted to assess how effective the gestures were in disambiguating the homonyms. In this pre-test, the videos were displayed to thirty German native speakers. At the offset of each video, the dominant and the subordinate target word were displayed on the screen. The participants had to select the target word that fit best into the previous video context. Gestures (and the corresponding homonyms) which did not elicit the selection of the correct target word in at least 50% of all subjects were excluded from the experimental set. In this final set of 42 homonyms, dominant gestures elicited a total of 88 % correct responses (*SEM* 2.23) whereas subordinate gestures elicited a total of 85 % correct responses (*SEM* 2.30). The difference was not significant ( $t(1,82) = 1.1, p > .27$ ). After a grooming video, participants selected the dominant meaning in 56% (*SEM* 4.44) of all cases. The meaning

selection after grooming was not significantly different from chance level ( $t(1,41) = 1.3, p > .19$ ).

#### Procedure

The experimental items were randomly divided into three blocks with the constraint that each homonym appeared only once within a block. Each block was then pseudo-randomized separately with the constraints that (1) no more than two consecutive videos belonged to the same condition and (2) the regularity with which one conditions followed another was matched. The experimental lists were assembled from all three blocks. All possible block orders were realized yielding a total of six experimental lists. These were distributed randomly across participants.

An experimental session consisted of three 10-minute blocks. Blocks consisted of an equal number of trials and a matched number of items from each condition. Each session contained 168 trials, consisting of 126 critical trials (42 x each critical condition) plus 42 null events, in which no stimulus was presented and the BOLD response was allowed to return to a baseline state.

The length of the trials in the critical conditions depended on the length of the video clip and ranged from 9.92 sec to 15.08 sec (mean 11.0 sec,  $SD$  0.55 sec). The length of the video clips did not differ significantly between the three experimental conditions ( $F(2,123) < 1$ ). Each trial started with the presentation of a video clip. Following this, two target words were presented visually for 3000 ms and cued participants to judge which of the two words fit better into the context of the previous video clip. Participants held a response box in their right hand and were requested to push one of two buttons depending on the relatedness of the target words. The side on the screen at which the related target word was presented (left or right) was randomly determined for each trial. Hence, participants could not anticipate during the video which button they were to press in the upcoming response phase. Participants were allowed 3 seconds to respond to the target words. Performance rates and reaction times were recorded. Following the presentation of the target words, the trial was ended by the presentation of a fixation cross for 4000 ms.

Null events consisted of a continuous presentation of a fixation cross for 10500 ms.

### FMRI Data Acquisition

Participants were placed in the scanner in a supine position. Visual stimuli (i.e., the videos and the subsequent target words) were presented on a computer screen outside of the scanner, which participants could see via mirror-glasses. Simultaneously with the videos, the corresponding sentences were presented via a set of specialized headphones (Resonance Technology Inc.) that attenuate the scanner noise about 30 dB. Furthermore, each participant wore ear plugs, which act as an additional low-pass filter. Before the experiment was conducted, the author tested whether the auditory sentences were clearly audible in the noisy scanner environment. Additionally, each participant was questioned after the experiment whether all the stimuli had been clearly audible and visible. Nobody reported any problems.

Eighteen axial slices (4 mm thickness, 1 mm inter-slice distance, FOV 19.2 cm, data matrix of 64 x 64 voxels, in-plane resolution of 3 x 3 mm) were acquired every 2 seconds during function measurements (BOLD sensitive gradient EPI sequence, TR = 2 seconds, TE = 30 ms, flip angle = 90, acquisition bandwidth = 100 Hz) with a 3 Tesla Siemens TRIO system. Prior to functional imaging T1-weighted MDEFT images (data matrix 256 x 256, TR 1.3s, TE 10 ms) were obtained with a non-slice-selective inversion pulse followed by a single excitation of each slice (Norris, 2000). These images were used to co-register functional scans with previously obtained high-resolution whole head 3D brain scans—128 sagittal slices, 1.5 mm thickness, FOV 25.0 x 25.0 x 19.2 cm, data matrix of 256 x 156 voxels (Lee et al., 1995).

### FMRI Data Analysis

The accuracy data was analyzed by means of a repeated-measure ANOVA with the factor GESTURE\_TYPE (dominant, subordinate) and BLOCK (1, 2, 3). The reaction time data was analyzed using a repeated-measure ANOVA with the factors MOVEMENT\_TYPE (dominant gesture, subordinate gesture, grooming) and BLOCK (1, 2, 3). Greenhouse-Geisser correction was applied where appropriate. In these instances, the uncorrected degrees of freedom, the correction factor  $\epsilon$  and the corrected  $p$  value are reported.

The functional imaging data was processed using the software package LIPSIA (Lohmann et al., 2001). Functional data were motion-corrected offline with the Siemens motion correction protocol (Siemens, Erlangen, Germany). Data were subsequently corrected for the temporal offset between slices acquired in one scan using a cubic-spline interpolation based on the Nyquist-Shannon-Theorem. Low-frequency signal changes and baseline drifts were removed

by applying a temporal highpass filter to remove frequencies lower than 1/120 Hz. A spatial gaussian filter with 8 mm FWHM was applied.

To align the functional dataslices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (3 rotational, 3 translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT (Norris, 2000) and EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference data set, which was acquired during a previous scanning session. The MDEFT volume data set with 160 slices and 1 mm slice thickness was standardized to the Talairach stereotactic space. The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system. The transformation parameters obtained from the normalization procedure were subsequently applied to the functional data. Voxel size was interpolated during co-registration from 3 x 3 x 4 mm to 3 x 3 x 3 mm.

Because the neural correlates of the interaction between gesture and speech was of special interest in Experiment 6, the disambiguation points of the gestures as determined in the gating experiment (Experiment 4) was modelled as an event. In the case of grooming the mean disambiguation point of the dominant and subordinate gesture of a sentence-triplet was used as event. The disambiguation point did not differ significantly between the three experimental conditions ( $F(2,123) < 1$ ). The design matrix was generated with a synthetic hemodynamic response function (Friston et al., 1998; Josephs, Turner, & Friston, 1997). The subsequent statistical analysis was based on a linear model with correlated errors (Worsley et al., 2002). Trials which were followed by an incorrect response were excluded from the statistical analysis.

For each participant three contrast images were generated: (1) Dominant gestures vs. Grooming, (2) Subordinate gestures vs. Grooming, (3) Subordinate gestures vs. Dominant gestures. Because individual functional datasets had been aligned to the standard stereotactic reference space, a group analysis based on the contrast images could be performed. Single-participant contrast images were entered into a second-level random effects analysis for each of the contrasts. The group analysis consisted of a one-sample t-test across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from zero. Subsequently, t-values were transformed into Z-scores. To

protect against false positive activation a double threshold was applied, by which only regions with a Z-score exceeding 3.09 ( $p < 0.001$ , uncorrected) and a volume exceeding 12 voxels ( $324 \text{ mm}^3$ ) were considered. This non-arbitrary voxel cluster size was determined by using the program AlphaSim (Ward, 2000) and is equivalent to a significance level of  $p < 0.05$  (corrected). Larger clusters of activation were checked for the existence of local maxima. A voxel was defined to be a local maximum if its z-value exceeded 3.09, if it was largest within a 12 mm radius and if the local volume of spatially contiguous activated voxels exceeded the cluster size threshold of  $324 \text{ mm}^3$ .

The time course of MR signal intensity was extracted for the most significant voxel of each cluster for each individual participant. Percent of signal change was calculated by dividing the MR signal by the constant of the linear model. Because the BOLD response typically peaks 6 seconds after stimulus onset, it was decided on the basis of the mean percent signal change between 4 and 8 seconds post stimulus onset whether a given activation difference was due to either a positive or a negative BOLD response.

#### 4.4.2 Results

##### Behavioral Results

Accuracy of responses and reaction times were recorded during the functional measurement. Here, first the accuracy data are reported, following by the reaction time data.

In general, participants reliably selected the intended target word after both the dominant as well as the subordinate gesture videos (see Figure 4.4). Differences in the performance of participants were analyzed in a repeated-measures ANOVA with the dependent variable performance rate and the independent variables GESTURE\_TYPE (dominant, subordinate) and BLOCK (1, 2, 3). The ANOVA yielded a significant main effect of BLOCK ( $F(2,32) = 5.4$ ;  $\epsilon = .83$ ;  $p < .05$ ) indicating that accuracy increased across the experimental run. The main effect of GESTURE\_TYPE ( $F(1,16) = 3.0$ ;  $p = .10$ ) and the interaction between GESTURE\_TYPE and BLOCK ( $F(2,32) < 1$ ) were not significant.

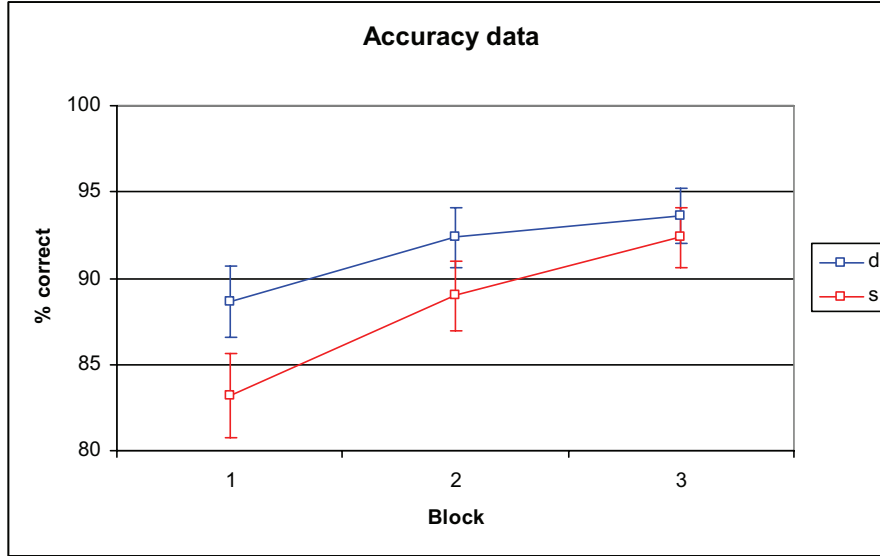


Figure 4.4: Accuracy Data for Experiment 6. Percentage of correctly selected target words for dominant and subordinate gestures. The blue line represents responses following dominant gestures, the red line responses following subordinate gestures. The error bars indicate the standard error of the mean.

Because there was no correct response possible after grooming videos, this data was analyzed separately. Overall, participants selected the dominant target word after 54.0 % ( $SEM\ 1.87$ ) of all grooming videos. The corresponding ANOVA indicated that the selection of dominant target word was significantly above chance level ( $F(1,16) = 4.4$ ;  $p = .05$ ). No other effects of this ANOVA were significant.

The reaction time data (see Figure 4.5) was analyzed in a repeated-measures ANOVA with the factors MOVEMENT\_TYPE (dominant gesture, subordinate gesture, grooming) and BLOCK (1, 2, 3). The ANOVA yielded a significant main effect of BLOCK ( $F(2,32) = 28.00$ ;  $\epsilon = .77$ ;  $p < .0001$ ), indicating that the reaction time decreased over the experimental run. Additionally, a significant main effect of MOVEMENT\_TYPE ( $F(2,32) = 35.24$ ;  $\epsilon = .79$ ;  $p < .0001$ ) was observed. The interaction between MOVEMENT\_TYPE and BLOCK was not significant ( $F(4,64) < 1$ ). Bonferroni-corrected post-hoc tests were performed to further investigate the main effect of MOVEMENT\_TYPE. These tests indicated that the reaction time was significantly shorter for the dominant gestures as compared to grooming ( $F(1,16) = 37.35$ ;  $p_{Bon} < .0001$ ) and also significantly shorter for the subordinate gestures as compared to

grooming ( $F(1,16) = 44.75$ ;  $p_{Bon} < .0001$ ). The difference between dominant and subordinate gestures was not significant ( $F(1,16) = 5.31$ ;  $p_{Bon} = .10$ ).

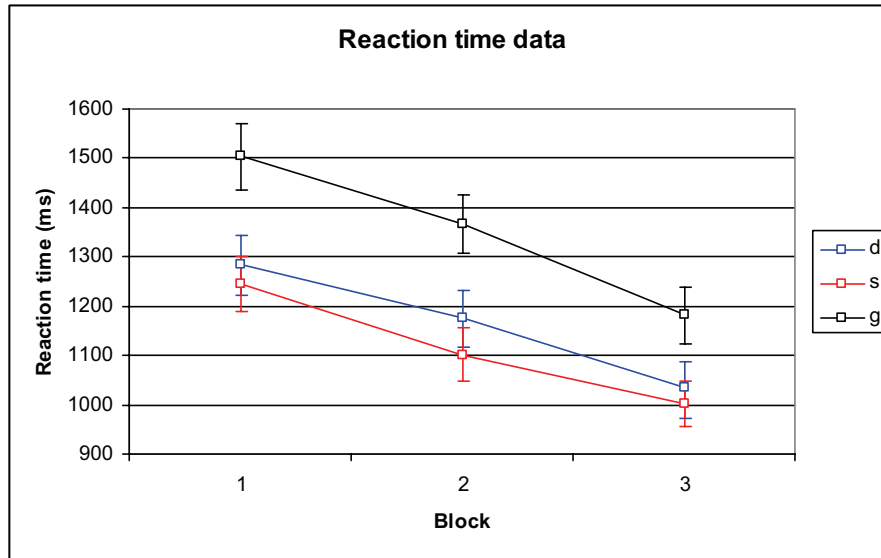


Figure 4.5: Reaction Time Data for Experiment 6. Mean reaction time in ms for dominant gestures, subordinate gestures and grooming.

### Imaging Results

#### *Dominant Gestures vs. Grooming*

The processing of dominant gestures vs. grooming elicited greater levels of activation in the left temporo-occipital cortex. The two local maxima of this activation were found in the posterior STS (see Table 4.2, Figure 4.6) and the lateral part of the middle occipital gyrus.

Increased levels of activation for dominant gestures as compared to grooming were also found in the inferior parietal lobule (BA 40) bilaterally and in the precentral sulcus bilaterally. Additionally, activations in the medial part of the left middle occipital cortex, the medial part of the left middle frontal gyrus (BA9), the right intraparietal sulcus and in the right fusiform gyrus were observed.

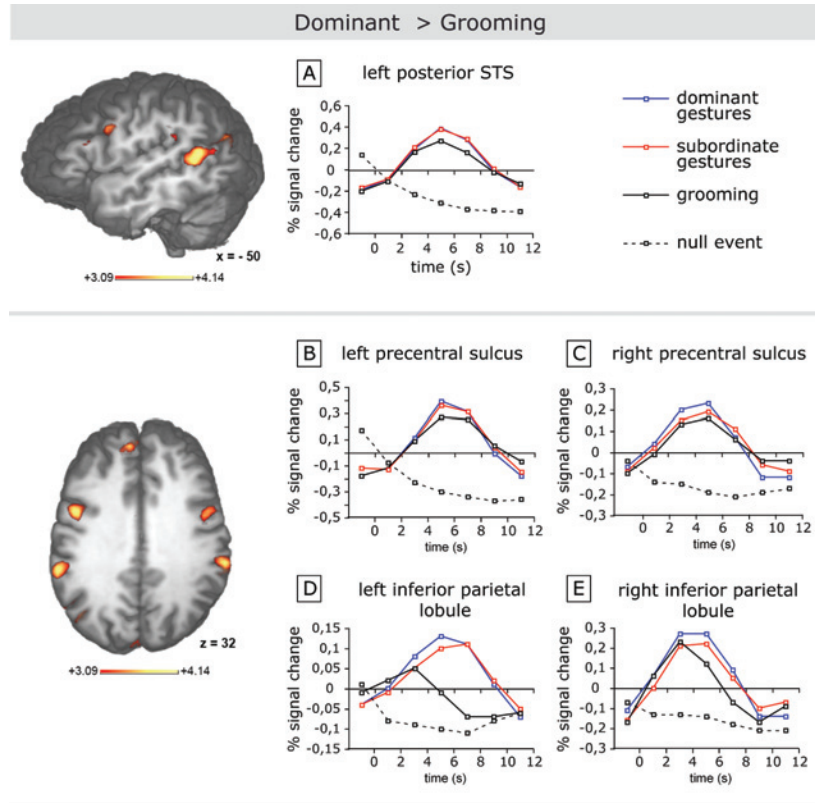


Figure 4.6: Imaging Data for Experiment 6: Dominant vs. Grooming. Illustration of brain regions showing an increased BOLD response to dominant gestures as compared to grooming. Time-courses are given for the most significant voxel of each cluster (for the Talairach coordinates of the voxels, see Table 4.2).

#### *Subordinate Gestures vs. Grooming*

The processing of subordinate gestures as compared to grooming was associated with increased activation in the left temporo-occipital cortex (see Table 4.2, Figure 4.7). The two local maxima of this activation were located in the posterior STS and the temporo-occipital junction. Additionally, increased activation was observed in the inferior parietal lobule (BA 40) bilaterally and the left fusiform gyrus. Upon reducing the activation threshold minimally ( $Z > 2.58$ ;  $p < .005$ ), it immediately became apparent that differences in the precentral sulcus bilaterally as well as the right fusiform gyrus were present in this contrast as well.

