

Christian Obermeier: Exploring the Significance of Task, Timing and Background Noise on Gesture-Speech Integration. Leipzig: Max Planck Institute of Cognitive Neuroscience, 2011 (MPI Series in Cognitive Neuroscience; 133)

---

Exploring the significance of task, timing and background noise  
on gesture-speech integration

## Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2011



Diese Arbeit ist unter folgender Creative Commons-Lizenz lizenziert:  
<http://creativecommons.org/licenses/by-nc/3.0>

Titelfoto: Christian Obermeier, 2011

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden

ISBN 978-3-941504-17-2

# **Exploring the significance of task, timing and background noise on gesture-speech integration**

Von der Fakultät für Biowissenschaften, Pharmazie und Psychologie  
der Universität Leipzig  
genehmigte

DISSERTATION

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

vorgelegt  
von Diplom-Psychologe Christian Obermeier  
geboren am 23.11.1976 in Kaufbeuren

Dekan: Prof. Dr. Matthias M. Müller

Gutachter: Prof. Dr. Angela D. Friederici  
Prof. Dr. Spencer D. Kelly

Tag der Verteidigung: 26.5.2011



## Acknowledgements

This thesis was only possible because of the support and contributions of many people. In particular I would like to thank:

Prof. Angela D. Friederici for giving me the opportunity to carry out the present dissertation at Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig and for agreeing to evaluate the thesis.

Thom Gunter for introducing me to the field of co-speech gestures and his excellent supervision and support throughout all phases of the dissertation; for his openness to critically discuss new scientific ideas.

Spencer Kelly for very interesting and helpful discussions about gesture-speech integration and for reviewing this thesis.

All the members of our little “gesture group”: Thom, Douglas, Henning, Thomas, Sanne, Arjan, Leonie, Iris, and Martin. Thanks for all the interesting discussions and the fun we had.

My wonderful colleagues at the MPI, in particular Douglas Weinbrenner, Corinna Bonhage, Manuela Missana, Patricia Garrido Vasquez, Maren Grigutsch and Jonas Obleser for help and discussions during various stages of this dissertation.

Heike Böthel, Kristiane Klein, Ina Koch, Cornelia Schmidt, and Sven Gutekunst for help with the EEG data acquisition.

Kerstin Flake, Andrea Gast-Sandmann, & Stephan Liebig for their support in creating the stimuli for the experiments as well as creating the graphical illustrations for this dissertation.

The members of the IT department for solving all the little and big soft- and hardware problems.

The library team, especially Thea and Anja, for providing me with all those rare gesture papers.

My friends and family for their continuous and unconditional support.



## Table of contents

<b>Preface .....</b>	<b>1</b>
 <b>Chapter 1: Introduction to iconic gestures .....</b>	 <b>7</b>
<b>1.1. Types of co-speech gestures .....</b>	<b>8</b>
1.1.1. Iconic gestures .....	8
1.1.2. Metaphorics .....	8
1.1.3. Deictics .....	9
1.1.4. Beats .....	11
1.1.5. Summary .....	12
<b>1.2. Specifying iconic gestures: structure and relation to speech .....</b>	<b>12</b>
1.2.1. Temporal/ Kinesic structure of an iconic gesture .....	13
1.2.2. The relation of iconic gestures to speech: Similarities and differences in the generation of meaning in both types of communicative input .....	16
1.2.3. The relation of iconic gestures to speech: Timing – a crucial but understudied Factor .....	18
<b>1.3. Summary .....</b>	<b>21</b>
 <b>Chapter 2: Research on iconic gestures .....</b>	 <b>25</b>
<b>2.1. Why do we gesture? – Findings from gesture-speech production research .....</b>	<b>25</b>
2.1.1. How do gestures aid the producer? .....	26
2.1.2. How do message content and situation influence gesture production? .....	28
<i><b>Excursus: Gesture production theories .....</b></i>	<i><b>30</b></i>
2.1.3. Summary .....	32
<b>2.2. Comprehension of iconic co-speech gestures .....</b>	<b>33</b>



2.2.1. Behavioral evidence .....	33
2.2.2. Neurophysiologic evidence for gesture comprehension .....	38
2.2.2.1. Event-related potentials (ERPs) of the EEG as a measure of gesture Comprehension .....	38
2.2.2.2. The nature of the human EEG .....	39
2.2.2.3. Event-related potentials (ERPs) .....	42
2.2.2.4. The N400 .....	44
2.2.2.5. ERPs as a correlate for gesture-speech integration .....	46
2.2.2.6. fMRI results on gesture-speech integration .....	53
<b>2.3. Summary of the literature on iconic co-speech gestures .....</b>	<b>57</b>
<b>2.4. The present dissertation: The significance of task, timing and background noise     on gesture-speech integration .....</b>	<b>59</b>
 <b>Chapter 3: Stimulus material .....</b>	 <b>65</b>
<b>3.1. The construction of the original full-length gesture material .....</b>	<b>65</b>
3.1.1. Sentence material (homonyms) .....	65
3.1.2. Recording of the original gesture videos .....	66
3.1.3. Pretests for the original gesture videos .....	67
3.1.4. Splicing .....	68
3.1.5. Rating of the gesture phases .....	71
<b>3.2. The construction of the gesture fragment stimulus material .....</b>	<b>72</b>
3.2.1. Gesture fragments .....	72
3.2.2. Gating .....	73
3.2.3. Stimuli: Gesture fragments .....	77
3.2.4. Pretests for the gesture fragment videos .....	79
3.2.4.1. Cloze procedure .....	79

3.2.4.2. Gesture fragment identification without speech context .....	80
3.2.5. Determining the identification points of the homonyms .....	80
<b>3.3. Summary .....</b>	<b>80</b>

## **Chapter 4: The impact of task on integration of gesture fragments with**

<b>speech .....</b>	<b>85</b>
<b>4.1. Experiment 1: Are we able to use gesture fragment information at all? .....</b>	<b>86</b>
4.1.1. Introduction .....	86
4.1.2. Methods .....	86
4.1.3. Results .....	90
4.1.4. Discussion .....	92
<b>4.2. Experiment 2: Is the integration of gesture fragment and speech</b>	
<b>task-independent? .....</b>	<b>94</b>
4.2.1. Methods .....	94
4.2.2. Results .....	95
4.2.3. Discussion .....	97

## **Chapter 5: The significance of timing for gesture-speech integration ..... 101**

<b>5.1. Experiment 3: The processing of synchronous gesture and speech information 101</b>	
5.1.1. Introduction .....	101
5.1.2. Methods .....	102
5.1.3. Results .....	104
5.1.4. Discussion .....	105
<b>5.2. Experiment 4: Is there a temporal window for gesture-speech integration? ..... 107</b>	
5.2.1. Methods .....	107
5.2.2. Results .....	109

5.2.3. Discussion .....	110
-------------------------	-----

## **Chapter 6: The impact of background noise on gesture-speech**

<b>integration .....</b>	<b>117</b>
<b>6.1. Experiment 5 .....</b>	<b>118</b>
6.1.1. Introduction .....	118
6.1.2. Methods .....	118
6.1.3. Results .....	121
6.1.4. Discussion .....	124

## **Chapter 7: General Discussion .....**

<b>7.1. Summary of the results .....</b>	<b>127</b>
7.1.1. The integration of gesture fragment and homonym .....	130
7.1.2. The effect of gesture-homonym integration at the target word .....	132
<b>7.2. A model for gesture-speech integration in comprehension .....</b>	<b>133</b>
7.2.1. <i>The Feature Integration Model (FIM) for gesture-speech comprehension</i> .....	134
7.2.1.1. The scope .....	134
7.2.1.2. The architecture of the <i>FIM</i> .....	134
7.2.1.3. The perceptual analysis .....	135
7.2.1.4. Feature extraction .....	138
7.2.1.5. The semantic integration of gesture and speech .....	140
7.2.1.6. How does the model account for the data of the dissertation .....	143
7.2.1.7. The model and other gesture types .....	144
<b>7.3. Open questions .....</b>	<b>144</b>
7.3.1. The relationship between local gesture-speech integration and the global message level .....	145

7.3.2. The integration of asynchronous gestures with speech – the role of working memory and attention .....	145
7.3.3. What modulates the integration process? – The role of the communicative context .....	147
7.3.4. The neural correlates of gesture-speech integration .....	148
7.3.5. The impact of gestures on speech comprehension in hearing impaired .....	149
<b>7.4. Concluding remarks .....</b>	<b>150</b>
 <b>References .....</b>	 <b>151</b>
<b>List of Figures .....</b>	<b>170</b>
<b>List of Tables .....</b>	<b>172</b>
<b>Appendix A: Sentence material used in Experiments 1 to 5 .....</b>	<b>173</b>



## Preface

Gestures are an integral part of every-day human communication and serve many different functions (e.g. they aid lexical retrieval, the learning of a second language or can be used as predictors for cognitive development). Even congenital blind people gesture when they talk to each other (Iverson & Goldin-Meadow, 1998). The interest in gesture and their function is nothing new and already originated in Ancient Greek and Roman times. At that time, as can be seen in the works by philosophers and scientists like Aristotle, Cicero and Quintilian, gesture was of interest because of its close connection to rhetoric. For example, in the first century AD, Quintilian in his *Institutio oratoria*, already noted the close relation between gesture and speech, which is one of the basic assumptions of modern gesture research: “the voice has our first claim on our attention, since even our gesture is adapted to suit it (XI, III, 14, Quintilian, trans. 1922). Up to the eighteenth century the rhetoric aspect of gesture has been dominant, until philosophers like Condillac (1756/1971) or Diderot (1751/1916), driven by the new idea that language was not god-made, began to reflect on the possible role of gesture as a predecessor for speech. The ensuing 19<sup>th</sup> century marks a first important shift in gesture research. The works of De Jorio (1832/2000), Tylor (1865), Mallery (1881/1972) as well as Wundt (1921/1973) led to a much better understanding of gestures, especially signs. For example, Wundt was the first using an experimental psychological approach for studying gestures (i.e. sign language) as well as the first to develop a “semiotic” classification of gestures.

After Wundt, gesture research received little attention until the 1930s, when scientists started to investigate gestures of everyday life, e.g. gestures during conversation (for details see Efron, 1941). This different approach to gestures benefited from new inventions like the slow-motion film, which for the first time allowed detailed studies about the relation of gesture and speech. However, over the next decades research on the interaction of gesture and speech faded again from the spotlight of psychological research. At the beginning of the 1970s, many scientists believed that body-language, including gestures, only reflects emotional states but has no communicational value. Exactly at that time, new interest in gesture research emerged with the early works of Kendon (1972) who introduced the notion that gestures are an integral part of language. However, it took almost another two decades,

until the growing interest from linguists and psychologists really was observable in a growing number of publications. In his seminal work, McNeill (1985, 1992) took up Kendon's idea that gesture and speech form language and further developed it into the first theoretical framework for gesture-speech production, the growth-point theory. In his view, gesture and speech are two parts of one single system, i.e. language. Since then, numerous studies have investigated different aspects of the interaction between gesture and speech, both in gesture-speech production as well as comprehension and from infant age up to elderly patients. For example, research on gesture production has shown that gestures aid lexical retrieval (Krauss, 1998), spatial memory (e.g. Morsella & Krauss, 2004), and generally ease cognitive load (e.g. Melinger & Kita, 2007). Furthermore, developmental research has shown that gestures provide insight into the cognitive development of children (e.g. Cook & Goldin-Meadow, 2006) and can also serve as early predictor for language development (Rowe & Goldin-Meadow, 2009a, 2009b). Some studies were able to show that gesture production depends on the context of the communicative situation (Holler & Stevens, 2007; Holler & Wilkin, 2009), thereby providing evidence for their communicative value. However, the most compelling evidence for a communicative effect of gesture stems from gesture comprehension research. It has been shown that gestures are rather vague in their meaning if seen without the accompanying speech (Hadar & Pinchas-Zamir, 2004). In the context of speech, however, one can glean additional information from the gestures, e.g. about size (Beattie & Shovelton, 1999b, 2002a; Holler & Beattie, 2002). Even stronger evidence comes from the field of neuroscience. Electroencephalography (EEG) and functional Magnet Resonance Imaging (fMRI) allow for a direct insight into the neural processes and structures involved in gesture-speech comprehension. EEG research has shown that addressees integrate gesture information with speech in a similar way as they integrate words into a preceding speech context (Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, 2007) and that one can use gestural information to disambiguate speech (Holle & Gunter, 2007). The process of integration is not completely automatic, but can be influenced by the amount of meaningful gesture information (Holle & Gunter, 2007) as well as the perceived intentionality of a speaker (Kelly, Ward, Creigh, & Bartolotti, 2007). fMRI research has identified the left inferior frontal cortex (IFG) and the bilateral superior temporal sulci and gyri (STSs/Gs) as putative regions where the integration of gesture and speech might take place (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Green et al., 2009; Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, 2008; Holle, Obleser, Rueschemeyer, & Gunter, 2010; Willems, Özyürek, & Hagoort, 2007, 2009).

Little, however, is known about the factors that impact gesture-speech integration. From a theoretical perspective, however, this is a very important aspect which has already attracted scientific interest early on (cf. Wundt, 1921/1973). To date, there has been no systematic, experimental approach that tried to shed light on this issue. Yet, there is little doubt, that identifying such factors presents a condition *sine qua non* en route to a full-fledged cognitive model of gesture comprehension. Note, that in contrast to gesture production research, there are no published theories or models on how gesture-speech comprehension might work<sup>1</sup>.

Using a disambiguation paradigm in combination with the event-related potentials (ERPs) of the EEG, the present series of experiments explores a number of factors that could play a crucial role in this process: task, timing and background noise. It is known, that task often affects how stimuli are processed (van Atteveldt, Formisano, Goebel, & Blomert, 2007) and that showing task-independence of stimulus processing is a good indicator for a rather automatic underlying process. Therefore, task is a very important factor if one wants to clarify the nature of gesture-speech integration. Timing was chosen as the second factor to look at, as theories on gesture production identify the temporal alignment between gesture and speech as a crucial factor for the whole integration process (McNeill, 1992, 2005). Moreover, timing is also known to be an important factor in multi-modal integration of any kind (e.g. Dixon & Spitz, 1980; van Atteveldt, Formisano, Blomert, & Goebel, 2007; van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2006a). The third factor tested was the influence of background noise on gesture-speech integration. Because the gestures of interest (i.e. iconic gestures) only become meaningful in connection with speech, changes in quality of the speech signal (e.g. being in a noisy bar) could have a substantial effect on gesture-speech integration.

In the following, I will first introduce the gestures of interest in this dissertation. I will contrast them to other types of co-speech gestures and provide information about the key characteristics of these gestures. Subsequently, I will review the literature on iconic gestures to date (Chapter 2), with a specific focus on gesture-speech comprehension. I will also introduce the paradigm (disambiguation paradigm) as well as the method used (ERPs), and end this section with introducing the basic research question of this dissertation. In Chapter 3, the stimulus construction as well as pretests will be described. After that, in the Chapters 4-6, the experiments will be explained in detail, including background, methods, results and short

---

<sup>1</sup> Note, however, that first steps towards a model of gesture-speech integration in comprehension have been taken (Holle, 2007; Kelly, Özyürek, & Maris, 2010).



discussion for each experiment. Specifically, Chapter 4 contains two experiments which target the question whether task plays a role in gesture-speech integration. Chapter 5 deals with the role of timing in this process, whereas Chapter 6 is concerned with the impact of background noise. Finally, in Chapter 7, I will summarize all the findings and relate them to the existing literature on gesture comprehension and multimodal integration. Furthermore, I will introduce a model for gesture-speech integration and conclude with some open issues that have to be addressed by future research.

# Chapter 1



## Chapter 1

### Introduction to iconic gestures

In everyday face-to-face communication, we do not only use speech to communicate with others, but also rely, amongst other things, on facial expressions, body posture and gesture. For example, imagine you meet an old friend from Cologne and you are talking about this year's carnival. He tells you how much he enjoyed it and how great he looked in his disguise as an 80s nerd: "*I was having a beard and wearing those glasses.*" Now this is perfectly understandable in itself. Yet, your friend additionally made some hand movements depicting a weird mustache and hilariously gigantic glasses simultaneously to the respective words. With these gestures he specified the form of both beard and glasses and thus gave you additional information which is strongly related to the content of speech but not available if you just heard his utterance. In other words, he used his hands to communicate.

In the example above the two gestures related to beard and glasses are called iconic gestures. They belong to the category of gesticulations, as do beats, deictics and metaphoric gestures. Because they are obligatorily accompanied by speech, these gestures are also termed co-speech gestures<sup>2</sup>. In the following, I will introduce the key characteristics of iconic gestures, which are the ones of interest for this dissertation, and outline how they differ from other types of co-speech gestures<sup>3</sup>. Then I will describe their specific temporal structure as well as their relation to speech. For the later part, I will specifically focus on the role of timing between iconic gestures and the accompanying speech

---

<sup>2</sup> Note that emblems, which are meaningful hand movements like the 'Ok' sign, can be accompanied by speech at times. Importantly and in contrast to co-speech gestures, they can also be understood without. They are conventionalized (and probably lexicalized like words), which is also quite distinct from co-speech gestures. Because of these considerable differences, I will not include them in the typology of co-speech gestures, but only introduce beats, deictics, metaphors and iconics.

<sup>3</sup> In his recent book (2005), McNeill suggested to move away from strict categories for gesture types and replace the categories with a dimensional approach (e.g. iconicity, metaphoricity, etc.), because most gestures represent a combination of multiple dimensions.

## 1.1. Types of co-speech gestures

### 1.1.1. Iconic gestures

As already stated above, iconic gestures are dynamic, spontaneous hand movements, which “bear a close formal relationship to the semantic content of speech” (McNeill, 1992, p. 12), and only occur in combination with speech (about 90 % of all iconic gestures occur during utterances, McNeill, 1985). They depict parts of the accompanying speech, e.g. they can represent objects (or concrete entities) as well as actions in an imagistic way. For instance, in the example given above, the gestures specify the shape of something (beard, glasses). In doing so, iconic gestures are both co-expressive as well as complementary to the accompanying speech. As a consequence, they allow an addressee to get additional information about a speaker’s mental imagery and thoughts that otherwise would be not available based on speech alone. Note, that iconic gestures would not be an effective communicational tool without speech. Previous research has shown that the meaning of an iconic gesture is rather vague unless it is combined with an utterance (e.g. Hadar & Pinchas-Zamir, 2004). A possible explanation for this finding is that iconics are spontaneously produced (often improvised) and therefore not very conventionalized.

### 1.1.2. Metaphorics

Metaphoric gestures have many similarities to iconics. The crucial difference between metaphoric and iconic gestures is that they illustrate some abstract concept whereas iconic gestures depict a concrete object or action. Like iconics, metaphorics reveal a speaker’s thought in a very pictorial way, i.e. they represent a concrete object as container for an image of something invisible, something abstract. For example, a person recounts a story to somebody else and first describes that the story is actually a cartoon (example taken from McNeill, 1992, p. 14). He says: “*It was a Sylvester and Tweety cartoon.*” Synchronously to the word “*Sylvester*” the person gestures an image of a bound container which reflects the abstract genre cartoon. By presenting a cartoon book with his hands, the narrator is able to put the abstract idea of cartoons into a concrete physical object. This object is usually termed *vehicle* as it transports the abstract image. Metaphoric and iconic gestures look similar on the surface level, i.e. they are imagistic depictions of objects. Both types of co-speech gestures are also quite speech dependent and unconventionalized, with metaphoric gestures probably

being a little bit more speech dependent and less conventionalized than iconic gestures due to their abstract content. Recent fMRI findings provide some proof for this assumption. Straube, Green, and Kircher (2010) have shown that there are some differences in the comprehension of both types. Whereas iconic gestures elicit increased activation within the bilateral STS/G region in contrast to baseline, metaphoric gesture processing additionally activates the left IFG, indicating that meaning generation may be more effortful for this type of gesture.

### 1.1.3. Deictics

A third kind of gesture are the so-called deictics or pointing gestures. Besides iconic gestures, pointing gestures are the second most used gestures in narratives and probably the most used in everyday conversation. The typical appearance of pointing is a hand with an extended index finger, but virtually every body part can be used for pointing. Two types of pointing can be distinguished: concrete and abstract pointing.

Concrete pointing refers to a function which is already implied by its name, i.e. indicating a reference in the real, concrete world. However, most pointing gestures in conversations are of the abstract type. These gestures do not aim at an existing physical target but rather use the indexed empty space as a projection area for an abstract concept. The following example illustrates how abstract pointing can be used: Two friends are talking to each other about one of them going on vacation to Las Vegas. The one who stays home asks: *“So, when will you go there?”* while pointing to the right at the end of the question. By pointing the speaker places his virtual Las Vegas right of his body. Thus, in this example pointing fills a seemingly empty space with a metaphorical image of a city. In general, abstract pointing assigns an abstract idea to an empty position in space similar to metaphoric gestures assign an abstract meaning to a container.

Pointing gestures are more dependent on speech than iconic gestures. For example, if someone simply is pointing at a well filled buffet without saying anything, you will probably have no clue about what she wants to communicate. Does she want to tell you *“Look at that!”*, *“This tastes fantastic!”* or *“Could you please get me one of these!”*. This is not surprising, as indicating a direction contains less semantic information than depicting the form of ball. But, if she accompanies her pointing with an utterance like, *“Could you get me some salad, please?”*, than it is pretty clear what she meant. With regard to the degree of conventionalization, deictics also take an intermediate position between iconic gestures and

another form of co-speech gestures, so-called beats, which are the least and insignificant looking of all co-speech gestures. The form of a pointing movement is conventionalized to some degree, because one or more fingers are aiming at a target space. Iconic gestures, however, are much more conventionalized and constraint based on their semantic content.

Pointing itself is especially interesting from a developmental perspective, because these gestures can already be seen very early in infants prior to speech onset. Some even describe them as an important step towards a social behavior (Tomasello, Carpenter, & Liszkowski, 2007) and stepping stone into language (Goldin-Meadow, 2007). Infants seem to be able to comprehend pointing at an earlier age than to produce pointing themselves. Showing pointings that were either directed at a previously presented object (congruent condition) or in the opposite direction (incongruent condition), Gredebäck, Melinder and Daum (2010) were able to show that 8-month old infants were aware of congruency between pointing and the location of an object as indicated by an enlarged N200 for the incongruent condition as compared to the congruent one. Thus, even 8-month olds are sensitive to the intention of concrete pointings. Four months later, at the age of 12 months, infants also begin to produce pointing gestures themselves. These early pointings give us some indications about the way social understanding and theory of mind might develop in children. E.g. children produce pointing at absent entities (e.g. pointing to the entrance door as a means for “pointing out” that dad is coming home soon (through the door)). They also use their shared experience with others to determine whether they should gesture or not. For example, infants point more often when they know that a person is ignorant, i.e. not knowing where a certain object is, than when a person is knowing, thus using common knowledge to determine whether a gesture is necessary or not (Liszkowski, Carpenter, & Tomasello, 2008). But they also use this shared common knowledge with another person in the interpretation of pointings (Liebal, Behne, Carpenter, & Tomasello, 2009). E.g. when infants cleaned up a room with an adult they interpreted the pointing at an object lying on the floor as a request to collect it. In contrast, they cannot infer the meaning of the pointing, if they haven’t shared the experience of cleaning up the room with the adult. All these findings suggest that infant pointing is very important from a social perspective and that such early communicative signals may serve as a basis for the later language development and cognitive development. In fact, pointings can also be used as an indicator for cognitive development. Cook and Goldin-Meadow (2006), for instance, showed that mismatches between gesture production and speech when solving mathematical equations are a reliable evidence for transition and acquisition of new problem

solving strategies in children. Additionally, Iverson and Goldin-Meadow (2005) discovered that early pointings in children of 14 months can serve as an early predictor for later vocabulary size. Although children are able to produce and comprehend pointing to some kind of degree, fully fledged abstract pointing may not be available to children prior to twelve years of age.

#### 1.1.4. Beats

Compared to the above mentioned types of co-speech gestures, beats look very insignificant at first glance. First described by Efron in 1941, beats or batons, as they were originally called (Efron, 1941; Ekman & Friesen, 1969), received their name from their physical properties and function, i.e. beats look like beating time in music. Typically, the hand moves in two phases (in / out; up / down) synchronously to the rhythmical structure of the accompanying speech. Note that all other co-speech gestures (e.g. iconic gestures) consist of three phases. Although these movements might often be not very striking as they look the same all the time they serve a very important purpose. Beats can reveal the underlying conception of discourse by marking words or sentences as being significant for the discourse-pragmatic content. Krahmer and Swerts (2007) were able to show that visual beats in comparison to a no-beat condition indeed enhance the perceived prominence of a co-occurring utterance by altering the F2 and F3 formants as well as increasing the duration of the speech signal. A beat, for example, can highlight the name of a person on introduction or put emphasis on words of special narrative value: A person might say “*George is always late, but **today** he was on time.*” while performing a beat simultaneously to uttering “*today*”. In this way, the speaker indicates the exception that George arrived in time. Note, that the semantic content is not decisive for the execution of beats but the overall discourse structure. Thus, from the combination of speech and beats, a listener can gain insight into the meta-level structure of discourse.

Beats are the most speech dependent and less conventionalized of all co-speech gestures. They can feature almost any kind of form without any conventions. For example, it does not make a difference if a person performs beats with flat hand or a fist. For iconic gestures it could make a difference if a speaker produces them with a fist or flat-handed as the form of the gesture can influence its meaning. Beats are much more speech-dependent than any other



type of co-speech gesture, although it has to be kept in mind that all mentioned gestures are obligatorily accompanied by speech.

#### **1.1.5. Summary**

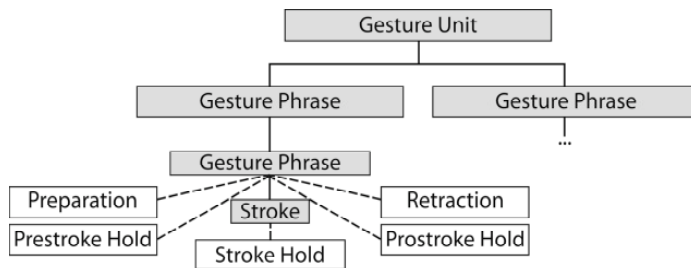
There are four types of co-speech gestures - iconics, metaphors, deictics and beats - which differ with regard to their form, degree of conventionalization and speech dependence. For example, the degree of conventionalization increases from beats, which are the least conventionalized gestures, across deictics and metaphors to the more conventionalized iconic gestures. The differences between the various gesture types are also reflected in the distinct functions they have in speech production. Beats are useful to structure utterances by highlighting important facts, deictics can give directional information (concrete pointing) or structure discourse (abstract pointing), while metaphors and iconic gestures provide an imagistic insight into the meaning and thinking of a speaker thereby providing additional information not present in speech.

### **1.2. Specifying iconic gestures: structure and relation to speech**

To get a better idea of how the interaction of iconic gestures and speech both in production and comprehension might work, it is necessary to take a close look at the structure of iconic gestures, because the form of a gesture has significant influence on how an addressee interprets its meaning. Furthermore, it is important to keep in mind that the attribution of meaning is only possible in the context of speech. Thus, to understand how both streams of information interact, it is also necessary to see the commonalities and differences in the way gesture and speech represent information. Last but not least, the temporal alignment between iconic gestures and speech might also be a very crucial factor in the whole process of meaning construction. In the following, all three points listed above will be addressed in detail.

### 1.2.1. Temporal/Kinesic structure of an iconic gesture

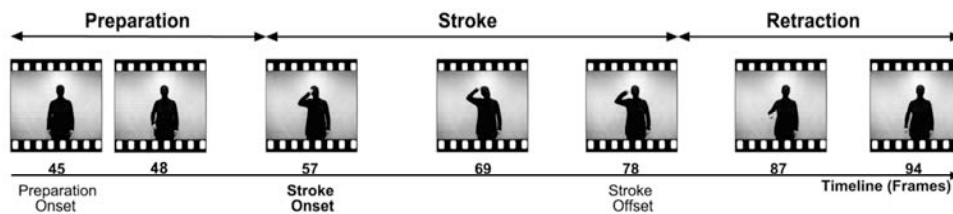
Kendon (1972, 1980) was the first to distinguish gestures into different parts. He identified three different temporal structures that enable a closer look into the dynamics of gestures: gesture units, gesture phrases and gesture phases. A gesture unit is the period between two resting phases of the arms. It can consist of one or more gesture phrases (see Figure 1.1). A gesture phrase in his terminology corresponds to what is normally referred to as “gesture”. Such a gesture phrase or gesture can consist of up to five different gesture phases, which I will describe below. Originally, Kendon (1972, 1980) only differentiated between three gestures phases: preparation, stroke (including stroke hold) and retraction (see Figure 1.2). Sotaro Kita (1990) reported two additional phases (pre- and poststroke hold).



**Figure 1.1** The temporal structure of an iconic gesture.

*Preparation* (optional): At the start of the preparation, the arms move away from the resting position towards the space where the stroke begins. There is no other reason for this movement than to prepare for the upcoming stroke. Kita, van Gijn and van der Hulst (1998) argue that hand-internal information like hand-shape or wrist location tends to emerge towards the end of the preparation phase, providing some anticipatory “phonological” information of what is to come next in the stroke phase. It is assumed that with the beginning of the preparation phase a common conceptual framework for both gesture and accompanying speech begins to form in the speaker’s cognition. Some support for this notion comes from a series of experimental studies by Levelt, Richardson and La Heij (1985). They investigated the temporal interdependency of speech and deictic gestures in planning and

execution by analyzing the time course of the pointing gestures and accompanying speech in various ways. Subjects were always required to point at referent lights while saying “*this / that light*”. The movements were registered using a Selspot opto-electronic system. Levelt and colleagues (1985) found that gesture and speech are independent in the phase of motor execution except for its initial part and that most of the temporal synchrony of gesture and speech is pre-established within the planning phase. Thus, it seems reasonable to suppose that there is one single underlying conceptualization for speech and accompanying gestures that is completely finalized at the beginning of the preparation phase.



**Figure 1.2** Typical time course of an iconic gesture (depicting a combing movement).

*Prestroke hold* (optional): This phase can occur just before the start of the stroke phase and marks a freeze in movement. The purpose of the prestroke hold is to delay the gestural movement in order to re-establish synchrony of gesture and the corresponding co-expressive speech unit. This also provides evidence for a single previously generated concept of gesture and speech. In order to hold a gesture, a speaker has to realize which units of gesture and speech should be aligned together in advance, which is only possible if planning of the utterance is somewhat finalized. The different modalities of gesture and speech may account for possible asynchronies following the mental conceptualization. These temporal differences are compensated by the prestroke hold.

*Stroke* (obligatory): The stroke is the most effortful, energetic and salient part of a gesture. Within this phase the meaning of a gesture is exhibited. Almost all the time (in 90% of the utterances) the stroke phase is synchronous to the accompanying co-expressive speech unit (Nobe, 2000; Valbonesi et al., 2001). Sometimes the stroke precedes this unit for a very short amount of time, but it rarely occurs after the co-expressive speech unit (Kendon, 1972).

Delay of the stroke beyond the co-expressive speech unit is almost always an indicator for neurological anomaly.

The stroke can also be accompanied by a so-called *stroke hold*: In principal, this phase is similar to the stroke phase, with the exception that meaning is not illustrated by a movement. During the stroke hold the hands are held motionless and it is exactly this effortful holding in place that bears meaning.

In contrast to all other gesture phases, the stroke is an essential part of every gesture, i.e. there is no iconic gesture without a stroke.

*Poststroke hold* (optional): Like the prestroke hold, the poststroke hold constitutes a cessation in movement. The hands are held in the position they adopted at the end of the stroke. This is again done to re-establish synchrony between gesture and speech in case the stroke is finished before the co-expressive speech unit ends. Thus, both pre- and poststroke holds are important factors for the synchrony of gesture and speech.

*Retraction* (optional): During the retraction phase the arms return to the resting position. In case the speaker produces two successive gestures, this phase can be omitted. The retraction is of some significance for the whole gesture phrase as it marks the end of it and thus clearly signals that the whole semantic content of the gesture has been displayed.

When analyzing all gestures phases it is striking that the whole gesture is organized around the stroke phase. McNeill (1992, 2005) suggests that the cognitive system tries everything (pre- and post-stroke holds) to hold up the synchrony of stroke and co-expressive speech unit. McNeill (2005) therefore concludes that temporal synchrony may play a crucial role in gesture-speech interaction (for more details about the temporal alignment of gesture and speech, see section 1.2.3. The relation of iconic gestures to speech: Timing – a crucial but understudied factor, p. 18).

Summing up, the stroke is the essential, meaningful part of an iconic gesture. The preparation phase can anticipate its meaning, whereas the retraction simply contains the return of the hands to the resting position. Any type of hold is produced to keep up a certain temporal alignment between gesture and co-occurring speech.

### 1.2.2. The relation of iconic gestures to speech: Similarities and differences in the generation of meaning in both types of communicative input

Gesture and speech are both integral part of our everyday communication. Using both allows us to communicate and comprehend messages more clearly and effectively by combining the unique way certain aspects of a message are represented in speech but not in gesture and vice versa. In the following, I will point out the differences between iconic gestures and speech in the way they convey meaning, which is in a large part related to the difference in modality. Language is *unidimensional* from a temporal perspective, e.g. each sentence develops word by word across time. It is characterized by *segmentation*, *linearization* and the use of *hierarchical* structures which are all typical for fully developed linguistic systems (e.g. this is also true for sign language). In language an event like somebody throwing a ball is formulated in a sentence like “*The boy throws the ball to the left.*”. The single segmented semantic units like “*the boy*”, etc. are linearly combined together to form the hierarchical structure of the sentence (in this case a simple subject-verb-object structure). Thus, the single meaningful units (e.g. the boy) determine the meaning of the sentence as a whole. This kind of meaning construction is termed *analytic*. Gestures convey meaning in a completely different way. In contrast to language, iconic gestures are *multidimensional*, *global-synthetic* and not *hierarchical*. *Global* refers to the fact that the meaning of the whole gesture has to be recognized first before single parts of it, e.g. a specific hand shape, can be interpreted. *Synthetic* signifies that a gesture can not only convey the meaning of a single word but also of a multiple words. For example, the sample sentence above can be depicted by one single throwing gesture (i.e. a throwing movement to the left). In that sense, gestures are also *multidimensional*, because they can combine multiple meanings or multiple aspects of one concept in time. In the case of the throwing example, the gesture contains both the throwing movement as well as the direction of it and the speed of the throwing movement. It is important to note that some information presented in speech is difficult to integrate into a gesture. For example, it is almost impossible to include the subject of a sentence into an iconic gesture (in the example depicted above, the ball could have also been thrown by a girl, an athlete, etc.). If one considers that a message is built upon the basic question “*Who is doing what to whom?*” it becomes immediately clear that iconic gestures are only useful in combination with speech.

In contrast to language, gestures are also assumed to be *noncombinatoric*. McNeill (1992, 2005) states, that two gestures cannot combine to form a new gesture, whereas different

words can form a sentence. However, there is some research that provides evidence against this stance. For example, Kita and Özyürek (2003) were able to show that it is the linguistic structure that determines whether, for instance, a movement expression (manner + trajectory) is realized in one or two gestures. In English, for example, it is easy to pack both manner and path into a single speech unit, which leads to the production of a single gesture. In Turkish or Japanese, however, both types of information are usually presented in separate speech units, which leads to the production of separate gestures for manner and for path. For example, if a Japanese speaker produces an utterance like “rolling down”, he will first perform a rotating gesture followed by a gesture depicting the trajectory of a downward movement.

Furthermore, gesture and speech can differ with regard to what McNeill (1992) calls non-linguistic properties, i.e. *standard of form* and *duality of patterning*. In contrast to speech, gestures do not have a *standard of form*. Whereas it is very simple to classify a sentence as non-German or non-English based on the knowledge of the linguistic properties of a language (grammar), this is impossible for gestures. However, this may be regarded as an advantage for gestures, as they can probably present aspects of meaning that language cannot (and vice versa), which might make them an important tool in second language learning until a person is proficient enough to speak that language fluently. *Duality of patterning* refers to the fact that in speech the form of word (phonetic structure) is arbitrarily linked with its meaning. For gestures, no such duality of patterning is found, because the form of a gesture determines its meaning. Therefore, both aspects are inseparable.

The previous section shows that gestures and speech appear to be very different on a structural level and how they express information. Yet, they refer to the same underlying concept (e.g. “growth point” in McNeill’s theory on gesture-speech production) and are thus co-expressive. The degree, to which gesture and speech express the same semantic content of an underlying concept, is termed co-expressiveness or semantic overlap. For example, a speaker may make a throwing movement while uttering, “*He threw the ball back.*” In this case both the speech and gesture are related to the same concept “*throwing*”, they are co-expressive. Sometimes the same aspects of an underlying concept are represented in gesture and speech but mostly there is no redundancy between gesture and speech<sup>4</sup> (McNeill, 2005). The following example shows, how iconic gestures additionally can provide information not conveyed in speech. The throwing movement in the given example might be fast or slow.

---

<sup>4</sup> Note, that this does not rule out semantic overlap as both iconic gesture and speech still refer to the same meaning.

This can only be inferred from the inspection of the gesture. In this example, the iconic gesture is not only co-expressive but also complementary, i.e. providing information not present in speech. The different way of representing semantic information in comparison with speech, i.e. *multidimensional, global-synthetic*, allows gestures to be especially effective in specifying the content of speech. Think of the example given above, where the direction of the throwing movement is incorporated in the same movement as the throwing.

To sum up, there are some major differences between iconic gestures and speech in the way they represent meaning. Whereas speech is *unidimensional, segmentized, linearized, hierarchical* and *analytic*, iconic gestures are *multidimensional, global-synthetic* and not *hierarchical*. Although they are very different or even because they are so different in the representation of a certain meaning, the combination of both gesture and speech provides us with a more complex and comprehensive image of a message. As stated above this complementary co-expressiveness is one of the key basic concepts in gesture-speech integration according to McNeill (2005). The second very important aspect for this process is the temporal alignment or synchrony between gesture and speech. In the next section, I will outline what is known so far about this issue on the basis of gesture production research.

### **1.2.3. The relation of iconic gestures to speech: Timing – a crucial but understudied factor**

For a long time there has been a controversy whether iconic gestures are anticipating the semantically linked speech part or are coincident with it. In the following I will first introduce the findings and concepts on the timing of iconic gestures and speech in production<sup>5</sup> and conclude that in fact there is no real “controversy” anymore with regard to this issue. Second, I will briefly explain what I mean with the “semantically linked speech part” by introducing the concepts of “co-expressive speech unit” and “lexical affiliate”.

#### *Gesture and Speech: Anticipation vs Coincidence – Anticipation and Coincidence?*

Butterworth and Beattie (1978) were the first to examine the temporal alignment of gesture and speech. Specifically, they looked at video recordings of college tutorials and identified

---

<sup>5</sup> I won't be including studies about timing between pointings and speech (Levelt et al., 1985), as well as timing between beats and speech (Treffner, Peter, & Kleidon, 2008).

when a gesture began (though they do not explicitly state it, they presumably looked at the onset of the preparation phase). They found that iconic gestures tended to be initialized more often during pauses prior to the semantically related speech than during speech itself. A similar finding was reported by Schegloff (1984) who studied discourse fragments. According to his view, iconic gesture and affiliated speech form a linear system, in which the gesture onset always precedes the respective word, thus the gesture cannot receive its meaning from speech. Moreover, he also states that at times gesture is not only initialized prior to speech but also terminated before it, leaving no temporal overlap between gesture and speech. Using an experimental setup, Morrel-Samuels and Krauss (1992) obtained results comparable to mentioned observation studies. They presented videos of photograph descriptions to participants, who had to underline the words or phrases related to meaning of the seen hand movements in a transcript of the videos. Afterwards, they analyzed the timing between gesture onset and these spoken items in the video material. Morrel-Samuels and Krauss (1992) found, that for all videos, the preparation onset occurred on average approximately 1 second prior to the onset of the related speech. The presented results seem to be supportive of theories on gesture production, which claim that gesture production facilitates lexical access (for more details on theories of gesture production, see Chapter 2 - Excursus: Gesture production theories, p. 30). Furthermore, they seemingly speak against McNeill's idea that gesture and speech are co-expressive and coincident (1985, 1992; also adapted by Nobe, 2000). McNeill (1985) states, based on his observations, that about 90 % of all iconic gestures are temporally aligned (i.e. overlapping) with the corresponding speech unit. If one takes a closer look at all the studies than it quickly becomes clear that both views are not incompatible but rather look at different phases of a gesture. Whereas the first three presented studies looked at the onset of the preparation phase in relation speech, McNeill (1985) examined the onset of the meaning bearing stroke phase in relation to it. This led McNeill (1992, 2005) to conclude that gestures can both anticipate and synchronize with speech. According to his view, the timing of gesture and speech is one of the most crucial aspects for understanding of gesture-speech integration. The preparation phase is the part of a gesture that is asynchronous with the corresponding word, i.e. it anticipates it both in time as well as on a semantic level. In contrast, the stroke phase is always synchronous with the co-expressive speech unit. It can be synchronous to it in various ways which McNeill (1992) formulates in three different synchrony rules. The phonological synchrony rule implies that the stroke of a gesture slightly precedes or ends at the phonological peak of an utterance (Kendon, 1980), thus the stroke is integrated into the phonology of speech. The semantic



synchrony rule signifies that the gesture stroke and the accompanying speech have to refer to the identical underlying idea at the same time. Additionally the pragmatic synchrony rule states that both gesture and corresponding speech can only relate to one, e.g. the same pragmatic referent (for details, see McNeill, 1992). As can be seen, iconic gestures and speech are closely synchronized with concern to their co-expressive units. It is important to note, that the co-expressive speech unit of a gesture does not necessarily have to be its lexical affiliate (and I will outline this below in this section).

Though the theoretical solution of the timing “controversy” by McNeill is very compelling, it cannot provide an explanation for the findings by Schegloff (1984), that gestures can occur even prior to its lexical affiliate, as well as similar results by Chui (2005), who examined everyday dyadic discourse and also found that sometimes gesture can be completely terminated before the corresponding word occurs. One thing that could partially contribute to this discrepancy (though in my opinion cannot satisfactorily explain it) are differences in coding (or identification) of the co-expressive speech unit and lexical affiliate in an utterance. In the following, I will define those two terms.

#### *Co-expressive speech unit vs. lexical affiliate*

The lexical affiliate, as it was first termed by Schegloff (1984), is the part of speech (one or more words) that most closely resembles the meaning represented in the gesture. Schegloff (1984) also observed that gestures show the tendency to precede the lexical affiliate in time. The lexical affiliate must not be confused with the co-expressive speech unit which is the part of speech that is synchronously produced with the gesture. Whereas the lexical affiliate is identifiable by comparing the semantic content of gesture and speech, the co-expressive speech unit can be discerned based on its co-occurrence with the stroke (McNeill, 2005). Sometimes both can be identical. An example which illustrates this distinction between co-expressive speech unit and lexical affiliate quite well is the following example (taken from Engle, 2000): A subjects tries to explain how a lock-and-key mechanism works. He says, “...lift them tumblers to a height, to the perfect height, where it **enables** the key to move,...”, while producing a gesture like turning a key simultaneous to “enable”. In this example “enables” is the co-expressive speech unit, while “key” is the lexical affiliate of the gesture, because it bears the strongest semantic relationship to the gesture.

Summing up, research on gesture production shows, that there is a tight temporal link between gesture and speech. Gesture both precedes and synchronizes with the co-expressive speech unit, which most of the time also is the lexical affiliate. An open question so far is whether the way gesture and speech are temporally aligned also has consequences for an addressee's information uptake? Some evidence that the timing might be important comes from a study by Woodall and Burgoon (1981), who investigated the effects of nonverbal synchronization on speech comprehension. They showed participants messages, in which kinesic cues (i.e. head, hand and body movements) were synchronized to the verbal message to different degrees. In a subsequent recall task, the authors found, that participants recalled significantly more of the messages when they had been presented with synchronous gesture and speech as compared to an asynchronous presentation. Thus, it seems clear that the timing of gesture and speech is a very important factor for the integration of both streams of information and therefore the communicational impact of gesture on speech.

### 1.3. Summary

There are four types of co-speech gestures: beats, deictics, metaphoric and iconic gestures, which all serve different purposes in gesture-speech production. For example, iconic gestures, which are the ones of interest for the dissertation, are used to specify certain aspects of objects or actions. They have a characteristic temporal structure, consisting of preparation, stroke and retraction phase. Both the preparation and the retraction phase are aligned around the stroke phase, which is said to be the meaning carrying part of a gesture. In contrast to the other phases, the stroke is assumed to be obligatory, i.e. without the stroke there is no gesture. Though iconic gestures also evolve along a temporal axis like speech, the way they represent information is quite different from the way the corresponding speech is doing that. Whereas speech is *unidimensional, segmentized, linearized, hierarchical* and *analytic*, iconic gestures are *multidimensional, global-synthetic* and not *hierarchical*. Despite, or maybe because of the differences in meaning representation, iconic gesture and speech go very well together in creating a new, more informative representation of the common underlying concept. Both the semantic as well as the temporal overlap seem to be important prerequisites for this effect. However, there is little knowledge on how timing really affects the comprehension of iconic gesture-speech combinations.

In the next chapter, I will briefly summarize what is known so far about the interaction of iconic gestures and their accompanying speech both with regard to production as well as comprehension research.

## **Chapter 2**



## **Chapter 2**

### **Research on iconic gestures**

In this chapter, I will present a review on iconic co-speech gesture research over the past two decades. For a substantial amount of time, this research was based around the controversy whether gestures are only a byproduct or epiphenomenon of speech or whether gesture and speech together form a single system, i.e. language. Researchers who favor the first idea assume that gestures are mainly produced for a speaker's benefit (Krauss, Dushay, Chen, & Rauscher, 1995), suggesting that gesture and speech are independent on a communicative, semantic level. In contrast, researchers who favor the second position believe that gestures convey substantial additional information not present in speech, and thus have a communicative value for an addressee (McNeill, 1992). By now it is clear, that iconic gestures can serve for many different purposes both for the speaker as well as the addressee. Not surprisingly, this has also been acknowledged both by Krauss (Morsella & Krauss, 2005) and McNeill (2005). In this chapter, I will first discuss the findings on gesture-speech production (to put it simply, the speaker side of gesture-speech communication). Then, I will focus on how gesture affects speech comprehension (the addressee side of gesture-speech communication). I will take an extensive look at both the existing behavioral and neurophysiologic data (EEG, fMRI), with the main focus on the results of EEG studies. Finally, I will present the main questions addressed in this dissertation.

#### **2.1. Why do we gesture? – Findings from gesture-speech production research**

Let's start off with two simple, everyday examples, to get an idea of why we produce gestures. Imagine you meet a good old friend and want to tell him about your recent success in fishing. You are so proud that you caught this very huge salmon. Rather than just relying on speech, because your friend is not an expert when it comes to fishing, you will most likely also produce a gesture that depicts the enormous size of the salmon you caught. In this

example, you use your gesture to specify an aspect of your message that may be underspecified in your speech, thus you use your gesture in a communicative way. Think of another situation: Everyone of us has talked to another person on the phone or has seen somebody else do so. Clearly, there is no visible communicative partner, but still we produce co-speech gestures. Thus, there has to be another reason than a gesture's communicative value as to why we do gesture in such a situation. Therefore, could it be that iconic gestures, beyond their communicative power, are also beneficial on a different level? – Over the past decades there has been a lot of research on gesture production that tried to clarify this issue. In the following, I will give an overview on the results so far. In the first part, I will focus on potential beneficial effects for the producer. Then, I will describe how the content of the message as well as the communicative situation can influence gesture production.

### **2.1.1. How do gestures aid the producer?**

One of the first questions with regard to iconic gesture production that caught researchers' attention was whether gesture production has a beneficial effect for the speaker. Based on the empirical observation that iconic gesture had the highest incidence during dysfluent phases of spontaneous speech, Butterworth and Hadar (Butterworth & Hadar, 1989; Hadar & Butterworth, 1997) proposed that the main function of iconic gesture production is to aid lexical access. There has been some support for this claim. For example, Rauscher, Krauss, and Chen (1996) varied whether a speaker could gesture or not in a cartoon retelling paradigm. They found, that speech, especially phrases with spatial content, was more fluent, when speakers could gesture as compared to when they were not allowed to. Thus, they concluded that gesture seems to facilitate lexical access. A similar finding has been reported by Frick-Horbury and Guttentag (1998), who examined the effects of restricting gesturing on the lexical retrieval and free recall of low frequency target words. Participants who were allowed to gesture both retrieved and recognized targets words to a higher degree. These results, together with the finding that gestures are usually initiated prior to the lexical affiliate (Morrel-Samuels & Krauss, 1992), led Krauss (1998) to the conclusion that gestures are produced to facilitate lexical retrieval in spontaneous speech. This stance is in line with Butterworth and Hadar (1989), who also suggested that gestures play a functional role in lexical retrieval. Further studies tried to specify these results by looking at gesture effects on memory functions. Wesp, Hesse, Keutmann, and Wheaton (2001) investigated the role of

gestures in spatial working memory performance. They had participants describe paintings which were either visually present or had to be recalled from memory. They proposed that if gestures help to sustain spatial imagery, gestures should occur more often when the paintings were not present. The results confirmed this hypothesis. Wesp et al. (2001) conclude that gestures facilitate the maintenance of spatial representations in working memory indirectly by enhancing lexical retrieval for spatial content. However, Morsella, and Krauss (2004) found that gestures can influence spatial working memory and lexical retrieval rather independently. Using a 2 x 2 x 2 design (present vs. absent object, codeable (concrete) object vs. uncodeable (abstract) object, gesturing restricted vs. gesturing allowed), they found that in the restricted condition participants showed the same amount of speech dysfluencies, no matter if spatial working memory was manipulated or not (present vs. absent object). Based on this finding, Morsella and Krauss (2004) concluded, that gesture can directly affect both lexical retrieval and spatial working memory by sustaining the semantic features of a to-be-retrieved word through feedback from motor commands (for more details about the so-called gestural feedback model see Morsella & Krauss, 2004).

Though there is a lot of evidence that gestures can enhance memory functions, in particular lexical retrieval, there are also some contradicting results. Beattie and Coughlan (1998) did not find significant changes in gesture production when speakers had to repeatedly produce a single narration, which is contrary to Butterworth and Hadar (1989), who assume that gesturing should decrease, as the lexical access gets easier with each repetition. Beattie and Coughlan (1998) therefore concluded that gesture production does not aid lexical access. However, this interpretation has to be treated with caution. Gestures may have been initially helpful and participants continued to produce them. Moreover, there is some new data by Sassenberg and Van der Meer (2010), which suggest that gesture production can also increase with reduced conceptual demands. Another study (Beattie & Shovelton, 2002b) also did not find any systematic relation between gesturing and lexical retrieval difficulties. Although gestures tended to occur more often with words of low transitional probability, speakers did not encounter any fluency problems. The authors suggest that gestures are not produced because of problems in lexical access, but for their role in the conceptualization of a communicative message, which is line with McNeill's growth point theory on gesture production (1992, see Chapter 2 - Excursus: Gesture production theories, p. 30). Indeed, the vast majority of papers on gesture production in the past decade focused on how the content



and structure of a message as well as the communicative situation influence the occurrence and form of gestures, and I will review their findings in the following part.

### 2.1.2. How do message content and situation influence gesture production?

Some of the earliest studies on gesture production tried to clarify whether certain expressions are more often associated with gesture production than others.<sup>6</sup> For example, Feyereisen and Havard (1999) asked their participants to either describe a visual image, motor image or mental / abstract image. They found that gesture production was highest in the motor condition, followed by the visual condition and lowest in the mental condition, leading them to the hypothesis that gesture production is influenced by the speech (message) content<sup>7</sup>. Using video analysis techniques, Beattie and colleagues (Beattie & Shovelton, 1999a, 2002b, 2006; Holler & Beattie, 2003) identified that highly imageable speech content as well as information about relative position and size in particular trigger gesture production (which is comparable to the findings of Feyereisen & Havard, 1999). Interestingly, highly relevant size information seems to be more likely to be encoded in gesture whereas less relevant size information is more likely to be encoded in speech (Beattie & Shovelton, 2006). This might be due to what Melinger and Levelt (2004) term the communicative intention of a speaker<sup>8</sup>. Using a picture description task, Melinger and Levelt (2004) investigated how speakers distribute the information across modalities. They found that speakers, who produced gestures depicting spatial relations, omitted more such information from speech in contrast to speakers who did not gesture. Melinger and Levelt (2004) argue that the gestures might serve as a common ground (or served knowledge) between speaker and listener, which once it is established should lead to a decrease of spatial information in speech. Evidence for this assumption comes from a study by Holler & Stevens (2007) who were able to show that speakers who shared common ground about the spatial relations in a picture with their addressees, tended to represent spatial information in speech only. In contrast, speakers, who did not share common knowledge with their communicative partner, predominantly represented spatial information in gesture or gesture and speech together (Holler & Stevens,

<sup>6</sup> All studies presented here share the basic assumption that gesture and speech are part of a single communicative system.

<sup>7</sup> Note, that Lausberg & Kita (2003) even found that the hand choice for gesture production is affected by the content of the message using an animation description task. Whether participants used the right or left hand to gesture depended on the relative spatial position of reference object. When they had to describe a spatial relationship between two objects they used both hands simultaneously.

<sup>8</sup> It is questionable at least whether communication is ever non-intentional.

2007). It seems that gesture is used to specify a message once speakers realize that pure verbal information is not sufficient or too ambiguous to be comprehended by an addressee in the way it is intended by the speaker. Two studies on the use of iconic gestures in verbal ambiguity resolution lend support to this notion. Both adults and children use gesture when they realize that there is a potential communication problem resulting from lexical ambiguity (Holler & Stevens, 2007; Holler & Wilkin, 2009). For example, when adults had to produce lexically ambiguous sentences like “*Her pupils were examined to detect potential illnesses*”, in which pupils can either refer to children or a part of the eye, they produce disambiguating gestures in almost half of the trials (Holler & Beattie, 2003). Alternatively (or even additionally) to the communicative intention of a speaker, there is also another explanation, why speakers may encode particular information in gesture: Distributing information across speech and gesture may reduce conceptual demands in message construction and thus allow speakers to communicate more effectively. For example, using a picture description task, Melinger & Kita (2007) showed that with increasing conceptual load (e.g. by increasing task difficulty) gesture production increases while speech remains comparable across the task conditions. Similar results have been found by Hostetter, Alibali and Kita (2007) as well as Kita and Davies (2009). Ping and Goldin-Meadow (2010) were able to show that the ability of gestures to reduce cognitive load does not only apply to the description of present objects, but also to the description of absent ones.

Interestingly, it is not only the task demands that affect the way information is encoded in gesture and speech but also cross-linguistic variation seems to play a role (Kita & Özyürek, 2003). In English, speakers use only one clause to encode manner and path, while Japanese and Turkish speakers use separate clauses. This difference can also be found in the gestures, as Japanese and Turkish were more likely to encode manner and trajectory separately. All these findings support the information packaging hypothesis (Kita, 2000), which states that gestures play a role in conceptualization of a spoken message, i.e. gesture production should increase with verbal conceptualization difficulty. There is, however, also some counterevidence against this assumption by Sassenberg and van der Meer (2010), who used a map description task, in which they manipulated task difficulty (e.g. new vs. already activated directions). Their results suggest that gesture production actually increases under lower conceptual demands (already activated condition) as compared to the high demand condition (new directions). They interpret their findings as in line with the Gesture-as-Simulated-Action framework (GSA, Hostetter & Alibali, 2008). According to the GSA,

which is based within the embodied cognition framework (e.g. Glenberg, 1997), gesture production is an epi-phenomenon of speech production. Both arise from the same mental representation, i.e. the mental imagery or simulated action underlying the speech. With increased strength of this representation, e.g. due to higher processing frequency (as it, for instance, is the case for repeated information), the likelihood of gesture production should increase. As predicted by GSA, Sassenberg and van der Meer (2010) found increased gesture production for already activated directions in contrast to new directions in a map direction description task.

### ***Excursus: Gesture production theories***

Besides the GSA framework there is number of different psychological and psycholinguistic models that try to explain how gesture production might work: the *sketch model* (de Ruiter, 2000), the *lexical gesture process model* (Krauss, Chen, & Gottesman, 2000), the *growth point theory* by McNeill (1992, 2005), and the *interface model* (Kita & Özyürek, 2003). Most models assume that gestures are originating from spatial representations, however all the models make different assumptions on how these representations are stored and on whether linguistic processes can influence gesture production. The earliest models like the *growth point theory* by McNeill (1992, 2005), the *lexical gesture process model* (Krauss et al., 2000) and the *sketch model* (de Ruiter, 2000) are clearly influenced by the early debate in gesture research whether gesture is simply a sidekick of speech or not. Both the *sketch model* and the *lexical gesture process model* state that gesture production is independent from linguistic factors. Yet, in the *sketch model*, gestures are generated from visuospatial imagery whereas sets of pure spatial representations form the basis for gestures in the *lexical gesture process model*. In contrast, McNeill (1992) formulates in his *growth point theory*, that gestures are generated from so-called growth points, which always present a combination of visuospatial imagery and corresponding linguistic representation. Thus, the *growth point theory* accounts for the influence of speech on gesture production. Kita and Özyürek (2003) took this idea one step further in their *interface model*. Similar to the *growth point theory*, they assume that a gesture originates from both linguistic and visuospatial factors. The speech part of an utterance is provided by the so-called message generator, whereas the gesture planning is done by the action generator which has access to visuospatial representations stored in working memory. The crucial part of the model is that action generator and message

generator are bi-directionally linked, allowing them to take into account both linguistic as well as visuospatial constraints in message generation. Thus, the *interface model* can, for instance, account the non-redundancy of gesture and speech in presenting size information (Beattie & Shovelton, 2006).

In contrast to the mentioned models, the *GSA* framework (Hostetter & Alibali, 2008) is not so much based on psycholinguistic assumptions, but on embodied cognition. In principal, the framework assumes that gesture production derives from motor or perceptual simulations which are based on mental imagery (motor and visual) and embodied language. In order to produce a gesture, the motor activation has to exceed the so-called gesture threshold. Three factors influence this process: a) the strength of action simulation, b) a speaker's individual gesture threshold (which is influenced by the communicative setting), and c) the simultaneous motor activations for speaking. Sassenberg and van der Meer (2010) were the first to provide experimental data in favor of the *GSA* framework. However, there are two caveats that work against the model. First, Hostetter and Alibali (2008) incorporated the social communicative situation as a factor that influences the level of the gesture threshold. While this factor may certainly play an important role in gesture production (as can be seen in the following section), it renders the model in its present form hard to test. For instance, if a speaker judges the communicative situation in such way that gesturing is helpful, his gesture threshold is lowered and more gestures are produced. If he judges the situation in such way that gesturing is not helpful, the gesture threshold will rise and thus fewer gestures will be produced. Thus, the factor social situation accounts for all communicative situations in which gestures may or may not be produced, which in turn makes it hard / impossible to come up with a design to test the model. Second, gesture production is somewhat described as an epiphenomenon of speech. The authors state that once the activation in premotor areas related to speech surpasses the threshold for production, it spreads out to motor areas and finally leads to an utterance. Simultaneously, the activation also spreads out to surrounding premotor areas which in turn lead to the elicitation of co-speech gestures. Hostetter and Alibali (2008) assume, that once speech is initialized, the inhibition of concurrent other activations may be very difficult. Thus, gesture production is a consequence or by-product of speech production. However, it is very hard to explain from this point of view why gesture production is initialized prior to the corresponding speech almost all the time, and why sometimes gestures are produced so much earlier, that they do not even overlap with the respective speech unit (Chui, 2005).

As noted above, the GSA extends other theories on gesture production by making the assumption that the communicative setting is an important factor that influences gesture production. So far, however, there are only a few studies which examined whether the social setting / communicative situation has an impact on gesture production. Bavelas and colleagues (2008) investigated, whether visibility and dialogue mode had an impact on gesture production using three different settings: face-to-face dialogue, telephone dialogue and monologue to tape recorder. Both visibility and dialogue manipulation had (independent) effects on gesturing. Participants gestured more in the face-to-face and telephone condition than in the tape-recorder condition, suggesting that knowledge about the conversational situation clearly affects rate of gesture production. Additionally, the produced gestures in both dialogue situations differed (visibility effect). Participants made more life-size gestures and more often incorporated information in their gestures that was not represented in their words in the face-to-face condition than in the telephone condition. Thus, they seemed to be quite aware of whether their gestures could actually be perceived by an addressee or not. Using a slightly different approach, Mol, Krahmer, Maes and Swerts (2009) let speakers retell cartoon stories to either an presumed audiovisual summarizer, an addressee in another room via webcam, or to an addressee in the same room. They found that speakers produced more and larger gestures towards human addressees than towards the computer system. Thus, gestures do not only show that speakers are aware of presence or absence of an addressee (Bavelas et al., 2008), but also whether they talk towards another human or not. In general, both studies show that gesture production is not fully automated, but that speakers account for their communicative setting when gesturing.

### **2.1.3. Summary**

Taken together, producing iconic gestures has a beneficial effect on lexical retrieval and memory, especially on spatial information. Speakers also seem to be aware of the gestures' communicative function. For example, gestures are more often produced in situations, in which a speaker wants to specify something that is underspecified in speech like size information, relative position information or the meaning of an ambiguous word. Interestingly, the information encoded in gesture and speech is often non-redundant, i.e. a certain feature is predominantly encoded in one modality. Whether and how such a feature is encoded in gesture may depend on the language a person is speaking, the difficulty of the to-

be-described concept (although there is also some counter evidence) and the communicative situation. Several theories try to account for the production of gestures, with the interface model (Kita & Özyürek, 2003) and the GSA framework (Hostetter & Alibali, 2008) being the most promising.

## **2.2. Comprehension of iconic co-speech gestures**

Although there are some good indications from gesture production research that iconic gestures might indeed have a communicative value, this evidence is rather indirect. In order to prove that gestures have a communicative value for an addressee it is necessary to look at the immediate effects of co-speech gestures on a person's comprehension. Some of the questions addressed by the gesture comprehension research are whether gestures are integrated at all with speech, what factors can influence this process, and what is the time-course and neural basis of the integration. For this purpose, behavioral, ERP, and fMRI methods can be used. In the next section, I will summarize the findings in gesture comprehension research with a specific focus on behavioral and ERP results.

### **2.2.1. Behavioral evidence**

In contrast to the behavioral research on iconic gesture production, the research on gesture comprehension is rather limited and has only been conducted since the mid-nineties (with one exception: Woodall & Burgoon, 1981, see Section 1.2.3. The relation of iconic gestures to speech: Timing – a crucial but understudied factor, p. 18). The basic question the first behavioral studies tried to clarify was whether addressees can glean additional information from watching the combination of gesture and speech in comparison to speech alone. Krauss, Dushay, Chen & Rauscher (1995), investigated the communicative effectiveness of iconic gestures in dyadic conversations. Participants had to describe an abstract stimulus either in a face-to-face situation (50 %) or via an intercom (50%). They gestured significantly more in the face-to-face situation. The descriptions were videotaped and presented either audio-visually or in an audio-alone condition to a new set of participants, who had to select the target item described from a list. All in all, there were four different conditions (face-to-face – audiovisual, face-to-face – audio only, intercom – audiovisual, intercom – audio only).

Interestingly, there was no significant effect between the audiovisual and the audio only conditions in response accuracy. This led Krauss and colleagues 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” (1995, p. 550). To date, however there is ample evidence against this conclusion. In one of the first studies to show that iconic gestures indeed are communicative, Cassell, McNeill & McCullough (1999) presented participants with short video clips of cartoon narrations that either contained gestures that matched the content of speech, mismatched it or contained no gesture at all. Note that in this study mismatch could either be an abstract pointing mismatch, a mismatch in perspective or a mismatch in the manner an action was performed. Mismatches in manner were not semantically incongruent to the speech, but rather provided additional information not present in speech. Participants had to retell the stories, which were analyzed with regard to whether information only presented in gesture would be incorporated in the retellings. The results show, that gestural information being it contradictory or supplementary is taken into account by the participants. For example, when they had heard “*and Granny whacked him one*” accompanied by a punching movement, they often told “*and Granny like punches him*”(original citations, Cassell et al., 1999, pp. 15-16). Cassell and colleagues (1999) took these so-called intrusions as an indication that gesture and speech form one tightly linked system and that iconic gestures are communicative. Support for this hypothesis comes from a whole series of experiments by Beattie and colleagues (Beattie & Shovelton, 1999a, 1999b, 2002a, 2005; Holler & Beattie, 2002) as well as from Kelly, Barr, Breckinridge Church and Lynch (1999). For instance, Kelly et al. (1999) presented very simple utterances like “My brother went to the gym.” either with or without gesture (in this case a shooting of a basketball). Participants had to recall these utterances in a paper-pencil task. In the gesture condition, Kelly et al. (1999) found significant intrusions of gestural information in the written responses, indicating that addressees integrated the gestures with speech, which is similar to the findings of Cassell et al. (1999). Moreover, Kelly et al. (1999) found that recall was significantly better in the gesture condition, indicating that gesture might have a positive effect on memory. Using a different approach, Beattie and Shovelton (e.g. 1999b) also provided evidence for the communicative abilities of gestures. They presented cartoon narrations with or without gesture and afterwards either used an interview or a questionnaire to test how well participants understood the stories. Additional gestural information was found to be particularly enhancing the addressees’ information about size and relative

position<sup>9</sup>. Thus, the authors concluded that iconic gestures were indeed communicative in a co-speech context. Applying a micro-analytic approach Holler and Beattie (2002) specified the effects of gesture on size and relative position comprehension. They were able to show that information between gesture and speech about these two categories was rarely overlapping and that gesture was especially used to encode the relative position between an agent and an object or instrument. Furthermore, Beattie and Shovelton (2002a) found that some iconic gestures can also communicate considerable information about size and relative position, even in the absence of speech. This result has to be interpreted with caution. Cartoon narrations might have triggered more pantomime-like gestures, which are communicative even in the absence of speech, because in cartoons, characters often show something like an over-acting in order to promote a joke or message. Moreover, the task included an immediate repetition of each stimulus and a very long response time (30 seconds), which might have triggered different strategies than used in normal gesture-speech processing. The result is also in contrast to the findings of Hadar & Pinchas-Zamir (2004). In their experiment, silent gesture clips were presented to untrained participants, who had to choose the right answer out of five possible ones (lexical affiliate, irrelevant distractor, visual distractor, semantic distractor, remote distractor) in a forced-choice paradigm. Participants chose the lexical affiliate in 40 % of all iconic gestures, but the irrelevant distractor was also chosen in 10 % of the cases. Based on these findings, Hardar and Pinchas-Zamir (2004) conclude that even competent communicators, who watch well-shaped gestural information, cannot unambiguously identify the meaning of a gesture. In their words, the interpretation of an iconic gesture is rather vague without context (something similar was also observed in a pre-test in this dissertation, see Section 3.2.4.2. Gesture fragment identification without speech context, p. 80). In fact, as Kelly, Özyürek and Maris (2010) have shown, the fit between gesture and speech has a substantial impact on how an addressee integrates both streams of information. They presented bimodal action primes (i.e. chop action accompanied by the utterance “chop”) followed by a bimodal target clip, containing a gesture and utterance. The target was presented in one of five conditions: Baseline (target gesture + verb correspond to action prime), weakly incongruent gesture (e.g. cutting gesture + verb “chop”), weakly incongruent speech (chopping gesture + verb “cut”), strongly incongruent gesture (twisting gesture + verb “chop”), strongly incongruent speech (chopping gesture + verb “twist”). They

---

<sup>9</sup> Interestingly, Beattie and Shovelton (2006) also showed in a production study, that relevant size information is more likely to be encoded in gesture than in speech. Thus, the combined production and comprehension data clearly show that gestural information is of high communicative value if size information is to be communicated.



found that participants were more accurate and faster in the congruent as compared to incongruent conditions. Additionally, the strength of incongruence affected the target processing, which suggests that contextual factors like semantic expectancy can modulate the integration of gesture and speech.