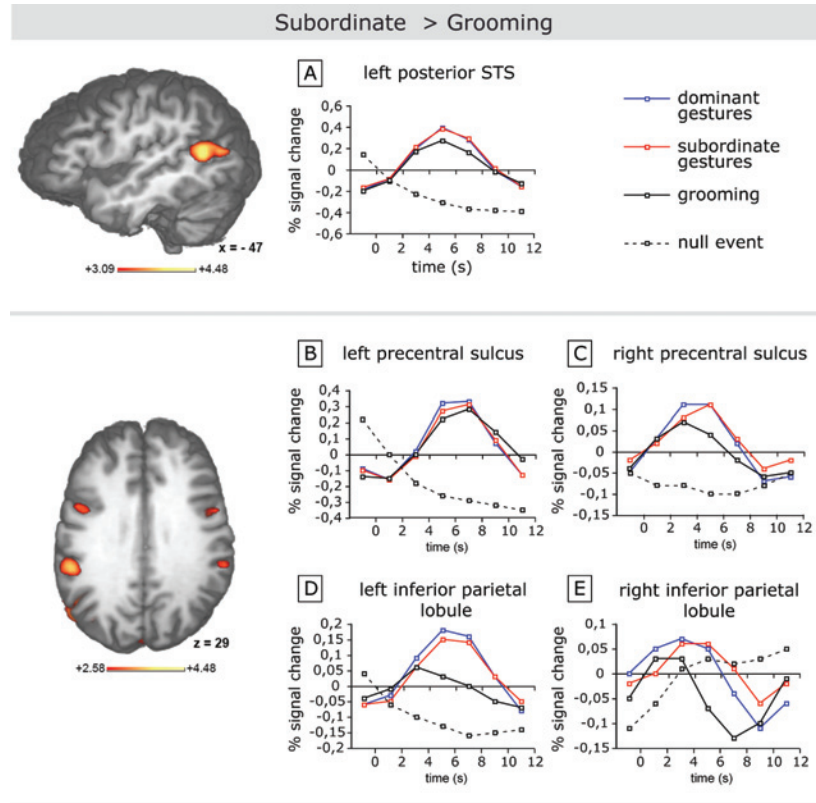


Figure 4.7: Imaging Data for Experiment 6: Subordinate vs. Grooming. Illustration of brain regions showing an increased BOLD response to subordinate gestures as compared to grooming.

#### *Subordinate Gestures vs. Dominant Gestures*

There was no increased activation for subordinate gestures vs. dominant gestures (Sub > Dom). However, in the reverse contrast (Dom > Sub), a significant activation difference in the left lateral middle frontal gyrus (BA9) was observed. The corresponding time-course analysis revealed that this difference was due to a negative BOLD response in the time range from 4 to 8 seconds which was stronger in the case of the subordinate gesture (see Figure 4.8).

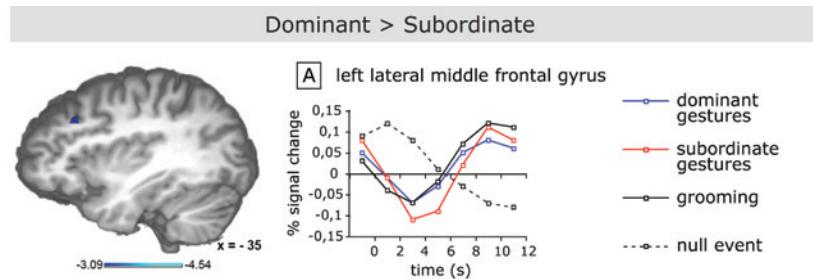


Figure 4.8: Imaging Data for Experiment 6: Dominant vs. Subordinate. Illustration of the brain region showing a less negative BOLD response to dominant gestures as compared to subordinate gestures.

Because the hypothesis for Experiment 6 specifically targeted the left IFG and the posterior STS, it was additionally checked whether there were activation differences at a reduced significance threshold present in these brain areas ( $Z > 2.58$ ,  $p < .005$ ). No differences were observed when directly contrasting the two gesture types (neither for Sub > Dom nor for Dom > Sub).

Additionally, it was checked whether the processing of gestures as compared to grooming yielded significant activation differences in the left anterior inferior IFG, because this brain area was previously suggested as global integration site of gesture and speech. The time course of MR signal intensity was extracted from a spherical ROI of 10 mm diameter around the center coordinate of the IFG activations reported in the Willems study (Talairach coordinates: -43, 11, 26 Willems et al., 2006, their Fig. 3b). The mean percent signal change between 4 and 8 seconds was analyzed as dependent variable in a repeated-measures ANOVA with the factor MOVEMENT\_TYPE (dominant, subordinate, grooming). The main effect of MOVEMENT\_TYPE was not significant ( $F(2,32) < 1$ ).

Table 4.2: List of significantly activated regions in Experiment 6

Contrast	Region	Z <sub>max</sub>	Extent (mm <sup>3</sup> )	x	y	z
D > G	Left medial middle frontal gyrus	3.95	1026	-8	45	36
	Right precentral sulcus	3.98	1458	49	3	36
	Left precentral sulcus	3.93	972	-47	3	33
	Right inferior parietal lobule	4.35	1728	58	-36	30
	Left inferior parietal lobule	4.16	1215	-59	-36	33
	Right intraparietal sulcus	4.03	351	34	-39	42
	Right fusiform gyrus	4.03	864	37	-48	-6
	Left temporo-occipital cortex		3699			
	Posterior STS	4.08		-50	-54	15
	Lateral middle occipital gyrus	4.2		-38	-75	24
	Left medial middle occipital gyrus	3.51	1107	-8	-96	9
	Left precentral sulcus *	3.17	594	-47	6	27
S > G	Right precentral sulcus *	3.19	594	43	0	27
	Right inferior parietal lobule	4.54	594	55	-24	39
	Left inferior parietal lobule	4.38	837	-56	-36	33
	Right fusiform gyrus *	3.29	513	40	-48	-9
	Left fusiform gyrus	3.96	378	-41	-51	-9
	Left temporo-occipital cortex		3861			
	Posterior STS	4.03		-47	-57	12
S > D	Occipito-temporal junction	4.66		-53	-72	12
	No significantly activated regions.					
D > S	Left lateral middle frontal gyrus↓	-3.45	378	-35	24	36
	White matter	-5.03	459	-20	42	12
	White matter	-4.28	2268	28	-60	24

Results of Experiment 6. Abbreviations: D = Dominant gesture; S = Subordinate gesture; G = Grooming; STS = Superior temporal sulcus. Significance threshold  $p < 0.001$  (uncorrected); cluster size threshold 324 mm<sup>3</sup>. Activations marked by \* are significant at a  $p < 0.005$  (uncorrected). The activation marked by ↓ was due to a negative BOLD response (see also Figure 4.8), all other activations were found to be due to positive BOLD responses (see also Figure 4.6 & Figure 4.7).

#### 4.4.3 Discussion

Experiment 6 investigated the neural correlates of the processing of co-speech gestures. Sentences containing an unbalanced ambiguous word were accompanied by either a meaningless grooming movement, a gesture illustrating the more frequent dominant meaning or a gesture illustrating the lesser frequent subordinate meaning. There were two specific hypotheses in mind when this experiment was designed. First, it was expected that the STS would be more involved in the processing of co-speech gestures than in the processing of co-speech grooming movements. Second, it was hypothesized that the processing of subordinate gestures would recruit the left IFG to a stronger degree than dominant gestures. While there was support for the first hypothesis, but the second hypothesis was not supported by the data. The main results are that when contrasted with grooming, both types of gestures (dominant

and subordinate) activated an array of brain regions consisting of the left posterior STS, the inferior parietal lobule bilaterally and the ventral precentral sulcus bilaterally.

#### Gesture vs. Grooming

##### *STS*

When contrasted with grooming, the processing of both gesture types (dominant and subordinate) elicited greater levels of activation in the left posterior STS (see Table 4.2, Figure 4.6, Figure 4.7).

The human STS is known to be an important audiovisual integration site (Beauchamp, 2005). For example, the McGurk-Effect is associated with increased levels of activation in the left posterior STS (Sekiya et al., 2003). The STS was also found to be crucial for the integration of letters and speech sounds (van Atteveldt, Formisano, Goebel, & Blomert, 2004), pictures and sounds (Beauchamp, Lee et al., 2004) as well as videos of tool actions and their corresponding sounds (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004a). This suggests a rather broad spectrum of audiovisual integration processes that recruit this brain area. In the present study, the local maxima in the posterior STS for dominant and subordinate gestures are in close proximity to those coordinates reported for the integration of lip movements and speech (Calvert et al., 2000; Sekiya et al., 2003). Given the interactive nature of iconic gestures (i.e., their dependency on a co-speech context in order to become distinctly meaningful), the increased activation for gestures vs. grooming observed in the left posterior STS is suggested to reflect the interaction of gesture and speech in comprehension. Because a gesture has to be interpreted to some extent before it can be associated with its co-speech unit, the interaction has to occur on a semantic-conceptual level. For instance, the processing of a repeated tapping movement of the index finger (see the stimulus example in Table 4.1) might initially activate an array of possible meanings (e.g., *mouse clicking* as well as *impatience*). This gestural information could then be related (i.e., it interacted) with the lexically ambiguous word, resulting in a selection of the appropriate word meaning. This multimodal matching process is suggested to yield increased activation in the posterior STS. In contrast, grooming did not interact in a meaningful way with the ambiguous sentence, hence the signal increase in the posterior STS is less pronounced.

Because the contrast is based on the comparison of different stimuli (gesture vs. grooming), it is in principle possible that the posterior STS activation primarily reflects differences in the

stimuli kinematics (e.g., amount of motion). There are some facts that speak against such an interpretation of the data. First, the average length of the videos did not differ between the three experimental conditions. Second, although the posterior STS has been found to be involved in the processing of biological motion, these activations have been characterized as being markedly right-lateralized (Pelphrey et al., 2003). In contrast, in the present study, greater levels of activation were found in the left posterior STS for gesture as compared to grooming, suggesting that the activation was not primarily driven by biological motion. Third, activation in the left posterior STS is not modified when observed limb motion is included in the analysis as a nuisance factor. This post-hoc analysis of the stimuli was conducted as follows: First, the position of the right hand was manually marked in each video frame. More precisely, the pixel coordinate of the junction point between index finger and thumb was marked (or estimated, if occluded from sight). Next, the procedure was repeated for the left hand. Subsequently, the euclidian distance between adjacent frames was calculated, yielding the mean amount of distance traveled by the hand for each video. The mean across both hands, for each video, was modeled as an epoch in the design matrix. When analyzing the parametric effect of arm motion, strong activations were observed in primary visual cortex as well as extrastriate cortical areas (including MT/V5, middle occipital gyrus, cuneus, precuneus). Thus, the variable seems to be a valid indicator for brain activity related to motion in video sequences (Dupont et al., 1997; Grill-Spector & Malach, 2004). When accounting for arm motion by introducing this variable as a nuisance factor to the design matrix, the STS activation was not affected, suggesting that it is not driven by kinematic differences between gesture and grooming.

Another possible explanation of the posterior STS activation is that it reflects the difference of meaningful vs. meaningless hand movements (cf. Allison, Puce, & McCarthy, 2000). However, as has been repeatedly stated in the literature, iconic gestures only become distinctly meaningful when accompanied by their co-speech context. Decontextualized iconic gestures convey little meaning to the listener (Feyereisen et al., 1988; Hadar & Pinchas-Zamir, 2004; Krauss et al., 1991), therefore it is rather unlikely that the STS activation reflects the processing of gesture meaning *per se*.

Finally, the greater levels of activation for gesture vs. grooming might partially reflect a less attentive processing of the grooming videos. Participants may have, as soon as they realized that the sentence was accompanied by a grooming movement, put less effort on processing

the stimulus and prepared themselves to respond at random. Such a strategy would result in shorter reaction times for grooming as compared to gesture. However, the reaction time after grooming was actually longer than after the gesture videos (see Results) suggesting that grooming videos were also attentively processed.

The processing of iconic gestures as compared to grooming did not elicit activation in the left anterior inferior IFG. Willems et al. (2006) have suggested that this brain area is involved in global integration processes at a sentence level. Although negative findings can happen for a variety of reasons, one possible explanation for the lack of activation in anterior inferior IFG in Experiment 6 is that the local integration of gesture and speech is anatomically distinct from global integration processes. The local integration of gesture and speech, presumably housed in the posterior STS, may be followed by an integration at the global level in the left IFG, where an amodal representation of the sentence meaning is assembled from the individual meaningful parts of the sentence. Of course, other factors like the employed design (mismatch vs. disambiguation) or the type of analysis (epoch-related vs. event-related) might also be a reason for the divergent findings between Experiment 6 and the study by Willems and co-workers. Clearly, further research is needed to determine the interplay of these two brain regions in the processing of co-speech gestures.

#### *Frontal and Parietal Activations*

When contrasted with grooming, both types of gestures elicited increased activation in the inferior parietal lobule (IPL, BA 40). Only dominant gestures additionally elicited greater levels of activation in the precentral sulcus (BA 6), extending anteriorly into BA 44. However upon reducing the activation threshold minimally ( $Z > 2.58$ ;  $p < .005$ ), it immediately became apparent that differences in this area were present for subordinate gestures bilaterally as well (see Table 4.2). Please note also that there were no significant differences in the right fusiform gyrus and the precentral sulcus observed when the two types of gesture were directly compared (Dominant vs. Subordinate) suggesting that the pattern of activation in the precentral sulcus and the right fusiform gyrus is not qualitatively different between dominant

and subordinate gestures. Because they fall within the specified area, the activation peaks in the precentral sulcus are henceforth referred to as ventral premotor cortex (vPMC)<sup>18</sup>.

The frontal and parietal brain regions in which the processing of co-speech gestures elicited increased levels of activation have been described in the literature as core components of the putative human mirror neuron system (Rizzolatti & Craighero, 2004). It has been demonstrated previously that planning as well as execution of transitive gestures (i.e., gestured movements involving an object) activates the left premotor cortex and left BA 40 (Fridman et al., 2006). Given that the majority of iconic gestures in the present dissertation re-enacted the actions described in the sentence, this system of fronto-parietal activations is interpreted as an involvement of the mirror neuron system in co-speech gesture comprehension. However, in which way might the mirror neuron system support the integration of gesture and speech? One recent theoretical suggestion is that the mirror neuron system determines the goal of observed actions through an observation-execution matching process (Iacoboni, 2005; Iacoboni & Wilson, 2006). Translated to the context of the present disambiguation paradigm, determining the goal is equivalent to finding the answer to the following question: “*Why did the speaker just produce this hand movement?*”. In the case of grooming, the answer would be “*because she wanted to scratch herself*”. In the case of gesture (e.g., the clicking mouse gesture), the answer would be “*because she wanted show how the touching was done*”. In both cases, there is a goal that can be attributed to the observed hand movement. However, the process leading to goal attribution might be more costly in the case of gesture. According the action-observation matching model (Iacoboni, 2005), the goal of an observed action has been determined when the predicted sensory consequences of the internal motor simulation matches the observed visual input. When there is no match, because the initial goal hypothesis was incorrect, a new goal has to be generated which entails a new simulation cycle. Iconic gestures are undeniably more complex hand movements than grooming and the meaning of these gestures is inherently vague. Because of this, the goal initially attributed to a gesture probably not always turns out to be the correct one, therefore the total number of simulation cycles needed for gesture is presumably larger

---

<sup>18</sup> The anatomical border between ventral and dorsal premotor cortex is still a matter of debate (see, for example, Schubotz, 2004). Here, we follow the suggestion from Rizzolatti et al. (2002), who locate the border at the upper limit of the frontal eye field, corresponding to  $z = 51$  in Talairach space.

than for grooming. Thus, the greater levels of activation in vPMC and IPL for gesture vs. grooming might reflect greater “simulation costs” for the processing of gestures.

An alternative explanation for the activation in the precentral sulcus might be that participants used a verbalization strategy for the gestures but not for grooming. The observed activation extended anteriorly into BA 44, an area known to be involved in verbalization processes (e.g., Nixon, Lazarova, Hodinott-Hill, Gough, & Passingham, 2004). However, this explanation is considered to be rather unlikely for two reasons: First, it is probably difficult to employ a verbalization strategy when the gesture is embedded in a co-speech context, because the phonological loop is already busy with the processing of the sentence (Baddeley, 2002). Second, a verbalization account would actually predict increased left IFG activation for grooming, because the meaning of an iconic gesture is often difficult to name (Feyereisen et al., 1988) and it is probably easier to use a verbalizing strategy for the grooming movements (e.g., “scratch”).

#### *Fusiform Gyrus*

Increased levels of activation were found in the right fusiform gyrus for dominant gestures vs. grooming. For subordinate gestures vs. grooming, significant activation in the fusiform gyrus was restricted to the left fusiform gyrus, however, at a lower significance threshold ( $Z > 2.58$ ,  $p < .005$ ), differences in the right fusiform gyrus were present for the subordinate gestures as well. It has been suggested that the fusiform gyrus supports the processing of complex visual stimuli for which visual expertise has been developed (Gauthier & Bukach, in press; Tarr & Gauthier, 2000). In this view, the activation of the fusiform gyrus during face observation (Kanwisher et al., 1997) indicates that we are all experts in face processing. Participants who are experts in other domains, such as recognition of types of birds or cars, exhibited increased levels of activation in the fusiform region for those stimuli (Gauthier, Skudlarski, Gore, & Anderson, 2000). Grooming movements tend to be very repetitive and most of them go unnoticed (Goldin-Meadow, 2003). The higher levels of activation for gestures vs. grooming in the right fusiform gyrus may therefore be due to the fact that participants have more expertise in the processing of gestures than in the processing of grooming movements.