The studies on gesture-speech comprehension presented so far clearly show that gestures are communicative but that they can exert their communicative power only in the presence of speech. Iconic gestures, however, cannot only be very useful in communication but can also have multiple other effects on the addressee. Whereas the early studies on gesture comprehension more or less all focused on the communicative value of iconic gestures, newer studies tried to identify whether iconic gestures serve other beneficial purposes and how the information uptake from gesture might function, i.e. for example, how overt visual attention influences this process. With regard to other beneficial effects, Feyereisen (2006) investigated the effects of representational (e.g. iconics) and non-representational gestures (e.g. beats) on sentence recall and recognition. He found that representational gestures enhance sentence recall in comparison to non-representational gestures. He also observed an enhancement of recognition for congruent representational gestures but not for incongruent. Thus, iconic gestures, as a member of the family of representational gestures, can aid verbal memory. However, more research is needed to clarify the nature of this effect. Besides the effect on verbal memory, it was recently discovered that gestures can also influence a more motor related memory component. Cook and Tanenhaus (2009) were able to show that observing a gesture also influences the way an addressee produces the identical gesture later on. In particular, the form and trajectory of the produced gestures resembled the form and the trajectory of the previously observed gestures. The authors conclude that iconic gestures reliably transfer perceptual-motor information that can be used by the communicative partner. This may be very helpful to build a communicative common ground.

Though gestures might be very helpful in information transmission, in easing communication and in aiding memory, they surely are not helpful for everything. To just give an example, Hirata and Kelly (2010) compared the beneficial effects of additional lip movements or / and hand movements on phonemic vowel length discrimination in Japanese. Their participants were native speakers of English who had not learned or been exposed to Japanese before. Whereas lip movements were very helpful in vowel length discrimination, the hand movements or combination of both did not have any beneficial effect. Thus, gestures may not serve for as additional learning help for the process of vowel discrimination. Multimodal

approaches for language learning, therefore, have to carefully select which kind of additional input they use in order to convey certain aspects of a language.

While all the studies presented so far investigated specific aspects in communication or learning that could be influenced by gesture, some newer studies follow a very different approach, i.e. they want to clarify what mediates the information uptake from gesture. Gullberg and Holmqvist (2006) investigated whether addressees overtly attend to gestures produced by a speaker in a communicative situation using eye tracking. In their study, they manipulated the social setting and the display size (live, life-size video, (computer) screen video). The gestures were annotated and coded with regard to the place of articulation in gesture space, gesture fixation by the speaker and presence / absence of gesture holds. They found that a speaker's gaze (social relevance) as well as gestural holds (impact on peripheral visual processing) attracts the visual attention of addressees. Whereas fixations on holds were not affected by the social setting, gaze had the largest effect in the live (face-to-face) condition, indicating that gaze is especially important in real-life communication as it serves a social function, e.g. joint attention. However, the most striking result of this study is that addressees focus on a speaker's face almost all the time (between 92.6 % (live) and 97 % (screen video)) and only very seldom directly fixate the gestures (maximum: 7.4 % (live)). This suggests that although addressees attend to gestures at times, the very small proportion of gesture fixations do not provide evidence that potential gestural effects on communication are mediated by overt visual attention. Two further eye tracking studies extend this finding. Gullberg and Kita (2009) examined the role of gaze (social factor), gesture's location in space and gesture holds (physical factors) on visual attention and information uptake from gesture. In principal, the results are quite similar to the previous study by Gullberg and Holmqvist (2006). Gullberg and Kita (2009) were also able to show that both gaze and holds, but not space affects an addressee's visual attention. Information uptake is only influenced by a speaker's gaze, but not overt gaze following by addressees (only 8% of all trials). The authors conclude that addressees were able to retain gestural information about direction when speakers gazed at their hands, suggesting that a speaker at first might have to mark the relevant gestural information to enable addressees to use it by covertly attending to it.

One result that sheds some light on this issue has been conducted by Beattie, Webster and Ross (2010). They also found only a small number of gestures to be fixated (2.1 % of all gestures), but when they specifically looked at the most meaningful part of the different gesture categories (i.e. the stroke phase) up to 26.5 % of the gestures were fixated by

addressees. Thus, they conclude that visual attention shifts unconsciously and very fast to the most informative parts of a gesture (i.e. covert attention). Certainly, more research is needed to get a better idea, when and what addressees look at with regard to gesture, and how this is all related to any beneficial effects of gesture on communication. For example, it is necessary to clarify the impact of a speaker's gaze on gesture comprehension in more detail, as there are many ERP studies that show an impact of gesture on speech comprehension even though the speaker's face is masked. This can, for instance, be done by investigating whether gestures are perceived as more prominent in such a situation or whether gaze becomes especially important in real-life or real-life-like communication (i.e. display size is life-size).

One thing that is quite clear already is that research using gesture videos provides a good and quite accurate way to assess the communicative impact of gestures in real life. Comparing three video condition (speech only, gesture only, gesture + speech) with a face-to-face condition (gesture + speech), Holler, Shovelton, and Beattie (2009) investigated the impact of the different conditions on the comprehension "size" and "relative position" information, as inquired with a questionnaire. Interestingly, the combination of gesture and speech was sometimes even more effective in the face-to-face condition than in the video condition, indicating that the use of video material to test the impact of gesture in comprehension underestimates the communicative effectiveness of gesture.

### **2.2.2. Neurophysiologic evidence for gesture comprehension**

#### **2.2.2.1. Event-related potentials (ERPs) of the EEG as a measure of gesture comprehension**

It is well known that speech comprehension is a very complex and fast process. Within just a few hundred milliseconds the comprehension system, i.e. the "parser", has to analyze all different kinds of information comprised by an utterance, e.g. phonetic, prosodic, syntactic and semantic information. If the timing and temporal order of these different sub-processes is to be investigated, one has to use a measure that has a high temporal resolution. ERPs with its excellent temporal resolution in the range of one millisecond seem to be especially well suited for investigating the nature and timing of speech comprehension. Like speech, iconic co-speech gestures are very complex and subject to rapid dynamic changes. In order to be integrated with the corresponding speech they have to be processed in parallel and at a comparable speed. Thus, ERPs also present a very adequate measure if one is interested in

the neuropsychological basis of gesture-speech interaction in comprehension. In the following section, I will briefly introduce the principles of the ERPs of the EEG (for more detailed information, see Handy, 2005; Luck, 2005), then I will focus on the ERP component of interest for the present experiments, i.e. the N400, and finally I will review the ERP literature on gesture-speech comprehension so far.

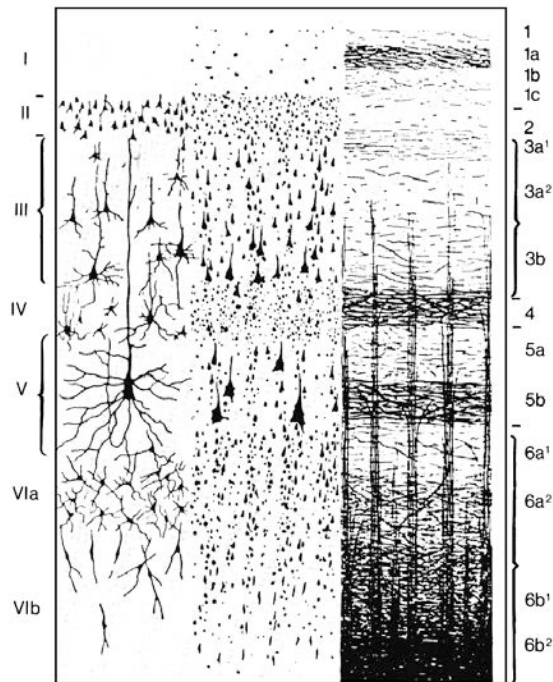
#### 2.2.2.2. The nature of the human EEG

The EEG itself is a rather “old” method, i.e. the first reported human recording already stems from 1929 (Berger, 1929). But it is still one of the main work-horses of neuropsychology due to its excellent temporal resolution. Since then, a lot of different EEG components, related to specific mental states and processes, have been identified. EEG oscillations consist of different frequency bands, with the range depending on settings of the EEG recording. For example, for most neuropsychological studies, all frequencies above 100 Hz are not of interest and thus filtered out.

In principal, the EEG measures electric activity (oscillations) from the human scalp, which is elicited by ionic currents in the brain (Gall, Kerschreiter, & Mojzisch, 2002). This is only possible because of the specific structural organization of the human cortex. Pyramid cells, which constitute 85 % of the neocortex neurons are the main generators of these currents (Nieuwenhuys, 1994). They are perpendicular to the cortex surface with their somata mostly located in layers III, IV and V of the cortex (see Figure 2.1).

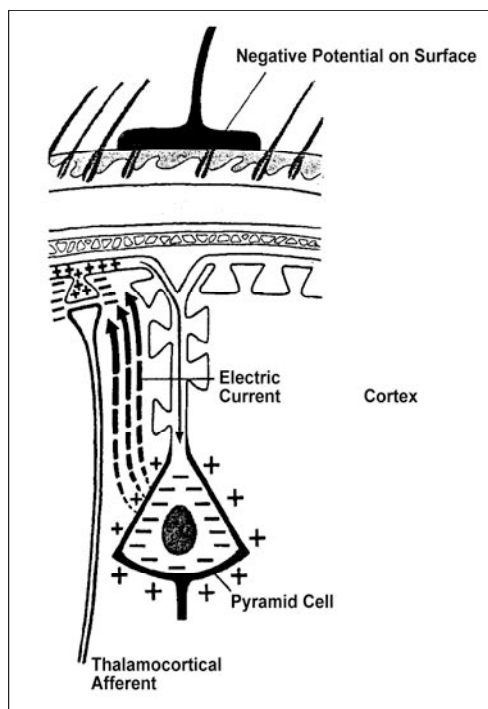
In the cortical layers I and II the dendrits of the pyramid cells form synapses with afferent neurons from the thalamus. At these synapses as well as at somata of the pyramidal cells, excitatory and inhibitory postsynaptic potentials (EPSPs & IPSPs) generate current flows which cause the neuron to act as an electric dipole (see Figure 2.2). In order to measure current changes with the EEG technique at the scalp surface, at least 1000 simultaneously firing pyramidal cells are needed to generate a sufficient electric potential (Rugg & Coles, 1995). Note that, for example, EPSPs at the apex and IPSPs at the soma of a neuron can cause voltage changes with the same polarity, which are reflected differently on the scalp surface, because of the asymmetry of the dipoles. Therefore, it is impossible to clearly determine the neural basis of a certain voltage change based on EEG data alone (Kandel, Schwartz, & Jessell, 2000). This is called the *inverse problem*. It means that there is not a

single dipole configuration that can explain a certain voltage distribution, but that an infinite number of dipole distributions can result in the identical distribution (Helmholtz, 1853; Nunez, 1981). Consequently, it is impossible to know which one of the possible configurations is the actual source of the observed distribution (for potential solution on how to overcome this problem, e.g. see Luck, 2005). As can be inferred from the *inverse problem*, spatial resolution is certainly not the strength of the ERP method.



**Figure 2.1** Layers of the human cortex. On the left, somata and dendrites of the pyramidal cells are visible, in the middle, the columnar organization of the somata can be seen, and on the right, the distribution of fibers is shown (illustration adapted from Thompson, 1990).

The EEG signal is usually measured as the difference potential between two electrodes, one scalp electrode and a reference electrode. In most of the cases the reference consists of a non-scalp electrode that is thought to be uninfluenced by brain activity, e.g. an electrode placed on the mastoids, nose or earlobes. To assure comparability across different experiments, guidelines for electrode placement have been established. In 1958, Jasper developed the classical 10-20 system, which included the positioning of 19 electrodes in a specific orientation on the scalp. As nowadays most studies use a larger set of electrodes a new system with adjusted guidelines was developed by the American Electroencephalographic Society (Sharbrough et al., 1991). According to these guidelines, a subset of 59 scalp electrodes was used in the present experiments (for an example, see Figure 2.3). Because the signal strength of the EEG is very small, it needs to be amplified. The amplified signal is digitized at a certain frequency (i.e. 500 Hz in the present case) during the recording.



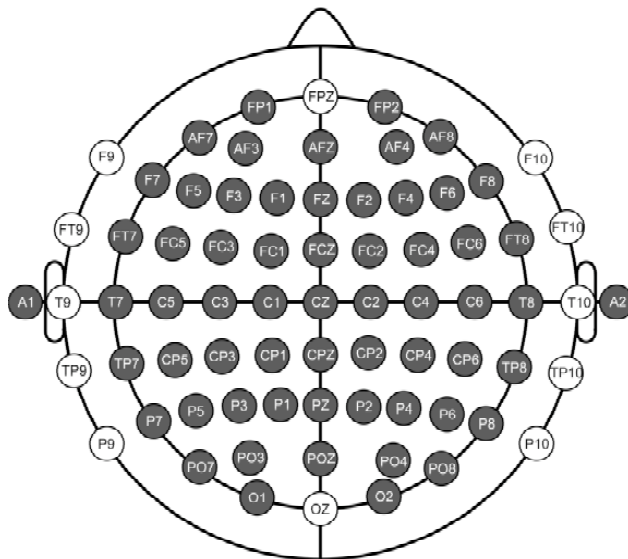
**Figure 2.2** Schematic illustration of the current flow after the excitation of an apical dendrit of a pyramidal cell (adapted from Birbaumer & Schmidt 1990).

It should be noted, that the digitization parameters determine the frequency range of the to-be analyzed data. According to the *Nyquist theorem*, an analogous signal can be digitized without any problems if the sampling rate is at least twice as high as the highest frequency of interest in the signal. Otherwise, a loss of signal information and aliasing (i.e. induced low frequency artifacts) will occur and disrupt the EEG signal quality. The brain activity recorded during the EEG measurement can be separated into two different components: spontaneous brain activity which mainly caused by rhythmic thalamic afferences and even-related activity related to the stimulus manipulation.

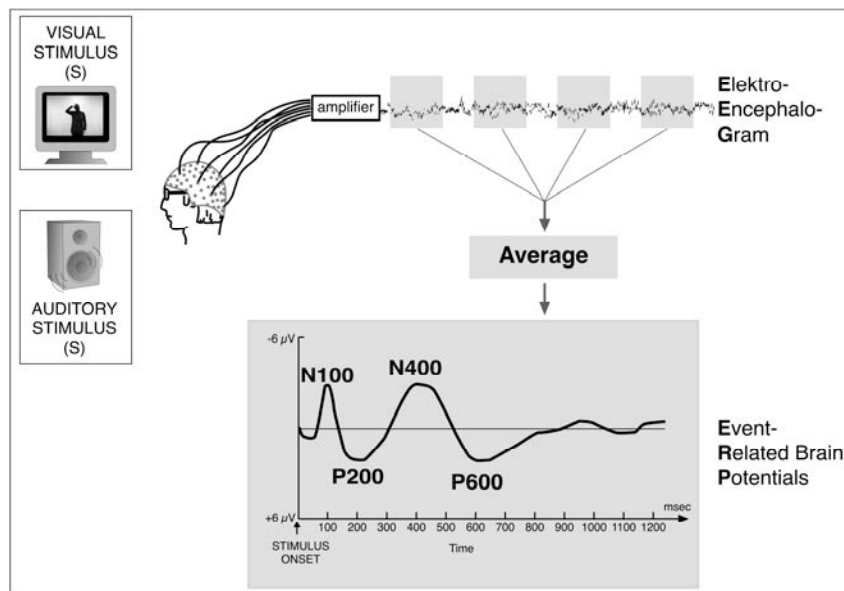
#### 2.2.2.3. Event-related potentials (ERPs)

ERPs constitute small voltages changes in continuous EEG which are time-locked to a certain external event, e.g. visual or auditory speech stimuli, pictures, etc. It is a widespread assumption, that the ERP is related to the neural processing of the event. In general, the amplitude of the ERPs is much smaller than that of the spontaneous voltage changes. Thus, it is necessary to apply an averaging procedure to minimize the random variation of the EEG (see Figure 2.4). The background for this procedure is that the time-locked ERPs always will have the same characteristic form whereas the spontaneous brain activity randomly varies. Therefore, by averaging the background activity is reduced leaving only the measure of interest for further analysis. In other words, averaging increases the signal to noise ratio in the measurement.

The resulting ERPs consist of several different positive and negative deflections which are typically observed with a specific process or experimental manipulation. These deflections are termed components and can vary considerably in latency, amplitude, duration, polarity and topography. Typically, they are measured in relation to a pre-stimulus baseline time window within which the average amplitude is zero per definition. Early components, i.e. those that occur up to 100 ms, are termed exogenous, as they occur mainly due to physical properties of the stimuli (e.g. the P1 varies with the luminance of a stimulus).



**Figure 2.3** Electrode setup used in Experiment 2 (in all other experiments the electrodes in row 1 and 2 were replaced with the electrodes in row 9 and 10 (standard EEG setup)).



**Figure 2.4** ERP averaging procedure (adapted from an illustration by Coles & Rugg, 1995).

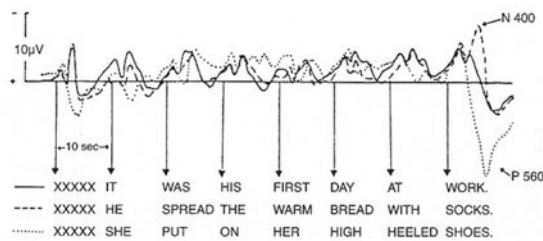


Deflections that occur after 100 ms do not exclusively depend on external factors, but rather depend on internal factors, e.g. the task a participant has to perform (P3). Therefore, the later components are called endogenous. It is important to bear in mind that the ERP method also has some limitations. The EEG signal corresponds to the summed potential of synchronously firing cells in the neocortex with a specific alignment to the cortex surface. Thus, it is only possible to measure a certain kind of brain activity with the EEG, which in turn means that a large amount of brain activity is not measurable with this method. It is also important to note, that the relation between the real neural activity and the components of the ERP is purely correlational, not causal. Therefore, one can never be a 100% sure whether a certain component really reflects the process of interest or just resembles some different process. The excellent temporal resolution of the ERPs, however, makes up for these and other disadvantages. For example, numerous ERP studies have been published over the past decades, which sought to identify subcomponents of language comprehension, including components for prosodic, syntactic and semantic processes (for an overview, see Friederici, 2004). One of these components, the so-called N400, has been particularly associated with the semantic processing of speech (e.g. Kutas & Hillyard, 1984), and even more importantly for the present work, the semantic integration of gesture and speech (e.g. Kelly et al., 2004). The properties of this component will be described in the following section.

#### 2.2.2.4. The N400

The N400 is probably the most extensively studied ERP component in language processing. It represents a negative waveform deflection peaking around 400 ms post stimulus onset was first reported by Kutas and Hillyard (1980). They visually presented sentences that either ended with a semantically correct word (e.g. *He spread the warm bread with **butter**.*) or a semantically incorrect word (e.g. *He spread the warm bread with socks.*) or a semantically correct, but physically incorrect word (e.g. *He spread the warm bread with **BUTTER**.*). An increased negativity compared to the ERP of the semantically correct word peaking about 400ms after the onset of semantically incorrect word was found in the semantically incorrect condition (see Figure 2.5). No such enhancement was found for the physically distinct, but semantically correct condition. Kutas and Hillyard (1980) labeled the deflection for semantically incorrect words compared to the correct words N400. This “classic” N400 in sentence processing has been replicated in numerous studies using different paradigms for

both visual speech stimuli (e.g. Bentin, McCarthy, & Wood, 1985; Holcomb, 1993; Van Petten & Kutas, 1987) and auditory speech stimuli (e.g. Connolly, Stewart, & Phillips, 1990; Friederici, Pfeifer, & Hahne, 1993). Not only semantically incorrect words can elicit a N400 but also semantically correct words which are highly improbable in a certain context (Kutas & Hillyard, 1984). Similar results were found for pseudowords (e.g. Holcomb & Neville, 1990). Using auditory and visual semantic priming (words were used as primes) in a lexical decision task, Holcomb and Neville (1990) found a larger N400 for pseudowords compared to related and unrelated words. Recent studies (Hagoort, Hald, Bastiaansen, & Petersson, 2004; van Berkum, Hagoort, & Brown, 1999) have shown that the N400 is influenced by sentence level meaning, discourse level meaning and word knowledge. Thus, the N400 not only reflects semantic anomaly but is rather an indicator of how well a word fits into a given context<sup>10</sup>.



**Figure 2.5** Example for an N400 effect (taken from Kutas and Hillyard, 1980).

Over the past few years, the N400 component has also been found for picture stimuli (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999), videos (Sitnikova, Kuperberg, & Holcomb, 2003) and gestures (e.g. Kelly et al., 2004). For example, Sitnikova et al. (2003) presented short movie clips of everyday situations (e.g. applying shaving cream), which either ended with an actor using an contextually congruent object (razor) or incongruent (tooth brush) object. They found, that the N400 varied with object congruency, with

<sup>10</sup> It should be noted that the N400 was also found in the studies that manipulated thematic role assignment (e.g. Bornkessel, 2002; Frisch & Schleewsky, 2001) and animacy (e.g. Frisch & Schleewsky, 2001; Weckerly & Kutas, 1999). However, the "thematic" N400 seems to be not only be functionally, but also physiologically different from the "semantic" N400 (Roehm, Schleewsky, Bornkessel, Frisch, & Haider, 2004). In the present series of experiments, the N400 is assumed to be a correlate of semantic integration.

incongruent objects eliciting a larger N400 than congruent ones, indicating that participants were capable of rapid on-line integration of real world events.

Similar to the study by Sitnikova et al. (2000), most studies which investigated the online integration of gesture and speech with ERPs were based on a mismatch paradigm (e.g. Kelly et al., 2004), with the exception of a study by Holle and Gunter (2007), who used a disambiguation paradigm. In the following, I will present the results of the ERP research on iconic gestures so far. I will first focus on studies which used a mismatch paradigm, before I will describe the disambiguation paradigm used by Holle and Gunter (2007) as well as their results.

#### 2.2.2.5. ERPs as a correlate for gesture-speech integration

As already stated above, there is good behavioral evidence that gesture are communicationally intended and used by addressees. Behavioral paradigms, however, can only provide a first glimpse of what might be going on in gesture-speech integration, because they cannot clarify the time-course and neural underpinnings of this process. ERPs are especially helpful in this respect, as they provide both excellent temporal resolution as well as constitute a correlate of neural activity. The interaction of gesture and speech also seems to be very well suited for ERP application. There are two paradigms, which have been used by researchers in order to address questions related gesture-speech comprehension: the mismatch paradigm and the disambiguation paradigm. First, I will review the studies using the mismatch paradigm. After that, I will describe the basics as well as the results of a series of experiments using the disambiguation paradigm in detail (Holle & Gunter, 2007).

#### *The mismatch paradigm*

In the mismatch paradigm a match condition (gesture and speech convey the same information) is compared with a mismatch condition (gesture and speech convey different information). In the first EEG study using this paradigm, which is in fact the first ever EEG study concerning iconic gesture-speech integration, Kelly et al. (2004) investigated whether gestures can influence speech comprehension and if so in what time-course they would influence it. Participants saw videos in which a male actor was sitting at a table behind a tall,

thin glass and a short wide dish. The actor made utterances with respect to salient dimensions of these objects, e.g. high. Each of these word was accompanied by one of four conditions of gesturing (*match*, *mismatch*, *complementary* and *no gesture*), resulting in four different gesture-speech relations. For example, in the match condition an actor said “*tall*” while indicating the tallness of a tall, thin glass via a gesture. The subjects had to respond to the gestures by pressing the response button corresponding to the referent of the gesture (which was either the tall, thin glass or the short, wide dish). ERPs were measured time-locked to the onset of the utterance for all conditions. The N400 was larger for gesture-speech mismatches (e.g. the actor says “*tall*” while gesturing the shortness of a short, wide dish) compared to gesture-speech matches, indicating that incongruent speech is more difficult to integrate into a gesture context than congruent speech. Kelly et al. (2004) also found some earlier, pre-semantic effects (N1, P1, P2) for gesture-speech integration with mismatching and complementary conditions eliciting larger deflections at each of the three components as compared to the matching and no gesture condition. Based on these results, the authors concluded that gesture does not only affect speech comprehension on a semantic level (N400) but also on various other more perceptual levels. Because other ERP studies do not report such early effects, these pre-semantic effects have to be interpreted with caution. They might be driven by the specific stimulus and / or task, which was in principal based on a pure perceptual matching of visual inputs. Nevertheless, the N400 results clearly support the view that gesture and speech are parts of a tightly integrated system (McNeill, 1992). In a series of ERP studies, Wu and Coulson (Wu, 2005; Wu & Coulson, 2005, 2007a, 2007b, 2010) assessed several different aspects of the semantic processing of iconic gestures. In her first study, Wu (2005) investigated whether gestures are similarly processed as picture probes when preceded by either a congruous or incongruous context. Participants were either presented with soundless videos of short dynamic gestures or with a picture extracted from the video. In both conditions, she found a larger N400 in the incongruent as compared to congruent context condition. Pictures additionally elicited a N300, a component typical for the semantic processing of static pictures (Holcomb & Mcpherson, 1994). Wu (2005) concluded that semantic understanding of iconic gestures is similar as the processing of pictures and words. The similarity in the semantic processing of words and gestures was the main focus in another study (Wu & Coulson, 2005). In two experiments participants watched cartoon clips paired with soundless videos of gestures which were either congruent or incongruent to the cartoon. Then a visually presented probe word followed. ERPs were measured time-locked either to the onset of the gesture or to the onset of the probe word. In

the first experiment, participants had to judge congruency of gesture video and cartoon, whereas they judged the relatedness of probe words and preceding cartoon-gesture pairs in the second one. Wu and Coulson (2005) found an N400-like component both for gestures and probe words which was significantly larger for the incongruent condition than for the congruent condition. Additionally, they found that congruency also had an effect on probe word comprehension. The N400 for related probe words following incongruent gesture-speech pairs was larger than following congruent pairs. No such effect was found for unrelated probes. This result clearly shows that gestures can affect the processing of related words. As in the first study by Wu (2005), the results indicate that gestures are semantically processed similar to the processing of other meaningful representations, e.g. pictures or words. Using a priming paradigm with gestures as prime, followed by related or unrelated probe words, Wu and Coulson (2007b) provided additional evidence for the tight semantic link between gesture and words. They found a larger N400 at semantically unrelated target words as compared to related ones indicating that gesture primed the meaning of the probe words. They also went beyond showing a semantic link between gesture and speech by presenting evidence that gesture may specifically aid speech comprehension. The authors demonstrated that gestures enabled addressees to better conceptualize the visuospatial aspects of an utterance, which goes hand in hand with the behavioral literature on the beneficial effects of gestures. Taken together, both the studies by Kelly et al. (2004) as well as Wu and Coulson (e.g. 2005) show that there is a semantic integration of speech and gesture and that both gesture and words elicit similar brain responses in mismatch paradigms. Whether the time course of this process, i.e. the semantic integration into a preceding sentence context, is practically identical for gestures and words was of interest in a study by Özyürek et al. (2007). These authors were interested in whether speech and gesture are integrated simultaneously or whether speech is integrated first and gesture later. Additionally, they wanted to know how the integration of gestures into a sentence context (global integration) compares to the integration of gesture with a co-expressive speech unit (local integration). In the experiment participants saw gesture videos accompanied by speech. In contrast to other studies, gestures started immediately with the stroke phase which was temporally aligned with the co-expressive speech part. The semantic fit of gesture and speech (critical verb) to the preceding context was manipulated, resulting in four different conditions: correct condition (both gesture and speech fit to context), language mismatch condition (only gesture fits to context, speech not), gesture mismatch condition (only speech fit to context, gesture not) and double mismatch condition (both gesture and speech do not fit into context, but semantically fit

together). Participants saw the gesture videos but did not have to accomplish any task. Özyürek et al. (2007) found similar N400 patterns as well as scalp distributions for all mismatch conditions (*no mismatch, gesture mismatch only, speech mismatch only, mismatch of both gesture and speech*). These results showed, once again, that semantic information cannot only be gleaned from speech but also from co-speech gestures. Yet more importantly, this study provides evidence that both gesture and speech are integrated within the same time-course, i.e. both are processed on-line. Interestingly, the double mismatch condition (gesture + speech) did not differ from the single mismatch, indicating that there is no additive effect of both mismatches. As possible explanation is, that since gesture and concurrent speech matched with regard to their meaning, they were locally integrated into a single meaningful representation. The new representation became a single global mismatch and elicited therefore the same ERP response as the “real” single mismatch. The study by Özyürek et al. (2007) therefore provides some evidence that semantic integration of gesture and speech into a sentence context relies on similar processes. However, the results do neither imply that gesture and speech are processed by a single system as proposed by McNeill (1992) nor provide sufficient evidence for the more general idea of a “single unification space” for the integration of all kinds of information with language (Hagoort, 2003, 2005).

Is the integration of gesture and speech also an automatic process as McNeill states (“*the point we wish to emphasize is the involuntary, automatic character of forming an idea unit out of the information from the two channels*”; McNeill et al., 1994, p. 236)? A small number of ERP studies give a first idea. For example, Kelly et al. (2007) showed participants short clips containing gesture and speech that were either congruous or incongruous. Half of the material were what they called intentionally coupled (i.e. gesture and speech were produced by the same person), the other half were not. This latter manipulation led to different distributions in the N400 effect between congruous and incongruous gesture-speech combinations. If gesture-speech integration was automatic as proposed by McNeill et al. (1994), the intention manipulation should not have affected the information uptake from gesture. The results from Kelly et al. (2007), however, have to be interpreted with caution. The authors told participants in the instruction that they would be presented with stimuli where gesture and speech belong together or not. Thus, not only the task manipulation, but also the explicit instruction might have influenced the findings. In a follow-up study, Kelly et al. (2010) were able to specify these earlier findings using a different task (gender discrimination instead of attending to certain speech items). Again there was an N400 effect

between congruous and incongruous gesture-speech combinations, but no differences in scalp distribution. Interestingly, this effect was larger when the gender of gesturer and speaker matched than when they did not, suggesting that gesture-speech integration is under some degree of cognitive control, and thus not automatic. Recently published data by Wu and Coulson (2010) point in the same direction, although their finding concerns an effect quite different from that reported by Kelly et al. (2007). Using a priming paradigm with multi-modal primes (gesture present vs. gesture absent (i.e. a stillframe of the stroke onset)) and pictures as targets (related, unrelated), the authors investigated the difference in the integration of gestures and still frames with the first and second content word in an utterance. After each target, participants had to perform an old-new-memory task either related to a word, video frame or picture. They found a reduced N400 triggered to the onset of the first content word for the integration of a concurrent gesture as compared to the integration of a still frame. No such difference was found at the second content word. Wu and Coulson (2010) conclude that gesture information is especially useful early in the discourse, when new information is introduced, but doesn't help at a later stage, when already known information is often repeated. Thus, gestural information does not impact speech at the same level all the time, but this process may rather depend on the amount of new, additional information provided by gesture. The findings by Wu & Coulson (2010), therefore, also clearly speak against the automaticity notion by McNeill et al. (1994).

Taken together, the mismatch studies showed that gesture and speech are tightly linked in language comprehension. They can affect speech on a semantic level and are semantically integrated into a preceding speech context similar to words (Özyürek et al., 2007). It should be noted, that this does not imply automaticity of gesture integration as some studies have already shown.

The mismatching paradigm has some disadvantages. First, the external validity is low, as in normal life there are usually no mismatches. Second, a mismatch can only show how gesture impairs language processing, but not how it can facilitate these processes (although (Feyereisen, 2006) claims this). The disambiguation paradigm, however, might be an interesting solution for investigating facilitative effects in a natural context.

*The disambiguation paradigm*

In everyday conversation, transmitting information is a very complex process and though communicative partners try their very best to be as precise as possible in their communicative effort, there are inevitably situations in which the communicative signal of a speaker is too unspecific or ambiguous. Nevertheless, we can cope with such situations in a very fast and efficient way by taking into account all the contextual information available to identify the correct meaning of an utterance (see also Twilley & Dixon, 2000). Ambiguity in speech occurs on many levels, e.g. syntactic, pragmatic, or lexical / semantic level. For example, a sentence like “*The man noticed the boxer.*” is lexically ambiguous, because it is not clear as to whether the man saw a boxer in the sense of a sportsman or a dog. Observational studies by Holler & Beattie (2003) as well as Kidd & Holler (2009) have shown that adults as well as children tend to produce gestures to clarify such lexical ambiguities in conversation.

The question that Holle & Gunter (2007) addressed in their experiments was whether addressees actually use the gestural information provided by a speaker for the disambiguation of temporally ambiguous sentences. The ambiguity in their material was provided by the use of an unbalanced homonym like ball or boxer. Homonyms are words that are orthographically and phonologically identical but have two distinct unrelated meanings, e.g. in the case of ball: sport or dance. In this regard, unbalanced means, that one of the two meanings is the more frequent, dominant one (in this case: sport) and the other one the less frequent subordinate one (dance). Research on homonym processing (for a review, see Twilley & Dixon, 2000) has shown that there are two types of information that allow listeners to select the correct meaning of a homonym: a) word meaning frequency and b) contextual constraints. There has been a lot of discussion how these factors interact in activating the meaning of a homonym. By now, it is clear that in the absence of any biasing context or in a neutral context, word meaning frequency alone determines the activation. In such cases, a listener always selects the dominant meaning, because it has a higher word meaning frequency than the subordinate meaning and is therefore more often the actual meaning of a homonym. This is in line with a number of findings in the homonym literature (Simpson, 1981; Simpson & Burgess, 1985; Vu, Kellas, & Paul, 1998).

If a homonym is preceded by a biasing context, the pattern of activations is more complex. Simpson (1981) as well as Martin and colleagues (1999) systematically investigated the effects of varying context on meaning selection. Context ranged from weak bias to strong



bias for either the subordinate or dominant meaning. Their findings indicate that the activation of the dominant meaning can only be modified by a strong context but not by a weak one. The activation is stronger when the homonym is preceded by a strong congruent (dominant) context as compared to a strong incongruent (subordinate) context. In contrast, the activation of the subordinate meaning always seems to be influenced by the preceding context, regardless whether it is strong or weak one. Congruent, subordinate context always enhances the activation of the subordinate meaning, whereas incongruent, dominant context always leads to a lower activation of the subordinate meaning (e.g. Simpson & Krueger, 1991; Vu et al., 1998). Thus, the activation of the subordinate meaning varies reliably dependent on context congruency.

Using ERPs, Holle & Gunter (2007) explored whether participants would use gestural information as a cue for homonym disambiguation and whether the context provided by gesture was a strong or weak one. In all of their experiments they used sentences containing an unbalanced homonym. The homonym was disambiguated downstream at a target word (either dominant or subordinate) occurring later in the sentence. Simultaneous to the homonym an iconic gesture was displayed to the participants. Iconic gestures could either be congruent to the meaning of the target word or not. For example, each homonym could be accompanied by the dominant or the subordinate gesture, which could be followed by either the dominant or the subordinate target. A dominant gesture that was followed by a dominant target was termed “congruent gesture” while a subordinate gesture followed by the dominant target word was termed “incongruent gesture” (and vice versa). Thus, a 2 x 2 factorial design (gesture: dominant vs. subordinate; target word: dominant vs. subordinate) was applied to investigate the effect of gesture-target congruency on homonym processing. ERPs were measured time-locked to the target word. In the first experiment, participants had to judge whether the gestures conveyed the same meaning as the speech. It was found that the N400 was smaller for congruent gesture-speech pairs than incongruent ones indicating that gestures can be used by listeners to disambiguate speech. In the second experiment, the results of Experiment 1 were replicated using a less explicit task (monitoring task) suggesting that using gestures to disambiguate speech is to some extent task-independent. Yet, the effects were a little bit smaller than in Experiment 1 indicating that the task may have exerted some influence on the extent to which gestures were used. In the Experiment 3, a third gesture condition was added, i.e. grooming or self-adaptors. Grooming constitutes movements like scratching the nose or rubbing your ear. These movements do not convey any important

semantic information and are therefore unrelated to the homonym and the target word. The same task as in Experiment 2 was applied. Only for the subordinate gestures an N400 effect was found, i.e. the N400 was significantly smaller for congruent subordinate gestures than incongruent dominant gesture or grooming. This result implies that gestures selectively can facilitate the processing of the subordinate meaning. The N400 for dominant gestures did not show any effect of congruency. Thus, it seems that adding grooming changed the influence of iconic gestures on speech comprehension. It is suggested, that the integration of gesture and speech is not an obligatory process but can be influenced by contextual factors such as the amount of meaningless hand movements.

#### *Iconic gesture and language learning*

Iconic gestures cannot only aid the understanding of our mother tongue, but also support the learning of a foreign language as has been shown by Kelly, McDevitt and Esch (2009). They measured ERPs to Japanese words that had been learned either with or without gesture in a three-day training. While they found no N400 effect, they observed a larger Late Positive Complex (LPC, an index for recollection) for words learned with gesture as compared to words learned without gestures, indexing that additional gesture information is indeed helpful in second language learning.

#### 2.2.2.6. fMRI results on gesture-speech integration

While ERP research has provided some very essential insights into gesture-speech processing, it is not very helpful for reasons mentioned above if one is interested in localizing brain structures involved in the integration of both streams of information. Only in the last few years, scientists in gesture research became interested in answering this question with the help of functional magnetic resonance imaging (fMRI). This method is especially suited to identify brain regions involved in cognitive processes. In contrast to EEG / ERPs, the fMRI technique has very low temporal resolution, but allows the exact localization of neural activity in the brain within the range of a few millimeters. Before I will summarize the results of the fMRI research on iconic gesture comprehension, I will introduce the so-called Blood-oxygenation-level dependent (BOLD), which is the dependent variable in fMRI measurements. This section is mainly based on Huettel, Song & McCarthy (2008).

Using properties of blood as a measure for brain activity is a concept that has been around for more than hundred years, but it took until 1990, when Ogawa and colleagues were the first to report such a measure. They acquired T2\*-weighted images<sup>11</sup> from rodents which either breathed 100% pure oxygen or normal air. In the later condition, the blood vessels in the brain were clearly visible on the brain images in contrast to the 100% pure oxygen condition. This effect, which later became known as the BOLD effect, occurs, because deoxygenated hemoglobin is paramagnetic in contrast to oxygenated blood, which in turn leads to local changes in the tissue surrounding blood vessels transporting the deoxygenated blood. These changes can be observed as greater signal loss (i.e. darker areas) in T2\*-weighted images of the respective region. It is important to note, that the BOLD response does not directly reflect neural activity, but is a rather indirect measure for it. Yet, the BOLD response is clearly correlated with neural activity. Interestingly, this correlation is not negative as one might expect (i.e. increased activity leads to a decrease of oxygen), but positive as many studies have shown (for review, see Nair, 2005). This can be explained by looking at the typical shape of the BOLD (hemodynamic) response. The response starts with an initial dip which is followed by a large rise in blood flow with a peak after a few seconds, the return to the baseline value and a so-called undershoot, i.e. the hemodynamic response drops below the baseline. Whereas, the initial dip might reflect something like deoxygenation, the more interesting part is the large increase in blood flow that follows, which is assumed to be due to the dilation of arteries in order to provide enough oxygen for the neurons engaged in a certain cognitive process. As already noted above the spatial resolution of the BOLD is excellent, however due to the nature of the hemodynamic response with a peak delay of several seconds the temporal resolution is very limited. When analyzing fMRI data, the brain activity for different experimental conditions can either be compared with a baseline condition or subtracted from each other in order to obtain brain regions that are specifically involved in the processing of a certain condition. If one is interested in identifying brain areas for gesture-speech integration, there are some special criteria for analysis that can be adopted from multisensory integration research (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Calvert & Thesen, 2004; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005): superadditivity, inverse effectiveness and response depression. For example, one can test if a potential multisensory integration site has superadditive properties or not. If the target region

---

<sup>11</sup> T2\*-weighted images give information about the relative T2\* value in a certain tissue. The T2\* value is related to the signal decay in transversal net magnetization due to inhomogenities in the local magnetic field as well as due to the mutual influence of the spins of the magnetically excited H-nuclei on each other (for more details about the basics of the fMRI techniques, see Huettel et al., 2008).

is really involved in multisensory integration, the multisensory condition should elicit a larger BOLD response in this region than the sum of both unisensory conditions, thereby showing superadditivity.

In the following, I will shortly sum up the results of the fMRI studies that used iconic gestures as a stimulus material. The main goal of most of these studies (e.g. Dick et al., 2009; Green et al., 2009; Holle et al., 2008; Holle et al., 2010; Willems et al., 2007, 2009) was to identify the brain regions involved in gestures-speech integration. All the mentioned studies differ considerably with regard to video material, task, experimental setups and statistical analysis and the definition of integration which makes it difficult to directly compare the results. For example, Willems et al. (2007) used a mismatch paradigm, whereas Holle et al. (2007) used a disambiguation paradigm and Dick et al. (2009) in turn simply presented video-taped narrations. Since these studies are so distinct, I will not go into detail about the specific research questions of each of them, but rather focus on the commonalities between the studies, i.e. to identify the neural correlates of gesture-speech integration. Therefore, in the following, I will only report the findings and assumptions of the different studies concerning this process (see Figure 2.6).

When looking at all the results and their interpretations, it immediately becomes clear that there are two regions in the brain that are especially sensitive to the combination of iconic gesture and speech: the Inferior Frontal Gyrus (IFG) and especially its left hemispheric part, and the bilateral Superior Temporal Sulcus / Gyrus (STS/G). However, there is some ongoing dispute with regard to which of these regions the integration of gesture and speech really takes place. To date, there are (at least) three distinct views on this issue. On the one hand, there are some researchers which claim that the left Inferior Frontal Gyrus (IFG), in particular Broca's area and its adjacent regions (BA45/47), are the putative region of gesture-speech integration (e.g. Willems et al., 2007, 2009). For example, Willems et al. (2009) assume that the left IFG is the sole area in the brain, where new, combined representations out of iconic gesture and speech information are constructed, whereas in their view the STS/G regions do not contribute in any way to this process. On the other hand, there is also a similar number of studies which show that bilateral Superior Temporal Sulci / Gyri (STSs/Gs) may function as the primary multimodal integration region for iconic gesture and speech (e.g. Green et al., 2009; Holle et al., 2008; Holle et al., 2010). These findings are in line with results from other multimodal integration research. For instance, Beauchamp, Lee, Argall & Martin (2004) demonstrated that the STS is also involved in the integration of animal pictures with the

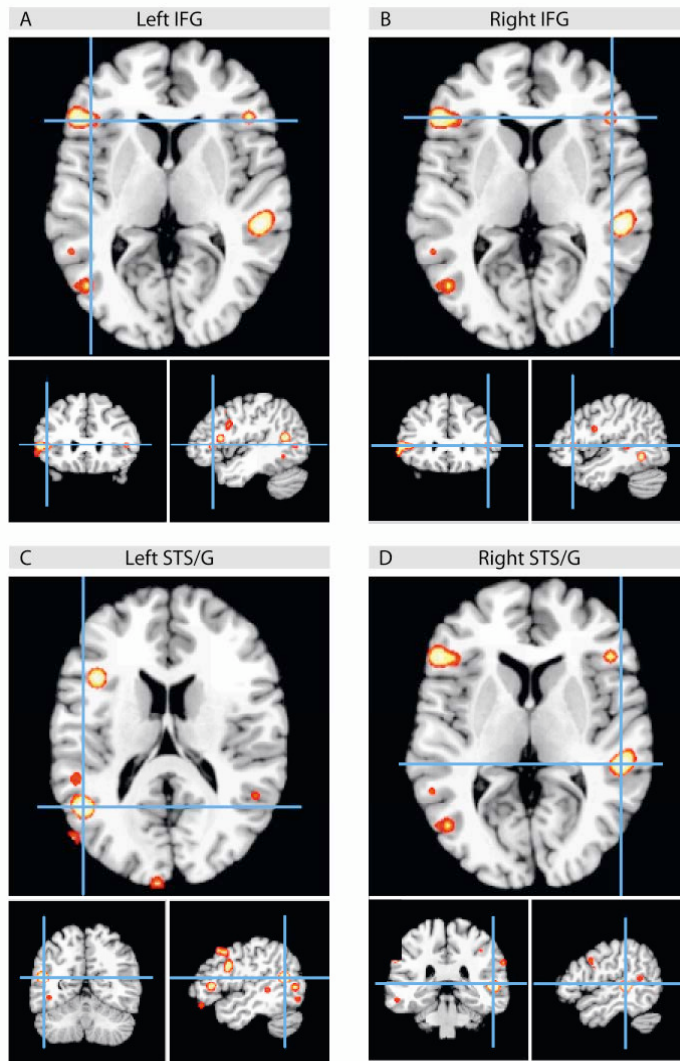
corresponding sounds. Especially, the left STS/G region has been found to be a very reliable correlate for gesture-speech integration, as it shows both superadditivity (e.g. Holle et al., 2008) and inverse effectiveness (e.g. Holle et al., 2010). It should be noted, that, both these studies also show some left IFG activations, which are however sub-threshold. Only recently, Dick et al. (2009) proposed a third idea how the different brain regions may contribute to gesture-speech integration. According to their view, both left IFG and bilateral STS/G regions are involved in the perceptual processing of the communicational input, whereas only the right IFG shows involvement in the semantic integration. The authors base their claim on the results of a fMRI study in which they presented participants with stories under three audiovisual (gesture and speech, self-adaptor (meaningless hand movement) and speech, still-frame and speech) and one auditory-only condition (speech only). For the bilateral posterior STS regions, they found stronger activations when speech was accompanied by gesture irrespective of the gestures' semantic relation to speech. Dick et al. (2009) suggest that the bilateral STSs might be involved in the general processing of biological motion but do not play a role in the semantic integration of gesture and speech. The bilateral anterior IFGs (pars triangularis) also showed stronger activations when speech was accompanied by hand movements than when not, which is in line with other work suggesting that the IFG is important for audiovisual integration (e.g. see Romanski, 2007 for a review). However, whereas the left IFG was not sensitive to semantic manipulation, the right IFG showed a significant stronger activation when speech was accompanied by a self-adaptor as compared to a gesture. Based on this finding, Dick et al. (2009) argue that the right IFG distinguishes meaningful from meaningless hand movements in gesture-speech integration. Summing up, it seems clear that both the IFG and STS/G brain regions play an important role in gesture-speech integration. However, much more work is needed to clarify the precise function of the different brain regions in this process.

Although most of the fMRI research on gesture-speech integration is concerned with the identification of the brain regions involved in this process, there is also a recent study that looked at a completely different aspect of gesture-speech processing. Macedonia, Müller, and Friederici (2010) were interested in identifying the neural basis of the beneficial effect of iconic gestures on foreign language learning. They had participants learn novel words either with gesture or meaningless hand movements. Once the training was finished, participants were subjected to a fMRI scan while accomplishing a word recognition task. The brain activity for words learned with a gesture was contrasted with the activity for words learned

with meaningless movements. Premotor cortex regions were more active when words were learned with gestures as compared to meaningless movements, whereas the reversed contrast resulted in a widespread network of activations related to cognitive control. The authors assume that the memory performance for words learned with gesture is driven by the motor imagery matching the semantic representation of the words.

### **2.3. Summary of the literature on iconic co-speech gestures**

Up to now, iconic gesture research has provided evidence for a very flexible interaction of gestures and corresponding speech. Gesture production results indicate that gesture aids lexical retrieval, strengthens memory performance and is used by the speaker to specify speech. Thus, gestures are used to boost cognitive processes as well as communication. The likelihood of conducting a gesture depends on the interpretation of a communicative situation by the speaker (e.g. “Is there ambiguity in my utterance?”) and on the communicative situation per se (speaking on the telephone vs. face-to-face). Gesture production research clearly shows that addressees glean additional information from co-speech gestures, thereby providing evidence that gestures are not only communicatively intended but also really useful in communication. In fact, gestures do not only provide additional information but are also helpful in word learning and memory retention. Gesture comprehension research has shown that addressees glean additional information from gesture, especially about size and relative position. Furthermore, gestures can help in second language learning and improve memory performance for verbal material. ERP results show that gesture-speech integration is a fast, online process which is comparable with the integration of words or pictures into a speech context. Both behavioral and EEG data clearly show that gesture are communicative and that listeners can benefit from the additional gestural information (e.g. for the disambiguation of homonyms). Brain regions located in inferior frontal (IFG) and superior temporo-parietal cortex (STS/G) have been identified as putative neural structures involved in the merging of both streams of information. The integration itself, however, does not seem to be an automatic process, but may be influenced by various factors like the amount of meaningful gesture information.



**Figure 2.6** The role of the bilateral IFG and STG/S areas in gesture-speech integration – a meta-analysis. The meta-analysis was based on the fMRI studies to date on gesture-speech integration regardless of gesture type (Dick et al., 2009; Hubbard, Wilson, Callan, Dapretto, 2009; Green et al., 2009; Holle et al., 2008; Holle et al., 2010; Kircher et al., 2009; Straube, Green, Weis, Chatterjee, & Kircher, 2009; Willems, et al., 2007, 2009). Only activations denoted as gesture-speech integration related in the respective papers were included in the meta-analysis. Special focus was put on the activations in the bilateral IFGs (panel A and B) and STSs/Gs (panel C and D), which are assumed to play an important role in gesture-speech integration.

## 2.4. The present dissertation: The significance of task, timing and background noise on gesture-speech integration

Research on gesture-speech comprehension so far has shown that addressees can benefit from additional gesture information in many ways, e.g. learning new vocabulary and most importantly, getting a better understanding of what our conversational partner wants to communicate. There is no doubt that gestures are an integral part of our everyday communication. However, little is known about the way gestures are processed in comprehension. For instance, there is no theory or model on gesture-speech integration in comprehension, whereas there are more than half a dozen different models for gesture production. If one wants to propose a model for gesture comprehension, there is one basic question that has to be addressed: What factors impact gesture-speech integration?

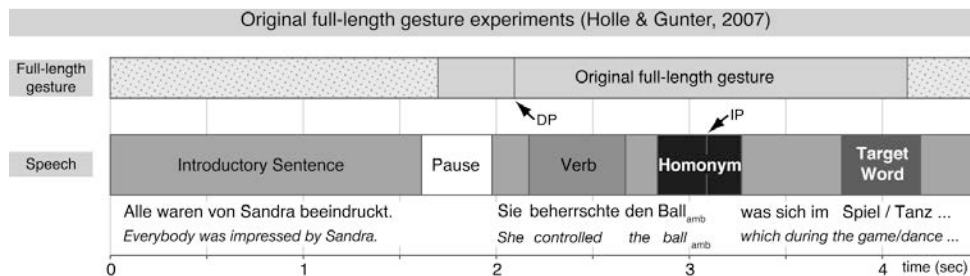
The aim of the present dissertation is to identify the significance of task, timing, and background noise on the integration of gestural information in sentence comprehension. Before going into detail about these factors, let us take one more look at the gesture literature on the nature of the integration of gesture and speech. According to McNeill et al. (1994), the integration of gestural and speech information is completely automatic, i.e. we cannot avoid to integrate them no matter if we want to or not (*“the point we wish to emphasize is the involuntary, automatic character of forming an idea unit out of information from the two channels”* McNeill et al., 1994, p. 236). Recent ERP data (Holle & Gunter, 2007; Kelly et al., 2007; Wu & Coulson, 2010), however, suggest, that this is not the case. Based on these recent findings, it is assumed that information uptake from gesture is not an automatic, but rather a very, flexible process, that may be influenced by various top-down and bottom-up factors to different degrees.

The setup for all the experiments was identical to the studies by Holle and Gunter (2007), i.e. a disambiguation paradigm was used. Similar to their experiments, participants were presented with sentences containing an unbalanced homonym (e.g. Ball / English: *ball*) which was disambiguated downstream in the sentence by a target word (dominant target: Spiel / *game*; subordinate target: Tanz / *dance*). In contrast to Holle and Gunter (2007), the homonym was not accompanied by a full-length iconic gesture which depicted either the dominant or subordinate meaning, but by a gesture fragment containing the minimal



necessary information to either cue the dominant or subordinate meaning of a homonym. The gesture fragments were determined with the use of a gating study.

The use of gesture fragments has several advantages over the use of full-length gestures as will be outlined in detail in Section 3.3. (p. 80). Gesture fragments have less variation with regard to their semantic content as compared to the full-length gestures and allow for a more precise investigation of timing between gesture and speech. The major advantage, however, is that the use of gesture fragments offers the unique possibility to investigate the direct integration of gesture fragment and homonym separately from the delayed disambiguating effect at the target word. In contrast, Holle and Gunter (2007) were only able to look at the later effect, because of the excessive temporal overlap of the full-length gestures with complete sentence of interest (see Figure 2.7). Thus, for all experiments of this dissertation, both the results for local integration as well as the global disambiguation at the target word will be reported. Before each of the experiments will be described in detail (Chapter 4-6), I briefly want to introduce the specific research questions addressed in each of the experiments.



**Figure 2.7** Timing of the original full-length gesture-speech material used by Holle and Gunter (2007).

### *Experiment 1 - Are we able to use gesture fragment information at all?*

Using an explicit congruency judgment task that requires participants to integrate both streams of information to solve the task, this experiment explores whether gesture fragments can be used at all in comprehension. Note, that due to stimulus construction, the gesture fragments ended about 1000 ms prior to point in time at which the homonym was identified. I.e. gesture fragment and speech were asynchronous. If participants make use of the minimal

gestural information, ERP effects for integration and disambiguation should be observed. More specifically, subordinate gesture fragments should elicit a larger N400 at the homonym as compared to dominant gesture fragments, due to the less frequent word meaning underlying the gestural representation. At the target word, one would expect a larger N400 for incongruent gesture cues as compared to congruent ones.

*Experiment 2 – Is the integration of gesture fragment and speech task-independent?*

In this experiment, a more shallow task (monitoring task) is used which does not require the integration of gesture and speech information. According to the two-process theories of information processing, automatic processes are characterized as being very fast, occurring without awareness and intention, and not tapping into limited-capacity resources (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). If the integration of the asynchronous gesture fragments with the homonyms was a completely automatic process, then it should be independent from task. If this was the case, similar effects as in Experiment 1 should be observed. Otherwise, an attenuation or even complete vanishing of any effect could occur.

*Experiment 3 – Is the integration of synchronously presented gesture fragments and speech obligatory?*

Due to the stimulus construction procedure, the gesture fragments used in Experiments 1 and 2 ended almost 1000 ms prior to the point in time where the meaning of the corresponding homonym was accessible. However, in his semantic synchrony rule, McNeill (1992) states that the same “idea unit” (p. 27) must occur simultaneously in both gesture and speech in order to allow proper integration. This assumption was tested by synchronizing gesture fragment offset and homonym identification point. Similar to Experiment 2, participants had to perform the monitoring task. We expect that even if there was no effect in Experiment 2, in which gesture and speech are asynchronous, the synchronization of both streams should result in observable effects for integration and disambiguation.

*Experiment 4 – Is there a temporal window for gesture-speech integration?*

Multimodal integration research has shown, that it is not so much exact synchrony that allows the integration of two streams of information, but that there is a so-called temporal window of integration, in which the combination of multisensory input is possible (e.g. results on the McGurk-Effect by van Wassenhove et al., 2007). Experiment 4 addresses this question.

Gesture fragments are presented at four different temporal alignments with regard to homonym identification point. This point was either prior (+120 ms), synchronous with (0 ms) or lagging behind the end of the gesture fragment (-600 ms / -200 ms). If the timing manipulations captured the temporal window of integration, one expects that this would especially show in the ERPs at the homonym position (immediate integration of gesture and speech).

*Experiment 5 – Is gesture information especially useful when speech is impaired?*

Gestural information may not be equally important in all situations. It is known, that communicators especially make use of gestural information when the speech signal quality is bad (e.g. Rogers, 1978). Using the identical experimental setup as Experiment 2, Experiment 5 tries to identify whether the processing of gesture fragment information and speech in a noisy environment differs from the processing of both streams of information in silence. We hypothesize that the information uptake is different in the sense that it might be more automatic in noise than in silence.

## **Chapter 3**



## **Chapter 3**

### **Stimulus material**

In this chapter, I will introduce the basic stimulus material used in all experiments of this dissertation. Whenever there is a specific stimulus manipulation that is only used in a single experiment, this particular modification is introduced in the stimulus section of the respective experiment. In the following, I will first describe the original gesture and speech material of Holle and Gunter (2007) and then provide information about the construction and testing of the iconic gesture fragments used in the present experiments.

#### **3.1. The construction of the original full-length gesture material**

##### **3.1.1. Sentence material (homonyms)**

For the present experiments a set of 48 unbalanced German homonyms derived from Holle and Gunter (2007) was used as stimulus material. These stimuli comprise a subset of the original 91 homonyms presented in an earlier study by Gunter, Wagner and Friederici (see also for more details on how the original homonym material was obtained, 2003). Each of the homonyms had a more frequent dominant and a lesser frequent subordinate meaning (e.g. Ball / *ball*: dominant meaning: Spielzeug / *toy*, subordinate meaning: Tanzball / *dance*). For both meanings of the original 91 homonyms a corresponding target word was assigned (e.g. Ball / *ball*: dominant target word: Spiel / *game*, subordinate target word: Tanz / *dance*). The relatedness of these target words had been assessed using a lexical decision task in the visual modality (see Wagner, 2002). For all target words the lexical decision time was significantly shorter than for unrelated items. In case one of the meanings of a homonym was very abstract, the homonym was removed from the stimulus set, resulting in a reduced set of 55 homonyms. For each of the remaining 55 homonyms, two sets of sentences were constructed, one for each target word. The sentences consisted of a short sentence introducing a character followed by a longer, more complex sentence describing an action of the 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 completely identical (see Table 3.1, p. 71; for the complete sentence material, see Appendix A, p. 173).

### 3.1.2. Recording of the original gesture videos

A professional actress was videotaped uttering the sentences. All videos showed the actress from a frontal view. The recording of the videos was accomplished in several steps. First, the actress had to memorize the sentences until she was able to utter them fluently. Subsequently, she was supposed to utter the two sentences while simultaneously performing a gesture supporting the complex sentence. None of the gestures was scripted, but spontaneously produced by the actress. She was instructed, however, to perform the gestures isochronously with the initial part of the complex sentence (containing the homonym). A typical gesture started with the actress holding her hands in the resting position (hands hanging down), followed by the gesture. Immediately, afterwards the actress returned her hands to the resting position (see Figure 3.1). Thus, every gesture item contained three phases: preparation, stroke and retraction<sup>12</sup>.

Two thirds of the gestures were re-enactments of the actions described in the sentences (e.g. gesture: typing on a keyboard; sentence: *Er vollendete den Brief... / He finished the letter...*). Most of them were performed from a first person perspective. The remaining third of the gestures depicted features of the objects (e.g. gesture: shape of a skirt; speech: *Er schilderte den Rock... / He described the skirt...*). To minimize mimic influences the actresses' face was covered with a nylon stocking. All gestures resembling emblems or directly depicting the target words were excluded. The remaining video sequences were edited using a commercial editing software (Final Cut Pro 5).

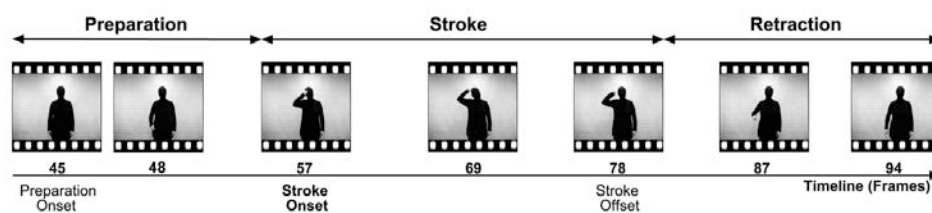
---

<sup>12</sup> The gestures did not contain any holds.

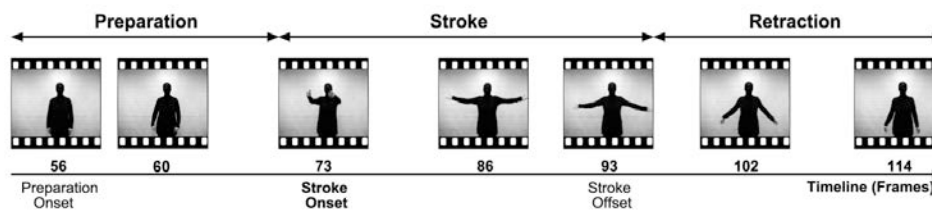
### 3.1.3. Pretests for the original gesture videos

A cloze procedure was performed to evaluate whether the gestures were able to successfully disambiguate the homonyms. The cloze procedure was first introduced into educational research by Taylor (1953). It is a "fill-in-the-blanks" task in which a large group of

#### a) dominant meaning



#### b) subordinate meaning



**Figure 3.1** Schematic illustration of the gesture phases for the (a) dominant and (b) subordinate gesture of the homonym "Kamm".

participants has to use contextual cues to fill in the deliberately removed words at the end of a sentence (for example: *She was in control of the ball which during the match at service showed* (original sentence). - *She was in control of the ball, which...* (example for a possible sentence in a cloze procedure)). The cloze probability of a word is defined as the proportion of subjects using that item to complete a particular sentence (Kutas & Hillyard, 1984). The response which is most likely to be chosen in a cloze task is called primary response (Fischler & Bloom, 1980). In general, increasing contextual constraints eases the identification of the correct sentence continuation and thus leads to a higher cloze probability, because the



increase in biasing information triggers a higher expectancy of the correct word (Bloom & Fischler, 1980; Shannon, 1948).

The ability to use contextual cues improves with age (see Doehring, 1976) and reading skill (Perfetti, Goldman, & Hogaboam, 1979). The cloze procedure can generally be used in all areas where language comprehension is to be investigated, including gesture-speech comprehension.

In the cloze procedure pretest, gesture videos were displayed to twenty German native speakers with sound disabled one word prior to the onset of the target word (e.g. Alle waren von Sandra beeindruckt. Sie kontrollierte den Ball, was sich im ... / *Everybody was impressed by Sandra. She controlled the ball, which during the ...*). Afterwards the participants had to choose the best fitting sentence continuation from a set of response alternatives. This set included the dominant sentence continuation (e.g. ...Spiel beim Aufschlag deutlich zeigte. / ... *game at the serve clearly showed.*) as well as the subordinate sentence continuation (e.g. ... Tanz mit dem Bräutigam deutlich zeigte. / ... *dance with the bridegroom clearly showed.*). The rationale of the cloze procedure was that participants should only be able to pick the correct sentence continuation if they previously had identified the semantic content of the gesture that accompanied the speech. This gesture was either related to the dominant or the subordinate meaning of the homonym. The percentage with which the correct sentence continuation was chosen corresponded to the cloze probability. Overall, the mean cloze probability was 93.7 %, which is significantly above chance ( $p < .01$ ), and did not differ significantly for dominant and subordinate gestures (paired  $t(1,47) = 0.69$ ,  $p > 0.4$ ). Only homonyms, which were disambiguated correctly by gesture information in at least 80% of all participants were kept, resulting in the final experimental set of 48 unbalanced German homonyms, which was both used by Holle and Gunter (2007) as well as in the present series of experiments.

#### 3.1.4. Splicing

The speech of the actress was re-recorded in a separate session to improve the sound quality of the videos. Because participants might use prosodic cues to disambiguate the homonyms, a cross-splicing procedure was performed (see Figure 3.2). The aim of this procedure was to keep the dominant and subordinate versions of a sentence identical up to the target word, and thus avoid potential confounding effects due to physical differences in the sentence material.

First, the best sounding recordings for each of the 96 speech files (48 “dominant” and 48 “subordinate” speech files) were selected. The initial part of the sentences up to the sub-clause of the complex sentence was retained unchanged. Secondly, the sub-clause itself was replaced by a recording of the other meaning for each single file. This procedure resulted in a spliced and an unspliced version of each speech file. This procedure resulted in a spliced and an unspliced version of each speech file. Spliced and unspliced sentences were equally distributed for both the dominant and the subordinate meaning across the experimental set of 48 homonyms, resulting in both 24 spliced and 24 unspliced speech files related to the dominant as well as subordinate meaning of the homonyms.

The cross-spliced speech files were then recombined with the gesture videos in a 2 x 2 design with Gesture (**D**ominant vs. **S**ubordinate) and Target word (**D**ominant vs. **S**ubordinate) as factors (see Table 3.1). All in all, there were 48 gesture videos for each condition (**DD**, **DS**, **SD**, **SS**), resulting in a stimulus set of 192 gesture videos.

**Original sentences:****Dominant**

Alle waren von Sandra beeindruckt. Sie beherrschte den Ball, was sich im Spiel beim Aufschlag deutlich zeigte.  
 Everybody was impressed by Sandra. She was in control of the ball which during the match at the service showed.

**Subordinate**

Alle waren von Sandra beeindruckt. Sie beherrschte den Ball, was sich im Tanz mit dem Bräutigam deutlich zeigte.  
 Everybody was impressed by Sandra. She was in control of the ball which during the dance with the bridegroom showed.

**Cross-spliced sentences (original dominant version better than original subordinate version):****Dominant**

Alle waren von Sandra beeindruckt. Sie beherrschte den Ball, was sich im Spiel beim Aufschlag deutlich zeigte.  
 Everybody was impressed by Sandra. She was in control of the ball which during the match at the service showed.

**Subordinate**

Alle waren von Sandra beeindruckt. Sie beherrschte den Ball, was sich im Tanz mit dem Bräutigam deutlich zeigte.  
 Everybody was impressed by Sandra. She was in control of the ball which during the dance with the bridegroom showed.

**Cross-spliced sentences (original subordinate version better than original dominant version):****Dominant**

Alle waren von Sandra beeindruckt. Sie beherrschte den Ball, was sich im Spiel beim Aufschlag deutlich zeigte.  
 Everybody was impressed by Sandra. She was in control of the ball which during the match at the service showed.





**Subordinate**

Alle waren von Sandra beeindruckt. Sie beherrschte den Ball, was sich im Tanz mit dem Bräutigam deutlich zeigte.  
 Everybody was impressed by Sandra. She was in control of the ball which during the dance with the bridegroom showed.



**Figure 3.2** Splicing procedure: Above are the original, unspliced sentences, below the cross-spliced sentences. During the splicing, the initial part of the dominant sentence is, for instance, combined with the end of the subordinate sentence and vice versa in order to reduce physical differences between the dominant and subordinate sentences. Original and spliced sentences together constituted the cross-spliced stimulus material.

**Table 3.1: Stimulus examples**

Introduction:		Alle waren von Sandra beeindruckt. <i>Everybody was impressed by Sandra.</i>	
gesture	target word	gesture / homonym	target word
D	D	Sie beherrschte 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 beherrschte 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 beherrschte 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 beherrschte 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>

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 is in italics. Cross-splicing was performed at the end of the main clause (i.e. in this case after the word "Ball").

### 3.1.5. Rating of the gesture phases

In order to get a more detailed understanding of the stimulus material and as a preparation for the present set of experiments, the onset of the gesture preparation as well as the on- and offset of the gesture stroke of the original gesture material was independently assessed by two persons. The phases of the gestures were determined according to their kinetic features described in the guidelines on gesture transcription by McNeill (p. 375f, 1992). To avoid a confounding influence of speech, the gesture videos were presented without sound for the rating procedure, as has been suggested in the Neuropsychological Gesture Coding System (NGCS, Lausberg & Sloetjes 2008; see also, Lausberg & Kita, 2003). First, the onset as well as the offset of the complete hand movement were determined, then the on- and offset of the

stroke phase were identified based on the change of effort in the movement, i.e. changes in movement trajectory, shape, posture and movement dynamics (for details see McNeill, 1992). The phase prior to the stroke onset was determined as the preparation phase, the phase after stroke offset as the retraction. The movements did not include any holds. Both raters highly agreed on the classification of the different gesture phases (e.g. inter-rater reliability (time of stroke onset)  $>.90$ ). In case there was dissent about the exact point in time of preparation onset, stroke onset or offset, raters afterwards discussed the results and chose the point in time they both felt appropriate. The values for the on- and offsets did not differ significantly across gesture conditions (all  $F(1,94) < 1$ ; see Table 3.2).