### Subordinate vs. Dominant Gestures

#### *Subordinate > Dominant*

It was hypothesized that the processing of subordinate gestures would elicit increased levels of activation in the left IFG than dominant gestures, because this brain area is known to be sensitive to semantic processing difficulties like word frequency (Fiebach et al., 2002; Fiez et al., 1999; Joubert et al., 2004). However, no significant differences in this brain area were observed. In light of the rather low amount of dominant target word selections after grooming videos (just above chance level), it is a possibility that word meaning frequency was not effectively varied in Experiment 6. Note, however, that in a number of previous experiments, the same set of homonyms elicited strong effects of word meaning frequency (Gunter et al., 2003, as well as Experiment 3 of the present dissertation). For instance, in Experiment 3, the same gestures were paired with sentences that were slightly longer and contained a disambiguating target word (e.g., *She touched the mouse, which the cat / computer ...*). The dependent variable in Experiment 3 was the N400 time-locked to the target words. Importantly, it was observed that following a grooming movement, the N400 time-locked to the target words was significantly larger at subordinate target words as compared to the dominant targets. This suggests that at the position of the target word, the subordinate word meaning was less active in working memory than the dominant meaning. Thus, in the absence of a gestural cue for meaning selection, word meaning frequency governed the selection process. Why did not a similar effect of word meaning frequency observed in Experiment 6, although the gestures and the homonyms were identical and the sentence structure highly similar? One explanation might be the nature of the task in Experiment 6 (two-alternative forced-choice) which contrasts with the much more subtle measure of N400 amplitude in Experiment 3. Frequency effects are generally considered to influence the lexical access of a word. However, in the case of Experiment 6, both word meanings of the homonym are explicitly presented to the participants (via the display of the two related target words during the response phase). Thus, there was no need for the participants to search their mental lexicon for the possible word meanings of the homonym and therefore little range for an effect of word meaning frequency to occur. Thus, there is some reason to assume that word meaning frequency was effectively manipulated in Experiment 6, although the behavioral data suggest the opposite. Please note also that the statistical modelling of the fMRI data was

performed at the disambiguation point during the gesture video (see Methods) and not during the delayed response.

#### Dominant > Subordinate

The only significant differences in the contrast of subordinate and dominant gestures were increased levels of activation for dominant gestures. The only significant activation difference between both types of gestures that was located in the grey matter was found in the left lateral middle frontal gyrus (BA 9), an area also known as the dorsolateral prefrontal cortex (DLPFC) which is crucially involved in cognitive control (Brass, Derrfuss, Forstmann, & Cramon, 2005; Hoshi, 2006). Experiment 3 as well as the study by Kelly et al. (2007) suggest that gesture-speech integration involves a considerable degree of controlled processes and is not entirely automatic. In Experiment 6, the time-course analysis for the DLPFC (see Figure 4.8) revealed that the observed signal difference in this brain area was due to more pronounced BOLD signal decreases in the time window from 4 to 8 seconds for the subordinate gestures as compared the dominant gestures. Although the functional significance of negative BOLD responses is still a matter of debate, several studies suggest that it might primarily reflect neuronal deactivation (Shmuel, Augath, Oeltermann, & Logothetis, 2006; Shmuel et al., 2002; Stefanovic, Warnking, & Pike, 2004). The DLPFC might control the degree of activation of the system underlying gesture-speech integration by means of phasic disinhibition of a tonic inhibitory connection. When a given stimulus input does not require a semantic integration of gesture and speech, the integration system may be tonically inhibited by the DLPFC. However, when speech is accompanied by a potentially meaningful hand movement, activation in the DLPFC is reduced, resulting in increased activation in the system underlying gesture-speech integration. Because of the bias of word meaning frequency, disambiguating the sentence towards the subordinate meaning requires more controlled effort which is suggested to be reflected in the stronger signal decrease of DLPFC in this condition. In contrast, disambiguating towards the dominant meaning requires less controlled effort; therefore the signal decrease of DLPFC is not as pronounced as it is for the subordinate meaning.

#### **4.4.4 Conclusion**

Experiment 6 investigated the neural correlates of co-speech gesture processing. The processing of speech accompanied by meaningful hand movements reliably activated the left

posterior STS which might reflect the multimodal semantic interaction between a gesture and its co-expressive speech unit. The processing of co-speech gestures additionally elicited a fronto-parietal system of activations in classical human mirror neuron system brain areas. The mirror neuron system is suggested to be involved in the decoding of the goal of observed hand movements through an observation-execution matching process.

## Chapter 5: General Discussion

This dissertation explored the processing of co-speech iconic gestures. The starting point was the ongoing controversy in the literature whether these gestures convey information to the listener at all (see Krauss et al., 1995; McNeill et al., 1994). Whereas the group around McNeill has maintained that the information from iconic gestures is routinely processed and combined with the co-expressive speech unit into an enriched unified representation, Krauss and colleagues have argued that these gestures are mainly an epiphenomenon of speech production processes and are only subject to minimal semantic analysis in comprehension.

This issue was addressed in Experiments 1 – 3, where participants listened to lexically ambiguous sentences containing an unbalanced homonym that were disambiguated at a later target word (e.g., *She touched the mouse, which the cat / computer ...*). Coincident with the initial part of the sentence, the speaker produced an iconic gesture illustrating either the more frequent dominant meaning (e.g., *animal*) or the less frequent subordinate meaning (e.g., *computer device*) of the sentence. The amplitude of the N400 of the ERPs time-locked the target word systematically varied as function of context congruency, i.e., the N400 was larger if it was preceded by an incongruent gesture and smaller if it was preceded by a congruent gesture (Experiments 1 and 2).

This result strongly suggests that listeners use the information from iconic gestures to disambiguate speech. Thus, it is evidence for the view that iconic gestures *do* communicate information to the listener. It is in line with other behavioral studies which have also shown that listeners are sensitive the information represented in these hand movements (Alibali et al., 1997; Beattie & Shovelton, 1999b; Cassell et al., 1999). Additionally, Experiments 1 and 2 are in line with previous ERP studies (Kelly et al., 2004; Wu & Coulson, 2005) in showing that an initial gesture context can modulate the processing of a subsequent target word. At the same time, Experiment 1 and 2 extend the previous ERP findings in showing that contextual effects of gesture are not restricted to the processing of isolated target words but can be generalized to auditory sentence processing.

Once meaningless grooming movements were added to the experimental paradigm, the impact of gesture was weakened, but not eliminated (Experiment 3). In this experiment, it was found that only the N400 at subordinate targets varied as a function of the preceding gesture. In contrast, the N400 at dominant target words no longer varied as a function of context

congruency in Experiment 3. A likely explanation of this effect is that once grooming was added, participants treated all hand movements (including the gestures) as less informative. This devaluation of gesture is suggested to be reflected in the fact that the gesture context exhibited the typical pattern of weakly biasing contexts, which are characterized by their inability to modulate the activation of the dominant word meaning (see also the General Discussion of Experiments 1 - 3, p. 44). Because apart from the addition of grooming, all other experimental details in Experiment 3 (including the task) were the same as in Experiment 2, the result of Experiment 3 was interpreted as reflecting that iconic gesture comprehension is not an entirely automatic process, but does also involve a considerable degree of controlled processes (for more on this, see below).

### **5.1 Requirements of a to-be-developed Theory of Iconic Gesture Comprehension**

While Experiments 1 – 3 have shown that listeners use the gestural information to disambiguate speech, Experiment 4 set out to investigate the earliest point in time at which a gesture becomes meaningful for an addressee. To this end, a modified gating paradigm (Grosjean, 1996) was used to determine the point in time at which the gesture information reliably contributed to selecting the appropriate meaning of the corresponding homonym. This time point (termed disambiguation point) was determined for all experimental gesture clips. After exploring where the disambiguation points occurred relative to stroke onset, it was found that the disambiguation points of 60 gestures were located prior to the onset of the stroke. This means that almost two thirds of all gestures in the experimental set enabled a meaning selection before the participants had actually seen the stroke, which has been suggested as the “phase that carries the gesture content” (McNeill, 1992, p. 84)<sup>19</sup>. The disambiguation point was validated in a co-speech context using a sentence completion task in Experiment 5.

When trying to discuss the theoretical significance of the findings of Experiment 4 and 5, one inevitably notices that there is no real theory of gesture comprehension, at least not one that is

---

<sup>19</sup> Note that this statement is based on an interpretative – and not an empirical - approach to gesture. McNeill has never tested whether addressees do also perceive the stroke phase as the content-bearing part of gesture.

based on an empirical approach to gesture. Most theories on gesture are concerned with production issues (de Ruiter, 1998; Kita & Özyürek, 2003; Krauss et al., 1996). In contrast, only little effort has been put so far into developing a coherent information processing theory of gesture comprehension. In the following, I will summarize the data that a to be developed theory of iconic gesture comprehension has to account for and identify the main theoretical issues. Where appropriate, I will explicitly refer to the few already existing theoretical suggestions, which have mainly been put forward by the McNeill group (Cassell et al., 1999; McNeill et al., 1994).

### 5.1.1 How Does Gesture Help the Listener?: Main Effect of Gesture or Gesture-Speech Interaction?

A first important issue that a to be developed empirical theory of iconic gesture comprehension has to account for is the actual cause of the communicative effect of iconic gestures: the main effect of gesture or the interaction between gesture and speech. The answer to this question is at the core of a theoretical understanding of the comprehension of these gestures. If these gestures do indeed form an integrated system with language in comprehension, as suggested by McNeill and colleagues (1994), then the observed communicative effects should be mainly due to an interaction between gesture and speech. If, however, iconic gestures derive their capacity for signification mainly through their form and independently of an accompanying speech context, the communicative effects should be mainly attributable to a main effect of gesture. As will become clear below, this issue cannot be considered as being settled yet.

The data of the present dissertation, in particular Experiments 1 – 3, can be construed as evidence for an interactive view of iconic gesture comprehension. When comprehending the experimental videos, listeners seem to have combined the gestural information with the spoken information (i.e., the homonym), resulting in a selection of the word meaning indicated by gesture, which in turn was reflected in the N400 effects observed at incongruent target words. An alternative explanation in terms of a simple priming effect of gesture, can, however, not be excluded. That is, the gestural information *per se* may have directly primed the subsequent target word. For instance, processing the mouse-clicking gesture *per se* may have directly primed the subsequent target word *computer*. While in this example, direct priming seems like a plausible explanation, most pairings of gesture and subsequent target

were arguably semantically much less associated, because gestures that seemed directly related to the target word were already excluded in the recording phase of the stimuli (see p. 24). The pattern of results from Özyürek et al. (2007) does also support the view that iconic gestures do only become meaningful in interaction with their accompanying speech, and not in a speech-independent manner. If they were, one would have expected in their experiment that when both gesture and speech were incongruent to the preceding sentence context (their *double mismatch condition*), this would yield a larger N400 effect than when only one information channel mismatched the preceding context (as in their *gesture mismatch* or *speech mismatch condition*). However, all three mismatch conditions of Özyürek et al. (2007) elicited N400 effects of similar amplitude and latency. Further evidence usually considered to support the interactive view of iconic gesture comprehension are studies that have investigated the degree to which information uniquely conveyed by gesture in audiovisual recordings “trickles” into speech, when participants are subsequently asked to retell the events of the previously seen recordings (Alibali et al., 1997; Cassell et al., 1999). For instance, Alibali et al. (1997) have found that one third of the information uniquely conveyed by gesture in the original recording was in the subsequent retelling expressed in the speech of the participants. This was interpreted to reflect that rather than maintaining separate representations for gesture and speech, the information from both domains is integrated into a unified representation.

However, at the same time, there is also evidence suggesting that iconic gestures can convey information independently of a co-speech context. In a recent ERP study, Wu & Coulson (2007) have found that completely decontextualized iconic gestures (i.e., presented without speech and any prior context) are nonetheless able to prime semantically related target words. In this experiment, it was found that the N400 to target words was larger, if the target word was preceded by an unrelated gesture and smaller, if the target word was preceded by a related gesture. Although interpreted otherwise, the data from Krauss et al. (1991) as well as Pinchas-Zamir and coworkers (2004) do also show that there is at least *some* information contained in decontextualized iconic gestures, because the observed accuracy rates in these studies were significantly above chance level. Furthermore, the studies by Alibali et al. (1997) as well as Cassell et al. (1999) cited above as evidence supporting the interactive view can at the same time be construed as evidence supporting a separate representation of gesture in comprehension. For example, in the Alibali study, two thirds of the information uniquely conveyed by gesture was also expressed by gesture in the subsequent retellings. Additionally,

in the Cassell study, over 40% of the instances where uniquely gestural information in the audiovisual recording had a detectable effect on the subsequent retelling, this was evident in the gesture domain only.

As can be seen from these conflicting findings, a definite answer to the question how iconic gestures derive their capacity for signification – through a main effect of gesture or the interaction of gesture and speech – cannot be given yet. This would require more studies with a design that allows the calculation of both the main effect of gesture as well as the interaction between gesture and speech. If iconic gestures convey meaning mainly through an interaction between gesture and speech, as suggested by McNeill and co-workers (Cassell et al., 1999; McNeill et al., 1994), then the statistical interaction between both factors should be significant and account for a substantial amount of variance of the dependent variable. Alternatively, if the communicative effects of gesture are predominantly independent of the co-speech context, the main effect of gesture should be significant and at the same time account for more variance than the interaction.

To sum up, the existing literature suggests that iconic gestures can convey meaning in both ways, i. e., through an interaction with speech as well as independently of a co-speech context. Accordingly, a to be developed theory of comprehension has to account for both ways of meaning transmission. The next section describes another issue that such a theory has to deal with, namely the temporal aspects of iconic gesture comprehension.

### **5.1.2 Temporal Aspects**

#### Early vs. Late Gesture Recognition

Iconic gestures are a dynamic signal that unfolds over time. Therefore, one obvious and important question is at what point in time the meaning of a such a hand movement is available for an addressee. Does one need to see the complete gesture in order to comprehend its meaning, or is it already available at an earlier time point? This question was addressed in Experiment 4.

Using a modified gating paradigm, the earliest point in time at which participants can reliably select a word meaning of a homonym on the basis of a gesture segment was determined. This so-called disambiguation point was determined for each experimental gesture clip. One finding was that participants did not need to see the complete gesture in order to make a

meaning selection. Instead, it was found that on average approximately 400 ms of gesture sufficed<sup>20</sup>.

A more informative evaluation of the disambiguation point becomes possible if it is analyzed in relation to the onsets of the different phases of gesture. On the basis of formal properties, an iconic gesture can be separated into distinct successive phases (see also McNeill, 1992). A frequently observed phase pattern of iconic gestures, that was also present in all experimental gesture clips, is the following tri-phasic pattern: preparation, stroke and retraction. When analyzing where the disambiguation points occurred relative to the onset of the stroke, it was found that almost two thirds of all experimental video clips enable a meaning selection before participants had actually seen the stroke. This was discussed in detail as possibly reflecting an early impact of gesture on speech disambiguation (see the discussion of Experiments 4 and 5, p. 64). In the following, it is speculated what might be the factors determining whether an observed hand movement is recognized as being meaningful either “early” (i.e., already during the preparation phase) or “late” (lateron, during the stroke phase).

Beattie and Shovelton (1999b) investigated what kind of information is conveyed by iconic gestures. In this experiment, participants were presented with visual, auditory or audiovisual recordings of speakers retelling a cartoon story. After each clip, participants responded to a number of questions, each relating to a different aspect of the scene. Typical questions included: What objects are depicted here (identity / number)? What are the objects doing (action / manner)? What shape are the objects (shape)? What size are the objects (size)? What is the position of the objects relative to anything else (relative position of objects, orientation, location of action, contact)? The authors observed significantly more correct responses for the audiovisual condition (including additional gestural information) than for either the audio-only or the video-only condition. More interestingly, the beneficial effect of gesture was not evenly distributed across the different dimensions. Instead, it was found that information about the relative size and position of objects seems to be especially well conveyed by iconic gestures. In a later study by the same authors, it was observed that gestures illustrating events from a character-viewpoint are a more effective means of communication than observer-viewpoint gestures (Beattie & Shovelton, 2002a). Returning the question of the timecourse of

---

<sup>20</sup> The gesture clips had a mean duration from preparation onset to stroke offset of 1260 ms.

iconic gesture comprehension, it is a plausible possibility that the information type a gesture predominantly conveys (e.g., orientation, manner,...) also influences how early the gesture is recognized as being meaningful. For instance, gestures primarily illustrating the relative size or position of objects might already become meaningful during the preparation phase, whereas iconic gestures illustrating other aspects are only recognized during the later stroke phase. Unfortunately, the stimulus set of the present study did not contain enough systematic variation along these dimensions. Therefore, it was not possible to test this hypothesis on the basis of the present data. However, it seems like worthwhile endeavor for future experiments to examine whether different types of iconic gestures also differ in their processing speed, as reflected by their recognition point.

#### Determining the Time Window of Gesture-Speech Integration

So far the discussion of the temporal aspects of iconic gestures has neglected the temporal synchrony between the two involved channels of information, i.e., the gesture channel and the speech channel. It has been suggested that iconic gestures and speech are synchronous and asynchronous at the same time (McNeill, 1992, 2005). They are asynchronous in that the onset of the preparation phase of gesture tends to precede the related speech unit (Morrel-Samuels & Krauss, 1992). They are synchronous in that the stroke and related speech tend to coincide (McNeill, 1992, p. 92).<sup>21</sup>

From the comprehension perspective, this raises the question whether stroke-speech synchrony is a necessary precondition for the successful decoding of the meaning of an iconic gesture. An analogy may be drawn from another instance of synchronized audiovisual information. Viewing lip movements producing /ga/, accompanied an auditory /ba/ typically results in the McGurk illusion, i.e., the perception of /da/ (McGurk & MacDonald, 1976). Importantly, participants do not perceive the McGurk illusion, if the auditory and visual

---

<sup>21</sup> Note that a strict synchrony between both domains (i.e. onset of stroke aligns with onset of related speech unit) has so far only been demonstrated for pointing gestures and speech (Levelt et al., 1985). With respect to iconic gestures, such *strict* evidence for synchrony is lacking. McNeill (1992, p. 92) has shown that 90 % of the stroke *phases* in his data corpus overlap with simultaneous speech, and 10 % of the stroke phases coincide with silence (e.g. speech pauses, hesitations, and so on). He did, however, not report the exact relationship between onset of stroke and onset of related speech unit.

information is too desynchronized. In order to specify the temporal window of integration underlying the McGurk effect, van Wassenhove et al. (2007) systematically manipulated the asynchrony between auditory and visual McGurk pairs. Asynchronies ranged from -467 ms (auditory lead) to +467 ms (auditory lag). The results were that participants more often reported the fusion instead of the auditory percept for asynchronies ranging from -30 ms to +170 ms. The authors interpreted this finding as evidence for a time window of 200 ms duration during which speech-related audiovisual information is integrated into a single percept. One intriguing question for future research on iconic gesture comprehension is certainly whether a similar time window of integration exists for iconic gesture and speech. For instance, one could manipulate the synchrony between speech and disambiguating gestures in order to determine the boundaries beyond which gestures lose their disambiguating influence.