**Table 3.2: Stimulus properties of the full-length gestures**

gesture	gesture stroke onset	gesture stroke offset
Dominant gesture	2.07 (0.46)	2.91 (0.48)
Subordinate gesture	2.17 (0.52)	3.01 (0.51)
Mean	2.12 (0.49)	2.96 (0.50)

Mean on- and offset values are in seconds relative to the onset of the introductory sentence (*SD* in parenthesis).

## 3.2. The construction of the gesture fragment stimulus material

### 3.2.1. Gesture fragments

As already stated above, the present series of experiments was set out to detail the timing issues related to gesture-speech integration during sentence processing. To do so, we were faced by an interesting challenge, namely the large overlap in time between the original gestures used by Holle and Gunter (2007) and speech. Such large overlaps with speech will make a precise measurement of speech-gesture integration (including synchrony issues) very difficult in a sentence context. As an illustration, the timing parameters of the stimulus material as used by Holle & Gunter (2007) are given in Figure 2.7 (p. 60). As can be seen clearly, the full length gesture completely overlapped with the first part of the second sentence, which makes it impossible to manipulate gesture-speech synchrony using full gestures without simultaneously changing the amount of gesture information that is available

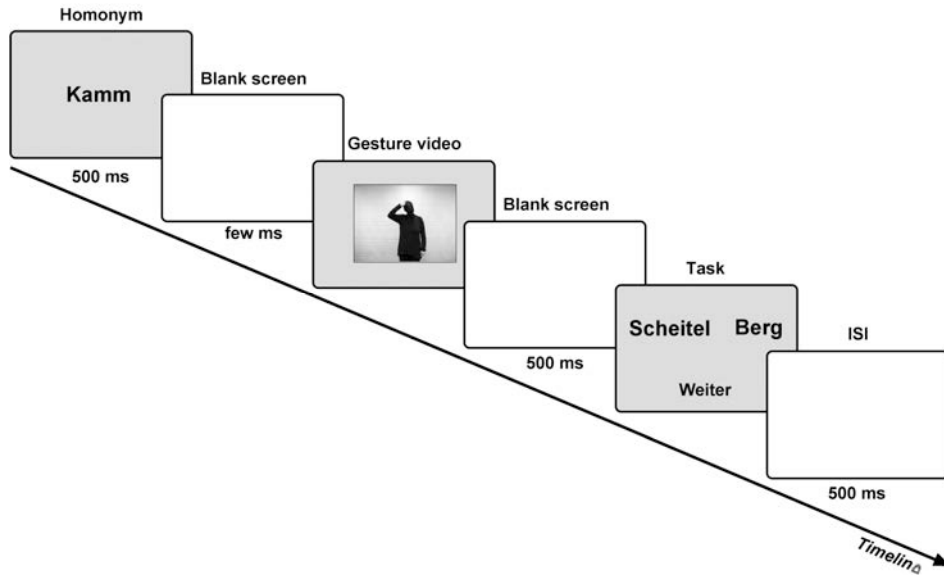
at a given point in time. To avoid such an undesirable confound between synchrony and gesture information, we decided to find a way to reduce the length of gesture information without changing its impact on speech substantially. Determining such gesture fragments would then enable us to investigate timing issues with much greater precision.

### 3.2.2. **Gating**

To determine the point in time at which a gesture can reliably disambiguate a homonym a context-guided gating procedure was applied to the gesture videos. Gating is a very popular paradigm in spoken word recognition (Grosjean, 1996). Its rationale is based on the assumption that spoken word recognition is a discriminative process, i.e. with increasing auditory information, the number of potential candidate words is reduced until only the correct word remains (cohort model, see Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987). The amount of information needed to identify a word without change in response thereafter is defined as the identification point. Although gating is most common in spoken word recognition, it can be used with virtually any kind of sequential material (e.g. ASL: Emmorey & Corina, 1990; music sequences: Jansen & Povel, 2004). The strength of the gating paradigm is that it can be adapted to suit the research question. Amongst others, it can differ with regard to increment size of the presented fragments (20 – 100 ms, cf. Walley, 1988; Grosjean & Hirt, 1996), presentation format (successive vs. individual, cf. Cotton & Grosjean, 1984; duration vs. blocked, cf. Walley, Michela, & Wood, 1995), type of response (written vs. oral, cf. Walley et al., 1995; free proposal, cf. Grosjean, 1980; fixed response set, cf. Allopenna, Magnuson, & Tanenhaus, 1998; Dahan & Gaskell, 2007), and context (without vs. with context, cf. Grosjean, 1980; Salasoo & Pisoni, 1985). Some of these variables can affect the outcome of the gating, e.g. gating with context leads to earlier isolation points than without a context (Salasoo & Pisoni, 1985). Because iconic gestures convey their meaning more clearly when produced with co-occurring speech, we employed a context-guided gating task to identify the isolation points of the gestures. Homonyms were used as context. This procedure has the advantage that the number of possible gesture interpretations is restricted to two, namely the dominant meaning and the subordinate meaning of the homonym. This allows the use of a fixed response set, as has been used for picture identification based on increasing word information by Dahan and Gaskell (2007). Using this response type as well as context may result in rather early isolation points for the gestures. However, based on the

literature at least some information of the stroke phase might be necessary to guess the correct meaning of a gesture.

Forty native German-speaking participants took part in the gating pretest. A gating trial started with a visual presentation of the homonym for 500 ms (e.g., Ball / English: *ball*), followed by the gated gesture video. 500 ms after the video offset, the participants had to determine whether the homonym referred to the dominant or the subordinate meaning based on gesture information. Three response alternatives were possible and simultaneously presented on the screen: (1) dominant meaning (e.g., the word Spiel / *game* was displayed on the screen), (2) subordinate meaning (e.g., Tanz / *dance*) and (3) "Weiter" / "*next frame*" (see Figure 3.3). Participants were instructed to choose the third response alternative until they felt they had some indication of which meaning was targeted by the gesture. The increment size was one video frame which corresponded to 40 ms, i.e. each gate 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). The gesture items were pseudo-randomly distributed across two experimental lists. Each of the lists contained 24 of the original dominant and 24 of the original subordinate gestures, resulting in a total of 48 gestures per experimental list. For each homonym, either the dominant or the subordinate gesture was presented within one list.



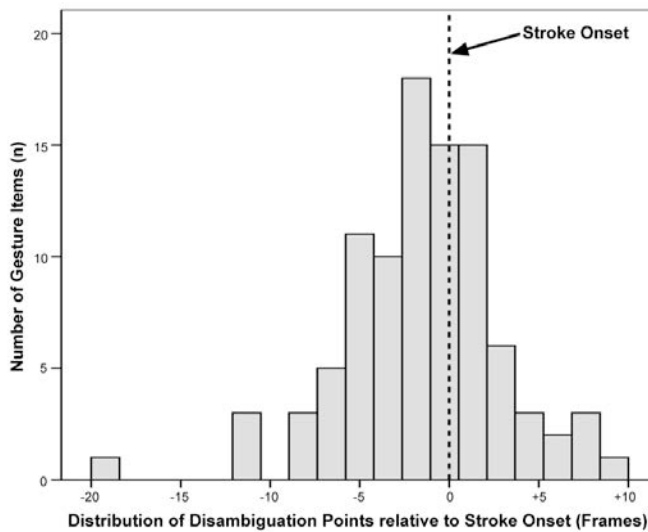
**Figure 3.3** Temporal sequence of a typical trial of the gating.

The dependent variable was the DP, which corresponds to 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. The mean DPs for the single items ranged from 2.22 to 19.63 frames ( $M = 9.88$ ;  $SD = 3.6$ ), calculated relative to preparation onset. Thus, on average the participants needed to see about 400 ms of gesture to disambiguate a homonym. An ANOVA with the factors word meaning frequency (2) and list (2) revealed that dominant gestures ( $M = 9.33$ ;  $SD = 3.6$ ) were identified earlier than subordinate gestures ( $M = 10.42$ ;  $SD = 3.58$ ) as indicated by the significant main effect of word meaning frequency ( $F(1,94) = 4.2$ ;  $p < .05$ ). This result indicates that more gesture information is needed to select the subordinate meaning.

When investigating the distribution of the DPs relative to the stroke onset, we found a surprising result. DPs ranged from almost 20 frames before the stroke onset to 9 frames past the stroke onset, with the DPs of 60 gestures being prior to the stroke onset (see Figure 3.4). This means that almost two thirds of all gestures enabled a meaning selection before the participants had actually seen the stroke. The difference between DP and stroke onset was found to be significantly smaller than zero across participants ( $t_t(1,39) = -4.7$ ;  $p < .001$ ) and



items ( $t_2(1,95) = -2.3; p < .05$ ). The corresponding  $\min F'$  statistic (Clark, 1973) was significant ( $\min F'(1,128) = 4.26; p < .05$ ) indicating that gestures reliably enabled a meaning selection before stroke onset.



**Figure 3.4** Distribution of the disambiguation points of all gestures relative to their stroke onset. As can be seen, more than half of all disambiguation points are prior to the stroke.

The DPs found in the context-guided gating might be considered as surprisingly early, given what McNeill (1992) has written about the meaning preparation phase. He suggests that the preparation phase is only optional, as the meaning of a gesture is represented in its stroke. Relative to the gesture phases as determined by our rating, most of the DPs actually occurred before stroke onset within the preparation phase of the gestures. It, therefore, seems that the preparation phase already suffices to select the appropriate meaning of a homonym. Although potentially intriguing, we have to be cautious in interpreting this result, since there are several methodologically-related explanations that may account for such an early effect.

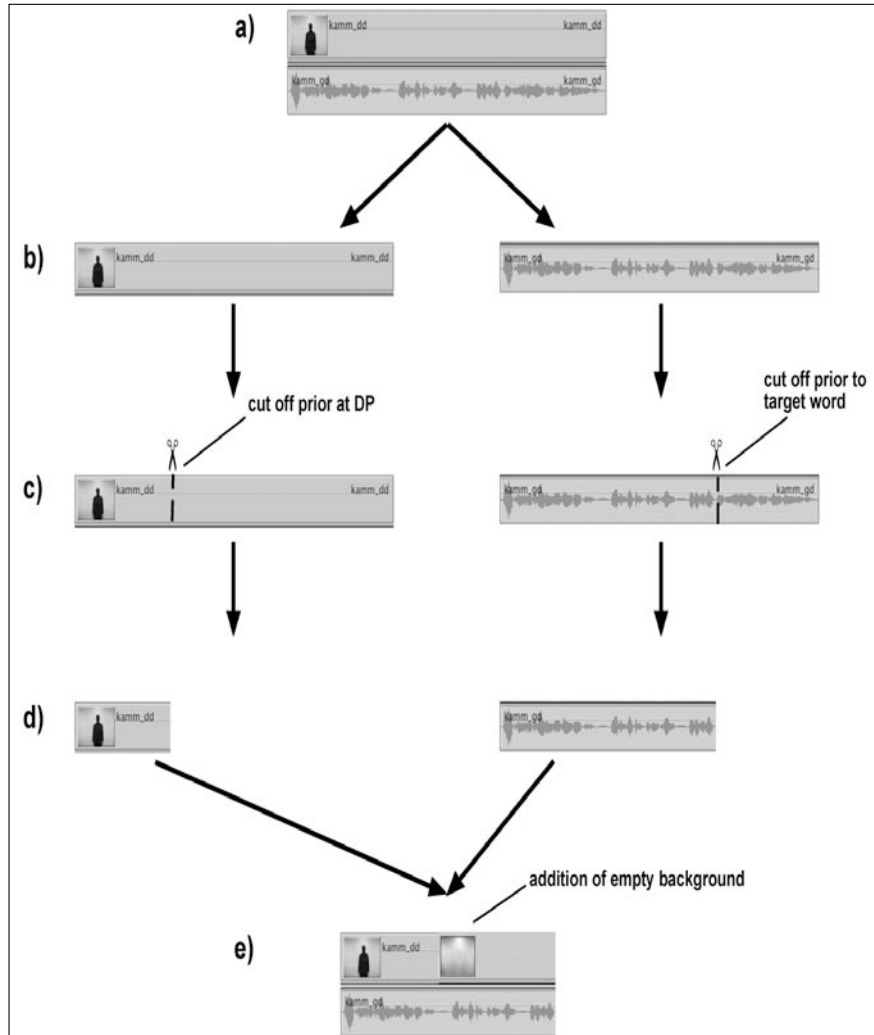
First, it is possible that the way we determined our stroke onset (with the sound turned off) may have resulted in later stroke onsets than a rating conforming entirely to the suggestions of McNeill (1992, p. 375f) would have. McNeill has suggested determining the phases of

gesture with sound turned on. This methodological difference makes it difficult to relate our finding to McNeill's claims about the preparation phase.

Second, an inherent feature of a gating procedure is the highly repetitive nature of the task. Such repetitions may have induced processing strategies different from those used in real-time speech comprehension. It is also well known from studies on spoken word recognition (e.g. Grosjean, 1980; Salasoo & Pisoni, 1985) that additional contextual information enables participants to identify words earlier than without context. Because iconic gestures are seldom, if ever, produced without context (i.e. accompanying speech) the meaning of a gesture may be accessible rather early. One could speculate that the earliness of meaning comprehension may depend on the degree of contextual constraint. For instance, the participants in the gating might have been able to decide upon the correct meaning more easily and faster, because they only had to choose between the two different meanings of a homonym. That is, gestures related to *Kamm*, which means either comb or crest, can be easily discriminated by hand shape. The preparation of the comb video contains the beginning of a one-handed gripping movement while there is an ascending and expanding two-handed movement in the crest video. This in line with Kita et al. (1998) who argue that hand-internal information like hand-shape or wrist location tends to emerge towards the end of the preparation phase, in other words the preparation anticipates features of the stroke (McNeill, 1992). Thus, it is not that surprising that the preparation phase is informing the recipient what type of stroke phase might be following. It is, however, a novel finding that a recipient can actively interpret and use such preparatory motor activity in a forced-choice situation. It is important to note that this meaning anticipation only seems to be possible within the context of speech (see section 3.2.4. Pretests for the gesture fragment videos, p. 79).

### 3.2.3. **Stimuli: Gesture fragments**

The gesture fragment videos for all present experiments were constructed as follows. First, the original gesture and speech streams were separated. Full-length gesture streams were then replaced with gesture streams cut at the DP. The duration of the gesture streams was adjusted to the duration of the speech streams by adding a recording of the corresponding empty video background (see Figure 3.5).



**Figure 3.5** Illustration of the video stimuli construction. The video stream and audio stream of the original videos (a) were separated (b). Afterwards the videos were cut off at the disambiguation points and the speech streams prior to the target word (c). Finally, the new video streams and speech streams (d) were merged again (e). Additionally, a still frame of the empty background was added, so both video and sound streams were of equal length.

This manipulation created the illusion of a speaker disappearing from the screen while the speech was still continuing for a short amount of time. Speech streams were recombined with both the clipped dominant as well as clipped subordinate gesture streams, resulting in the identical 2 x 2 design as used by Holle and Gunter (2007) with Gesture (**D**ominant & **S**ubordinate) and Speech (**D**ominant & **S**ubordinate) as within-subject factors (see Table 3.1,

p. 71). Each of the four conditions (**DD**, **DS**, **SD** and **SS**) contained 48 items, resulting in an experimental set of 192 items (for details about the gesture parameters see Table 3.3).

**Table 3.3: Stimulus properties of the gesture fragments**

gesture	speech	gesture preparation onset	disambiguation point (DP)	gesture stroke onset	gesture stroke offset	homonym onset	homonym identification point (IP)	target word onset	target word offset
D	D	1.72 (0.41)	2.10 (0.44)	2.07 (0.46)	2.91 (0.48)	2.84 (0.40)	3.09 (0.41)	3.78 (0.38)	4.16 (0.38)
D	S	1.72 (0.41)	2.10 (0.44)	2.07 (0.46)	2.91 (0.48)	2.84 (0.40)	3.09 (0.41)	3.80 (0.38)	4.17 (0.38)
S	D	1.68 (0.50)	2.10 (0.51)	2.17 (0.52)	3.01 (0.51)	2.84 (0.40)	3.09 (0.41)	3.78 (0.38)	4.16 (0.38)
S	S	1.68 (0.50)	2.10 (0.51)	2.17 (0.53)	3.01 (0.51)	2.84 (0.40)	3.09 (0.41)	3.80 (0.38)	4.17 (0.38)
Mean		1.70 (0.45)	2.10 (0.47)	2.12 (0.49)	2.96 (0.50)	2.84 (0.40)	3.09 (0.41)	3.79 (0.38)	4.17 (0.38)

Mean onset and offset values are in seconds relative to the onset of the introductory sentence (SD in parentheses).

### 3.2.4. Pretests for the gesture fragment videos

#### 3.2.4.1. Cloze procedure

Similar to the pretest for the full-length gestures a cloze-procedure was performed to evaluate whether the gesture fragments were able to disambiguate homonyms. The mean cloze probability for the gesture fragments was 78 %, which is significantly above chance (all  $p < .01$ ) and did not differ significantly for dominant and subordinate gestures (paired  $t(1,47) = 1.24$ ,  $p > 0.22$ ). Gestures fragments, however, did differ from the complete gestures with respect to their cloze probability. Complete gestures elicited a mean cloze probability of 93.7 %, while trimmed gestures elicited only a probability of 78 %. This difference proved to be significant (paired- $t(1,95) = 8.75$ ,  $p < .01$ ). This indicates that gesture fragments may pose a weaker context for homonym disambiguation than the complete gestures used by Holle and Gunter (2007).

#### 3.2.4.2. Gesture fragment identification without speech context

An additional behavioral study, in which 9 participants had to guess the meaning of the gestures clipped at DP without any context was performed. The aim was to test whether the gesture fragments themselves carried a certain meaning or depended on the homonym context to be identified. Participants were presented with silent clips of the gesture fragments and had to write down a free proposal of the meaning of the fragments. Only 7 % of all gestures were identified correctly (i.e. semantically related to the meaning of the target word or homonym) and again only 7 % of these correct responses included the actual target word (i.e. 0.5 % overall). This result shows that participants were not able to get the correct meaning of the gesture fragments without context and is in line with other studies which showed that the meaning of an iconic gesture is rather imprecise in absence of speech (Hadar & Pinchas-Zamir, 2004; Krauss, Morrel-Samuels, & Colasante, 1991). When a context is given, however, most of our gesture fragments are able to disambiguate by means of displaying solely pre-stroke information.

#### 3.2.5. Determining the identification points of the homonyms

In order to investigate the online integration of the gesture fragments with the homonym with as much temporal precision as possible, we also determined the earliest point in time at which the homonym is identified. In a gating paradigm, spoken words fragments of increasing duration (increment size: 20 ms) were presented to 20 participants who did not participate in any of the experiments reported here. The identification point (IP) was determined as the gate where the participants started to give the correct response without any change in response thereafter. On average, participants were able to identify the homonyms after 260 ms. The homonym IPs were used as triggers for the ERPs that dealt with the direct integration of the gesture fragments with the homonyms.

### 3.3. Summary

The use of gesture fragments as stimulus material provides a unique opportunity to test factors involved in gesture-speech integration. First, gesture fragments are very short and comparable with regard to their semantic content, i.e. at the DP the minimal necessary

information to disambiguate the homonym is present in all gestures. Second, the disambiguating power of the gesture fragments is less strong as in the full-length gestures, indicating that they probably are a weak biasing context in terms of homonym disambiguation (e.g. Simpson, 1981; Martin et al., 1999; Holle & Gunter, 2007). Finally, all gestures end prior to the homonym IP and in contrast to the original full-length material do not overlap with almost the complete complex sentence (including verb, homonym and target). This allows to very precisely investigate the integration of the gesture fragments with the homonym at various levels of asynchrony, without changing the amount of available gesture information at a certain point in the sentence, i.e. in our case the homonym IP. In this context, it is important to note that the gesture fragments only get their disambiguating power in combination with the homonym. Of course, gesture fragments have one major disadvantage in comparison to full-length gestures. They are not natural, i.e. one would not encounter such fragments in everyday discourse. This disadvantage is, however, more than compensated by the fact that, from an experimental perspective, gesture fragments represent a highly controlled and very flexibly usable stimulus material, which is especially important, if one is interested in identifying factors that can influence the integration process of gesture and speech.



## **Chapter 4**





## **Chapter 4**

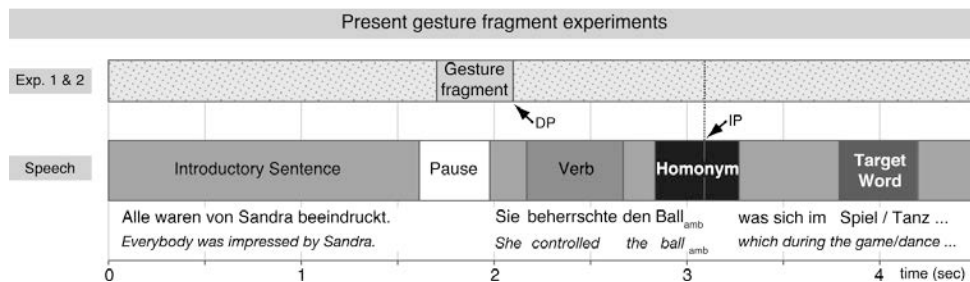
### **The impact of task on the integration of gesture fragments with speech**

This chapter contains two ERP experiments using the disambiguation paradigm as introduced by Holle and Gunter (2007). The main goal of both experiments is to address the role of task in the integration of gesture fragments with speech. Importantly, the gesture fragments did not share any temporal overlap with the corresponding speech unit, i.e. the homonym. Previous research on multi-modal integration has shown that in such a situation, the processing can differ if participants have to perform an explicit task in contrast to an implicit task (van Atteveldt, Formisano, Goebel, et al., 2007). Experiment 1 was conducted to clarify whether asynchronous gesture fragments can be integrated with the corresponding homonym and used as cues for disambiguation of the homonym. Participants had to perform a congruency judgment task, which required them to integrate gesture and speech in order to solve it (explicit task). Using the identical task, Holle and Gunter (2007) found the largest disambiguation effects for full-length gestures. If participants make use of the minimal gestural information, ERP effects for integration and disambiguation should be observed. More specifically, subordinate gesture fragments should elicit a larger N400 at the homonym as compared to dominant gesture fragments, due to the less frequent word meaning underlying the gestural representation. At the target words, one would expect a larger N400 for incongruent gesture cues as compared to congruent ones, similar to the results by Holle and Gunter (2007). In Experiment 2, participants were not required to integrate gesture and speech. They only had to perform a very shallow monitoring task (Did you see this movement? - Did you hear that word?) in a very small amount of trials. This task was introduced to control for participants' attention during the experiment. If gesture-speech integration was task-independent, similar effects as in Experiment 1 should be expected in Experiment 2.

## 4.1. Experiment 1: Are we able to use gesture fragment information at all?

### 4.1.1. Introduction

Experiment 1 serves as a basis for all the other experiments. It tests whether gesture-fragments presented up to their DP are able to disambiguate speech. As described in Chapter 3, the DPs were assessed using a context-guided gating with the original gesture material of Holle and Gunter (2007). The big advantage of using gesture fragments made out of this particular material is that we can measure the brain activity of our participants with great precision at two positions in time. At the homonym position, speech-gesture integration can be directly measured, whereas a few words downstream the target word position gives us direct online evidence whether this integration led indeed to a successful disambiguation. As can be seen in Figure 4.1, the gesture fragment was presented earlier than the homonym. That is, there was a clear gesture-speech asynchrony.



**Figure 4.1** Timing of the gesture fragments and speech in Experiments 1 and 2.

### 4.1.2. Methods

#### *Participants*

Thirty-nine native German-speaking participants were paid for their participation and signed a written informed consent. Seven of them were excluded because of excessive artifacts. The remaining 32 participants (16 female; 20-28 years, mean 23.8 years) were right-handed (mean laterality coefficient 94.3, Oldfield, 1971), had normal or corrected-to-normal vision, no known hearing deficits and had not taken part in the pretest of the stimulus material.

*Stimuli*





The complete experimental set of 192 gesture fragment videos was used as stimulus material (for a detailed description see Chapter 3.2.3., p. 77ff). The set contained 48 videos of each of the 4 possible conditions resulting from a 2 x 2 design with gesture fragment meaning (**Dominant & Subordinate**) and speech / homonym (**Dominant & Subordinate**) as within-subject factors (see Table 4.1). The items were pseudo-randomly distributed to four blocks of 48 items, ensuring that (i) each block contained 12 items of all four conditions and (ii) each block contained only one of the 4 possible gesture-speech combinations for each homonym.

*Procedure*

Participants were seated in a dimly-lit, sound proof booth facing a computer screen. They were instructed to attend both to the movements in the video as well as the accompanying speech. After each item, participants judged whether gesture and speech were compatible. Note that in order to perform this task, participants had to compare the meaning indicated by the homonym–gesture combination with the meaning expressed by the target word (see Table 4.1). A trial started with a fixation cross, which was presented for 2000 ms, followed by the video presentation. The videos were centered on a black background and extended for 10° visual angle horizontally and 8° vertically. Subsequently, a question mark prompted the participants to respond within 2000 ms after which feedback was given for 1000 ms.

The experiment was divided into four blocks of approximately 9 minutes each. For all blocks, the presentation order of the items was varied in a pseudo-randomized fashion. Block order and key assignment was counter-balanced across participants, resulting in a total of eight different experimental lists with 192 items each. One of the eight lists was randomly assigned to each participant. Thus, each experimental list was presented to four participants. An experimental session lasted approximately 45 min.

**Table 4.1: Stimulus examples for Experiments 1 to 5**

Introduction:		Alle waren von Sandra beeindruckt. <i>Everybody was impressed by Sandra.</i>	
gesture	target word	gesture / homonym	target word
D	D	Sie beherrschte 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 beherrschte 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 beherrschte 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 beherrschte 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>

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 is in italics. Cross-splicing was performed at the end of the main clause (i.e. in this case after the word "Ball").

#### ERP recording

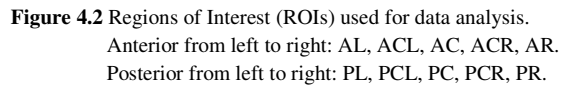
The EEG was recorded from 59 Ag/AgCl electrodes (Electrocap International). It was amplified using a PORTI-32/MREFA amplifier (DC to 135 Hz) and digitized at 500 Hz. Electrode impedance was kept below 5 kΩ. The left mastoid served as a reference. Vertical and horizontal electro-oculogram (EOG) was measured for artifact rejection purposes.

*Data Analysis*

Participants' response accuracy was assessed with a repeated measures ANOVA with Gesture (D, S) and Target Word (D, S) as within subject factors. EEG data were rejected offline by applying an automatic artifact rejection using a 200 ms sliding window on the EOG ( $\pm 30 \mu\text{V}$ ) and EEG channels ( $\pm 40 \mu\text{V}$ ). All trials followed by incorrect responses were also rejected. On the basis of these criteria, approximately 33 % of the data were excluded from further analysis. Single-subject averages were calculated for every condition both at the homonym and target word position.

In the analyses at the homonym position, epochs were time-locked to the IP of the homonyms and lasted from 200 ms prior to the IP to 1000 ms afterwards. A 200 ms pre-stimulus baseline was applied. Ten Regions of Interest (ROIs) were defined: anterior left (AL): AF7, F5, FC5; anterior center-left (ACL): AF3, F3, FC3; anterior center (AC): AFZ, FZ, FCZ; anterior center-right (ACR): AF4, F4, FC4; anterior right (AR): AF8, F6, FC6; posterior left (PL): CP5, P5, PO7; posterior center-left (PCL): CP3, P3, PO3; posterior center (PC): CPZ, PZ, POZ; posterior center-right (PCR): CP4, P4, PO4; posterior right (PR): CP6, P6, PO8. Based on visual inspection (see Figure 4.2), a time window ranging from 100 to 400 ms was used to analyze the integration of gesture and homonym. A repeated measures ANOVA using Gesture (D, S), ROI (1, 2, 3, 4, 5), and Region (anterior, posterior) as within-subject factors was calculated. Only effects which involve the crucial factor Gesture will be reported.

In the target word analysis, epochs were time-locked to the target word and lasted from 200 ms prior to the target onset to 1000 ms post target. A 200 ms pre-stimulus baseline was applied. The identical ten ROIs as in the previous analysis were used. The standard N400 time window ranging from 300 to 500 ms after target word onset was selected to analyze N400 effects. A repeated measures ANOVA using Gesture (D, S), Target Word (D, S), ROI (1, 2, 3, 4, 5), and Region (anterior, posterior) as within-subject factors was performed. Only effects which involve the crucial factors Gesture or Target Word will be reported. In all statistical analyses the 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. Prior to all statistical analyses the data were filtered with a high-pass filter of 0.2 Hz. Additionally, a 10 Hz low-pass filter was used for presentation purposes only.

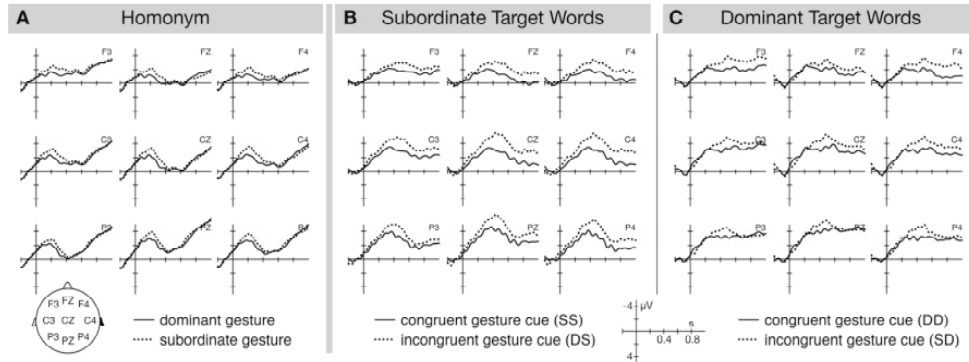


### *Behavioral Data*

The response accuracy was adequate across the different congruency conditions (congruent gesture-speech pairings: 77%; incongruent gesture-speech pairings: 73%). A significant main effect of congruency (paired  $t(31) = 2.30$ ;  $p < .05$ ) indicated that response accuracy was better for the congruent than the incongruent pairings. Congruent pairings also showed a faster reaction time (RT; congruent: 450 ms; incongruent: 474 ms; paired  $t(31) = -2.04$ ;  $p = .05$ ). Note that because the response occurred with some delay, the RT data should be treated with caution. Overall, the behavioral data suggest that speech comprehension is enhanced when gesture fragment and speech are congruent compared to when they are incongruent.

*ERP data - Homonym*

In Figure 4.3 A, an early, enhanced negativity can be observed when the homonym is preceded by subordinate gesture fragments as compared to dominant gesture fragments. Although this effect seems very early, this negativity is likely to be a member of the N400 family when considering its scalp distribution. The early onset can be explained by the use of the IP as the onset trigger of the averages. A repeated measures ANOVA revealed a main effect of Gesture ( $F(1,31) = 17.61$ ;  $p < .0002$ ), indicating that the integration of a subordinate gesture fragment with the corresponding homonym is more effortful than the integration of a dominant gesture fragment.



**Figure 4.3** ERPs as found in Experiment 1. The left panel (A) shows the ERPs time-locked to the identification point of the homonyms. The solid line represents when the ERP when the homonym was preceded by a dominant gesture fragment. The dotted line represents ERP when the homonym was preceded by a subordinate gesture fragment. The middle (B) and left (C) panel represent the ERPs time-locked to the onset of the target word. The solid line represents the cases in which gesture cue and subsequent target word were congruent. The dotted line represents those instances where gesture cue and target word were incongruent.

*ERP data – Target Word*

As can be seen in Figure 4.2 B & C, the ERPs show an increased negativity starting at about 300 ms for incongruent gesture-target word relations (DS, SD) in comparison to the congruent ones (DD, SS). Based on its latency and scalp distribution, the negativity was identified as an N400. The analysis of the 300-500 ms time window showed a significant two-way interaction of Gesture and Target Word ( $F(1,31) = 16.33$ ;  $p < .0005$ ) as well as a significant two-way interaction of Target Word and Region ( $F(1,31) = 4.79$ ;  $p < .05$ ).



On the basis of the Gesture and Target Word interaction step-down analyses were computed to assess the main effect of Gesture for both target word conditions. At dominant target words, the N400 was larger after a subordinate gesture compared to a dominant gesture ( $F(1,31) = 10.14$ ;  $p < .01$ ). In contrast, the N400 at subordinate target words was larger when being preceded by a dominant gesture ( $F(1,31) = 12.16$ ;  $p < .01$ ). Thus, incongruent gesture context elicited a larger N400 at both target word conditions. Yet, the effect was slightly larger for subordinate (Cohen's  $f^2 = 0.38$ ) than dominant targets (Cohen's  $f^2 = 0.32$ ).

#### 4.1.4. Discussion

Experiment 1 addressed the question whether gestures clipped at the DP suffice as disambiguation cues in online speech comprehension using a congruency judgment task. The observed ERP effects at the homonym and at the target word position indicate that indeed these short gesture fragments can be used for disambiguation, even though there was an asynchrony of 970 ms between the end of a gesture fragment and the corresponding homonym IP. In the following, the results at the homonym and target word position will be discussed in more detail.

*ERPs at the homonym position.* The ERPs elicited at the IP position of the homonym showed a direct influence of gesture type. Subordinate gestures elicited a more negative ERP compared to dominant gestures. Although its onset was very early (probably due to the use of the IP as trigger point) we would like to suggest, on the basis of its scalp distribution, that this component belongs to the N400 family. The data therefore suggests that the integration of the homonym with the subordinate gesture fragment is probably more effortful than the integration with the dominant gesture fragment. A more extended discussion of this effect will be given in the general discussion. For the moment it is enough to know that the gesture fragments had a direct and differential impact during the processing of the homonym. The next question relates to whether this impact leads to a disambiguation of the homonym, influencing sentence processing further downstream. Such an effect would indicate that the gesture fragments indeed contained disambiguating information.