To sum up, there is so far very little data available on the temporal aspects of iconic gesture comprehension. The findings of Experiments 4 and 5 suggest that a substantial amount of information is already conveyed in the “early” preparation phase of gesture, at least when a forced-choice task is used. More research is needed in order to determine whether this effect generalizes to online processing situations as well to situations where the task does not explicitly require to take gesture into account. On a more general level, this calls for a theory that predicts how gestures acquire meaning over the course of the successive gesture phases. Additionally, more research is needed in order to determine the factors influencing how early a hand movement is recognized as being meaningful. Potential factors have been discussed above. Finally, how the temporal synchrony between gesture and speech affects comprehension is a completely unexplored issue.

### **5.1.3 The Potential Automaticity**

One important theoretical aspect of iconic gesture comprehension is the degree to which the process is considered to be automatic. Two-process theories of information processing (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) state that automatic processes occur very fast, without intention or awareness, and do not tap into limited-capacity resources. In comparison, controlled processes are slower, cannot operate without intention or awareness, and use limited-capacity resources. Thus, if the interactive processes between gesture and speech involved in the comprehension of iconic gestures are

indeed automatic, as it was strongly suggested by McNeill et al. (1994), the impact of gesture in comprehension should not depend on any of the following factors: Type of employed task, experimental context, strategic factors.

The result of Experiment 3 of the present dissertation suggests that iconic gesture comprehension is not an entirely automatic process, but does also involve a considerable degree of controlled processes. Although the gestures and the task in Experiment 3 were identical to Experiment 2 (the only difference being the addition of the grooming movements in Experiment 3), it was found that the gestures were unable to modulate the activation of the dominant word meaning in Experiment 3. On the basis of the literature of homonym processing, this was interpreted as indicating that the gestures in Experiment 3 constituted (only) a weak contextual cue, whereas in Experiments 1 and 2, the gestures had elicited a pattern typical for strongly biasing contexts. One possible explanation is that in a experimental context where the task does not explicitly require an integration of gesture and speech (as in Experiments 2 and 3), listeners put less weight on gestural information as soon as the gestures are intermixed with meaningless grooming movements (for a more detailed account of this explanation, as well as other possible explanations for the different pattern of results between Experiments 2 and 3, please see the General Discussion of Experiments 1 - 3 pp. 44).

Whatever the underlying mechanism, the results of Experiment 3 provide clear evidence that co-speech iconic gesture comprehension does not operate in an entirely automatic fashion, as it was suggested by McNeill et al. (1994). Instead, contextual factors, such as the amount of meaningless hand movements, do also influence the degree to which listeners take gestural information into account. This interpretation is in keeping with a recent ERP study by Kelly et al. (2007). In this experiment, it was found that the N400 effect at target words which were preceded by incongruent vs. congruent gestures is modulated by what the authors call the “intentional coupling” of gesture and speech. That is, a gesture could be followed either by a target word spoken by the same person or by a target word spoken by a different person. 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.

Because automaticity cannot be captured as an all-or-none phenomenon, the individual aspects associated with automaticity should be investigated separately (see also, Schupp et al., 2007). The data available so far already suggests that listeners can to some extent control the degree to which gestural information is taken into account. Therefore, future research should aim to identify further situational factors that promote or impede the influence of gesture on language comprehension. It might turn out to be the case that the gesture channel can be turned on or off in a flexible manner, depending on the situational demands. For instance, the impact of gesture might be higher in a noisy environment and almost absent, if the addressee knows that a given speaker produces a lot of unintended arm movements (imagine observing a gesturing speaker suffering from the Huntington disease).

## **5.2 Functional Neuroanatomical Correlates of Iconic Gesture Comprehension**

Experiment 6 sought after the neural correlates of iconic gesture comprehension. Participants listened to lexically ambiguous sentences (e.g., *She touched the mouse*) that were accompanied by one of three hand movements: (1) a gesture illustrating the more frequent dominant meaning, (2) a gesture supporting the lesser frequent subordinate meaning or (3) a meaningless grooming movement. Because the main interest of Experiment 6 was to identify brain areas involved in the local interaction of gesture and speech, an event-related analysis was performed. That is, the disambiguation point, as it was determined in Experiment 4, was modeled as the event of interest. The main results are that when contrasted with grooming, both types of gestures (dominant and subordinate) elicited greater levels of activation in the ventral premotor cortex bilaterally, the inferior parietal lobule bilaterally and the left posterior STS. The activations in the premotor and the inferior parietal cortex were interpreted as arguably reflecting an involvement of the putative human mirror neuron system in iconic gesture comprehension, in terms of increased “simulation costs” for gesture as compared to grooming. The activation in the left posterior STS was suggested as possibly reflecting the multimodal interaction of the meaning inferred from gesture with the meaning of the ambiguous sentence. The literature supporting these interpretations, as well possible alternative accounts, were already discussed in detail earlier (see p. 100). Here, an attempt is made to put the findings of Experiment 6 into a larger context.

Action and cognition have traditionally been studied separately, with the underlying assumption that mind and body are fundamentally distinct entities. This strict separation is also known as the Cartesian view of the mind-body problem, named after a strong advocator of this position (Rene Descartes). The Cartesian view holds that the mind is primarily an abstract information processor, whose connections to the outside world are of little theoretical importance. Following this tradition, language, which is an important sub-domain of cognition, has been conceptualized as a manipulation of symbols by rules (Chomsky, 1965). The symbols are considered to be stripped of all perceptual and motor content, thus becoming amodal and abstract (cf. Glenberg, 2006).

This contrasts starkly with the “embodied cognition” viewpoint, where cognitive processes are suggested to be deeply rooted in the body’s interactions with the world (Gallese & Lakoff, 2005; Glenberg, 2006). According to this view, humans have evolved from creatures whose neural resources were devoted primarily to perceptual and motoric processing and whose cognitive abilities were mainly concerned with online interaction with the surrounding world. Therefore, human cognition, rather than being centralized and abstract, may instead have deep roots in sensorimotor processing (cf. Wilson, 2002).

Embodied cognition has gained considerable popularity in the cognitive neurosciences in recent years, which was partly fueled by the discovery of the mirror neurons in the macaque brain (Gallese et al., 1996, 2002; Umiltà et al., 2001). The more radical advocates of embodied cognition have argued on the basis of the putative human mirror neuron system that the meaning of certain action words is actually represented in action schemas located in the motor system (Glenberg, in press; Pulvermüller, Shtyrov, & Ilmoniemi, 2005). The assertion made by these researchers is that comprehending action-related language necessitates activation in the relevant parts of one’s own motor system. Note that this is a very strong claim and there is several data suggesting that it overemphasizes the undeniably existing link between action and language. For instance, Aziz-Zadeh et al. (2006) found that visually presented utterances containing action verbs only elicited premotor activation when the verb was used in a literal sense (e.g., *grasping the pen*), but not when it was used in a metaphorical sense (e.g., *grasping the idea*). Similarly, Rüschemeyer et al. (2007) found activation in motor areas for action verbs presented in isolation (e.g., *greifen, to grasp*), but not when the action verb was the stem of a complex abstract verb (e.g., *begreifen, to comprehend*).

In light of such findings, Willems and Hagoort (2007) have argued for a more balanced account of the neural basis of cognition as the dynamic interplay between several cognitive domains. Large scale neural networks are suggested to be formed dynamically, involving those parts of the cortex that are needed by the specific task at hand for the organism (Fuster, 2003; Mesulam, 1990, 1998). This requires of course a high degree of flexibility.

In order to identify large scale networks suggested to be involved in the interaction of language and action, one needs a suitable testing area, that allows distinguishing between effects primarily driven by action, effects primarily driven by language and, crucially, effects driven by the interaction between both domains. The comprehension of co-speech iconic gestures may turn out to become such a suitable testing arena. Iconic gestures are a prime example of actions recruited in the context of language (cf. Özyürek et al., 2007). Together, they form a “composite signal” consisting of semantically and temporally deeply intertwined gestural and spoken information. Appropriate paradigms may allow to disentangle the influence of gestural and spoken information in comprehension, as well as determine the interaction between both domains.

Together with the study by Willems et al. (2006), Experiment 6 is one of the first studies that have aimed to identify regions involved in the interaction of gesture and speech. One important result of Experiment 6 was that the processing of gestures as compared to grooming elicited greater levels of activation in the ventral premotor cortex bilaterally and the inferior parietal lobule bilaterally. This was suggested to reflect an involvement of the human mirror neuron system in the comprehension of co-speech iconic gestures. On the basis of the action-observation matching model (Iacoboni, 2005), this result was interpreted as reflecting increased simulation costs for the processing of gestures as compared to grooming. Most of the gestures used in the present dissertation (i.e., more than two thirds of all gestures) were re-enactments of actions. Therefore, it is after all not very surprising that the processing of these gestures activated brain regions known to be associated with action comprehension (i.e., in the form of the human mirror neuron system). If the brain network for iconic gesture comprehension indeed flexibly and dynamically recruits different cortical areas depending on the features of the stimulus at hand, as suggested by Willems and Hagoort (2007), then an interesting test for future research would be to contrast iconic gestures primarily illustrating objects with iconic gestures primarily illustrating actions. If the network is flexibly recruited, object-related gestures should activate brain regions associated with object recognition in the

ventral stream (e.g., the lateral occipital complex, fusiform gyrus, extrastriate body area, see Downing et al., 2001; Grill-Spector, 2003; Kanwisher et al., 1997) whereas action-related gestures should activate inferior frontal and inferior parietal brain areas, as it was observed in Experiment 6. If, however, an involvement of the mirror neuron system in iconic gesture comprehension is mandatory, then both gesture types (action and object-related) should activate inferior frontal and inferior parietal brain areas. Such an experiment could also yield insights for a better theoretical understanding of how iconic gestures derive their capacity for signification; either in a fixed manner by always primarily providing links to actions, or in a flexible manner, by either providing links to actions or links to objects, depending on the type of gesture that is being processed.

Behavioral data (Alibali et al., 1997; Beattie & Shovelton, 1999b; Cassell et al., 1999) as well as several ERP studies including those of the present dissertation (Kelly et al., 2004; Özyürek et al., 2007; Wu & Coulson, 2005) suggest that there is a considerable degree of interaction between information derived from gesture and information derived from speech. Although alternative explanations cannot be excluded (see section 4.4.3), the results of Experiments 6 are compatible with the suggestion the left posterior STS might be the brain basis of the multimodal interaction between information conveyed by gesture and speech. Via the arcuate fasciculus, the human STS has prominent fiber connections with the inferior parietal lobule and the inferior frontal gyrus (Catani, Jones, & ffytche, 2005). In the monkey, the cortex immediately surrounding the STS has also been termed superior temporal polysensory area (STP), because of its polymodal capabilities (Puce & Perrett, 2003). For instance, Hikosaka et al. (1988) investigated the polymodal capabilities of the neurons in the posterior monkey STS. Out of the 200 neurons tested, 40 neurons showed a multimodal response with 21 responding specifically to audiovisual stimulation. In the human brain, a considerable number of studies implicate an important role of the STS in the integration of speech-related audiovisual information, ranging from simple mappings on the form level (e.g., the McGurk effect, Sekiyama et al., 2003) to complex interactions at the semantic-conceptual level (e.g., pictures and their corresponding sounds, Beauchamp, Lee et al., 2004; Saygin et al., 2003).

As was argued above, during the processing of multimodal stimuli such as co-speech gestures, the brain may flexibly recruit brain areas depending on the properties of the multimodal stimulus. One can speculate that the functional role of the posterior STS in this process is to link together the contributions from each modality. A similar argument was put

forward by Beauchamp et al. (2004), who suggested that “the anatomical location of pSTS/MTG between high-level auditory and visual cortices (as well as the response properties of temporal neurons) renders it well situated to make links between auditory and visual object features” (2004, p. 818). Because this brain area is additionally discussed as the convergence site of the ventral and dorsal visual stream, the authors speculated that the STS “may serve as a general-purpose association device, both within and across modalities” (2004, p. 819). The results of Experiment 6 are compatible with this notion of the posterior STS as a multi-modal association device. However, because alternative explanations cannot yet be excluded, future research should aim to unambiguously determine the role of the posterior STS in the processing of gesture and speech.

### **5.3 How Iconic Gestures Contribute to Language Comprehension: A Tentative Model**

As was mentioned at various points throughout this General Discussion, there is no model that makes testable predictions about how and when gesture and speech interact during the processing of co-speech iconic gestures. In the following, an attempt is made towards such a model.

#### **5.3.1 Scope of the Model**

The model aims to explain how gestures interact with their co-expressive speech unit in comprehension. Hence, the focus is on the *local* integration of gesture and speech. More precisely, the theory aims to explain how the meaning of a gesture is integrated with the meaning of a single co-expressive word<sup>22</sup>, when preceded by a neutral sentence context.<sup>23</sup>

---

<sup>22</sup> Note that sometimes the speech units related to a gesture contain more than a single word (e.g., a complete verb phrase). However, because this would introduce syntax as an additional complexing factor into the model, this initial version of the model focuses on single words as the speech units to be integrated with gesture.

<sup>23</sup> Although the effect of a biasing preceding sentence context is not of primary interest in this initial version, it should be easy to implement it into the model, for instance by assuming that prior contextual information pre-activates the relevant (verbal or non-verbal) concept nodes (see Figure 5.1).

### 5.3 How Iconic Gestures Contribute to Language Comprehension: A Tentative Model 123

The model is in keeping with Occam's razor, in that it aims to account for the existing data on iconic gesture comprehension without postulating unnecessary theoretical entities. As will be outlined below, it makes – in contrast to the Growth Point Theory (McNeill, 1992, 2005) – some testable predication which can be subjected to a rigorous empirical test.

The theoretical considerations are based on an information processing approach. That is, it is assumed that the brain does its job by processing information (see also de Ruiter, 1998). Comprehending a co-speech gesture is considered to involve a sequence of processing steps. The processing steps operate upon internally stored retrievable information entities (the so-called representations).


The purpose of the model is first and foremost to specify the computations arguably involved in the comprehension of gestures and their interaction with their co-expressive speech on a *conceptual* level. Nonetheless, it is also important to make the link from such cognitive “boxologies” to the functional neuroanatomy of the brain. Therefore, after the architecture and the processing mechanisms of the model have been introduced, the potential neural correlates of the postulated processes are briefly discussed.

#### 5.3.2 Stimulus Example

In order to illustrate the processes and representations involved, one example of the present stimulus set is used. The chosen example describes the subordinate meaning of the homonym *mouse* (e.g., *computer device*, see Table 5.1). The gesture-speech synchrony described here is as it was naturally produced by the actress in the initial video recording session (for more details on the recording scenario, see section 2.6.2). The temporally overlapping segments of gesture stroke and speech, roughly corresponding to the utterance of the ambiguous word *Maus*, are marked by a dashed box (see Table 5.1). The model described in the next section focuses on the processing and integration of this isolated stroke-speech segment indicated by the dashed lines (i.e., the local integration of gesture and speech). In order to keep matters as

simple as possible for illustration purposes, it is initially assumed that there is no disambiguating gesture information before stroke onset.<sup>24</sup>

Table 5.1: The iconic co-speech gesture used to illustrate the model

Korinna streckte die Hand aus. Sie berührte die		Maus,	die den Computer steuerte.
			
Preparation / Pre-Stroke Hold	Stroke	Retraction	
Korinna reached out her hand. She touched the		Mouse,	which controlled the computer.

Sentence describing the subordinate word meaning of *mouse*. English translation in italics. The preparation phase started at the offset of the word *aus* and lasted approximately until the onset of the verb *berührte*, where a pre-stroke hold occurred. Following the pre-stroke hold, the stroke was performed simultaneously with the word *Maus*. The stroke consisted of two rapid downward movements of the index finger which re-enacted a double-clicking movement. The dashed box indicates the local integration between gesture and speech.

### 5.3.3 Model Architecture and Involved Processes

Figure 5.1 depicts how the model accounts for the processing of the example stroke-speech unit. Circles refer to representations, and oblique boxes refer to processes operating upon these representations. The assumed cognitive hierarchy is reflected in the Figure. It ranges from an initial low-level analysis of the physical input at the form level (both for gesture and speech), an intermediate lemma level (only for speech) to a higher conceptual level (again

<sup>24</sup> This simplification is of course incompatible with the findings of Experiments 4 and 5, which have shown there *is* disambiguating information before stroke onset. For an implementation of these “early” disambiguating effects of gesture into the model, see below.

both for gesture and speech). The additional intermediate lemma level for the processing of spoken words is needed, because unlike iconic gestures, the form of a spoken word does not resemble the denoted thing. In essence, the lemma serves to “fuse” together the arbitrarily combined form-meaning pairings of spoken words (for similar three-layered models of spoken language, see Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999).<sup>25</sup>

A major decision for any model on co-speech gesture comprehension is whether to assume one conceptual knowledge store or multiple stores. Some authors assume that there is only one conceptual store, where knowledge is stored in completely amodal and abstract verbal entities (e.g., in the form of propositions, Anderson, 1981). However, on the basis of evidence for a double-dissociation in the recognition of verbal and nonverbal stimuli in patients with left vs. right-hemispheric lesions (Coltheart, 1980; Fujii et al., 1990; Seliger et al., 1991; Warrington, 1982), researchers have argued for the existence of multiple conceptual stores (Coltheart et al., 1998; Sadoski & Paivio, 2001; Warrington & Crutch, 2004). Using a sample of healthy participants, a recent fMRI study by Thierry and Price (2006) also demonstrated a double dissociation between verbal and nonverbal processing at the conceptual level. Therefore, the present model borrows the idea from the Dual Coding Theory (DCT, Sadoski & Paivio, 2001) that there are two distinct sets of representations, one system specialized for language (i.e., the verbal system) and one system specialized for dealing with non-verbal objects, actions and events (i.e., the non-verbal system). Within the verbal system, concepts are defined through their set of connections to other related concept nodes. The activation of one concept node results in the activation of related concepts through spreading of activation.

---

<sup>25</sup> The models cited here are concerned with language production, rather than language comprehension. This is because (to the best of the author’s knowledge) there is no information processing model of spoken word comprehension that includes a separate conceptual level which is (arguably) needed to model the interaction between iconic gestures and speech.

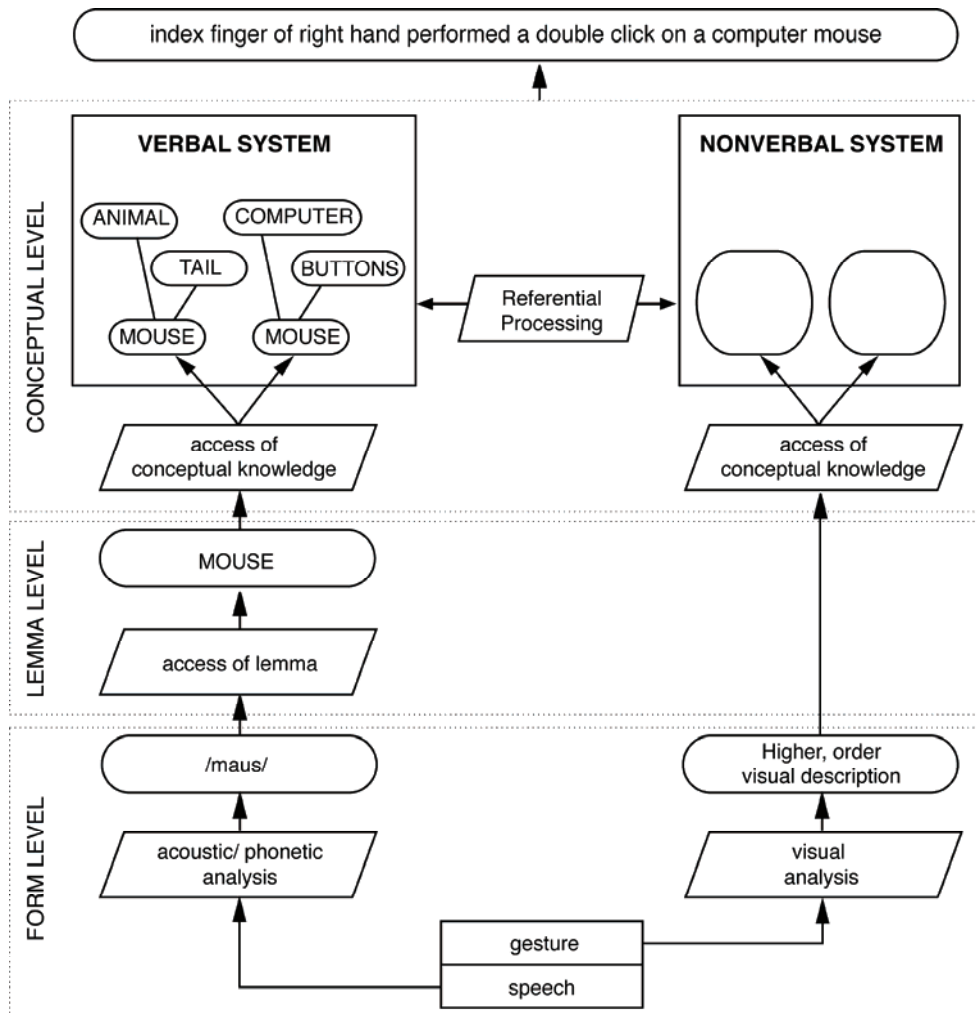


Figure 5.1: Illustration of the proposed model. The physical input (bottom-most box) consists of the spoken word *Maus*, the gesture is the double-clicking movement (see Table 5.1). For other details, see text.

#### Processing Within the Verbal System

Figure 5.1 illustrates how the processing of spoken words within the verbal system is conceptualized in the model. Initially, there is an acoustic-phonetic analysis of the input, resulting in an access of the phonetic form (i.e., the phoneme sequence /*maus*/). On the basis of this phonetic form, the lemma *Maus* can be accessed. Besides their important function as a hinge between phonetic form and conceptual meaning, it has been suggested that the nodes in the lemma level store abstract word information, such as gender and word category

information (Jescheniak, 2002). Because *Maus* is a homonym, this lemma is linked to two distinct concept nodes, one related to the animal meaning of *Maus*, the other related to the computer-device meaning of *Maus*.<sup>26</sup> Both word meanings of the homonym are initially activated (Onifer & Swinney, 1981; Swinney, 1979, 1991) and in the absence of other contextual information, word meaning frequency determines which word meaning is eventually selected (see Experiment 3, as well as Simpson, 1981; Simpson & Burgess, 1985; Simpson & Krueger, 1991; Vu et al., 1998). This frequency effect on meaning selection can be modelled by assuming a stronger connection weight between the lemma *Maus* and the concept node representing the dominant word meaning than between the lemma and the concept node representing the subordinate word meaning.

#### Differences Between the Verbal and the Non-Verbal System

Before giving a description of the processing steps specific to gestural information, it is important to acknowledge the differences between the way verbal and non-verbal conceptual knowledge is learned and arguably stored. In their description of the DCT (Sadoski & Paivio, 2001), the authors stress that language, and in particular spoken language, is of a sequential processing nature. In contrast, non-verbal information (e.g., objects, actions, gestures) induces more parallel processing strategies. For instance, a picture of an object conveys many information types simultaneously, such as size, color, texture, orientation and so on. Similarly, one iconic gesture can convey several information types simultaneously, such as the manner and the trajectory of a movement. Sadoski and Paivio (2001) argue that the differences in the way verbal and nonverbal information is conveyed are reflected in the representation format. Many types of linguistic information are stored in a sequential format, as for instance indicated by the poor performance in backward spelling. In contrast, non-verbal representations are less sequential, and often take the form of “nested sets”, where one representation (e.g., of an eye) can be a sub-part nested into a larger representation (e.g., of a face). Rather than being photo-realistic, non-verbal representations are suggested to take the form of templates or prototypes (Kiefer, 1999). The templates are specified only as far as they need to be (de Ruiter, 1998). For instance, in the action representation of “throwing a ball”,

---

<sup>26</sup> The captions of the concept nodes are in CAPITALS in order to underline the fact that these are semantic concepts and not word-like entities.

only the defining characteristics are fully specified in the template (e.g., a rapid open-handed forward movement), while number of hands (1 hand vs. 2 hands) or type of throwing (from the body center vs. over the head) are free parameters. These free parameters are set according to the concrete observed visual perception.

#### Processing Within the Non-Verbal System

The first stage in the processing of a gesture is an initial visual analysis of the physical input. The output of this process is a higher-order visual description. These higher-order visual descriptions are then used to access those representations in the non-verbal system, which trigger conditions are most closely matched by the observed visual input. There are at least three possible mechanisms through which gestures may access the non-verbal system. First, iconic gestures that are re-enactments of actions may activate the corresponding action schema. For instance, the observation of the example gesture might activate the action schema for *mouse clicking*. The suggested mechanism through which such action-related iconic gestures are able to access the corresponding action schema is through an internal simulation process (observation-execution matching, for a detailed description, see section 4.3). Second, iconic gestures primarily depicting salient features of objects may access the corresponding object representations. For instance, observing a speaker outlining a circle with his hands may entail the activation of all object representations that have *roundness* as a defining characteristic (e.g., *football*, *head*, *sun*). A third way through which gestures can have access to the non-verbal system is through direct access of an existing representation. Gestures that are very conventionalized (the emblems) are suggested to have such a direct representation in the non-verbal system. Note that otherwise it would be impossible to understand the meaning of emblems with little or opaque iconicity, such as the “rude finger”.

#### The Interaction Between the Verbal and the Non-Verbal System

According to the DCT (Sadoski & Paivio, 2001), the interaction between the verbal and the non-verbal system occurs via referential connections between related representations in the two systems. Like all other connections at the conceptual level, these referential connections are established through associative learning. For instance, repeatedly observing a thumbs-up hand posture accompanied by the spoken utterance *Well done* results in the creation of a referential connection between the corresponding verbal and non-verbal concept nodes. An important feature of these referential connections is their bidirectional nature: After the

connection has been formed, processing the verbal phrase in isolation automatically results in the activation of the emblem representation through spreading of activation. Conversely, processing the thumbs-up posture in the absence of a speech context also activates the corresponding verbal concept node.

Sadoski and Paivio (2001) also stress that verbal and non-verbal representations are not connected in a one-to-one fashion but rather in a one-to-many fashion. That is, one verbal concept may be referentially connected to many non-verbal representations, and one non-verbal representation to many verbal representations.

During the processing of the spoken word *mouse* and the accompanying double-clicking gesture, all of the verbal and non-verbal representations whose trigger conditions are fulfilled will initially be activated. That is, processing the spoken word *Maus* will activate both the verbal concept of COMPUTER(MOUSE) and the concept of ANIMAL(MOUSE). Similarly, the processing of a gesture will activate the action schema *mouse-clicking*, as well as other potentially appropriate non-verbal representations (e.g., the emblem *impatience*, expressed by repeatedly tapping with the fingers). Note, however, that the verbal concepts (COMPUTER)MOUSE and BUTTONS as well as the non-verbal representation *mouse-clicking* will eventually receive the strongest activation, because of their additional referential connections. The nodes which activation value surpasses a certain threshold after a certain amount of time will be selected as being the most likely representation of the meaning of the observed audiovisual stimulus. The combined meaning of these activated nodes is then integrated into the sentence context. For instance, the combined meaning of gesture and speech in the example might be that the subject touched a computer mouse with the index finger of the right hand, and the touching was actually a double-clicking movement.

#### 5.3.4 How the Model Accounts for the Existing Data

In Experiments 1 – 3 of the present dissertation, it was shown that listeners use gestural information to disambiguate speech. According to the proposed model, this occurs because the verbal concept node representing the appropriate word meaning of the homonym receives additional activation via referential connections with the non-verbal system. On a more general level, the nature of the interaction between gesture and speech can be described as process of mutual disambiguation. Not only can gesture have a disambiguating influence on speech (as in the present dissertation), but speech has also a disambiguating effect on gesture.

Iconic gestures are not considered to have consistent form-meaning pairings (see also Cassell et al., 1999). Instead, one observed iconic gesture initially activates an array of possible meanings. The appropriate gesture meaning is selected in interaction with the speech context. This bidirectional influence between the verbal and the non-verbal system is realized through the referential connections.

On the basis of findings that gestures can convey additional information that is not found in speech (e.g., Beattie & Shovelton, 1999b), researchers have argued that listeners integrate the information of gesture and speech into one unified, enriched representation (Alibali et al., 1997; McNeill et al., 1994). This “enrichment” through gesture is explained by the model through the assumption that the *integral* of activated concept nodes (across both the verbal and the non-verbal system) is integrated into the sentence context, not only the most strongly activated verbal concept.

Another consistent finding in behavioral research on gesture comprehension is that there is large variability in the meaning addresses attribute to decontextualized iconic gestures (Feyereisen et al., 1988; Hadar & Pinchas-Zamir, 2004; Krauss et al., 1991). This can be explained straightforward by the model. In this case, speech is not present, therefore the verbal system cannot exert its disambiguating influence on the non-verbal system via the referential connections.

Incompatible combinations of gesture and speech have an interfering effect on language comprehension, as demonstrated in a number of behavioral (Cassell et al., 1999; McNeill et al., 1994) and ERP studies (Kelly et al., 2004; Kelly et al., 2007; Özyürek et al., 2007; Wu, 2005; Wu & Coulson, 2007). An example of such a gesture-speech mismatch would be gesturing *knock* while saying *write*. In this case, the processing of the auditory input would entail the activation of the corresponding phonetic form, the lemma and finally the concept WRITE and other associated verbal concepts (LETTER, COMPUTER,...). Processing the gesture would result in the activation of the action schema *knocking*. Because this action schema has also associatively learned referential connections with verbal concepts (e.g., DOOR, KNUCKLES,...), a second cluster of verbal concepts becomes activated. Hence, the processing of a gesture-speech mismatch resulted in the activation of two unrelated clusters of verbal concepts, which of course entails same processing difficulties within the verbal system that are reflected in the interference effect. Because the referential connections between the

verbal and the non-verbal are assumed to be bidirectional, a gesture-speech mismatch should also elicit a detectable interference effect in the non-verbal system. However, to the best of the author's knowledge, such a study has not been conducted yet.

The results of Experiments 3 as well as the study by Kelly et al. (2007) suggest that gesture-speech integration is not an entirely automatic process, but also involves a considerable degree of controlled processes. With respect to the proposed model, this means that the connection weight of the referential connections can be modulated by situational factors, such as the amount of observed meaningful hand movements (see Experiment 3) and the "intentional coupling" of gesture and speech (see Kelly et al., 2007).

Finally, the results from Experiments 4 and 5 indicate that listeners do not have to see a complete gesture before it is recognized as being meaningful. This suggests that the proposed processing stages of gesture comprehension do not follow each other in a discrete serial fashion, i.e., it is not the case that the visual analysis of a gesture has to be completed before access of the conceptual knowledge begins. Instead, it is proposed that processing occurs in a serial-cascading fashion (see also van den Brink, Brown, & Hagoort, 2006). The onset of lower-level processes precedes the onset of higher-level processes, but the onset of a higher level process does not depend on the completion of a lower level process. In analogy to a popular theory on spoken word comprehension (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987), it may be fruitful to think of gesture comprehension in terms of a gesture cohort. All non-verbal representations that are positionally compatible with the gestural input are activated (the gesture cohort). As more input is being processed, more and more interpretations can be discarded, until a final (and relatively small) set of possible interpretations remains. A preceding context may serve to pre-activate certain conceptual nodes, and thereby accelerate the comprehension process.

### 5.3.5 Some Testable Predictions

One prediction of the model is that the access of gesture is exhaustive, i.e., all meanings compatible with a gestural input should be initially activated, regardless of the accompanying speech unit. Cross-modal priming experiments may serve to test this issue. Extending the idea of exhaustive gestural access to the temporal domain, the model predicts that some form of a gesture cohort should exist.

Another test that could be performed concerns the suggested indirect interaction between the verbal and non-verbal system. It is assumed that a spoken word can access the non-verbal system only indirectly, after the corresponding verbal concept(s) have been accessed. Conversely, a gesture can also access the verbal system only in an indirect way, after the corresponding non-verbal representations have been accessed. A task that taps into processes at the conceptual level in the verbal and non-verbal system could be used to test this hypothesis. For instance, participants could perform a lexical decision task to spoken words, which are either accompanied or not by a gesture. Similarly, they could perform a “gestural decision task” (i.e., is the observed hand movement meaningful or not) to gestures, which are either accompanied by spoken words or not. If it turns out that in both situations, the reaction times are shorter in the bimodal than in the unimodal stimulation, the claim about the indirect interaction between the two systems is falsified.

Finally, the model does not assume top-down modulation, i.e., the output of higher-level processes is not fed back to lower-level processes. Possible ways to test this hypothesis would be to use tasks that arguably tap into lower level processes (e.g., phoneme decision task) and see whether the response is modulated by the presence vs. absence of a gesture.

### **5.3.6 Potential Neural Correlates**

As has been stated in the introduction to this section, the main purpose of suggesting a model was to spell out the putative processes involved in the processing of co-speech iconic gestures. Nonetheless, it is important to see whether the processes suggested in the model are compatible with what is known about the processing mechanisms of the brain. Because the potential functional significance of the brain areas identified in Experiment 6 has already been discussed in detail above, the interpretations are here only briefly recapitulated and translated into the context of the model.

The inferior frontal gyrus (including the ventral premotor cortex) and the inferior parietal lobule are suggested to be involved in retrieving the appropriate meaning of observed gestures, through an observation-execution matching process. The input of this simulation process is a higher-order visual description; the output is an array of activated action schemas. An open question is whether all gestures are processed via such a sensorimotor mechanism, or whether it is specific to action-related iconic gestures.

The posterior STS might be involved in the processes mediating between the verbal and the non-verbal system. In particular, activation flow between connected verbal and non-verbal representations is suggested to yield increased activation in the posterior STS. Future research should elucidate whether this brain area is also involved in the formation of new connections between verbal and non-verbal representations (for instance, when learning new words with the help of gesture, or when learning new movements with the help of verbal instruction). Note that there is some initial evidence (Meyer, Baumann, Marchina, & Jancke, 2007) suggesting that this brain area is not only involved in the retrieval of existing audiovisual associations, but also in the formation of new associations.

With respect to the neural correlates of the processes within the verbal system, Heim (2005) suggested in his review that middle temporal and inferior temporal areas are involved in storing the semantic aspects of the mental lexicon, whereas left inferior frontal areas are more involved in the selection of information from the lexicon. If the posterior STS is indeed providing the link between verbal and non-verbal information, this brain area should show an increased functional connectivity to areas storing the semantic aspects of language. Future experiments with a factorial manipulation of the relationship between gesture and speech should be able to shed further light on this issue.

## **5.4 Concluding Remarks**

In sum, the present dissertation added to the existing evidence in showing that listeners benefit from the additional information conveyed by iconic gestures in comprehension. It extends the existing literature, because it was not only shown that iconic gestures can facilitate language comprehension, but also identified one mechanism through which gesture can impact online language comprehension, namely through the disambiguation of ambiguous speech. Contrary to the existing beliefs (McNeill et al., 1994), Experiment 3 suggests that the processes underlying iconic gesture comprehension are not entirely automatic, but do also involve a considerable degree of controlled processes.

In Experiment 4, the earliest point in time at which the iconic gesture start to exert their disambiguating influence (the so-called disambiguation point) was determined in a modified gating paradigm. The disambiguation points were validated in a co-speech context in Experiment 5. The main result of these two Experiments was that almost two thirds of all gestures enabled a meaning selection before participants had seen the stroke of the gesture.