*ERPs at the target position.* The ERP data on the target word showed clearly that the gesture fragments were used to disambiguate the homonym. When a target word was incongruent with how the gesture fragments disambiguated the homonym a larger N400 was

elicited compared to when targets were congruent with the preceding gesture-driven disambiguation. Interestingly, both types of target words showed this effect suggesting that the activation of both meanings of a homonym varied reliably as a function of the preceding gesture context. Note, however, that the N400 effect was larger for the subordinate target words suggesting a larger sensitivity towards gesture influence than dominant targets. Such a finding may indicate that gesture fragments are a relatively weak disambiguating context (see Section 2.2.2.5 ERPs as a correlate for gesture-speech integration (pp. 51-53; see also Martin et al., 1999; Simpson, 1981))

It is important to note that in Experiment 1, participants were explicitly asked to compare the semantic content of a gesture fragment-homonym combination with the subsequent target word in order to solve the task. Thus, the task forced them to actively combine and integrate both sources of information. Due to the large distance between the end of the gesture fragment and the homonym IP (about 970 ms, see Figure 4.1, p. 86), it is, on the one hand, not unrealistic to assume that gesture-speech integration in this particular case is an effortful memory-related process, because the gestural information has to be actively kept in working memory until the homonym is encountered. On the other hand, there are many suggestions in the literature that speech-gesture integration should occur more or less automatically and therefore effortless (Kelly et al., 2004; Özyürek et al., 2007). Automatic processes are characterized as being very fast, occurring without awareness and intention, and not tapping into limited-capacity resources (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). If the integration of the gesture fragments with the homonyms is an automatic process, as suggested for gesture-speech integration in general by McNeill and colleagues (1994), it should be independent from experimental context and task.

To explore the underlying nature of the integration of a gesture fragment with a homonym, we used a more shallow memory task in Experiment 2 and examined whether participants would still use the gesture fragments as disambiguation cues even when the task did not require them to do so. As in Experiment 1, there was an asynchrony between gesture and speech, in that the gestures fragments ended about 970 ms before the IP of the homonyms. The rationale of the task was as follows: After a random number of trials a task prompt sometimes indicated participants that they were now being asked whether they had seen a certain movement or heard a certain word in the previous video. No reference was made to the potential relationship between gesture and speech in the task instructions. Thus, participants had to pay attention to both gesture and speech, but were not required to actively

combine both streams to solve the task. Holle and Gunter (2007), who used the same shallow task to investigate whether the integration of full gestures is automatic, found an N400 effect for both target word conditions. Based on that study, we hypothesized that the shortened gestures used in the present study should also modulate the N400 at the position of the target word under shallow task conditions. Additionally, we also expected an enhanced negativity for the integration of the subordinate as compared to dominant gesture fragments at the position of the homonym, as it was observed in Experiment 1.

## **4.2. Experiment 2 - Is the integration of gesture fragment and speech task-independent?**

### **4.2.1 Methods**

#### *Participants*

Thirty-four native German-speaking participants were paid for their participation and signed a written informed consent. Two of them were excluded because of excessive artifacts. The remaining thirty-two participants (16 female, age range 21-29 years, mean 25.6 years) were right-handed (mean laterality coefficient 93.8), had normal or corrected-to-normal vision and no known hearing deficits. None had taken part in any of the previous experiment.

#### *Stimuli*

The same stimuli as in Experiment 1 were used.

#### *Procedure*

Presentation of the stimuli was identical to Experiment 1. Participants were, however, performing a different, shallower task and 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 (see Figure 4.4). After 6.25 %

of all videos, the prompt cue indicated to prepare for the movement task. A short silent video clip was presented as a probe. The probes consisted of soundless full-length gesture videos. After the offset of each probe video, a question mark prompted the participants to respond whether the probe contained the movement of the previous experimental item. Feedback was given if participants answered incorrectly or if they failed to respond within 2000 ms after the response cue. After the remaining 6.25 % of the videos, the prompt cue informed the participants that the word task had to be carried out. Participants had to indicate whether a visually presented probe word had been part of the previous sentence. The probe words were selected from sentence-initial, -middle and -final positions of the experimental sentence. Response and feedback were identical to the movement task trials.

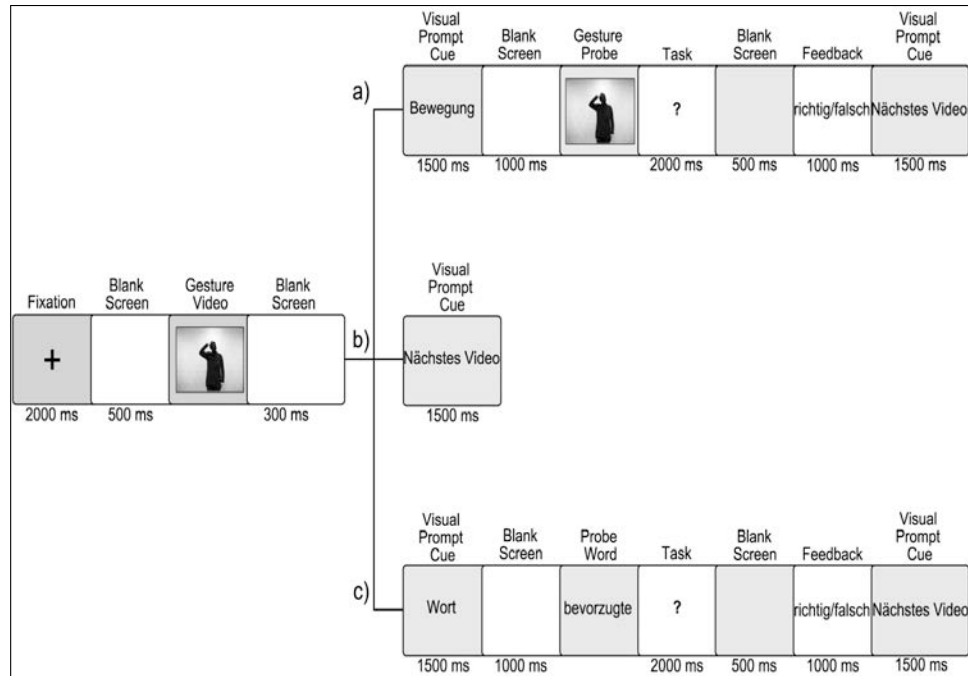
#### *ERP recording and data analysis*

The parameters for the recording, artifact rejection, and analysis were the same as in Experiment 1. The amount of behavioral data obtained in the present experiment is quite small (24 responses overall) with half of them originating from the movement task and the other half of the word task. Therefore, we decided to not use the behavioral data as a rejection criterion for the ERP analyses. Approximately 22 % of all trials were rejected for the final analysis of both the homonym as well as target word position.

### **4.2.2. Results**

#### *Behavioral Data*

Overall, participants gave 87 % correct answers, indicating that although the task in Experiment 2 was rather shallow, participants nonetheless paid attention to the stimulus material. Performance was less accurate in the movement task (79 % correct) than in the word task (96 % correct; Wilcoxon signed-rank test;  $z = -4.72$ ;  $p < .001$ ).



**Figure 4.4** Schematic illustration of the temporal sequence of a trial in Experiment 2. (a) shows the movement task condition, (b) shows the no task condition and (c) the word task condition.

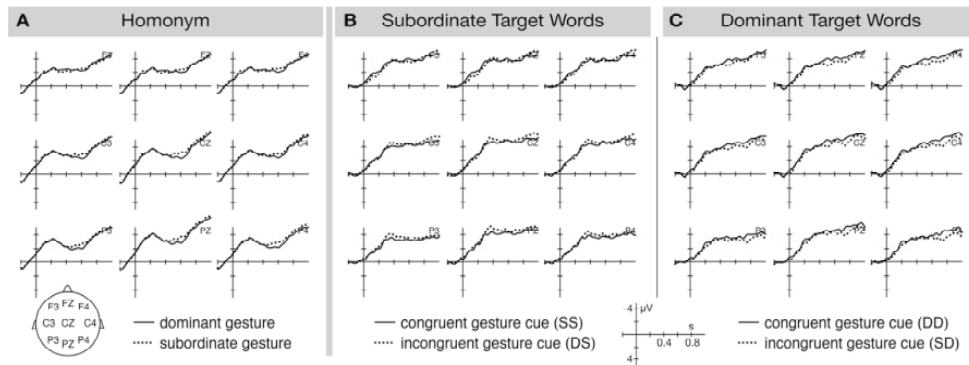
#### ERP data - Homonym

Figure 4.5 A shows no visible difference between subordinate and dominant gesture fragments at the homonym position. The corresponding ANOVA indicated no statistically significant differences (all  $F_s < .53$ ;  $p > .49$ ).

#### ERP data – Target Word

As can be seen in Figure 4.5 B & C, there is barely a visible difference between the congruent and incongruent gesture cues for both target word conditions. The repeated measures ANOVA confirmed this impression by yielding no significant 4-way interaction of Gesture, Target Word, ROI and Region ( $F(4, 124) = .32$ ;  $p > .69$ ;  $\epsilon = .42$ ), nor any other significant interaction involving the crucial factors of Gesture or Target Word (all  $F_s < 1.28$ ; all

$ps > .28$ ). That is, there was no significant disambiguating influence of gesture on speech in the data.



**Figure 4.5** ERPs as found in Experiment 2.

#### 4.2.3. Discussion

Experiment 2 dealt with the question of whether gesture fragments are integrated with speech when a shallow task was used. Both at the homonym as well as on the target words, no significant ERP effects were found. Thus, gesture fragments do not influence the processing of co-expressive ambiguous speech when the task does not explicitly require an integration of gesture and speech. One way to interpret this finding is to suggest that the integration of gesture fragments is not an automatic process. Such a conclusion would, however, contradict the literature that indicates that gesture-speech integration is more or less automatic in nature (McNeill et al., 1994). It is therefore sensible to look more carefully at Experiment 2 and see whether a more parsimonious hypothesis can be formulated. Using the identical experimental setup, Holle & Gunter (2007) found a disambiguating effect of the original full-length gestures under shallow task conditions. One crucial difference between the gestures used by Holle & Gunter (2007) and the gesture fragments used here is whether the gesture overlaps with its corresponding co-speech unit, i.e., the homonym. Whereas complete gestures span over a larger amount of time and have a significant temporal overlap with the homonym, no such temporal overlap is present between the gesture fragments and the homonyms (see Figure 4.1, p. 86). Remember that due to the clipping procedure the gesture fragments end on average 970 ms prior to the homonym IP. Thus, at the time the gesture fragment ends there

is no co-expressive speech unit with which it can be integrated. When effortful processing is induced by the task (Experiment 1), this time lag does not seem to be problematic. If, however, the task does not explicitly require participants to actively combine gesture and speech as in Experiment 2, the time lag between gesture and speech may be problematic, probably because the minimal amount of information present in the gesture fragments gets lost over time. Thus, an alternative explanation is that automatic integration of gesture fragments does not occur when a gesture and its corresponding speech unit do not have a sufficient amount of temporal overlap. It is important to note that such an alternative explanation is also in accordance with McNeill (1992), who suggested that it is the simultaneity between gesture and speech that enables a rather automatic and immediate integration of gesture and speech. Note that simultaneous gesture and speech presentation usually results in a temporal overlap between gesture and speech. Therefore, it is necessary to clarify whether synchrony, temporal overlap or both play important (independent?) roles in gesture-speech integration. In his semantic synchrony rule, McNeill (1992) states that the same “idea unit” (p. 27) must occur simultaneously in both gesture and speech in order to allow proper integration. In other words, he suggests that if gestures and speech are synchronous, they should be integrated in a rather automatic way. So far, however, there has been little empirical work on the effects of gesture-speech synchronization in comprehension (but see Treffner et al., 2008). The aim of the next chapter is to fill this gap in gesture comprehension research.

## **Chapter 5**





## **Chapter 5**

### **The significance of timing for gesture-speech integration**

Chapter 5 comprises two ERP experiments that explore the significance of timing for gesture fragment speech integration in sentence comprehension. The first Experiment (Experiment 3), addresses the question whether synchronous gesture and speech information is integrated when participants have to perform the shallow monitoring task similar as in Experiment 2. If integration takes place, this should be reflected in the ERPs at the homonym and target word position. In the second experiment (Experiment 4), the timing between gesture DP and Homonym IP was varied, to identify whether there is a temporal window for gesture-speech integration, similar to those found for other types of multimodal integration so far. For instance, van Wassenhove et al. (2007) found that participants perceive the McGurk effect within a time range of -170 to + 30 ms for the onset asynchrony of lips and voice. For this purpose, four different temporal alignments (-600 ms, -200 ms, 0 ms, -120 ms) between gesture fragment offset and homonym identification point were explored.

#### **5.1. Experiment 3: The processing of synchronous gesture and speech information**

##### **5.1.1. Introduction**

As stated above (see Section 2.2.2.5 ERPs as a correlate for gesture-speech integration, pp. 49-50), McNeill (1992; McNeill et al., 1994) suggested whenever gestures and speech are synchronous (and / or are overlapping in time (own assumption)), their integration should be rather automatic. In Experiment 3, we explored this synchrony / temporal overlap hypothesis. We synchronized the gesture fragments with the homonyms in such a way that the DPs of the gestures were aligned with the IPs of the homonym. Note that the way synchrony is defined differs from the definition usually used in other multimodal integration studies (e.g. van

Wassenhove et al., 2007), which determine synchrony in terms of stimulus onset asynchrony (SOA) between visual and auditory information. I.e. audiovisual information is in synchrony if the SOA is 0 ms. In this dissertation, synchrony between gesture and speech is defined relative to the identification point of the homonym. The logic is that gestures can only become meaningful in the presence of the homonym and thus only when an addressee knows the meaning of a homonym, i.e. at the homonym identification point

Besides the synchrony manipulation, Experiment 3 was exactly the same as Experiment 2. Thus, again, the shallow task was used. If, as suggested by the temporal overlap hypothesis, synchronization is playing a crucial role during speech-gesture integration, one would predict ERP effects similar to those observed in Experiment 1, both in the immediate context of the homonym as well as further downstream at the target word. In contrast, if the integration of gesture fragments is impossible independent of timing under shallow task conditions, no effects as in Experiment 2 should be observed.

### 5.1.2. Methods

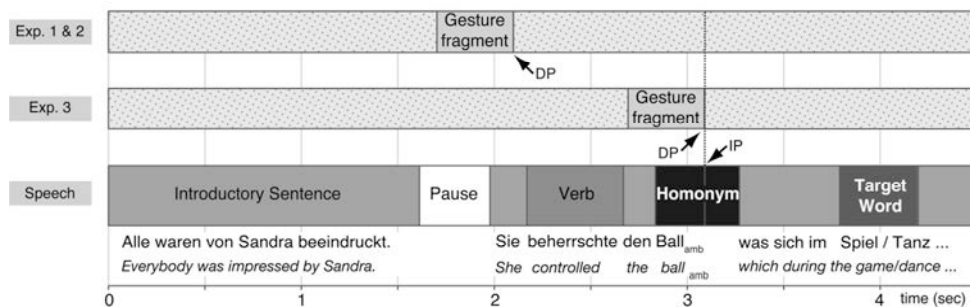
#### *Participants*

Thirty-eight native German-speaking participants were paid for their participation and signed a written informed consent. Six of them were excluded because of excessive artifacts. The remaining thirty-two participants (15 female, age range 19-30 years, mean 25.5 years) were right-handed (mean laterality coefficient 93.9), had normal or corrected-to-normal vision, no known hearing deficits and did not participate in any of the previous experiments.

#### *Stimuli*

The original 96 gesture fragment videos (48 containing dominant gesture fragments, 48 containing subordinate gesture fragments) used in Experiments 1 and 2 constituted the basis for the stimuli of Experiment 3. In order to establish a temporal synchrony between a gesture fragment and the corresponding speech unit, the DP of gesture was temporally aligned with the IP of the homonym, i.e. the point in time at which the homonym was clearly recognized by listeners (see Figure 5.1). The IPs had been determined previously using a gating paradigm (see Section 3.2.5. Determining the identification points of the homonyms, p. 80).

Interestingly, the onset of the preparation phase of the synchronized gesture fragments still precedes the onset of the homonym by an average of 160 ms. Thus, the gesture onset is still preceding the onset of the co-expressive speech unit as it is usually observed in natural conversation (McNeill, 1992; Morrel-Samuels & Krauss, 1992).



**Figure 5.1** The temporal alignment between gesture fragments and speech in Experiment 3 in contrast to Experiments 1 & 2.

#### *Procedure, ERP recording and data analysis*

The procedure as well as parameters for ERP recording, artifact rejection and analysis were identical to Experiment 2. Behavioral data were not used as rejection criterion. Overall, 25 % of the trials were excluded from further analysis. Based on visual inspection, separate repeated measures ANOVAs with Gesture (D,S), Target Word (D,S), ROI (1,2,3,4,5) and Region (anterior, posterior) as within-subject factors were performed for time window of the homonym (100 to 400 ms) and the target word (300 to 500 ms). These time windows were identical to those used in Experiment 1 and 2. Additionally, an ANOVA was performed for an earlier time window at the position of the homonym (50 to 150 ms) based on visual inspection.

### 5.1.3. Results

#### *Behavioral Data*

Similar behavioral results as in Experiment 2 were observed. Participants responded correctly in 82 % of all test trials. Again, the movement task was carried out less efficient (74 % correct) than in the word task (90 % correct; Wilcoxon signed-rank test;  $z = -3.80$ ;  $p < .001$ ).

#### *ERP data –Homonym*

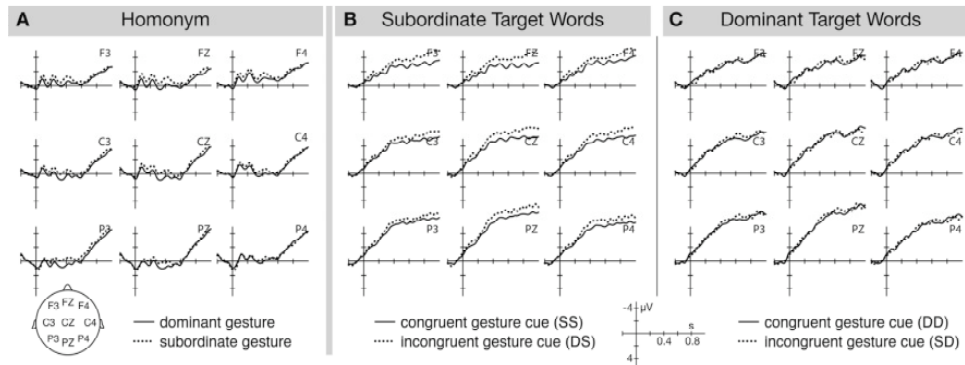
As can be seen in Figure 5.2 A, an increased negativity is elicited when the homonym is preceded by a subordinate gesture fragment as compared to a dominant one<sup>13</sup>. As in Experiment 1, the earliness of the effect can be explained by the use of the homonym IPs as a trigger for the averages. A repeated measures ANOVA yielded a significant main effect of Gesture ( $F(1,31) = 6.09$ ;  $p < .05$ ), a two-way interaction of Gesture and ROI ( $F(4,124) = 8.09$ ;  $p < .001$ ) as well as a significant three-way interaction of Gesture, Region and ROI ( $F(4,124) = 4.07$ ;  $p < .05$ ;  $\epsilon = .46$ ). These results suggest that the integration of a subordinate gesture fragment with a homonym is more difficult than the integration of a dominant one. Further step-down analyses revealed that the main effect of Gesture was strongest over fronto-central sites ( $F(1,31) = 10.46$ ;  $p < .001$ ).

#### *ERP data –Target word*

No significant Gesture Target Word interaction was found in the early time window (Figure 5.2 B & C; all  $F_s < 1.05$ ; all  $p_s > .34$ ). For the N400 time window, however, the ANOVA revealed a significant interaction of Gesture and Target Word ( $F(1,31) = 7.72$ ;  $p < .01$ ). Based on this interaction, the simple main effects of Gesture were tested separately for the two Target Word conditions. At subordinate target words, the N400 was significantly larger after a dominant gesture compared to a subordinate one ( $F(1,31) = 6.63$ ;  $p < .05$ ). No such effect of Gesture Target Word congruency was found at dominant target words ( $F(1,31) =$

<sup>13</sup> In contrast to Experiment 1 and 2, clear early ERP-components can be seen in both gesture conditions in Experiment 3. These components are due to the offset of the gesture fragment and relate to the physical properties of the stimulus (cf. Donchin, Ritter, & McCallum, 1978). For the present purpose, only the negative modulation of the ERP is of importance.

0.33;  $p = .57$ ). Thus, when gesture fragments and speech are synchronized, the integration of both sources of information seems to be more automatic / less effortful, at least for the subordinate word meaning.



**Figure 5.2** ERPs as found in Experiment 3.

#### 5.1.4. Discussion

In Experiment 3, gesture fragments were presented in synchrony with the homonyms to test the synchrony / temporal overlap hypothesis based on McNeill (1992). This manipulation led to a robust enhanced negativity for the subordinate gestures at the homonym in comparison to the dominant gesture indicating that participants integrated the gesture fragments with the homonym. For the disambiguation, a significant N400 effect at subordinate target words was found, but not at the dominant target word. This result is a bit puzzling in light of the results of Experiment 1, and needs further explanation. Previous research on homonym disambiguation has shown that weak contexts only affect the subordinate meaning but not the dominant meaning of a homonym (e.g. Holle & Gunter, 2007; Martin et al., 1999; Simpson, 1981). The findings of Experiment 3, where a shallower task was used, resemble these earlier results. When participants are not pushed by the task to integrate gesture and speech, the meaning of the fragments seems to be treated as weak context. In Experiment 1, however, the task required the participants to actively combine the information from the two domains. Because the task demands modified the perceived importance of the semantic relationship between gesture and speech, it also changed the weak gesture context into a strong one. Summing up, the ERP data of Experiment 3 therefore suggest that when gesture and speech

are in synchrony, they are integrated interaction is obligatory. When both domains are not in synchrony, one could speculate that effortful gesture-related memory processes are necessary to be able to combine the gesture fragment and speech context in such a way that the homonym is disambiguated correctly.

In general, the findings of Experiment 3 provide good evidence in favor of the synchrony / temporal overlap hypothesis. When the iconic gesture fragments were synchronized with their co-expressive speech unit (i.e., the homonym), an effect of immediate gesture-speech integration was found at the position of the homonym. This result also gives rise to the assumption that processing of synchronous gesture and speech information is obligatory as proposed by Kelly et al. (2010). Participants were not required to take the gesture fragments into account. If, however, there is no overlap and immediate integration is not feasible (as in Experiment 2), the information within the gesture fragments probably gets lost over time and cannot be integrated with the homonym. Thus, our results confirm McNeill's suggestion (1992) that temporal synchrony of gesture and speech is a crucial factor for a proper (i.e. in his view automatic) integration of both streams of information.

Temporal synchrony or overlap is not only important in gesture-speech integration, but also has a significant impact on all other types of multimodal integration. Most multi-modal integration studies, do not find a single specific temporal alignment between the auditory and visual stream that triggers integration, but rather identify a time window in which integration is feasible (e.g. McGurk-Effect, van Wassenhove et al., 2007). In general, the literature on multimodal integration suggests that the more complex a signal is, the larger the temporal window of integration is. I.e. a greater amount of temporal asynchrony is tolerated for complex material (e.g. Dixon & Spitz, 1980; Grant, van Wassenhove, & Poeppel, 2004; Vatakis & Spence, 2006a, 2006b). Simple audiovisual stimulus material (Zampini, Shore, & Spence, 2003) is already perceived as asynchronous, if the stimulus onset asynchrony (SOA) exceeds 60-70 ms. For more complex material, the visual signal can start up to 200 ms prior to the auditory signal or start up to 100 ms after the auditory signal (Dixon & Spitz, 1980; Grant et al., 2004; Vatakis & Spence, 2006a, 2006b). Based on all these findings, it is very likely that there is also a time window for gesture-speech integration. Without doubt, the combined gesture fragment and speech signal is probably more complex compared to, for instance, lip movements and syllables (van Wassenhove et al., 2007). Thus, it is likely that a

potential temporal window of integration is in the range of -200 ms (auditory lag) to +100 ms (visual lag<sup>14</sup>). As already noted above, synchrony was not defined with regard to the SOA, but based on the occurrence of the gesture DP in relation to the homonym IP. In order to test whether there is a temporal window of integration for gesture fragments and homonym, we used four different temporal alignments between the homonym and the gesture fragment: the uniqueness point of the noun was either prior (+120 ms), synchronous with (0 ms, replication of Experiment 3) or lagging behind the end of the gesture fragment (-600 ms / -200 ms). As in all previous experiments, both direct integration of gesture fragment and homonym and the delayed effect at the subsequent target word were analyzed.

## 5.2. Experiment 4: Is there a temporal window for gesture-speech integration?

### 5.2.1. Methods

#### *Participants*

Forty-one native German-speaking participants were paid for their participation and signed a written informed consent. Nine of them were excluded because of excessive artifacts. The remaining 32 participants (16 female; 22-29 years, mean 25.4 years) were right-handed (mean laterality coefficient 95.4, Oldfield, 1971), had normal or corrected-to-normal vision, no known hearing deficits and had not taken part in any previous experiments using the identical stimulus material.

#### *Stimuli*

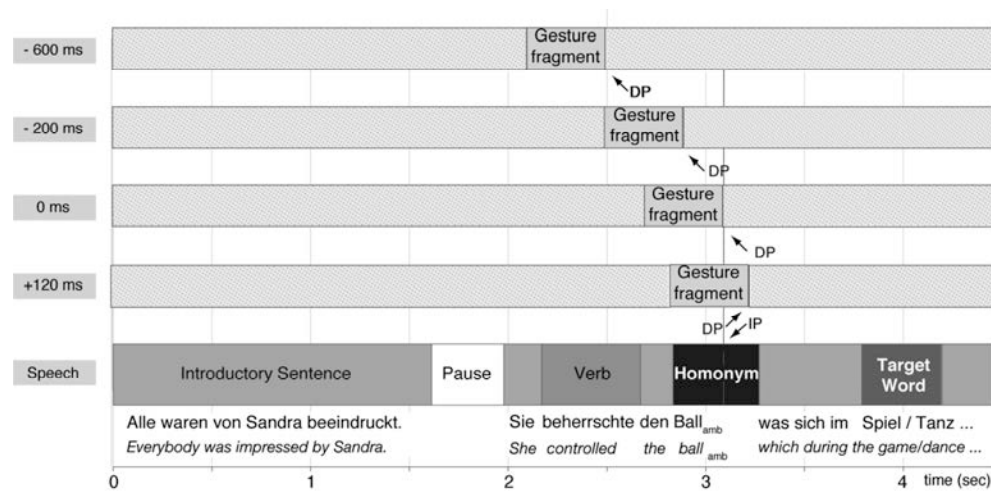
Again, the original 96 gesture videos constituted the basis for the stimuli of Experiment 4. In order to establish the four different levels of temporal synchrony between a gesture fragment and the corresponding speech unit (-600 ms, -200 ms, 0 ms (synchronous, as in Experiment 3, + 120 ms – relative timing of gesture offset to the homonym IP), the DP of a gesture was

---

<sup>14</sup> Auditory lag means that, for example, the onset of a word occurs later than the gesture onset, whereas visual lag means the opposite.



shifted with regard to the corresponding homonym IP (see Figure 5.3). The additional timing manipulation led to a 4 x 2 x 2 design with Timing (-600 ms, -200 ms, 0 ms, +120 ms), Gesture (**D**ominant vs. **S**ubordinate) and Target word (**D**ominant vs. **S**ubordinate) as within subject factors, resulting in an experimental set of 768 stimuli. The stimulus material was pseudo-randomly distributed in such way that all factors were balanced across the different experimental lists, leading to a total set of 16 lists for each of the two sessions of the experiment.



**Figure 5.3** The temporal alignment between gesture fragments and speech in Experiment 4.

#### *Procedure, ERP recording and data analysis*

For each participant, there were two experimental sessions separated by 7 to 14 days. The procedure itself as well as the parameters for ERP recording, artifact rejection and analysis were identical to Experiments 2 and 3. Behavioral data were not used as rejection criterion. Overall, 21 % of the trials were excluded from further analysis. Based on visual inspection, separate repeated measures ANOVAs with Timing (-600 ms, -200 ms, 0 ms, +120 ms), Gesture (D,S), Target Word (D,S), ROI (1,2,3,4,5) and Region (anterior, posterior) as within-subject factors were performed for time window of the homonym (200 to 500 ms) and the target word (300 to 500 ms). These time windows were identical to those used in

Experiments 1-3. Main effects of synchrony are not reported, as they are difficult to interpret due to the considerable differences in movement parameters between the synchrony conditions at the position of the homonym.

## 5.2.2. Results

### *Behavioral Data*

The behavioral results were slightly better than in Experiments 2 and 3, probably due to the higher number of trials and repetitions. Participants responded correctly in 93 % of all test trials. The movement task was carried out marginally less efficient (92 % correct) than in the word task (94 % correct;  $F(1,31) = 2.63$ ;  $p > .13$ ).

### *ERP data –Homonym*

As can be seen in Figure 5.4 A, an increased negativity is elicited when the homonym is preceded by a subordinate gesture fragment as compared to a dominant one for some of the four timing conditions. As in Experiment 1 and 3, the earliness of the effect can be explained by the use of the homonym IPs as a trigger for the averages. A repeated measures ANOVA yielded a significant main effect of Gesture ( $F(1,31) = 5.64$ ;  $p < .05$ ), and a significant three-way interaction of Timing, Gesture, Region and ROI ( $F(12,372) = 2.63$ ;  $p < .05$ ;  $\epsilon = .50$ ). Step-down analysis revealed the following: There was no significant effect of Gesture in the -600 ms condition (all  $F$ s  $< 2.16$ ; all  $p$ s  $> .12$ ). In the -200 ms, the additional analyses revealed a significant main effect of gesture. However this effect was limited to the posterior two left-most ROIs (both paired  $t$ s(31)  $> 2.22$ ; both  $p$ s  $< .05$ ). In the 0 ms condition, the analyses revealed a broadly distributed main effect of gesture (all paired  $t$ s(31)  $= 2.22$ ; all  $p$   $< .05$ ), whereas in the +120 ms condition, there was only an anterior main effect of gesture (paired  $t$ (31)  $= 2.60$ ;  $p < .05$ ). Thus, integration of gesture and homonym only seems to be possible in the -200 ms, 0 ms and +120 ms conditions. As in Experiment 1 and 3, the results suggest that the integration of a subordinate gesture fragment with a homonym is more difficult than the integration of a dominant one.

*ERP data –Target word*

As can be seen in Figure 5.4 B and C, there is a clearly enhanced negativity for incongruent as compared to congruent gesture cues at the subordinate target word, seemingly independent of the synchrony manipulation. In the classic N400 time window, the ANOVA revealed significant main effect of Gesture ( $F(1,31) = 8.13$ ;  $p < .01$ ), a significant main effect of

Target Word ( $F(1,31) = 16.42$ ;  $p < .001$ ), as well as a significant interaction of Gesture and Target Word ( $F(1,31) = 21.09$ ;  $p < .0001$ ). Based on this interaction, the simple main effects of Gesture were tested separately for the two Target Word conditions. At subordinate target words, the N400 was significantly larger after a dominant gesture compared to a subordinate one (paired  $t(31) = 5.17$ ;  $p < .0001$ ). No such effect of Gesture-Target Word congruency was found at dominant target words (paired  $t(31) = 1.09$ ;  $p > .28$ ). Thus, independent of our timing manipulation gestures exerted a disambiguating effect at the subordinate target word.

**5.2.3. Discussion**

Experiment 4 was set out to clarify whether there is a temporal window for the integration of gesture fragments with speech. To do so, the temporal alignment of the gesture fragments was varied with regard to the homonym IP. The offset (DP) of the gesture fragments either ended prior (-600 ms or -200 ms) to the homonym IP, was synchronous with it (0 ms), or ended after the homonym IP (+120 ms). For the integration of the gesture fragment with the homonym, the ERP results only show effects in the -200 ms, 0 ms and +120 ms condition, with the strongest effects in the synchronous condition. Thus, there seems to be a time window for gesture fragment and speech integration as it has been found for a lot of other types of multimodal integration (e.g. Dixon & Spitz, 1980; Grant et al., 2004; Soto-Faraco & Alsius, 2007; Vatakis & Spence, 2006a). The range of this time window is from -200 ms (audio lag) to at least +120 ms (gesture lag). No impact of timing was found for the disambiguation effect at the target word. In all conditions, the incongruent dominant gesture cues elicited a larger N400 at subordinate target words than congruent subordinate gesture cues. Thus, integration of gesture and homonym and the processing target word are assumed to be distinct. Why there is this differential impact of timing and whether local integration of gesture and speech is a necessary prerequisite or not for the global effect at the target, cannot be clarified on the basis of the present data. The results of Experiment 4 also fit to a recently

published paper by Habets et al. (2010), who specified the significance of timing for the disambiguation of a gesture by an accompanying word. Although they used a SOA manipulation in contrast to the present study, it is possible to compare their findings with those of Experiment 4, as they also determined a gesture identification point. Habets et al. (2010) presented their participants with gesture clips at three different SOAs with regard to a target word. The gestures either started 360 ms prior, 160 ms prior or synchronous to the word. Participants had to answer a post-test questionnaire about the gestures and words. The authors found effects of gesture-speech integration for the SOA of 160 ms (gesture IP is 200 ms after the onset of the corresponding word) and in the synchronous condition, but not for the SOA of 360 ms (gesture IP end before the onset of the corresponding word). In Experiment 4, the gesture DPs in the -200 ms condition also overlap with the homonym in time, whereas the +120 ms condition is similar to the synchronous condition in the Habets et al. study (2010). Thus, both studies find quite similar results for the time window of gesture-speech integration although Habets et al. (2010) did use gestures comprising the whole stroke and retraction phase in contrast to the gesture fragments used in the present study. Note, however, there are a few caveats to the study. First, their gestures started immediately with the stroke phase, because the preparation phase was omitted. The present results, however, show that the preparation phase can contain important information. Second, the gestures were not cut at their identification points but continued and completely overlapped with the corresponding words. Thus, in light of previous research (e.g. Holle & Gunter, 2007), it is very surprising that Habets et al. (2010) did not find any effect in the SOA 360 ms condition. Importantly, Wu and Coulson (2007b), using a comparable paradigm have shown that SOAs of up to 1000 ms do not necessarily have to present a problem for the integration of gesture and speech. Habets et al. (2010) argue, that the gestures they used are much more ambiguous and that in the SOA 360 ms condition the process of meaning assignment to the gesture is already finished before the word starts (they assume that this process is completed after 360 ms). Based on the gating literature, the meaning of a word is seldom identified prior to 250 ms of presentation. Thus, the authors' argumentation seems to be correct with regard to the SOA 360 ms condition, however it causes some problems for the results of the SOA 160 ms condition. According to the authors, the meaning assignment to the gesture in this condition should be completed after 200 ms of the corresponding word have been heard. This would rather contradict the gating literature on word meaning identification (e.g. Grosjean (1980) reports a mean isolation time for English words of about 330 ms). Therefore, if the authors' argumentation for the 360 ms condition was correct they also should not find any effects in



Figure 5.4 ERP results of Experiment 4.

the SOA 160 ms condition. The determination of the IPs for the words they used could clarify this issue.