This theoretical significance of this finding can, however, not be evaluated yet, because there is so far no information processing model of iconic gesture comprehension, let alone one that predicts how gestures acquire meaning over time. The general discussion tried identify some of the main points that such a model has to account for, and proposed an initial version.

Finally, Experiment 6 explored what brain regions are involved in the processing of co-speech gestures. In this experiment, it was found that the processing of co-speech iconic gestures as compared to grooming elicited greater levels in the left posterior STS, the ventral premotor cortex bilaterally as well as the inferior parietal lobule. The potential functional significance of these activations was discussed in reference to current hypotheses about how action and language interact at the level of brain activity.

Admittedly, this dissertation has raised probably more questions than it has answered. It is, however, only fair to note that the field of gesture comprehension has just recently started to attract the interest from empirically motivated researchers. The previous neglect is reflected in the fact there is yet no information processing model in the field allowing for some testable predictions. Future research should therefore aim to advance the understanding of iconic gesture comprehension both on a theoretical as well as on a neurocognitive level.

## List of Figures

Figure 1.1:	Gesture phases of the „and he bends it way back“ gesture..	6
Figure 2.1:	Schematic drawing of an extended 10-20 system montage.	15
Figure 2.2:	Accuracy Data for Experiment 1..	31
Figure 2.3:	ERP Data for Experiment 1..	32
Figure 2.4:	ERP Data for Experiment 2.....	36
Figure 2.5:	ERP Data for Experiment 3.....	42
Figure 3.1:	Results of Experiment 4. ....	60
Figure 3.2:	Results of Experiment 5 .....	63
Figure 3.3:	Example of “early” disambiguation .....	65
Figure 4.1:	The basic principle of MR signal generation .....	71
Figure 4.2:	Spin nutation .....	74
Figure 4.3:	EPI-BOLD sequence .....	79
Figure 4.4:	Accuracy Data for Experiment 6.....	95
Figure 4.5:	Reaction Time Data for Experiment 6. ....	96
Figure 4.6:	Imaging Data for Experiment 6: Dominant vs. Grooming.....	97
Figure 4.7:	Imaging Data for Experiment 6: Subordinate vs. Grooming. ....	98
Figure 4.8:	Imaging Data for Experiment 6: Dominant vs. Subordinate.....	99
Figure 5.1:	Illustration of the proposed model. ....	126



## List of Tables

Table 2.1:	Experiment 1 Stimulus Examples. ....	26
Table 2.2:	Experiment 1 Stimulus Properties. ....	28
Table 2.3:	Examples of additional stimuli used in Experiment 3.....	41
Table 2.4:	Properties of additional stimuli in Experiment 3. ....	41
Table 4.1:	Stimulus example for Experiment 6.....	90
Table 4.2:	List of significantly activated regions in Experiment 6 .....	100
Table 5.1:	The iconic co-speech gesture used to illustrate the model .....	124



## References

- Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture - do teachers have the upper hand. *Journal of Educational Psychology*, 89(1), 183-193.
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4(7), 267-278.
- American Electroencephalographic Society: Guidelines for standard electrode position nomenclature. (1991). *Journal of Clinical Neurophysiology*, 8, 200-202.
- Anderson, J. R. (1981). Concepts, propositions, and schemata: What are the cognitive units? In J. Flowers (Ed.), *Nebraska Symposium on Motivation*. Lincoln, Nebraska: University of Nebraska Press.
- Arciuli, J., & Cupples, L. (2004). Effects of stress typicality during spoken word recognition by native and nonnative speakers of English: Evidence from onset gating. *Memory & Cognition*, 32(1), 21-30.
- 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.
- Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, 7(2), 85-97.
- Banati, R. B., Goerres, G. W., Tjoa, C., Aggleton, J. P., & Grasby, P. (2000). The functional anatomy of visual-tactile integration in man: a study using positron emission tomography. *Neuropsychologia*, 38(2), 115-124.
- Barber, H., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: evidence from ERPs. *Neuroreport*, 15(3), 545-548.
- Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123(1-2), 1-30.
- Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language & Social Psychology*, 18(4), 438-462.
- Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication. *Gesture*, 1(2).
- Beattie, G., & Shovelton, H. (2002a). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*, 93(Part 2), 179-192.
- Beattie, G., & Shovelton, H. (2002b). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology*, 41(Part 3), 403-417.
- Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15(2), 145-153.

- 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., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004b). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Supplementary Online Material. *Nature Neuroscience*, 7(11), 1190-1192.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5), 809-823.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57(1), 289-300.
- Binkofski, F., & Buccino, G. (2006). The role of ventral premotor cortex in action execution and action understanding. *Journal of Physiology-Paris*, 99(4-6), 396-405.
- Brass, M., Derrfuss, J., Forstmann, B., & Cramon, D. Y. v. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences*, 9(7), 314-316.
- Buxton, R. B., Uludag, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, 23 Suppl 1, S220-233.
- Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine*, 39(6), 855-864.
- 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.
- 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., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*, 14(2), 427-438.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology, Paris*, 98(1-3), 191-205.
- Cassell, J., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition*, 7(1), 1-33.
- Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1), 8-16.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. London: MIT Press.
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3), 274-285.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior* Vol 12(4) Aug 1973, 335-359.

- Coleman, J. S., & Keith, B. (2006). Design Features of Language. In *Encyclopedia of Language & Linguistics* (pp. 471-475). Oxford: Elsevier.
- Coltheart, M. (1980). Deep dyslexia: A right hemisphere hypothesis. In M. Coltheart, K. Patterson & J. C. Marshall (Eds.), *Deep dyslexia*. London: Routledge & Kegan Paul.
- Coltheart, M., Inglis, L., Cupples, L., Michie, P., Bates, A., & Budd, B. (1998). A semantic subsystem of visual attributes. *Neurocase*, 4(4-5), 353-370.
- Craig, C. H., & Kim, B. W. (1990). Effects of time gating and word-length on isolated word-recognition performance. *Journal of Speech and Hearing Research*, 33(4), 808-815.
- Craig, C. H., Kim, B. W., Rhyner, P. M. P., & Chirillo, T. K. B. (1993). Effects of word predictability, child-development, and aging on time-gated speech recognition performance. *Journal of Speech and Hearing Research*, 36(4), 832-841.
- de Ruiter, J. P. (1998). *Gesture and Speech Production (Ph.D. thesis)*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4), 801-838.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118.
- 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.
- Dupont, P., De Bruyn, B., Vandenberghe, R., Rosier, A. M., Michiels, J., Marchal, G., et al. (1997). The kinetic occipital region in human visual cortex. *Cerebral Cortex*, 7(3), 283-292.
- Ekman, P. (1999). Emotional and conversational nonverbal signals. In L. Messing & R. Campbell (Eds.), *Gesture, Speech and Sign* (pp. 45 - 55). London: Oxford University Press.
- Elston-Guttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory & Language*, 52(2), 256-283.
- Emmorey, K., & Corina, D. (1990). Lexical recognition in sign language: effects of phonetic structure and morphology. *Perceptual and Motor Skills*, 71(3 Pt 2), 1227-1252.
- Feyereisen, P., & de Lannoy, J. D. (1991). *Gestures and speech: Psychological investigations*. New York, NY: Cambridge University Press.
- 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.
- Fiebach, C. J., Friederici, A. D., Muller, K., & von Cramon, D. Y. (2002). fMRI Evidence for Dual Routes to the Mental Lexicon in Visual Word Recognition. *Journal of Cognitive Neuroscience*, 14(1), 11-23.
- Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron*, 24(1), 205-218.

- 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 in Medicine*, 33(5), 636-647.
- Fox, P. T., Raichle, M. E., Mintun, M. A., & Dence, C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, 241(4864), 462-464.
- Fridman, E. A., Immisch, I., Hanakawa, T., Bohlhalter, S., Waldvogel, D., Kansaku, K., et al. (2006). The role of the dorsal stream for gesture production. *Neuroimage*, 29(2), 417-428.
- Friederici, A. D. (2004). Event-related brain potential studies in language. *Current Neurology and Neuroscience Reports*, 4(6), 466-470.
- 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.
- Fujii, T., Fukatsu, R., Watabe, S. I., Ohnuma, A., Teramura, K., Kimura, I., et al. (1990). Auditory sound agnosia without aphasia following a right temporal lobe lesion. *Cortex*, 26(2), 263-268.
- Fuster, J. M. (2003). *Cortex and mind*. New York: Oxford University Press.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(Part 2), 593-609.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action*. New York: Oxford University Press.
- 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.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language & Cognitive Processes*, 12, 613-656.
- Gauthier, I., & Bukach, C. (in press). Should we reject the expertise hypothesis? *Cognition*.
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191-197.
- Glenberg, A. M. (2006). Naturalizing cognition: the integration of cognitive science and biology. *Current Biology*, 16(18), R802-804.
- Glenberg, A. M. (in press). Language and action: Creating sensible combinations of ideas. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.
- Goldin-Meadow, S. (2003). *Hearing Gesture - How Our Hands Help Us Think*. Cambridge, Massachusetts, and London, England: The Belknap Press of Harvard University Press.
- 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.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.

- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, 13(2), 159-166.
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27, 649-677.
- Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, 16(2), 240-246.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267-283.
- Grosjean, F. (1996). Gating. *Language & Cognitive Processes*, 11(6), 597-604.
- Gunter, T. C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters*, 372(1-2), 52-56.
- Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: a difficult case for activation theories. *Journal of Cognitive Neuroscience*, 15(5), 643-657.
- Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture - Implications for gesture classification and function. *Journal of Language & Social Psychology*, 23(2), 204-214.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11(2), 194-205.
- Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, 26(4), 579-596.
- Heim, S. (2005). The structure and dynamics of normal language processing: insights from neuroimaging. *Acta Neurobiologiae Experimentalis (Wars)*, 65(1), 95-116.
- Hikosaka, K., Iwai, E., Saito, H., & Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *Journal of Neurophysiology*, 60(5), 1615-1637.
- Hockett, C. (1960). The origins of speech. *Scientific American*, 203, 89-96.
- Holcomb, P. J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain and Language*, 35(1), 66-85.
- Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, 3(2), 127-154.
- Hoshi, E. (2006). Functional specialization within the dorsolateral prefrontal cortex: A review of anatomical and physiological studies of non-human primates. *Neuroscience Research*, 54(2), 73-84.
- Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sunderland: Sinauer.
- Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15(6), 632-637.
- Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, 7(12), 942-951.

- Iacoboni, M., & Wilson, S. M. (2006). Beyond a single area: Motor control and language within a neural architecture encompassing Broca's area. *Cortex*, 42(4), 503-506.
- Iadecola, C. (2004). Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nature Reviews Neuroscience*, 5(5), 347-360.
- Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396(6708), 228.
- Jansen, E., & Povel, D. J. (2004). The processing of chords in tonal melodic sequences. *Journal of New Music Research*, 33(1), 31-48.
- Jescheniak, J. D. (2002). *Sprachproduktion: Der Zugriff auf das lexikale Gedächtnis beim Sprechen*. Göttingen: Hogrefe.
- Jezzard, P., Matthews, P. M., & Smith, S. M. (2001). *Functional MRI: An Introduction to Methods*. Oxford: Oxford University Press.
- Josephs, O., Turner, R., & Friston, K. (1997). Event-related fmri. *Human Brain Mapping*, 5(4), 243-248.
- Joubert, S., Beaugard, M., Walter, N., Bourgouin, P., Beaudoin, G., Leroux, J. M., et al. (2004). Neural correlates of lexical and sublexical processes in reading. *Brain & Language*, 89(1), 9-20.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 17(11), 4302-4311.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain & Language*, 89(1), 253-260.
- Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain & Language*, 101(3), 222-233.
- Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177 - 210). New York: Pergamon Press.
- Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207 - 227). The Hague: Mouton and Co.
- Kendon, A. (1981). Introduction: current issues in the study of "nonverbal communication". In A. Kendon (Ed.), *Nonverbal communication interaction and gesture* (pp. 1 - 53). The Hague: Mouton and Co.
- Kendon, A. (1997). Gesture. *Annual Review of Anthropology*, 26, 109 - 128.
- Kiefer, M. (1999). *Die Organisation des semantischen Gedächtnisses*. Bern: Huber.
- 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.
- Koyama, S., Nageishi, Y., & Shimokochi, M. (1992). Effects of semantic context and event-related potentials: N400 correlates with inhibition effect. *Brain and Language*, 43(4), 668-681.

- Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 389 - 450). New York: Academic Press.
- Krauss, R. M., Dushay, R. A., Chen, Y. S., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, 31(6), 533-552.
- Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do Conversational Hand Gestures Communicate? *Journal of Personality & Social Psychology* November, 61(5), 743-754.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463-470.
- 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., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2 ed., pp. 659-724). New York: Elsevier.
- Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, 166(3-4), 289-297.
- Lee, J. H., Garwood, M., Menon, R., Adrian, G., Andersen, P., Truwit, C. L., et al. (1995). High contrast and fast three-dimensional magnetic resonance imaging at high fields. *Magnetic Resonance in Medicine*, 34(3), 308-312.
- Levelt, W. J. M., Richardson, G., & la Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24(2), 133-164.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38; discussion 38-75.
- Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, 10(9), 873-888.
- Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al. (2001). LIPSIA - a new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging & Graphics*, 25(6), 449-457.
- MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., et al. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *Neuroimage*, 22(4), 1605-1618.
- Magistretti, P. J., & Pellerin, L. (1999). Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philos Trans R Soc Lond B Biol Sci*, 354(1387), 1155-1163.
- Mandeville, J. B., Marota, J. J., Ayata, C., Zaharchuk, G., Moskowitz, M. A., Rosen, B. R., et al. (1999). Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J Cereb Blood Flow Metab*, 19(6), 679-689.

- Mansfield, P., & Maudsley, A. A. (1977). Medical imaging by NMR. *British Journal of Radiology*, 50(591), 188-194.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71-102.
- Martin, C., Vu, H., Kellas, G., & Metcalf, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *Quarterly Journal of Experimental Psychology A*, 52(4), 813-839.
- Matthews, P. M. (2001). An Introduction to fMRI of the brain. In P. Jezzard, P. M. Matthews & S. M. Smith (Eds.), *Functional MRI: An Introduction to Methods*. Oxford: Oxford University Press.
- McAllister, J. M. (1988). The use of context in auditory word recognition. *Perception & Psychophysics*, 44(1), 94-97.
- 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, Illinois, and London, England: The University of Chicago Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago and London: University of Chicago Press.
- McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, 27(3), 223-237.
- Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol*, 28(5), 597-613.
- Mesulam, M. M. (1998). From sensation to cognition. *Brain*, 121 ( Pt 6), 1013-1052.
- Meyer, M., Baumann, S., Marchina, S., & Jancke, L. (2007). Hemodynamic responses in human multisensory and auditory association cortex to purely visual stimulation. *BMC Neuroscience*, 8, 14.
- Molnar-Szakacs, I., Kaplan, J., Greenfield, P. M., & Iacoboni, M. (2006). Observing complex action sequences: The role of the fronto-parietal mirror neuron system. *NeuroImage*, 33(3), 923-935.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18(3), 615-622.
- Morris, D. (1979). *Gestures: Their Origins and Distributions*. London: Cape.
- Nair, D. G. (2005). About being BOLD. *Brain Res Brain Res Rev*, 50(2), 229-243.
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The Inferior Frontal Gyrus and Phonological Processing: An Investigation using rTMS. *Journal of Cognitive Neuroscience*, 16(2), 289-300.
- Norris, D. G. (2000). Reduced power multislice MDEFT imaging. *JMRI-Journal of Magnetic Resonance Imaging*, 11(4), 445-451.
- Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med*, 14(1), 68-78.

- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *Vol.*, 9(1), 97-113.
- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225-236.
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2006). Conceptual priming for environmental sounds and words: an ERP study. *Brain Cogn*, 62(3), 267-272.
- 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.
- Paul, S. T., Kellas, G., Martin, M., & Clark, M. B. (1992). Influence of contextual features on the activation of ambiguous word meanings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(4), 703-717.
- Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *Journal of Neuroscience*, 23(17), 6819-6825.
- Penny, W. (2006). fMRI model specification. In T. F. M. Group (Ed.), *SPM5 Manual*. London: Functional Imaging Laboratory (<http://www.fil.ion.ucl.ac.uk/spm/>).
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.
- Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond B Biol Sci*, 358(1431), 435-445.
- Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17(6), 884-892.
- 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.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, 12(2), 149-154.
- Rugg, M. D., & Coles, M. G. H. (1995). Event-related brain potentials: an introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford: Oxford University Press.
- Rüschemeyer, S. A., Brass, M., & Friederici, A. (2007). Comprehending Prehending: Neural Correlates of Processing Verbs with Motor Stems. *Journal of Cognitive Neuroscience*, 19(5), 855-865.
- Sadoski, M., & Paivio, A. (2001). *Imagery and Text: A Dual coding Theory of Reading and Writing*. London: Lawrence Erlbaum Associates.
- Saygin, A. P., Dick, F., Wilson, S. M., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds: evidence from aphasia. *Brain*, 126(Pt 4), 928-945.