It is much more difficult to compare the present findings with results from multimodal integration research, as there is usually in contrast to gestures no meaning in multimodal stimuli and timing is always manipulated on a SOA basis (e.g. Dixon & Spitz, 1980; Grant et al., 2004; Soto-Faraco & Alsius, 2007; Vatakis & Spence, 2006a). Therefore, in order to compare the present data with previous findings in multimodal integration, the timing manipulations in Experiment 4 (-600 ms, -200 ms, 0 ms, +120) were recalculated in terms of a SOA manipulation (-860 ms, -360 ms, -160 ms, -40 ms). On a SOA basis the temporal window of gesture fragment and speech integration ranges from -360 ms to -40 ms. Whereas the left boundary (-360 ms) of the time window might be accurate, the right boundary (-40 ms) has to be further specified in terms of a potential visual lag. The SOA-based temporal window of integration of gesture fragments and speech allows to directly compare the present results with the findings of Habets et al. (2010). The time windows of integration are very similar. In fact, the time-window is even larger in Experiment 4 than in their study. Clearly, some more work is needed to get a better idea about the actual size of the temporal window of integration for gesture and speech.

If one compares the time window of integration for gesture and speech (an especially the left boundary) with other multimodal integration research (-200 ms to +100 ms), then it becomes clear that the gesture-speech integration system can cope with asynchronies that are almost twice as large as in other audiovisual integration experiments. In the literature, it is hypothesized, that the temporal window integration increases with complexity of the to-be-merged information (Vatakis, Navarra, Soto-Faraco, & Spence, 2007). If one assumes that gesture-speech integration is more complex (additional semantic level) than previously studied multimodal stimuli (e.g. syllables and lip movements, van Wassenhove et al., 2007), it is not surprising that the gesture-speech integration system can cope with larger asynchronies. Alternatively, addressees might be able to cope with larger asynchronies between gesture and speech, because gesture and speech are produced with some asynchrony most of time (e.g. McNeill, 1992). Morrel-Samuels and Krauss (1992), for example, found that gestures are initiated approximately 1000 ms prior to the lexical affiliate. In contrast, the temporal coupling between lip and mouth movements (as it is used in the McGurk effect) is much stronger, most likely because it is based on pure physical features in contrast to gesture and speech, which are also semantically linked. Therefore, it could be that we are habituated

to a certain degree of asynchrony between gesture and speech and are thus able to cope with larger asynchronies.

## **Chapter 6**





## Chapter 6

### The impact of background noise on gesture-speech integration

The results so far indicate that the integration of gesture and speech in comprehension may only be an obligatory process within a certain temporal alignment (Experiments 2 - 4) and otherwise be modulated by situational factors such as the amount of observed meaningful hand movements (Holle & Gunter, 2007) and task (Experiments 1 - 2). It is known, however, that in natural conversation, gesture and speech can sometimes completely lack temporal overlap (Chui, 2005). Considering this finding, one could ask whether there are certain natural situations or situational factors which promote the online integration of gesture and speech, even if both streams of information are temporally non-overlapping.

For example, think of yourself sitting in a crowded and noisy bar talking to a friend. In such a situation you will probably take her gestures more into account compared to when the conversation would take place in a quiet place (cf. Rogers, 1978). It is unknown, whether this use of gesture information is done automatically or not. It is very well possible when you are not that interested in what she is saying you do not take her gestures into account because you process her communicative signals very shallow.

There is consent among gesture researchers that an impoverished speech signal represents a situation in which co-speech gestures occur with a higher frequency (e.g. Hoskin & Herman, 2001; Kendon, 2004). It is very well possible that persons who are either temporally (like in the bar example) or chronically exposed to such a suboptimal situation (say noise on the work floor or when a person has hearing problems), take gestures always (and possibly automatic) into account because, as the Rogers (1978) study already showed, gestures are a very relevant and helpful cue. Thus, in Experiment 5, the question of interest was whether normal hearing persons exposed to a suboptimal hearing situation would process gesture information differently, i.e. more automatically, than in silence (optimal hearing situation). The out-of-sync situation in our stimulus material gives us an interesting case to explore the automaticity of gesture-speech integration when a suboptimal communicative situation is present. As

Experiments 1 and 2 have shown, participants do not integrate the asynchronous gesture and speech unless the situational factor (i.e. the task) pushes them to do so. Note, however, that in contrast to a task-manipulation, impoverished speech signal does not pose an externally evoked top-down process but can be seen as an internally generated one.

In order to explore the influence of an impaired speech signal on gesture-speech processing, the gesture-speech integration system of the participants was put to the test by presenting our stimuli embedded in multi-speaker babble noise. Assuming that due to the use of the interfering babble noise the perceiver is looking for as much cues as possible in a more or less automatic fashion, it is hypothesized that participants integrate highly asynchronous gesture and speech information even when they have to perform a shallow task.

## **6.1. Experiment 5**

### **6.1.1. Introduction**

In Experiment 5 we explore whether asynchronously presented gesture-fragments are used to disambiguate babble distorted speech in a more or less automatic fashion. The same materials and experimental setup as in Experiment 2 was used. Participants had to take part in two sessions. In one of the sessions the participants received the original stimulus material whereas in the other session speech was embedded in multi-speaker babble noise. Because the silence session is a direct replication of Experiment 2, it is expected that the participants will not show any sign of gesture-speech integration. In contrast, since gestures are an important cue in a noisy environment, it is hypothesized that in the babble noise condition participants will both integrate gesture and speech and show clear effects of disambiguation at the target word position later in the sentence.

### **6.1.2. Methods**

#### *Participants*

Thirty-four native German-speaking participants were paid for their participation and signed a written informed consent. Six of them were excluded because of excessive artifacts. The remaining 28 participants (14 female; 19-32 years, mean 25.7 years) were right-handed

(mean laterality coefficient 94.1), had normal or corrected-to-normal vision, no known hearing deficits and had not taken part in an experiment using the same stimulus material.

### *Stimuli*

The stimuli were identical to Experiments 1 and 2. The multi-speaker background noise used in this experiment was constructed as follows.

#### *Multi-speaker babble speech*

Multi-speaker babble tracks (MSB) were created by overlaying speech streams from different commercially available German audio books. A total of 10 different speakers were used (5 female). First, intros, music and silence (intensity  $< -70.9$  dB, duration  $> 140$  ms) were removed. The tracks were cut to a length of 15 minutes and normalized to 65 dB using the Praat 4.5.1.0 software (Boersma & Weenink, 2005). Each track was duplicated and the resulting twenty tracks were aligned with a randomly varying temporal onset (between 0 and 30 ms). The first 1.5 minutes as well as the final 3.5 minutes were removed before the mix-down. This procedure resulted in a “dense” 10-minute babble track, which made it impossible to identify individual speakers or words. The 10-minute babble track was extended to a length of 60 minutes to allow the use of one continuous babble speech stream throughout the experiment. Finally, the babble stream was normalized to five different sound intensities (62.5 dB, 65 dB, 67.5 dB, 70 dB and 75 dB) while speech was normalized to 65 dB. In order to obtain the optimal speech-to-babble ratio (S/B ratio) for the actual ERP experiment, several pretests were carried out.

In a first pretest, participants were presented with speech embedded in babble at the five different signal-to-babble ratios (S/B ratio: +2.5 dB, 0 dB, -2.5 dB, -5 dB, -10 dB) and had to orally reproduce the sentences. They were completely unable to reproduce them at an S/B ratio of -5 and -10 dB. Thus, both of these S/B ratios were excluded from further testing. In a second pretest, nine right handed participants (5 female, age range 21-32 years, mean 24.2 years, mean laterality coefficient 95.9) had to listen to sentences at the remaining three S/B ratios (+2.5 dB, 0 dB, -2.5 dB). Their task was to write down the sentences they heard. Sentences were identical to those used in the previous experiments. Response accuracy at the homonym, target word and complete sentence level was investigated. The response accuracy

was quite high for homonyms for S/B ratios of +2.5 dB and 0 dB (97 % and 92 %), target words (98 % and 88 %) and complete sentences (88 % and 72 %) but considerably worse at a S/B ratio of -2.5 dB (homonym: 75 %; target word: 68 %; complete sentences: 36 %). Repeated measures ANOVA with S/B ratio (-2.5 dB, 0 dB, 2.5 dB) as within-subject-factor revealed a significant main effect of S/B ratio for homonyms, target words and complete sentences (all  $F_s(2,8) > 44.30$ ; all  $p_s < .001$ ; all  $\varepsilon = .54$ ). Paired t-tests for the different sentence components revealed, that participants were able to understand both the homonym, target word and the complete sentences significantly better at a S/B ratio of 0 dB or + 2.5 dB than at - 2.5 dB (all paired  $t_s(8) > 6.67$ ; all  $p_{sBon} < .001$ ). The 0 dB and the +2.5 dB condition did not differ significantly for homonyms (paired  $t(8) = 2.58$ ;  $p_{Bon} > .10$ ) but for target words and complete sentences (all paired  $t_s(8) > 4.75$ ; all  $p_{sBon} < .001$ ). Because of its low response accuracy, the -2.5 dB S/B ratio condition seemed to be inappropriate for the ERP experiment. In contrast, participants were able to understand almost every homonym and target word at an S/B ratio of +2.5 dB making this condition also unusable for our purposes because it resembles too much the processing of speech without background noise. We therefore opted to use the intermediate 0 dB S/B ratio condition. Subjects comprehended speech quite well in this condition, and thus should be able to integrate gesture and speech. Yet, they did not understand everything and should therefore benefit from additional gestural input. Previous research concerning multi-sensory integration supports this stance, as it has been found that the speech recognition system benefits the most from additional visual input at intermediate signal-to-noise levels (Holle et al., 2010; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Consequently, the babble noise as well as speech were both presented at an intensity of 65 dB.

### *Procedure*

The procedure was identical to Experiments 2. The experiment, however, consisted of two sessions, one in which participants saw the stimuli in the babble noise condition, and another one in which participants were presented with the silence condition. The sessions were separated by 7 to 14 days and the order of the noise conditions was counterbalanced across the participants.

*ERP recording*

The EEG was recorded from 61 Ag/AgCl electrodes (Electrocap International). It was amplified using a REFA 8, 72 channel amplifier (DC to 135 Hz) and digitized at 500 Hz. Electrode impedance was kept below 5 k $\Omega$ . The left mastoid served as a reference. Vertical and horizontal electro-oculogram (EOG) was measured for artifact rejection purposes.

*Data Analysis*

The artifact rejection and analysis were similar to Experiment 2 - 4. Behavioral data were, as in Experiment 2 - 4, not used as rejection criterion. Overall, 21 % of the data were excluded from further analysis. Based on visual inspection, a time window ranging from 100 to 250 ms was used to analyze the integration of gesture and homonym. A repeated measures ANOVA using Noise (babble, silence), Gesture (D, S), ROI (1, 2, 3, 4, 5), and Region (anterior, posterior) as within-subject factors was calculated. Only effects which involve the crucial factors Noise and Gesture will be reported.

For the target word analysis, the standard N400 time window ranging from 300 to 500 ms after target word onset as well as a time window from 550 to 800 ms was chosen. A repeated measures ANOVA was performed using Noise (babble, silence), Gesture (D, S), Target Word (D, S), ROI (1, 2, 3, 4, 5), and Region (anterior, posterior) as within-subject factors. As in all previous experiments, only effects which involve the crucial factors noise, gesture or speech will be 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. Prior to all statistical analyses the data were filtered with a low-pass filter of 20 Hz. Additionally, a 10 Hz low-pass filter was used for presentation purposes only.

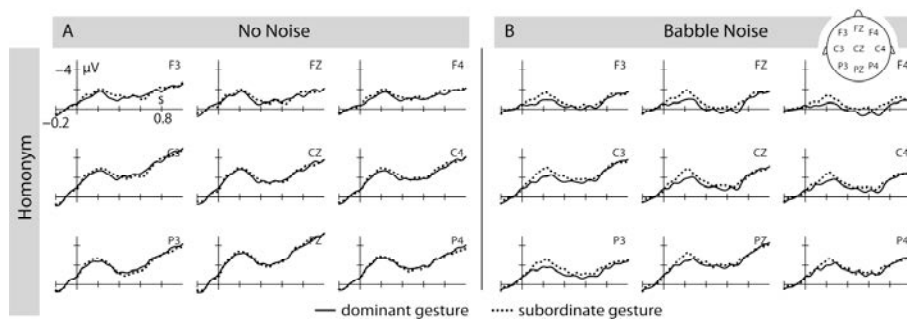
**6.1.3. Results***Behavioral Data*

Overall, participants gave 90 % correct answers, indicating that although the task in the experiment was rather shallow, participants nonetheless paid attention to the stimulus material. A repeated measures ANOVA with the factors noise (babble, silence) and task (movement task, word task) revealed a significant interaction of both factors ( $F(1,27) = 4.80$ ,

$p < .05$ ). Step-down analyses revealed that this interaction was driven by the significantly worse performance in the word task during babble noise (91% correct responses) as compared to the word task in the silence condition (99 %; paired  $t(27) = 3.66$ ,  $p = .001$ ). No such difference between the noise conditions was found for the gesture task (babble: 85%, no-noise: 84%; paired  $t(27) = -.47$ ,  $p > .64$ ). This data indicates that, as could be expected, the noise manipulation only affected the processing acoustic stimuli.

#### ERP data – Homonym

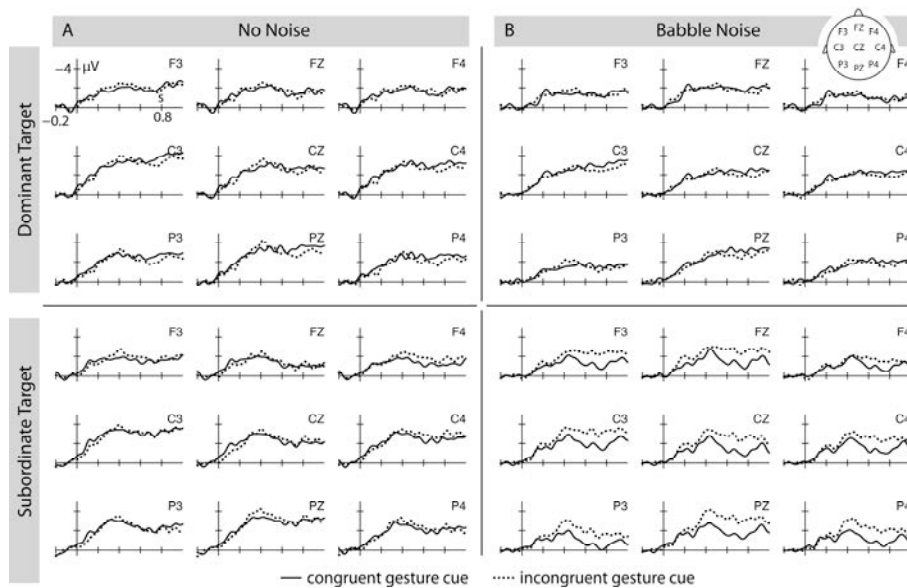
In Figure 6.1, an early, enhanced negativity can be observed when the homonym is preceded by subordinate gesture fragments as compared to dominant gesture fragments in the babble condition (Figure 6.1 A). No such difference is visible in the silence condition (Figure 6.1 B). A repeated measures ANOVA revealed a significant interaction of the factors noise, gesture and ROI ( $F(4,108) = 3.20$ ,  $p = .05$ ,  $\eta^2 = .49$ ). Subsequent step-down analyses yielded a main effect of gesture for all ROIs in the babble condition (all paired  $t_s(27) > 2.74$ , all  $p_s < .05$ ), but no significant effects in the silence condition. This finding not only indicates that it is more effortful to integrate a subordinate gesture with the corresponding homonym during babble noise than a dominant one, but also that gesture-speech integration only occurred in the babble but not in the silence condition.



**Figure 6.1** ERPs time-locked to the identification point of the homonyms as found in Experiment 5. The left panel (A) shows the ERPs in the silence condition whereas the right panel (B) represents the ERPs in the babble condition. The dotted line represents ERPs when the homonym was preceded by a subordinate gesture fragment, the solid line those cases when it was preceded by a dominant gesture fragment.

## ERP data – Target Word

Figure 6.2 A shows an extended negative deflection starting around 400 ms for incongruent dominant gestures cues in comparison to congruent subordinate gesture cues at subordinate target words. Though a bit late in time, the negativity was identified as a member N400 based on the experimental manipulation and scalp distribution. The analysis of the 300 to 500 ms time window showed no significant effects or interactions (all  $F$ s < 2.76, all  $p$ s > .11). The later time window from 550 to 800 ms, however, yielded a significant three-way interaction of the factors noise, gesture and speech ( $F(1,27) = 4.26$ ,  $p < .05$ ). Step-down analyses for the subordinate targets revealed that only in the babble noise condition the N400 was significantly larger after a dominant compared to a subordinate gesture (paired  $t(27) = 3.92$ ,  $p < .001$ ). No such effect was found at dominant target words (Figure 6.2 B; paired  $t(27) = 1.01$ ,  $p > .32$ ). Thus, when gesture fragments and speech are presented under adverse listening conditions, the integrated information from both sources is used for homonym disambiguation, at least for the subordinate word meaning.



**Figure 6.2** ERPs time-locked to the onset of the target words as found in Experiment 5. The upper panel represents the ERPs at dominant target words whereas the lower one represents the ERPs at subordinate target words. The left panel (A) shows the ERPs in the silence condition whereas the right panel (B) represents the ERPs in the babble condition. The solid line represents the cases in which gesture cue and subsequent target word were congruent. The dotted line represents those instances where gesture cue and target word were incongruent.



**6.1.4. Discussion**

Experiment 5 explored whether normal hearing persons take gesture more or less obligatorily into account when speech is embedded in multi-speaker babble noise. We therefore presented our participants with gesture fragments which were out of synchrony with their lexical affiliate (homonym) and gave participants a shallow task. In one of the sessions speech was embedded in multi-speaker babble noise, in the other session speech was noise free. The data show that when speech was embedded in babble noise, gesture information was integrated with the homonym which led to a disambiguation later in the sentence. When speech was processed in total silence the participants did not take the gesture information into account and did not show any disambiguation effects. The ERP data in the silence condition are a direct replication of the null-effect found in Experiment 2. Experiments 1-4 have shown that gesture-speech integration at the homonym only takes place under two conditions. When gesture and speech occur in a certain temporal relation or when active gesture-related memory processes are used in the case both streams of information are asynchronous. The results of Experiments 5 show that in difficult communicative situations the information uptake from asynchronous gestures becomes more obligatory. The crucial difference to Experiment 1 is that background noise leads to an internally generated change in the default processing of gesture and speech. Most likely and similarly to the processing under explicit task conditions (Experiment 1), the utilization of some kind of gesture related memory might be crucially involved in the integration of gestures with speech in a noisy environment. To sum up, the results of Experiment 5 suggest that the gesture-speech integration of asynchronous materials becomes rather obligatory in a noisy environment. The default processing strategy (in contrast to the strategy used under optimal conditions) now additionally seems to involve the active use of memory.

## **Chapter 7**



## **Chapter 7**

### **General Discussion**

#### **7.1. Summary of the results**

The present series of experiments explored gesture-speech integration in sentence comprehension and the degree in which integration depends on task, temporal synchrony and background noise (speech signal quality). For this purpose, a disambiguation paradigm was used, i.e. participants were presented with sentences containing an unbalanced homonym which was later on disambiguated by a target word. In order to enhance the precision in measuring temporal aspects of gesture-speech integration, we presented our participants with gesture fragments. To do so, we first assessed the minimally necessary amount of iconic gesture information needed to reliably disambiguate a homonym using a context-guided gating task. The gesture fragments were presented as disambiguating cues with the corresponding homonyms. Both the direct integration of gesture fragment and homonym was analyzed as well as the disambiguating effect of the gesture-homonym combination at the target word.

In Experiment 1, where gesture fragment and homonym were asynchronous and an explicit task was used, the ERPs triggered by homonym IPs revealed a direct influence of gesture during the processing of the ambiguous word. Subordinate gesture fragments elicited a more negative deflection compared to dominant gesture fragments, indicating that the integration of subordinate gesture fragments with the homonym is more effortful. The ERP data at the target words showed that the gesture fragments were not only integrated with speech, but were also used to disambiguate the homonym. When a target word was incongruent with the meaning of the preceding gesture-homonym combination, a larger N400 was elicited as compared to when this meaning was congruent. The target word effect is similar to the findings by Holle & Gunter (2007) who used full-length gestures instead of fragments. Thus, the results of Experiment 1 revealed that participants were able to use the gesture fragments for disambiguation both at local homonym as well as global sentence level.

In order to explore the nature of the gesture-speech interaction, a shallower task was used in Experiment 2. If gesture-speech integration was an automatic, task-independent process, this task manipulation should have resulted in similar ERP patterns as found in Experiment 1, where participants were explicitly asked to judge the compatibility between gesture and speech. This was, however, not the case, since no significant ERP effects were observed in Experiment 2. One possible interpretation is that gesture-speech integration is task-dependent and thus not automatic according to the two-process theories of information processing (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The data of Experiment 3, however, suggests a different reason for the null finding of Experiment 2.

In Experiment 3, the identical task as in Experiment 2 was used. The crucial difference was that the gesture fragments and homonyms were synchronized (to the homonym IP) this time. This synchrony manipulation led to a robust negativity for the subordinate gestures at the homonym as well as to significant N400 effects at subordinate target words. Thus, the information uptake from gesture seemed to be rather obligatory when both streams of information were synchronous or shared some temporal overlap, which was not the case in Experiments 1 and 2. This finding can be linked to McNeill's statements on the role of timing in gesture production (1992) and comprehension (1994). He suggests that synchrony is a very crucial aspect in gesture production in order for gesture and speech to form a single idea unit (1992). When a synchronous gesture-speech combination is encountered by an addressee, McNeill (1992) proposes that the information from both sources is automatically and involuntarily integrated.

Findings from other studies on gesture production suggest, however, that the timing between gesture and speech varies considerably (e.g. Morrel-Samuels & Krauss, 1992). It has been shown for other multimodal integration processes, that the exact synchrony is not decisive for the integration. The human perceptual system is able to integrate to some degree asynchronous audio-visual input, i.e. it can compensate a certain degree of asynchrony (e.g. between visual and auditory speech information: Dixon & Spitz, 1980; Soto-Faraco & Alsius, 2007; van Wassenhove et al., 2007; Vatakis et al., 2007). This is very useful, for instance, when talking via a chat client. It is assumed that the so-called temporal window of integration varies with regard to stimulus complexity, i.e. it is larger for more complex stimuli (see Vatakis et al., 2007). For example, for audiovisual speech onset asynchronies within a range of -200 ms (visual lead) to +100ms (auditory lead) can be compensated.

In Experiment 4, the effect of synchrony on gesture-speech integration was explored in more detail. The temporal alignment of gesture and speech was varied in order to explore whether there was also some kind of temporal window of integration for gesture and speech. To be precise, there were four different levels of synchrony between gesture and speech. The identification point of the homonym was either prior (+120 ms), synchronous with (0 ms)<sup>15</sup> or lagging behind the end of the gesture fragment (-600 ms / -200 ms). The ERPs triggered to the homonym IP showed integration of gesture fragment and speech only in the -200ms, 0ms and +120 ms conditions. In these conditions, the subordinate gesture fragments elicited a larger N400 than the dominant gesture fragments. The disambiguating influence of the gesture-homonym combination at the target word was unaffected by the timing manipulation. For all timing conditions, subordinate target words preceded by an incongruent gesture-speech context elicited a larger N400 than those preceded by a congruent context. Thus, based on the present results, the temporal window for the integration of gesture fragment and speech seems to be somewhere between -200 ms (audio lag) and +120 ms (visual lag).

Experiment 5 addressed the question whether there are situations in real life, where even completely asynchronous / non-overlapping gesture information is taken into account in speech processing. In other words, is there a situation in which addressees would integrate gestures with speech in which they normally would not use them (cf. Experiment 2)? It is known, that gestures are especially useful when the speech signal is impaired, for example, when being in a noisy environment like a bar (Rogers, 1978). In Experiment 5, such a situation was reproduced by embedding speech in multi-speaker babble noise. The experimental setup was identical to Experiment 2. Participants were carrying out the experiment both in silence as well as in a babble noise condition. In the silence condition, the null result of Experiment 2 was replicated. In contrast, the ERPs in the babble noise condition showed both clear effects of gesture-homonym integration and disambiguation of the target word. Thus, participants must have switched their default processing strategy of gesture from “no integration due to asynchrony” to ‘integration to compensate speech difficulties’.

The combined ERP data therefore suggest that when gesture and speech are in synchrony, their interaction is more or less obligatory. When both domains are not in synchrony, effortful gesture-related memory processes are necessary to be able to combine the gesture fragment and speech context in such a way that the homonym is disambiguated correctly. These

---

<sup>15</sup> This condition was identical to Experiment 3.

processes can either be triggered externally via task / instruction or generated internally by the addressee her- / himself as in the case of an impaired speech signal. In the following I will address some aspects of the integration and disambiguation process in more detail and will also provide a blueprint for a model of gesture-speech integration. Then, I will deal with some questions raised by the results of this dissertation and how they can be addressed. Importantly, all these questions can be directly related to the processing stages of the proposed model.

### 7.1.1. The integration of gesture fragment and homonym

To date, there is only little theoretical background on gesture-speech integration in comprehension (Holle, 2007; Kelly, Özyürek, et al., 2010; McNeill et al., 1994). For instance, McNeill and colleagues (1994) claim that the information uptake from gesture and its merging with speech is so highly automatic, that addressees simply cannot avoid combining both sources of information. The experimental results of this dissertation show that this stance is not correct, because gesture-speech integration depends on various factors like task, timing and quality of the speech signal. Some papers have already shown that it also depends on the quality of the gesture signal (Holle & Gunter, 2007) and the communicative context (Kelly et al., 2007; Wu & Coulson, 2010).

The data of Experiments 2 to 4 indicate that timing is a very dominant factor in the whole integration process of gesture and speech. As long as gesture DPs and speech IPs are aligned within a certain time window (-200 ms to + 120 ms for DP in relation to IP) the integration seems to be rather obligatory<sup>16</sup>, which is in line with the *Integrated Systems Hypothesis* by Kelly et al. (2010).

If we consider gesture and speech not only from a linguistic perspective, but also as parts of a crossmodal integration process, then there is also a lot of literature that is in line with present results (van Wassenhove et al., 2007; Vatakis & Spence, 2006a). Typically, audiovisual asynchronies of up to 200 ms can be compensated by an addressee such that she / he still perceives visual and auditory information as a single entity, which is quite similar to the temporal window of gesture-speech integration found in Experiment 4<sup>17</sup>. Van Atteveldt,

<sup>16</sup> The integration of synchronous gesture and speech information is *not* automatic. For example, adding meaningless grooming movements to the gestural information alters the impact of gesture on speech (Holle & Gunter, 2007).

<sup>17</sup> Note, that synchrony was defined differently in this dissertation in contrast other studies investigating the impact of timing on crossmodal integration (for details see Section 5.1.1. Introduction, p. 101)

Formisano, Goebel et al. (2007) call this form of crossmodal integration the “perceptual state”-type of integration. I.e. different modalities are judged to belong together according to stimulus-related information (e.g. temporal alignment & content). Synchronous and congruent stimuli are consequently integrated with each other in a bottom-up driven way (e.g. Driver & Spence, 2000; Lalanne & Lorenceau, 2004).

When the asynchronies between gesture and speech surpass the limits set by the temporal window of integration the local integration of gesture and speech breaks down and bottom-up, perceptual-state processing is no longer successful. Thus, the integration of asynchronous gesture and speech is not obligatory anymore. However, outside the temporal window of integration, top-down processes can be activated to enable addressees to integrate asynchronous gesture information. These top-down processes can either be triggered by task instruction (see Experiment 1) or be the result of an input analysis process as in Experiment 5, where the diminished speech input quality results in enhanced integration of gesture information. It is known from research on language processing in deaf people and normal hearing in a noisy environment (Sumbly & Pollack, 1954), that visual cues become a very important source of information. Hearing impaired, for instance, will usually try to use facial and lip movements in face-to-face communication to gain a better understanding of speech. This ability is termed lip- or speech-reading. Since gestures also seem to be an important source of information when speech is difficult to understand, there might also be something like gesture-reading, i.e. addressees try to use every available bit of gestural information as a default.

With regard to task induced top-down effects in audiovisual integration, van Atteveldt, Formisano, Goebel et al. (2007) were able to show using letters and sounds as stimuli, that the timing between the visual and auditory information has no impact once the task requires participants to actively match both streams of information (similar to Experiment 1). Thus, previous crossmodal research as well as the present results shows that top-down processes can clearly override the perceptual-state type of crossmodal integration. This change in the default processing of incoming gesture-speech information may be accompanied by changes in the distribution of attention on gesture and speech as well as the additional recruitment of a potential gesture working memory resources. However, more research is needed to clarify these issues.



### 7.1.2. The effect of gesture-homonym integration at the target word

Gesture fragments and homonyms are obligatorily integrated within the temporal window of integration (-200 ms to +120 ms). This section addresses the question whether the local integration at the homonym has consequences for the processing of the homonym and relates the findings of the dissertation with the results of Holle and Gunter (2007). The main focus will be on those experiments that used the monitoring task and a noise-free presentation of speech, as these conditions most closely resemble our everyday conversational situation. For complete gestures, Holle and Gunter (2007) found that the combined information from gesture and homonym served as a strong disambiguating context at the target word level. I.e. both the subordinate as well as the dominant target word were more difficult to integrate when the preceding context was incongruent than when it was a congruent one. In contrast, the results of Experiments 2 to 4 only reveal a significant effect at the subordinate target words. According to theories on homonym processing, the finding in Experiments 2 to 4 is typical for a weakly biasing context in homonym disambiguation (Martin et al., 1999; Simpson, 1981). In a weak biasing context, contextual information is only used for the disambiguation of the less frequent, subordinate meaning, whereas the dominant meaning is unaffected by context and solely processed based on word meaning frequency.

There are two possible explanations for this difference in ERP patterns. First, the complete gestures used by Holle and Gunter (2007) share some temporal overlap with the target word, which may have amplified the disambiguating impact by allowing participants to directly integrate the gesture with the target.

This explanation, though, is unlikely as there are also results that show that full-length gestures which share temporal overlap with the target word sometimes do not elicit an ERP effect for gesture congruency at the position of the target (see the results at the dominant target word in Experiment 3, Holle & Gunter, 2007). Second, full-length gestures contain more communicative information than gesture fragments. When looking at the results of the cloze pretests for both types of stimuli (See sections 3.1.3 [p. 67] & 3.2.4.1[p. 79]), it becomes evident that there is a difference in communicative value which may account for the differences in the disambiguating power of full-length gestures and fragments.

It is important to note that the communicative impact of full-length gestures is not always the same. For example, when the value of the gesture channel is diminished by adding meaningless movements (thereby decreasing the percentage of meaningful gesture) even full-

length gesture are treated as weak context (Holle and Gunter (2007)). Similarly, the communicative impact of gesture fragments depends on the communicative situation. If they provide strongly needed information, e.g. to solve a task, gesture fragment information can also become a strong context.

## **7.2. A model for gesture-speech integration in comprehension**

Despite more than two decades of systematic and increasing research, there is no published model / theory on how gesture and speech might be processed during integration. A first tentative model by Holle (2007) is based on the Dual Coding Theory by Sadoski and Paivio (2001), which states that there are separate conceptual representations for verbal and non-verbal input. In his model, Holle (2007) assumes that the incoming gestural and speech information is separately analyzed on a pure perceptual level (i.e. the so-called Form Level), before it is compared and identified via the already stored verbal and non-verbal representations (Conceptual Level)<sup>18</sup>. Only on the Conceptual Level, gesture and speech interact to form a combined meaningful concept. Holle (2007) already noted, that there might be some modulating component that could have a significant impact on the whole integration process. So far, however, little was known about factors that can affect this process. This dissertation shed some light on how task, timing and background noise influence the integration of gesture in speech. On the basis of these new findings it is therefore possible to restructure and specify the model sketch by Holle (2007) in such way that the new model also accounts for the modulating effects of, for example, the temporal synchrony between gesture and speech. In the following, I will first introduce some general aspects of the model, like the scope and architecture, before going more into detail about the single processing steps. Finally, I will relate the model to various findings in gesture research and try to make some claims about the neuro-anatomical basis of the model.

---

<sup>18</sup> On the side of the verbal system there is another intermediate step between those two levels, the so-called Lemma Level, as it is assumed that there is no direct link between phonetic information and word meaning. Yet, the phonetic information can be used to access the lemma of a certain phoneme sequence, which contains information such as word category and gender (e.g. Levelt, 1992; Jescheniak, 2002), which in turn can be used to identify the conceptual meaning.

### 7.2.1. The *Feature Integration Model* for gesture-speech comprehension

#### 7.2.1.1. The scope

The model aims to explain how iconic gestures are integrated with their co-expressive speech unit in sentence comprehension. It is primarily constructed to account for the local integration of synchronously (i.e. within the temporal window of integration) perceived gesture and speech information. Nevertheless, the model is also able to explain the integration of asynchronous gestures with speech input as well as the global effects of the integration on the sentence level.