- Schneider, W., & Shiffrin, R. M. (1977). Controlled and Automatic Human Information-Processing .1. Detection, Search, and Attention. *Psychological Review*, 84(1), 1-66.
- Schubotz, R. I. (2004). *Human premotor cortex: Beyond motor performance*. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.
- Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology-Human Perception and Performance*, 29(6), 1155-1178.
- Schupp, H. T., Stockburger, J., Bublatzky, F., Junghofer, M., Weike, A. I., & Hamm, A. O. (2007). Explicit attention interferes with selective emotion processing in human extrastriate cortex. *BMC Neurosci*, 8, 16.
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3), 277-287.
- Seliger, G. M., Lefever, F., Lukas, R., Chen, J., Schwartz, S., Codeghini, L., et al. (1991). Word deafness in head injury: Implications for coma assessment and rehabilitation. *Brain Injury*, 5, 53-56.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and Automatic Human Information-Processing .2. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review*, 84(2), 127-190.
- Shmuel, A., Augath, M., Oeltermann, A., & Logothetis, N. K. (2006). Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nature Neuroscience*, 9(4), 569-577.
- Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P.-F., Adriany, G., Hu, X., et al. (2002). Sustained Negative BOLD, Blood Flow and Oxygen Consumption Response and Its Coupling to the Positive Response in the Human Brain. *Neuron*, 36(6), 1195-1210.
- Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning & Verbal Behavior*, 20(1), 120-136.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 28-39.
- Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, 30(6), 627-643.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, 25(1), 76-89.
- Song, A. W., Huettel, S. A., & McCarthy, G. (2006). Functional Neuroimaging: Basic Principles of Functional MRI. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional Neuroimaging of Cognition*. Cambridge, MA: MIT press.
- Stefanovic, B., Warnking, J. M., & Pike, G. B. (2004). Hemodynamic and metabolic responses to neuronal inhibition. *NeuroImage*, 22(2), 771-778.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, Massachusetts: MIT press.

- Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The Visually Responsive Neuron and Beyond - Multisensory Integration in Cat and Monkey. *Progress in Brain Research*, 95, 79-90.
- Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain & Language*, 86(2), 326-343.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior*. Vol, 18(6), 645-659.
- Swinney, D. A. (1991). The Resolution of Interdeterminacy During Language Comprehension: Perspectives on Modularity in Lexical, Structural and Pragmatic Process. In G. B. Simpson (Ed.), *Understanding Word and Sentence*. Amsterdam; New York: North Holland Elsevier Science Publ.
- Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, 27(3), 324-340.
- Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, 49(2-3), 161-167.
- Talairach, J., & Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging*. New York, NY: Thieme Medical Publishers.
- Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8), 764-769.
- Thierry, G., & Price, C. J. (2006). Dissociating verbal and nonverbal conceptual processing in the human brain. *Journal of Cognitive Neuroscience*, 18(6), 1018-1028.
- Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review*, 7(1), 49-82.
- Tyler, L. K., & Wessels, J. (1985). Is gating an online task - evidence from naming latency data. *Perception & Psychophysics*, 38(3), 217-222.
- Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et al. (2001). I know what you are doing. a neurophysiological study. *Neuron*, 31(1), 155-165.
- 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 den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32(2), 364-372.
- Van Petten, C. K., & Kutas, M. (1987). Ambiguous words in context: An event-related potential analysis of the time course of meaning activation. *Journal of Memory and Language*, 26(2), 188-208.
- Van Petten, C. K., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang*, 97(3), 279-293.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.
- Vu, H., Kellas, G., & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory & Cognition*, 26(5), 979-1001.

- Wagner, S. (2003). *Verbales Arbeitsgedächtnis und die Verarbeitung lexikalisch ambiger Wörter in Wort- und Satzkontexten (Ph.D. thesis)*. Leipzig: Max-Planck-Institute for Cognitive Neuroscience.
- Walley, A. C., Michela, V. L., & Wood, D. R. (1995). The gating paradigm - effects of presentation format on spoken word recognition by children and adults. *Perception & Psychophysics*, 57(3), 343-351.
- Ward, B. D. (2000). AlphaSim - Simultaneous Inference for FMRI Data. Milwaukee, WI: Biophysics Research Institute, Medical College of Wisconsin.
- Warrington, E. K. (1982). Neuropsychological studies of object recognition. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 298(1089), 15-33.
- Warrington, E. K., & Crutch, S. J. (2004). A circumscribed refractory access disorder: A verbal semantic impairment sparing visual semantics. *Cognitive Neuropsychology*, 21(2-4), 299-315.
- West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Brain Res Cogn Brain Res*, 13(3), 363-375.
- Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain & Language*.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2006). When Language Meets Action: The Neural Integration of Gesture and Speech. *Cerebral Cortex*, bhl141.
- Wilson, M. (2002). Six views of embodied cognition. *Psychon Bull Rev*, 9(4), 625-636.
- Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., et al. (2002). A general statistical analysis for fMRI data. *Neuroimage*, 15(1), 1-15.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4, 58 - 73.
- Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, 13(10), 1034-1043.
- Wu, Y. C. (2005). Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures. *The newsletter of the Center for Research in Language*, 17(2).
- Wu, Y. C. (2006). *Coordinated Minds: How Iconic Co-speech Gestures Mediate Communication (Ph.D. thesis)*. San Diego, CA: University of San Diego.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654-667.
- Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain & Language*, 101(3), 234-45.

## Appendix A: Sentence Materials

### Experiments 1 – 3

Fritz hatte sich schon um eine Stunde verspätet. Er vollendete den **Absatz**, damit der **Text** endlich abgeschickt werden konnte.

Fritz hatte sich schon um eine Stunde verspätet. Er vollendete den **Absatz**, damit der **Schuh** endlich ausgeliefert werden konnte.

Michaela war beschäftigt. Sie bearbeitete den **Anbau**, weil beim **Haus** dringend der Putz erneuert werden musste.

Michaela war beschäftigt. Sie bearbeitete den **Anbau**, weil beim **Reis** die Erntezeit angebrochen war.

Paul hatte alle überrascht. Er sorgte für einen **Auflauf**, weil die **Sensation** sich in Windeseile herumgesprochen hatte.

Paul hatte alle überrascht. Er sorgte für einen **Auflauf**, weil die **Nudeln** dringend verwertet werden mussten.

Veronika sorgte für die nötigen Änderungen. Sie passte den **Aufsatz** an, damit beim **Heft** das Layout stimmte.

Veronika sorgte für die nötigen Änderungen. Sie passte den **Aufsatz** an, damit beim **Schrank** nichts klemmte.

Manuela musste bei den beiden aufpassen. Sie kontrollierte die **Aussprache**, um den **Streit** zu vermeiden.

Manuela musste bei den beiden aufpassen. Sie kontrollierte die **Aussprache**, um den **Dialekt** zu verbergen.

Alle waren von Sandra beeindruckt. Sie beherrschte den **Ball**, was sich im **Spiel** beim Aufschlag deutlich zeigte.

Alle waren von Sandra beeindruckt. Sie beherrschte den **Ball**, was sich im **Tanz** mit dem Bräutigam deutlich zeigte.

Alle Augen waren auf Tim gerichtet. Er beschrieb einen **Bogen**, welcher der **Kurve** ungefähr folgte.

Alle Augen waren auf Tim gerichtet. Er beschrieb einen **Bogen**, welcher dem **Pfeil** angemessen war.

Dietmar hatte ein klares Ziel. Er imitierte den **Boxer**, um den **Sport** lächerlich zu machen.

Dietmar hatte ein klares Ziel. Er imitierte den **Boxer**, um den **Hund** lächerlich zu machen.

Karl war mit der Bestellung zufrieden. Ihm gefiel die **Brause**, weil die **Cola** im Vergleich zu süß war.

Karl war mit der Bestellung zufrieden. Ihm gefiel die **Brause**, weil die **Dusche** einen Massagestrahl hatte.

Petra war nicht ganz bei der Sache. Sie entdeckte die **Bremse**, als das **Fahrrad** schon auf den Abhang zurollte.

Petra war nicht ganz bei der Sache. Sie entdeckte die **Bremse**, als das **Insekt** schon auf ihrer Schulter saß.

Sonja musste es ihrem Kollegen deutlich machen. Sie zeigte den **Eingang**, weil die **Tore** alle gleich aussahen.  
Sonja musste es ihrem Kollegen deutlich machen. Sie zeigte den **Eingang**, weil die **Briefe** sich auf ihrem Schreibtisch stapelten.

Bernd fiel etwas auf. Er bemerkte die **Fahne**, die dem **Staat** viel Geld gekostet haben musste.  
Bernd fiel etwas auf. Er bemerkte die **Fahne**, die dem **Bier** geschuldet war.

Thomas musste den Job zuende bringen. Er arbeitete an der **Fassung**, die für den **Artikel** vorgesehen war.  
Thomas musste den Job zuende bringen. Er arbeitete an der **Fassung**, die für die **Lampe** vorgesehen war.

Kerstin machte ihre Arbeit gründlich. Sie prüfte die **Feder**, weil der **Vogel** für den Export vorgesehen war.  
Kerstin machte ihre Arbeit gründlich. Sie prüfte die **Feder**, weil der **Hebel** defekt war.

Hubert war total genervt. Er beseitigte die **Fliege**, die ihn wie die **Mücke** um den Schlaf brachte  
Hubert war total genervt. Er beseitigte die **Fliege**, die ihn wie die **Krawatte** am Hals würgte.

Sebastian war beeindruckt. Er staunte über den **Flügel**, der dem **Klavier** überlegen war.  
Sebastian war beeindruckt. Er staunte über den **Flügel**, der dem **Papagei** etwas Exotisches gab.

Andreas machte sich nützlich. Er bereitete das **Futter** vor, weil der **Trog** schon von Schweinen umringt war.  
Andreas machte sich nützlich. Er bereitete das **Futter** vor, weil der **Mantel** schnell fertiggestellt werden sollte.

Martin passierte ein Missgeschick. Er wählte den falschen **Gang**, weil im **Flur** ein Licht defekt war.  
Martin passierte ein Missgeschick. Er wählte den falschen **Gang**, weil im **Auto** keine Automatik eingebaut war.

Marcos Entscheidung war eindeutig. Er bevorzugte den **Kamm**, weil der **Scheitel** sich so leichter in Form bringen ließ.  
Marcos Entscheidung war eindeutig. Er bevorzugte den **Kamm**, weil der **Berg** sich hier von seiner schönsten Seite zeigte.

Jens war immer pflichtbewusst. Er widmete sich der **Kapelle**, um den **Dirigenten** zu vertreten.  
Jens war immer pflichtbewusst. Er widmete sich der **Kapelle**, um der **Kirche** zu dienen.

Ulrike war vollauf beschäftigt. Sie kämpfte mit dem **Kater**, weil dem **Tier** das Herumtollen so viel Spass machte.  
Ulrike war vollauf beschäftigt. Sie kämpfte mit dem **Kater**, weil dem **Wein** noch so viele Schnäpse gefolgt waren.

Hannes war bekannt für seinen guten Manieren. Er erwartete die **Kundschaft**, weil der **Laden** neu eröffnet wurde.

Hannes war bekannt für seinen guten Manieren. Er erwartete die **Kundschaft**, weil die **Nachricht** alles verändern könnte.

Ina ging auf Nummer sicher. Sie probierte die **Linse**, weil die **Suppe** seltsam aussah.

Ina ging auf Nummer sicher. Sie probierte die **Linse**, weil die **Brille** noch nicht repariert war.

Paula freute sich. Sie hatte die **Lösung** gefunden, weil das **Rätsel** sehr einfach war.

Paula freute sich. Sie hatte die **Lösung** gefunden, weil die **Säure** mit dem Metall reagierte.

Achim handelte schnell. Er steckte das **Magazin** ein, weil der **Kiosk** unbeaufsichtigt war.

Achim handelte schnell. Er steckte das **Magazin** ein, weil die **Pistole** geladen werden musste.

Korinna streckte die Hand aus. Sie berührte die **Maus**, die die **Katze** vor die Tür gelegt hatte.

Korinna streckte die Hand aus. Sie berührte die **Maus**, die den **Computer** steuerte.

Anna hatte ihre Gründe. Sie profitierte von der **Messe**, weil die **Wirtschaft** kräftig investiert hatte.

Anna hatte ihre Gründe. Sie profitierte von der **Messe**, weil die **Kirche** ihr Trost spendete.

Christina war es peinlich. Sie blamierte sich mit der **Note**, obwohl das **Zeugnis** sonst gut war.

Christina war es peinlich. Sie blamierte sich mit der **Note**, obwohl das **Lied** leicht zu singen war.

Gustav war voller Stolz. Er präsentierte den **Orden**, weil die **Ehrung** für ihn wichtig war.

Gustav war voller Stolz. Er präsentierte den **Orden**, weil das **Kloster** sein Lebensinhalt war.

Maren war aufgeregt. Sie berührte den **Ordner**, als das **Stadion** betreten wurde.

Maren war aufgeregt. Sie berührte den **Ordner**, als das **Papier** herausfiel.

Oliver wollte behilflich sein. Er half bei dem **Pflaster**, weil der **Arzt** darum gebeten hatte.

Oliver wollte behilflich sein. Er half bei dem **Pflaster**, weil der **Asphalt** eine Umrandung benötigte.

Peter war ein gründlicher Mensch. Er begutachtete die **Probe**, um die **Musik** des Orchesters zu beurteilen.

Peter war ein gründlicher Mensch. Er begutachtete die **Probe**, um die **Biologie** des Bodens zu bestimmen.

Nina suchte gründlich. Sie schaute nach der **Quelle**, weil dem **Bach** eine große Bedeutung zugeschrieben wurde.

Nina suchte gründlich. Sie schaute nach der **Quelle**, weil dem **Zitat** eine große Bedeutung zugeschrieben wurde.

Christian erklärte eindringlich den Sachverhalt. Er schilderte den **Rock**, der die **Hose** ersetzen sollte.

Christian erklärte eindringlich den Sachverhalt. Er schilderte den **Rock**, der die **Disko** auszeichnete.

Susanne war zufrieden. Sie bekam die **Rolle**, weil der **Schauspieler** optisch so gut zu ihr passte.  
Susanne war zufrieden. Sie bekam die **Rolle**, weil der **Schneider** das Garn nicht mehr benötigte.

Meike war sehr vorsichtig. Sie entfernte die **Schale**, weil beim **Kristall** kleinste Erschütterungen zum Bruch führen können.

Meike war sehr vorsichtig. Sie entfernte die **Schale**, weil beim **Apfel** einige Stellen dreckig waren.

Benjamin war unzufrieden. Er klagte über die **Schicht**, weil vom **Arbeiter** in der Fabrik zu viel verlangt wurde.  
Benjamin war unzufrieden. Er klagte über die **Schicht**, weil vom **Erz** wenig zu sehen war.

Nicole war überrascht. Sie wunderte sich über den **Schimmel**, weil kein **Pferd** bisher so gut gewesen war.  
Nicole war überrascht. Sie wunderte sich über den **Schimmel**, weil kein **Käse** so schnell schlecht werden sollte.

Yvonne war sprachlos. Sie war beeindruckt von dem **Schloss**, bis der **König** ihr den Hof zeigte.  
Yvonne war sprachlos. Sie war beeindruckt von dem **Schloss**, bis der **Schlüssel** steckenblieb und abbrach.

Zum Glück war Michael aufmerksam. Er entdeckte die **Spalte**, obwohl die **Zeitung** sonst nur Werbung enthielt.  
Zum Glück war Michael aufmerksam. Er entdeckte die **Spalte**, obwohl die **Schlucht** als ungefährlich galt.

Nadine war die Freude anzumerken. Sie bestaunte die **Spitze**, weil der **Gipfel** mit Schnee bedeckt war.  
Nadine war die Freude anzumerken. Sie bestaunte die **Spitze**, weil der **Stoff** handgeklöppelt war.

Tanja machte einen Fehler. Sie veranschaulichte den **Stamm**, ohne auf **Afrika** näher einzugehen.  
Tanja machte einen Fehler. Sie veranschaulichte den **Stamm**, ohne auf **Baum-** oder **Blattform** näher einzugehen.

Karin merkte es sofort. Ihr fiel die **Stärke** auf, welche die **Schwäche** in anderen Bereichen aber nicht wettmachte.  
Karin merkte es sofort. Ihr fiel die **Stärke** auf, welche den **Kuchen** sehr fest machte.

Beate war begeistert. Sie freute sich über den **Strauss**, weil der **Vogel** so schnell rennen konnte.  
Beate war begeistert. Sie freute sich über den **Strauss**, weil die **Blumen** so schön dufteten.

Tom schaffte es. Er erzeugte den **Ton**, der für die **Musik** zentral war.  
Tom schaffte es. Er erzeugte den **Ton**, der für die **Vase** vorgesehen war.

Marta machte eine unangenehme Entdeckung. Sie sah die **Wanze**, weil der **Agent** das Telefon nicht zugeschraubt hatte.  
Marta machte eine unangenehme Entdeckung. Sie sah die **Wanze**, weil der **Käfer** direkt darauf zugelaufen war.

Ulrich übernahm die Verantwortung. Er gab alles für die **Zeche**, weil die **Kneipe** so teuer war.  
Ulrich übernahm die Verantwortung. Er gab alles für die **Zeche**, weil das **Bergwerk** sein Lebensinhalt war.

Nadja war aufgebracht. Sie verfluchte den **Zirkel**, weil der **Kreis** mit dem defekten Gerät nicht gelingen wollte.

Nadja war aufgebracht. Sie verfluchte den **Zirkel**, weil die **Gruppe** sie verraten hatte.

The first sentence always indicates the dominant meaning, the second sentence the subordinate meaning.

Ambiguous word and target word in **bold**.

## Experiment 6

Fritz musste schnell fertig werden. Er vollendete den **Absatz**.

Michaela war beschäftigt. Sie bearbeitete den **Anbau**.

Paul hatte alle überrascht. Er sorgte für einen **Auflauf**.

Veronika sorgte für die nötigen Änderungen. Sie passte den **Aufsatz** an.

Alle waren von Sandra beeindruckt. Sie beherrschte den **Ball**.

Alle Augen waren auf Tim gerichtet. Er beschrieb ausführlich den **Bogen**.

Karl war mit der Bestellung zufrieden. Ihm gefiel die **Brause**.

Petra reagierte schnell. Sie entdeckte die **Bremse**.

Nikolas war offensichtlich beschäftigt. Er stellte die **Dichtung** fertig.

Bernd fiel etwas auf. Er bemerkte die **Fahne**.

Thomas musste den Job zu Ende bringen. Er arbeitete an der **Fassung**.

Kerstin machte ihre Arbeit gründlich. Sie prüfte die **Feder**.

Hubert war total genervt. Er beseitigte die **Fliege**.

Sebastian war beeindruckt. Er staunte über den **Flügel**.

Andreas machte sich nützlich. Er bereitete das **Futter** vor.

Martin passierte ein Missgeschick. Er wählte den falschen **Gang**.

Marcos Entscheidung war eindeutig. Er bevorzugte den **Kamm**.

Jens war immer pflichtbewusst. Er widmete sich der **Kapelle**.

Ulrike war vollauf beschäftigt. Sie kämpfte mit dem **Kater**.

Nicola traf eine Entscheidung. Sie übernahm die **Leitung**.

Ina ging auf Nummer sicher. Sie probierte die **Linse**.

Paula freute sich. Sie hatte die **Lösung** gefunden.

Achim handelte schnell. Er steckte das **Magazin** ein.

Korinna streckte die Hand aus. Sie berührte die **Maus**.

Annas Vorteil war offensichtlich. Sie profitierte von der **Messe**.

Christina war es peinlich. Sie blamierte sich mit der **Note**.

Gustav war voller Stolz. Er präsentierte den **Orden**.

Björn hatte so etwas noch nie gesehen. Er betrachtete den **Pass**.

Oliver wollte behilflich sein. Er half bei dem **Pflaster**.

Peter war ein gründlicher Mensch. Er begutachtete die **Probe**.

Nina suchte gründlich. Sie schaute nach der **Quelle**.

Christian erklärte eindringlich den Sachverhalt. Er schilderte den **Rock**.

Meike war sehr vorsichtig. Sie entfernte die **Schale**.

Nicole war überrascht. Sie wunderte sich über den **Schimmel**.

Yvonne war sprachlos. Sie war beeindruckt von dem **Schloss**.

Nadine war die Freude anzumerken. Sie bestaunte die **Spitze**.

Karin merkte es sofort. Ihr fiel die **Stärke** auf.

Tanja machte es allen deutlich. Sie veranschaulichte den **Stamm**.

Beate war begeistert. Sie freute sich über den **Strauss**.

Marta machte eine unangenehme Entdeckung. Sie entdeckte die **Wanze**.

Ulrich übernahm die Verantwortung. Er gab alles für die **Zeche**.

Nadja war aufgebracht. Sie verfluchte den **Zirkel**.

Ambiguous word is in **bold** letters.

## Curriculum Vitae

Name	Henning Holle
Date of birth	July 12 <sup>th</sup> , 1974
Place of birth	Thuine / Emsland
Country of birth	Germany

### Education, Qualifications and Professional Experience

2005 – present	Ph.D. student at the Max Planck Institute for Human Cognitive and Brain Sciences
2003 – 2004	Scientific Coordinator at the Center for Advanced Studies, University of Leipzig
1998 – 2003	Diplom in Psychology from the University of Trier, Germany
1995 – 1998	State-approved educator ( <i>staatlich-anerkannter Erzieher</i> ) from the Vocational School St. Franziskus ( <i>Fachschule St. Franziskus</i> ), Lingen
1994 – 1995	Alternative National Service ( <i>Zivildienst</i> ), DRK Lingen
1994	A-levels ( <i>Abitur</i> )



## Bibliographic Details

Holle Henning

The comprehension of co-speech iconic gestures: Behavioral, Electrophysiological and Neuroimaging studies

Universität Leipzig, Dissertation

161 pages, 205 references, 18 Figures, 6 Tables

**Paper** The present dissertation investigated the comprehension of co-speech iconic gestures using behavioral techniques, event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI). Three general questions were addressed: (1) Do iconic gestures convey additional information to the listener? (2) If so, what is the earliest point in time at which the meaning of gesture is accessible? (3) What brain areas are involved in the interaction of gesture and speech during comprehension? In general, the impact of gesture on language comprehension was tested by means of a disambiguation paradigm, where spoken lexically ambiguous sentences were accompanied either by disambiguating gestures or meaningless grooming movements.

The ERP results suggest that listeners use the additional information provided by iconic gestures to disambiguate speech. In addition, these experiments provided evidence that the integration of gesture and speech during comprehension is not entirely automatic, but also modulated by contextual factors such as the amount of observed meaningful hand movements. Using a gating paradigm, Experiment 4 determined the earliest point in time at which gesture starts to exert its disambiguating influence. The disambiguation points of many gestures were found to be remarkably early, with almost two thirds of all gestures enabling a meaning selection before participants had seen the segment considered to be the most meaningful, i.e., the stroke phase. Finally, in the fMRI Experiment, it was found that the processing of co-speech gestures elicited activation in cortical regions previously associated with action comprehension and audiovisual integration.

In the general discussion, a model was proposed on the basis of the existing data. It is suggested that comprehending a co-speech iconic gesture is a two-step process, where gesture

first elicits activation of the corresponding non-verbal representations, followed by an interaction between verbal and non-verbal information at the conceptual level.

**Referat** In einer Serie von 6 Experimenten wurde die Verarbeitung sprachbegleitender ikonischer Gesten untersucht. Methodisch kamen dabei sowohl behaviorale und bildgebende Verfahren, als auch Ereignis-Korrelierte Potentiale (EKPs) zum Einsatz. Die Arbeit gliedert sich in drei Hauptfragestellungen: (1): Übertragen ikonische Gesten zusätzliche Informationen an den Rezipienten? (2): Wann ist der frühestmögliche Zeitpunkt, an dem die Bedeutung einer solchen Geste extrahiert werden kann? (3): Welche Hirnareale sind an der Interaktion von gestischer und sprachlicher Information während der Verarbeitung beteiligt? Grundsätzlich wurde dabei der gestische Einfluß auf die Sprachverarbeitung durch ein Disambiguationsparadigma operationalisiert, in dem gesprochene lexikalisch ambige Sätze entweder mit disambiguierenden Gesten oder bedeutungslosen Kratzbewegungen kombiniert wurden.

Die EKP Ergebnisse legen nahe, daß Rezipienten die zusätzliche Informationen der ikonischen Gesten nutzen, um Sprache zu disambiguieren. Außerdem scheint die Integration von Gestik und Sprache kein vollständig automatischer Prozeß zu sein, sondern ebenfalls kontrollierte Anteile zu enthalten. Mit Hilfe eines Gating-Paradigmas wurde dann der frühestmögliche Zeitpunkt bestimmt, an dem Gesten beginnen, ihre disambiguierende Wirkung zu entfalten. Diese sogenannten Disambiguationspunkte vieler Gesten lagen bemerkenswert früh. Beinahe zwei Drittel aller Gesten des Stimulussets erlaubten eine Bedeutungsauswahl, bevor die Probanden das Segment der Gesten gesehen hatten, daß in der Literatur als das bedeutungstragende beschrieben worden ist (i.e., die Stroke-Phase). Im letzten Experiment wurde schließlich registriert, daß die Verarbeitung sprachbegleitender ikonischer Gesten Aktivierungen in Hirnarealen hervorruft, die jeweils eng mit der Verarbeitung beobachteter Handlungen bzw. audiovisueller Integration assoziiert sind.

In der allgemeinen Diskussion wurde auf Basis der vorliegenden Daten ein Modell entwickelt, in dem vorgeschlagen wird, daß die Verarbeitung einer sprachbeleitenden ikonischen Geste in zwei Schritten erfolgt. Zunächst bewirkt die Gestenverarbeitung eine Aktivierung der entsprechenden non-verbalen Repräsentationen, bevor verbale und non-verbale Informationen auf konzeptueller Ebene miteinander interagieren.

## **Selbständigkeitserklärung**

Hiermit erkläre ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde und dass aus fremden Quellen direkt oder indirekt übernommene Gedanken in der Arbeit als solche kenntlich gemacht worden sind.

Henning Holle

Leipzig, 15. Juni 2007



## MPI Series in Human Cognitive and Brain Sciences:

- 1 Anja Hahne  
*Charakteristika syntaktischer und semantischer Prozesse bei der auditiv Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien*
- 2 Ricarda Schubotz  
*Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion*
- 3 Volker Bosch  
*Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale*
- 4 Jorge Jovicich  
*An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain*
- 5 Rosemary C. Dymond  
*Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations*
- 6 Stefan Zysset  
*Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie*
- 7 Ulrich Hartmann  
*Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes*
- 8 Bertram Opitz  
*Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße*
- 9 Gisela Müller-Plath  
*Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen*
- 10 Thomas Jacobsen  
*Characteristics of processing morphological structural and inherent case in language comprehension*

- 11 Stefan Kölsch  
*Brain and Music*  
*A contribution to the investigation of central auditory processing with a new electrophysiological approach*
- 12 Stefan Frisch  
*Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen*
- 13 Markus Ullsperger  
*The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis*
- 14 Martin Koch  
*Measurement of the Self-Diffusion Tensor of Water in the Human Brain*
- 15 Axel Hutt  
*Methoden zur Untersuchung der Dynamik raumzeitlicher Signale*
- 16 Frithjof Kruggel  
*Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie*
- 17 Anja Dove  
*Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselfaradigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie*
- 18 Karsten Steinhauer  
*Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache*
- 19 Silke Urban  
*Verbinformationen im Satzverstehen*
- 20 Katja Werheid  
*Implizites Sequenzlernen bei Morbus Parkinson*
- 21 Doreen Nessler  
*Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition*

- 22 Christoph Herrmann  
*Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse*
- 23 Christian Fiebach  
*Working Memory and Syntax during Sentence Processing.  
A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging*
- 24 Grit Hein  
*Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten*
- 25 Monica de Filippis  
*Die visuelle Verarbeitung unbeachteter Wörter.  
Ein elektrophysiologischer Ansatz*
- 26 Ulrich Müller  
*Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen*
- 27 Kristina Uhl  
*Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information*
- 28 Ina Bornkessel  
*The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation*
- 29 Sonja Lattner  
*Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen*
- 30 Christin Grünewald  
*Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie*
- 31 Annett Schirmer  
*Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women*
- 32 André J. Szameitat  
*Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben*

- 33 Susanne Wagner  
*Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten*
- 34 Sophie Manthey  
*Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie*
- 35 Stefan Heim  
*Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words*
- 36 Claudia Friedrich  
*Prosody and spoken word recognition: Behavioral and ERP correlates*
- 37 Ulrike Lex  
*Sprachlateralisierung bei Rechts- und Linkshändern mit funktioneller Magnetresonanztomographie*
- 38 Thomas Arnold  
*Computergestützte Befundung klinischer Elektroenzephalogramme*
- 39 Carsten H. Wolters  
*Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain*
- 40 Ansgar Hantsch  
*Fisch oder Karpfen? Lexikale Aktivierung von Benennungsalternative bei der Objektbenennung*
- 41 Peggy Bungert  
*Zentralnervöse Verarbeitung akustischer Informationen  
Signalidentifikation, Signallateralisation und zeitgebundene Informationsverarbeitung bei Patienten mit erworbenen Hirnschädigungen*
- 42 Daniel Senkowski  
*Neuronal correlates of selective attention: An investigation of electro-physiological brain responses in the EEG and MEG*

- 43 Gert Wollny  
*Analysis of Changes in Temporal Series of Medical Images*
- 44 Angelika Wolf  
*Sprachverstehen mit Cochlea-Implantat: EKP-Studien mit postlingual ertaubten erwachsenen CI-Trägern*
- 45 Kirsten G. Volz  
*Brain correlates of uncertain decisions: Types and degrees of uncertainty*
- 46 Hagen Huttner  
*Magnetresonanztomographische Untersuchungen über die anatomische Variabilität des Frontallappens des menschlichen Großhirns*
- 47 Dirk Köster  
*Morphology and Spoken Word Comprehension: Electrophysiological Investigations of Internal Compound Structure*
- 48 Claudia A. Hruska  
*Einflüsse kontextueller und prosodischer Informationen in der auditorischen Satzverarbeitung: Untersuchungen mit ereigniskorrelierten Hirnpotentialen*
- 49 Hannes Ruge  
*Eine Analyse des raum-zeitlichen Musters neuronaler Aktivierung im Aufgabenwechselparadigma zur Untersuchung handlungssteuernder Prozesse*
- 50 Ricarda I. Schubotz  
*Human premotor cortex: Beyond motor performance*
- 51 Clemens von Zerssen  
*Bewusstes Erinnern und falsches Wiedererkennen:  
Eine funktionelle MRT Studie neuroanatomischer Gedächtniskorrelate*
- 52 Christiane Weber  
*Rhythm is gonna het you.  
Electrophysiological markers of rhythmic processing in infants with and without risk for Specific Language Impairment (SLI)*
- 53 Marc Schönwiesner  
*Functional Mapping of Basic Acoustic Parameters in the Human Central Auditory System*

- 54 Katja Fiehler  
*Temporospatial characteristics of error correction*
- 55 Britta Stolterfoht  
*Processing Word Order Variations and Ellipses: The Interplay of Syntax and Information Structure during Sentence Comprehension*
- 56 Claudia Danielmeier  
*Neuronale Grundlagen der Interferenz zwischen Handlung und visueller Wahrnehmung*
- 57 Margret Hund-Georgiadis  
*Die Organisation von Sprache und ihre Reorganisation bei ausgewählten, neurologischen Erkrankungen gemessen mit funktioneller Magnetresonanztomographie – Einflüsse von Händigkeit, Läsion, Performanz und Perfusion*
- 58 Jutta L. Mueller  
*Mechanisms of auditory sentence comprehension in first and second language: An electrophysiological miniature grammar study*
- 59 Franziska Biedermann  
*Auditorische Diskriminationsleistungen nach unilateralen Läsionen im Di- und Telenzephalon*
- 60 Shirley-Ann Rüschemeyer  
*The Processing of Lexical Semantic and Syntactic Information in Spoken Sentences: Neuroimaging and Behavioral Studies of Native and Non-Native Speakers*
- 61 Kerstin Leuckefeld  
*The Development of Argument Processing Mechanisms in German. An Electrophysiological Investigation with School-Aged Children and Adults*
- 62 Axel Christian Kühn  
*Bestimmung der Lateralisierung von Sprachprozessen unter besondere Berücksichtigung des temporalen Cortex, gemessen mit fMRT*
- 63 Ann Pannekamp  
*Prosodische Informationsverarbeitung bei normalsprachlichem und deviantem Satzmaterial: Untersuchungen mit ereigniskorrelierten Hirnpotentialen*

- 64 Jan Derrfuß  
*Functional specialization in the lateral frontal cortex: The role of the inferior frontal junction in cognitive control*
- 65 Andrea Mona Philipp  
*The cognitive representation of tasks  
Exploring the role of response modalities using the task-switching paradigm*
- 66 Ulrike Toepel  
*Contrastive Topic and Focus Information in Discourse – Prosodic Realisation and Electrophysiological Brain Correlates*
- 67 Karsten Müller  
*Die Anwendung von Spektral- und Waveletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale*
- 68 Sonja A.Kotz  
*The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging*
- 69 Sonja Rossi  
*The role of proficiency in syntactic second language processing: Evidence from event-related brain potentials in German and Italian*
- 70 Birte U. Forstmann  
*Behavioral and neural correlates of endogenous control processes in task switching*
- 71 Silke Paulmann  
*Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations*
- 72 Matthias L. Schroeter  
*Enlightening the Brain – Optical Imaging in Cognitive Neuroscience*
- 73 Julia Reinholz  
*Interhemispheric interaction in object- and word-related visual areas*
- 74 Evelyn C. Ferstl  
*The Functional Neuroanatomy of Text Comprehension*

- 75 Miriam Gade  
*Aufgabeninhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen*
- 76 Juliane Hofmann  
*Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German*
- 77 Petra Augurzky  
*Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing*
- 78 Uta Wolfensteller  
*Habituelle und arbiträre sensomotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen*
- 79 Päivi Sivonen  
*Event-related brain activation in speech perception: From sensory to cognitive processes*
- 80 Yun Nan  
*Music phrase structure perception: the neural basis, the effects of acculturation and musical training*
- 81 Katrin Schulze  
*Neural Correlates of Working Memory for Verbal and Tonal Stimuli in Nonmusicians and Musicians With and Without Absolute Pitch*
- 82 Korinna Eckstein  
*Interaktion von Syntax und Prosodie beim Sprachverstehen: Untersuchungen anhand ereigniskorrelierter Hirnpotentiale*
- 83 Florian Th. Siebörger  
*Funktionelle Neuroanatomie des Textverstehens: Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten*
- 84 Diana Böttger  
*Aktivität im Gamma-Frequenzbereich des EEG: Einfluss demographischer Faktoren und kognitiver Korrelate*

- 85 Jörg Bahlmann  
*Neural correlates of the processing of linear and hierarchical artificial grammar rules: Electrophysiological and neuroimaging studies*
- 86 Jan Zwickel  
*Specific Interference Effects Between Temporally Overlapping Action and Perception*
- 87 Markus Ullsperger  
*Functional Neuroanatomy of Performance Monitoring: fMRI, ERP, and Patient Studies*
- 88 Susanne Dietrich  
*Vom Brüllen zum Wort – MRT-Studien zur kognitiven Verarbeitung emotionaler Vokalisationen*
- 89 Maren Schmidt-Kassow  
*What's Beat got to do with it? The Influence of Meter on Syntactic Processing: ERP Evidence from Healthy and Patient populations*
- 90 Monika Lück  
*Die Verarbeitung morphologisch komplexer Wörter bei Kindern im Schulalter: Neuropsychologische Korrelate der Entwicklung*
- 91 Diana P. Szameitat  
*Perzeption und akustische Eigenschaften von Emotionen in menschlichem Lachen*
- 92 Beate Sabisch  
*Mechanisms of auditory sentence comprehension in children with specific language impairment and children with developmental dyslexia: A neurophysiological investigation*
- 93 Regine Oberecker  
*Grammatikverarbeitung im Kindesalter: EKP-Studien zum auditorischen Satzverstehen*
- 94 Şükrü Barış Demiral  
*Incremental Argument Interpretation in Turkish Sentence Comprehension*