Similar to the model by Holle (2007), this new model is based on an information processing approach. The basic assumption of this approach is that the brain does its job by processing information (see de Ruiter, 1998). I.e. the brain sequentially operates upon stored information (so-called representations) to generate the desired output (e.g. integrate gesture and speech information). However, in contrast to the model by Holle (2007), the *Feature Integration Model (FIM)* also tries to account for the influence of contextual factors (e.g. task).

Furthermore, the *FIM* makes some additional assumptions regarding the integration process. Based on the literature on crossmodal audiovisual integration (e.g. Werner & Noppeney, 2010; Schroeder & Foxe, 2005), it is assumed that the interaction of gesture and speech already starts at the basic perceptual level. Note that Holle (2007) assumes that there is only interaction on a high conceptual (lexical) level.

In the next section, the basic structure of the model as well as the processing steps involved in the integration of gesture and speech will be introduced.

#### 7.2.1.2. The architecture of the *Feature Integration Model (FIM)*

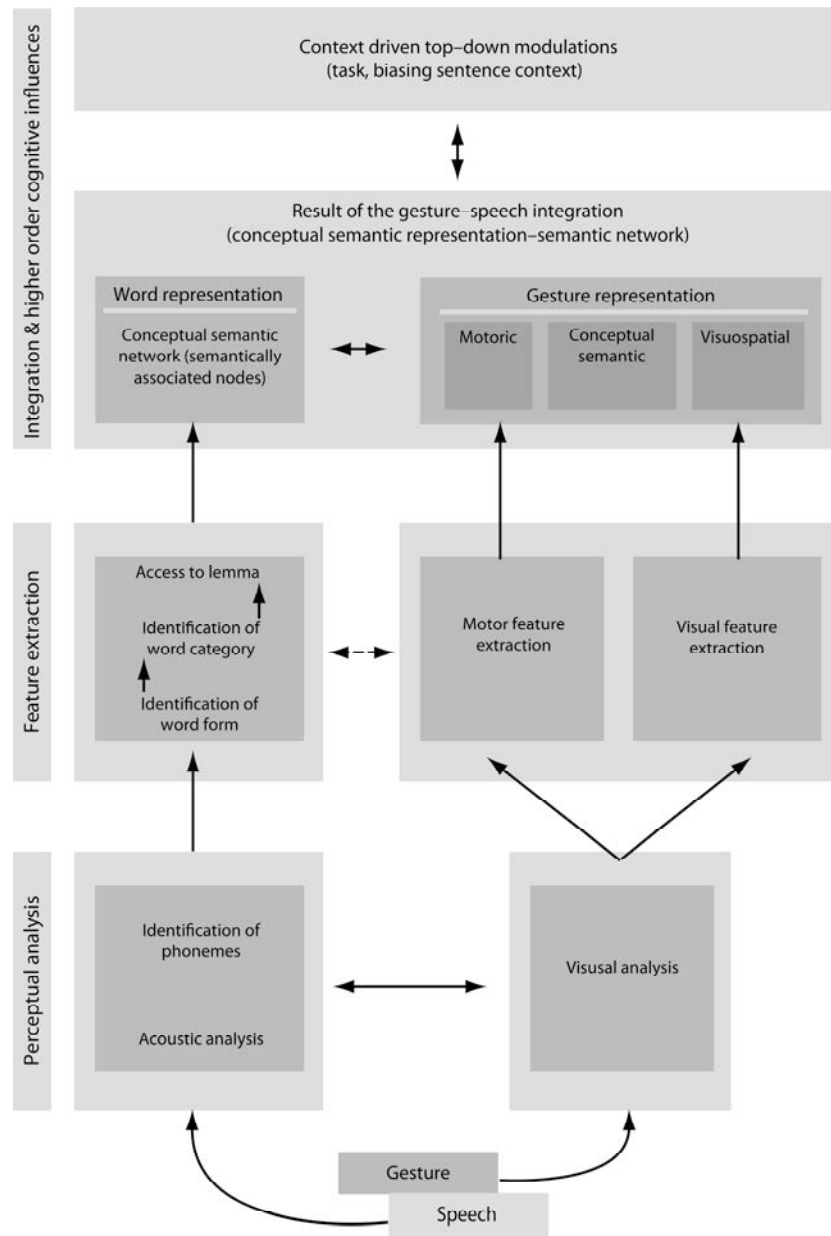
Prior to proposing any model on gesture-speech integration, there are a few principal decisions that have to be made. First, it is important to determine whether there is one single, amodal, abstract store that contains both verbal and nonverbal information (e.g. as so-called propositions, Anderson, 1981; in gesture research: Wagner, Nusbaum & Goldin-Meadow, 2004) or whether verbal and non-verbal information are stored separately. There is evidence for a double-dissociation of verbal and nonverbal material for the recognition of verbal and nonverbal stimuli in patients with left vs. right-hemispheric lesions (e.g. Coltheart, 1980;

Fujii et al., 1990; Seliger et al., 1991), which has served as the basis for the argument that there are multiple stores (Coltheart et al., 1998; Sadoski & Paivio, 2001; Warrington & Crutch, 2004). Further support for this assumption comes from a fMRI study with healthy participants, where a double dissociation between verbal and nonverbal material at a conceptual level was found (Thierry & Price, 2006). Similar to the model by Holle (2007) the *FIM* adopts the view proposed in the Dual Coding Theory by Sadoski and Paivio (2001), that there is a verbal system, specialized for speech, and a nonverbal system, processing the gestural information. Within the verbal system, processes like auditory analysis, identification of phonemes, word form, word category and lemma, etc. take place as proposed by many models of language comprehension, e.g. the *Neurocognitive Model of Auditory Sentence Processing* by Friederici (2002). Second, it is necessary to define the processing steps for both verbal as well as nonverbal input. The *FIM* proposes three levels of processing for both types of information, i.e. a perceptual analysis level, a feature extraction level and the integration / meaning generation level (see Figure 7.1). Importantly, the *FIM* assumes that verbal and nonverbal representation can interact and influence each other on each of the three levels. I.e. there is a mutual influence of gesture and speech as it has been proposed in the Integrated Systems Hypothesis (Kelly, Özyürek, et al., 2010) on all levels. In the following the processing steps and the interaction between verbal and gesture information will be described for each of the three levels.

#### 7.2.1.3. The perceptual analysis

##### *Auditory and visual analysis*

As a first step in the integration process of gesture with speech the perceptual input has to be analyzed. The visual input is analyzed with regard to physical, visual properties. More or less in parallel, the auditory input is subjected to an acoustic / phonetic analysis that results in the phonetic representation of a word, i.e. the phoneme sequence of that word. The quality of the visual and acoustic signal may influence their analysis at the perceptual stage. For example, when encountering a speech in a noisy environment, the reduced quality in the auditory stream renders it more difficult to identify the phoneme sequence of a word. Note, if a gesture onset considerably precedes the corresponding word, the processing of the gesture will be ahead of the processing of the speech input throughout all levels of the *FIM*. In this case, a gesture can become some kind of context for the whole speech processing.



**Figure 7.1** The architecture of the Feature Integration Model.

*The interaction of the auditory and visual stream at the perceptual level*

Numerous studies in monkeys and humans have shown that multisensory integration already takes place at low, unisensory levels (e.g. Fu et al., 2003; Molholm et al., 2002). For example, Werner and Noppeney (2010) have shown that the simple presence of a visual stimulus affects the processing of auditory input in the primary auditory areas of the middle superior temporal gyrus. Another, and in the light of the results of this dissertation, even more important observation is that perception of audiovisual asynchronies leads to changes in the brain activity in primary auditory and visual brain areas, which are both directly and indirectly connected (Noppeney, 2010).

Based on these findings, it is assumed that gesture and speech also already interact on the perceptual level. This interaction can occur in various ways. First, the simple presence of a gesture may impact the auditory processing of a word. Noppeney and colleagues (2008) suggest that prior visual information (remember that gestures almost all the time start prior to the co-expressive speech unit) influences the phonology of a word. Therefore the presence of a co-speech gesture should alter the processing of a spoken word in contrast to when there is no gesture. Krahmer & Swerts (2007) found that participants perceived words more prominently when they were accompanied by speech in comparison to when they were not, suggesting that there might indeed be an early perceptual influence of gesture on the processing of phonology in the auditory stream. Second, the presence of an additional visual input could lead to an enhancement of the auditory processing. This is especially important, for example, in a noisy environment, where vocalizations alone might not suffice for an unambiguous identification of the auditory input. Previous research (Sumbly & Pollack, 1954; Ross et al., 2007) has shown that multisensory enhancement is larger if the vocalization is very hard to understand. There are two theoretical concepts how auditory processing can benefit from additional visual input like gestures. They may add additional spatial precision (cf. Rauschecker, Guard, Phan, & Su, 2000), thereby adding new information (features) not present in speech (e.g. in the case of pointings) or gesture may lead to a re-set in auditory processing (cf. Schroeder and Foxe, 2005). I.e. previous auditory information is attenuated whereas the response to subsequent auditory input is enhanced. Both concepts can explain the phonological effect found by Krahmer and Swerts (2007). Third, asynchrony detection between gesture and speech should already take place at the perceptual level (in accordance to findings by Noppeney, 2010).

If gesture and speech are accessed within the time window of integration, the next levels of the *FIM* (feature extraction and semantic integration) follow subsequently. If gesture and speech are not accessible within the time window of integration, gesture and speech information are processed separately and can still be merged under certain circumstances (e.g. explicit task) at a higher level, the semantic integration level.

#### 7.2.1.4. Feature extraction

##### *Visual*

After the basic visual analysis, the gestural input is further processed in two ways. First, as for every type of observed movement, motor-driven (or action-based) features are extracted from the gesture. Whether this process is based on simulation, part of the matching of a movement with a previously stored template (Iacoboni, 1999) or part of a common coding process (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 2005) is not of relevance for the model. Second, gestural features like handshape, trajectory, etc., which are characteristic for a gesture, are extracted.

##### *Auditory*

The phonetic representations obtained through the perceptual analysis are used to identify the word form, word category and eventually the lemma of a word (for more details about the lemma, e.g. see Levelt, 1992; Levelt, Roelofs, & Meyer, 1999). The lemma represents an important link between the perceptual, phonetic input on the one hand and the semantic / conceptual content of a word on the other hand. In language, the phonetic structure per se is arbitrary, i.e. it does not give rise to the meaning of a word. The meaning can only be precisely assigned with the help of the lemma of a certain word. Each lemma is a package of syntactic word information, which is not specified for phonological form. It is determined by so-called diacritic parameters (or features). For verbs, for instance, the diacritic parameters include number, person, tense, and mood (for nouns, diacritic parameters include, for example, gender and number). It is obligatory, that the diacritic parameters are valued. Each lemma is also related to an unique lexical / semantic concept. I.e. using a simple statistical mapping of lemma and lexical concept allows a very precise allocation of meaning to a

lemma and thus the original phonetic input. Thus, meaning can be precisely assigned to a certain phonological input.

*The interaction of the auditory and visual stream at the feature extraction level*

So far, nothing is known about any interactions of gesture and speech on this intermediate level. Yet, there is a study which investigated the relation between word category and the neuronal activity in motoric as well as visual brain regions (e.g. Pulvermüller, Preissl, Lutzenberger and Birbaumer, 2005). Participants had to read nouns and verbs while EEG was recorded. Verbs elicited stronger 30 Hz activity over motor cortices, whereas nouns elicited stronger activities over visual cortices. The authors assume that their results reflect the conscious processing of motor and visual associations of verbal material and conclude that nouns and verbs have distinct neuronal generators. In a recently published study, Rueschemeyer, Lindemann, van Rooij, Van Dam, & Bekkering (2010) were able to show that the execution of an action selectively affects the semantic processing of words. The authors assumed that actions and action-related verbs activate identical areas in the neural motor cortex. Based on such findings one could give word category an active role in the feature extraction from gesture. For example, a verb might ease or even enhance the extraction of motor features from a gesture while a noun could enhance the extraction of visuospatial features from the gestural information. Note that in either case the other type of features is also extracted from the gesture, but is less important on the level of conceptual semantic integration. In turn, an action related gesture might influence the identification of the word category in such way that it is more likely to be identified as a verb instead of a noun. In contrast, a more form-based gesture might rather trigger a noun than a verb. Think about the German word *fang* / *Fang* (catch / catch), which can either be a verb or a noun. Seeing a catching gesture might be helpful to resolve the ambiguity with regard to the word category information. Thus, a gesture might be already helpful in the process of lemma identification / selection for the accompanying word.



## 7.2.1.5. The conceptual semantic integration of gesture and speech

*Visual*

The extracted motor and visual information is compared with already stored motor and visual representations. The idea that there are separate memories for different types of knowledge has received some support by various studies who found distinct brain regions for different types of conceptual knowledge (e.g. functional versus perceptual knowledge, see Cappa, Perani, Schnur, Tettamanti, Fazio, 1998). Identical or very similar stored motor representations are activated. Importantly, these stored motor representations can already (but do not necessarily have to) be associated with a conceptual semantic representation (e.g. the representation of a throwing arm movement is directly connected with the concept of “*throwing*” and indirectly with the concept of “*ball*”). The extracted visual features are mapped onto already stored visuospatial representations, which can also be associated with conceptual semantic representations (e.g. a gestured spherical form is directly associated with concepts “*ball*” or “*round*” and indirectly with the concept “*throwing*”).

*Auditory*

As already stated above, the feature extraction of the auditory input results in the unique lemma of a word. Analogous to Levelt’s theory on speech production (e.g. Levelt, Roelofs, & Meyer, 1999), which states that there is a unique lemma for every lexical concept, it is assumed within the framework of the *FIM* that there is a specific lexical concept which is related to each lemma extracted during speech comprehension. On the basis of a statistical mechanism, the conceptual representation best fitting to the lemma is activated. This related lexical concept is represented by a node within the whole conceptual network. Once a certain conceptual node is activated by the lemma input, the activation also spreads from the initial node to semantically related nodes. Each connection is characterized by a specific semantic association between the nodes. For example, a word like “*ball*” is connected to “*dance*” (in this case the connection would be: *a ball is a dance*), which is both a synonym and a specification, but it is also connected to “*spheric*” (*a ball is spheric*), which describes a feature of “*ball*”. The connections between the nodes are weighted based on the frequency of co-occurrence. E.g. a word like *ball* is more often used in the context of *game* than *dance* and thus has a stronger connection with *ball* than *dance*. Thus, on a conceptual level the auditory

input, e.g. a word, is defined by a unique network of activated conceptual nodes with a specific level of activation for each node as well as specific connection weights between the nodes.

#### *The semantic integration of gesture and speech*

The basic assumption for the complete semantic integration process of gesture and speech is that they are connected in a one-to-many fashion, as it has been generally proposed for audiovisual integration by Sadoski and Paivio (2001). I.e. one verbal representation is connected with many non-verbal representations and vice-versa. In the following, I will describe how integration of gesture and speech might work according to the *FIM*. This process has some similarities to the feature-based account of semantic memory by Warrington and Shallice (1984). The authors assume that conceptual knowledge is represented by large networks representing semantic features of all kind (visual, auditory, action, etc., Allport, 1985).

The basic idea for the integration process of the *FIM* is that information from the semantic-conceptual, motor and visuospatial storages interacts and is connected via a network of nodes and links. The nodes in the motor and visuospatial subsystems are activated by the gestural information and associated with nodes in the conceptual semantic storage, which can either be activated by the gestural or verbal input. The interaction of gesture and speech information works in a one-to-many fashion and can occur in two distinct ways. First, gesturally induced input, irrespective of whether it is motor, visuospatial or semantic, can directly (semantic content) or indirectly (motoric or visuospatial) activate conceptual nodes in the conceptual semantic storage. If the nodes are already activated by the verbal input, this process may result in a strengthening of identical nodes and links. Otherwise, new nodes and links are added to conceptual semantic network elicited by the speech. For example, a gesture can depict the speed of a movement, which often is not encoded in speech. More generally speaking, gestures often specify certain aspects of a word that cannot be found in speech and this is done by adding new nodes and connections to the conceptual representation of the word. Thus, the combination of gesture and speech leads to an enriched representation. The intrusion effects found by Cassell et al. (1999) and the word meaning frequency effects at the homonyms in this dissertation provide evidence for this assumption.

The whole integration process is timing dependent, i.e. if gesture and speech are not accessed within the temporal window of integration the merging process does not take place unless gesture aids the principal speech understanding (as in a noisy environment) or an addressee is required to use gesture and speech (e.g. by task instruction, see next section *The role of communicative context*).

#### *The role of communicative context*

The data of this dissertation show that addressees are able to integrate gesture and speech even when they are presented outside the time window of integration (Experiment 1) or not locally integrated with the homonym (Experiment 4: -600 ms condition). The model accounts for this results by introducing *communicative context* as an additional level on top of the local integration. Communicative context can directly influence the identification of a word or gesture as well as the semantic integration of gesture and speech via top-down modulation. The context can be influenced by task or experience but is also shaped by prior speech (sentence) or gesture context. In contrast to the mentioned top-down influences, noisy environment (Experiment 5), amount of meaningful gesture information (Holle and Gunter, 2007) or gesturer-speaker mismatches (Kelly et al., 2007) are not considered to be top-down influences as they already affect the early perceptual processing stage. For example, explicit task instructions like in Experiment 1 (“compare the hand movement with the sentence you hear”) affect the storage process of the visual input by increasing the duration of working memory storage time e.g. through rehearsal. Other factors, like common ground, previous communicational experience, etc. might also elicit similar influences, but more research is needed to get a better an idea about potential top-down modulators.

Additionally, context plays another important role in cases where the local integration cannot be completed or the result is ambiguous<sup>19</sup>. For example, the -600 ms condition of Experiment 4 does not show an effect of local integration at the homonym, but still an effect at the subordinate target word. While the result at the homonym is easily explained by the temporal window of integration, the result at the target word needs some explanation. The gesture fragments in the -600 ms condition share overlap with speech, i.e. the verb of the main clause and thus the perceptual requirements for integration are met. Thus, perceptual integration is

---

<sup>19</sup> Note that overt mismatches do not fall into this category, as they are clearly not fitting the speech.

initialized. On the level of semantic integration, the additional gesture information cannot be integrated respectively it can be integrated but does not provide useful information to the verb. As the gesture remains as ambiguous as before and does not provide useful information. With the additional incoming biasing sentence information up to the target word, the message of the sentence becomes clearer. As a result, the ambiguous gesture can be unambiguously interpreted. A second attempt to integrate the gesture with speech is made, this time not on the local word level, but on the global sentence level. However, such a process can only be initialized when prior local integration has taken place. For instance, this is not the case in Experiment 2 where there is no overlap with speech and thus no local integration is possible.

#### 7.2.1.6. How does the model account for the data of the dissertation

Having already described how the model can account for some findings of this dissertation, I will now describe how it can explain the rest of the results.

When gesture and speech are perceived within the temporal window of integration (Experiments 3 & 4), the integration process is more or less automatically initialized on the perceptual level and continued until the local integration of gesture and homonym is finished. The resulting new extended conceptual semantic representation is then integrated into the sentence context, which leads to the global disambiguating effect at the target word. In principal the same mechanism also holds true for Experiment 5, where the continuous babble stream (remember that the gesture originally falls into the pause between two sentences) triggers the perceptual integration of gesture and speech. As already stated above, an increased multisensory enhancement is likely to take place, as vocalizations alone do not suffice for an unambiguous identification of the auditory input. This can be done via a re-set in auditory processing (cf. Schroeder and Foxe, 2005). The increased attention on the auditory input in combination with the already integrated gesture information results in the effect of local integration of gesture and homonym as well as the global effect at the target word.

The results of Experiment 1 are clearly top-down driven as the task forced participants to maintain the gestural information in memory as well as to compare it with every word of the sentence in order to perform the congruency judgment. As the gesture information did not overlap with speech, there should be no integration at the perceptual level, but gesture and

speech are more or less independently processed up to the semantic integration level, where the task exerts its influence and semantic integration takes place.

The *FIM* can account for all existing EEG data, e.g. the mismatch experiments by Kelly et al. (2004) or Özyürek et al. (2007), but also for the behavioral data on gesture comprehension, e.g. Casell et al. (1999) or Beattie & Shovelton (1999a).

#### 7.2.1.7. The model and other gesture types

Although the model is especially constructed to account for the local integration of iconic gestures with speech, it is also able to account for the effects of all other types of co-speech gestures (beats, deictics, metaphors), emblems and pantomimes. For example, beats, which do not contain any semantic meaning, are only integrated on the perceptual level, which would allow them to highlight speech or influence the prosody of speech at best, which is in line with findings by Krahmer and Swerts (2007). In contrast, emblems, for instance, which are meaningful, lexicalized gestures, can be processed purely visually according to the *FIM* until the semantic level, where the visual information is then transformed into the verbal representation. Note that in the case of emblems the verbal processing stream can be completely omitted, which is supported by the architecture of the *FIM*.

### 7.3. Open questions

As can be seen from the summary of the results of this dissertation, the findings on the significance of task, timing and background noise on gesture-speech integration provide new, interesting insight into the processes underlying the interaction of both streams of information. The findings also led to a new model describing the whole integration process, the *FIM*. Yet, both the results of the experiments as well as the architecture of the model generate more, new questions. Some of these new questions have been already addressed in the section about the *FIM*. There are, however, many more questions that are of interest on the way to. In the following, I will present some open issues which are of relevance for the gesture-speech comprehension research in general as well as target certain aspects of the *FIM*.

### **7.3.1. The relationship between local gesture-speech integration and the global message level**

The results of Experiment 4 indicate that the integration of gesture and homonym and the effect at the target word are different with regard to their timing. Whereas there is only an effect of direct integration in a time window between -200 ms and +120 ms, the effect at the target word is already present at an asynchrony of -600 ms. Is the later, global message effect at the target therefore independent from the local integration of gesture fragment and homonym? – It is possible, because we did not use epochs that covered the whole sentence range thus might have missed some very slow potentials. However, there is also an alternative and more plausible explanation (as already explained in the section *The role of communicative context*). The -600 ms condition does not share any temporal overlap with the homonym, but more or less is presented parallel to the verb. It is therefore possible, and in line with *FIM*, that the gesture fragment is perceptually integrated with the verb, which however does not lead to a disambiguation of gesture and homonym. Yet, on a more global message level addressees can still access the integrated information to identify the contextually appropriate target word. Note again that this is not possible in the -1000 ms condition, as these gestures are presented in the pauses between the introductory and the main clause. Thus, there is no possibility to locally integrate gesture and speech and consequentially use it on a message level. Clearly, more research is needed to substantiate this idea. For example, one could use stimuli where local integration is possible at two different positions (e.g. verb and object or subject and object) with a delayed target later on in the sentence (to measure the global effect). Additionally, the material could also be constructed in such way that is also possible to test the role of sentence boundaries in gesture-speech integration. I.e. present a gesture in the intro sentence and test whether it has an effect in the following sentence. Such an experiment would help to clarify the role of local gesture-speech integration for the global integration of gesture information into a larger context.

### **7.3.2. The integration of asynchronous gestures with speech – the role of working memory and attention**

Experiments 1 and 5 provide evidence that addressees can integrate asynchronous, non-overlapping gesture information with speech. This is of communicational relevance as speakers sometime produce non-overlapping gesture and speech that refers to the same

semantic idea unit. In order to do so, gesture information, which only gets its specific dominant or subordinate meaning in the presence of the homonym has to be stored in working memory until the meaning of the homonym is accessed. This obligatory use of working memory resources is part of a top-down moderated change in the default processing of gesture and speech. An unsolved question relates to the kind of working memory system used to store gestural information. A pretest of the gesture fragment has shown that this information is rather vague ruling out a purely semantic / verbal memory buffer. Gesture research so far has revealed mixed results with regard to the representational format of gestures. Wesp et al. (2001) assume that the storage is visuospatial. Feyereisen and Havard (1999) suspect that gesture may also trigger action-related memory processes, whereas Wagner et al. (2004) suggest that the memory representation is not visuospatial, but propositional. I.e. the representation combines similar visuospatial and verbal information, which is in contrast to (Morsella & Krauss, 2004) who state that gestures influence both types of representations independently. Thus, it is not clear by now whether the working memory representation of a gesture consists of one or a combination of the mentioned representations. Using a dual-task paradigm, this issue can either be behaviorally or neurophysiologically tested with the setup of Experiment 1 by applying a secondary task, which is either verbal, visuospatial or action-related. One could speculate that the type of representation might depend on the speech context. For example, verbs may trigger more action-related gestures, whereas nouns may trigger more visual-spatial gestures. The gesture-speech combination, in turn, could then determine to which degree a more visuospatial or action related memory representation is generated. Thus, the impact of the secondary task could vary depending on the type of the gesture-speech combination. There are at least some indications from Feyereisen et al. (1999) that would justify such a hypothesis. To test this assumption, the factor gesture type could be additionally varied. Such an experiment would also serve as a test for the assumption that a single gesture contains several different, functionally separated types of information (motor, visual, semantic).

Besides the working memory, attention could also influence the integration of asynchronous gesture and speech information. For example, the explicit task as well as the addition of background noise could also have triggered an increased attention towards the gesture fragments, which could also have contributed to the findings of Experiments 1 and 5. Participants in Experiment 5, however, reported that they focused more on speech, because it was so hard to understand. Thus, it seems unlikely that they turned their attention more

towards the gesture. To test whether there is an attentional modulation of gesture processing, the ERPs triggered to the gesture onsets in the silence and babble noise condition could be compared with regard to differences in early ERP components (N1, P2).

### **7.3.3. What modulates the integration process? –The role of the communicative context**

Gesture-speech integration is not an automatic process as the data of this dissertation as well as previous research (Holle & Gunter, 2007; Kelly et al., 2007; Wu & Coulson, 2010) have clearly shown. Instead, gesture information uptake seems to be more or less obligatory depending on various factors like timing, visual and auditory signal quality. During the integration process, all this information is analyzed very fast. Based on the outcome, the default processing mode of gesture and speech is adapted in an incredibly flexible and highly adaptive way, taking into account and comparing all the available information from the perceptual up to the integration level. Although this dissertation provides a step forward to a comprehensive model of gesture-speech integration, some of the most important aspects of everyday communication that could modulate this process have not been studied so far. In particular, the impact of context and context strength has not been studied with neuropsychological methods. This is of high importance for the *FIM*, as communicative context can serve as a top-down modulator for the whole integration process of gesture and speech. Kelly et al. (2010) have shown that differences in context strength affect the processing of gesture and speech using a behavioral priming paradigm. Maybe, it is also possible to find such graded differences in the online integration of gesture and speech by means of ERPs. One step further would be not to manipulate the currently present context in the stimulus material, but to manipulate the previously experienced context, for instance, by manipulating common ground that is related to the later presented stimuli (cf. Holler & Stevens, 2007; Holler & Wilkin, 2009). Another, very interesting context manipulation would be to familiarize participants with different gesturing styles of speakers prior to the ERP measurement (e.g. from only informative gestures up to a minimal amount of informative gestures). This manipulation would allow to investigate whether the previous experience with a speaker affects the gesture-speech integration when encountering a new communicative situation with the respective speaker. The advantage of such an experiment would be that only meaningful gesture could be compared with regard to their impact on comprehension. I.e. it would be very naturalistic.



#### 7.3.4. The neural correlates of gesture-speech integration

The data of this dissertation as well as the architecture of the *FIM* point to a complex interplay of top-down and bottom-up processes in the construction of a joint message unit out of gestural and speech information.

One important question that cannot be answered by the ERP method is which brain regions are responsible for the various processes involved gesture-speech integration. For instance, where does the identification and compensation of temporal asynchronies between gesture and speech or the top-down modulation by task take place? These questions can be addressed by means of fMRI. For example, one could present the gesture fragments at various temporal alignments to the homonym based on the results of the ERP data (e.g. -1000 ms - no integration, - 200 ms - little integration, 0 ms - clear integration)<sup>20</sup>. If one is not only interested in the effects of timing on gesture-speech integration but also wants to investigate the interaction of top-down factors with timing in meaning construction, one could additionally add task (congruency judgment vs. monitoring) as an independent variable similar to a fMRI experiment by van Atteveldt, Formisano, Goebel et al. (2007), in which they looked at the impact of timing and task on letters and sounds. Though the second design allows gaining more insight into gesture-speech integration, it is much less feasible to test, because of its complexity. Both experiments, however, may be instrumental in clarifying the role of these brain areas with regard to gesture-speech integration, which has been a matter of debate ever since the publication of the first fMRI studies on gesture comprehension. Various fMRI studies (e.g. Dick et al., 2009; Green et al., 2009; Holle et al., 2008; Holle et al., 2010; Willems et al., 2007, 2009) have tried to identify brain regions involved in gestures speech integration. However, this issue has not been solved in a satisfying manner despite the use different kinds of materials & tasks. In general there are two distinct views of where in the brain both sources of information are integrated. On the one hand, there are some studies which claim that the left Inferior Frontal Gyrus (IFG), in particular BA45/47, is the putative region of gesture-speech integration (Willems et al., 2007, 2009). On the other hand, there is also a similar number of studies which state that bilateral Superior Temporal Sulci / Gyri (STSs/Gs) function as multimodal integration sites for gesture and speech (Dick et al., 2009; Holle et al., 2008; Holle et al., 2010). Keeping this in mind, the present study may also be

---

<sup>20</sup> Importantly, both gesture fragments and speech also have to be presented unimodally in order to identify regions of crossmodal integration of gesture and speech based on the concept of superadditivity (e.g. Beauchamp, Argall, et al., 2004). I.e. regions of integration should elicit larger brain activities than the sum of the brain responses elicited by the unimodal presentations.

instrumental in clarifying the role of these brain areas with regard to gesture-speech integration. The above mentioned brain regions (IFG, STSs/Gs) will be the predominantly analyzed based on the literature on gesture-speech integration. Only recently, Dick et al. (2009) proposed that the right IFG is the area of semantic integration and that the left IFG and the bilateral STS/G regions are only there for perceptual processing. Thus, fMRI studies that vary the perceptual input (e.g. by a timing variation) or even vary perceptual input and top-down processes involved in gesture-speech integration (e.g. by task) may provide useful new evidence about the functional role of inferior frontal and temporo-parietal regions found in previous fMRI studies. Specifically, the STS/G should show variations depending on the timing between gesture and speech (cf. Macaluso, George, Dolan, Spence, & Driver, 2004; van Atteveldt, Formisano, Goebel, et al., 2007). Based on Noppeney (2010), one could also speculate that the middle region of the STS/G should be especially sensitive to variations of timing, as it represents more perceptual audiovisual processing (which corresponds to the perceptual level of the *FIM*). The posterior part of the STS/G, however, should be rather insensitive to the synchrony manipulation as it already presents higher order audiovisual integration (as it takes place in the semantic integration level of the *FIM*). There is also some data, that the insula might be involved in the timing analysis of multisensory input (Bushara, Grafman, & Hallett, 2001). In contrast to the temporo-parietal regions, the IFG may be more involved in the construction of the message meaning and may in combination with the inferior frontal sulcus (IFS, see van Atteveldt, Formisano, Goebel, et al., 2007) be sensitive to the task manipulation.

### 7.3.5. The impact of gestures on speech comprehension in hearing impaired

Experiment 5 clearly demonstrated that a less optimal communicative situation leads to an increased information uptake from gesture. Although such situations clearly exist in real life (those of you who regularly visit crowded areas like train stations, concerts, or bars know what is meant) it seems relevant to explore the impact of gestures on persons who experience such suboptimal situations in their daily life, e.g. elderly people or hearing impaired. It is known that hearing impairment, similar to a noisy background in normal hearing, leads to the increased use of visual cues in speech comprehension, i.e. the so-called lip-reading (e.g. Sumby & Pollack, 1954). If hearing impaired individuals compensate their hearing loss by incorporating as much visual cues as possible, one would expect to find effects of gesture-

speech even in a noise-free environment (cf. Grant, Tufts, & Greenberg, 2007). These effects should be comparable to those found for the normal hearing participants in the noise condition of Experiment 5. Using the identical setup as in Experiment 5, the impact of gesture fragments on homonym disambiguation could be tested with a group of hearing impaired adults as well as age-matched controls. For the hearing impaired, the ERP data should show effects both for the local integration of gesture fragment and homonym as well as the result of this disambiguating process at the target word. No such effects should be found in the control group, thereby replicating the results of Experiments 1 and the silence condition of Experiment 5. First results seem to confirm these hypotheses, indicating that hearing impaired rather obligatorily use gestural information to compensate their hearing deficits.

#### **7.4. Concluding remarks**

In sum, the present dissertation provided a first insight into the complexity and flexibility of gesture-speech integration by showing the significance of task, timing and background noise for this process. In particular, the temporal alignment between gesture and the corresponding proofed to be crucial for a rather automatic integration of both streams information. Based on the present results, a time window ranging from -200 ms (auditory input lags visual input) to +120 ms (visual input lags auditory input) was determined for the local, immediate integration of gesture and corresponding speech unit. The present work also provides first evidence that integration of gesture and speech is also beyond the temporal window of integration. This process, however, is no longer “automatic” but driven by additional executive, top-down (e.g. task) influences as well as situational, bottom-up influences (e.g. noisy environment).

The results of this dissertation represent a first step to a more precise determination of the interaction of gesture and speech and thus, to a first, neuroscientific model of gesture-speech integration.

## References

- Allport, D. A. (1985). Distributed memory, modular subsystems, and dysphasia. In S. K. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 32-60). Edinburgh: Churchill Livingstone.
- Anderson, J. R. (1981). Concepts, prepositions, and schemata: What are the cognitive units? In J. Flowers (Ed.), *Nebraska Symposium on Motivation*. Lincoln, Nebraska: University of Nebraska Press.
- Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58(2), 495-520. doi: DOI 10.1016/j.jml.2007.02.004
- Beattie, G., & Coughlan, J. (1998). Do iconic gestures have a functional role in lexical access? An experimental study of the effects of repeating a verbal message on gesture production. *Semiotica*, 119(3-4), 221-249.
- 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 and Social Psychology*, 18(4), 438-462.
- 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, 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, 403-417.
- Beattie, G., & Shovelton, H. (2005). Why the spontaneous images created by the hands during talk can help make TV advertisements more effective. *British Journal of Psychology*, 96, 21-37.
- Beattie, G., & Shovelton, H. (2006). When size really matters. *Gesture*, 6(1), 63-84.
- Beattie, G., Webster, K., & Ross, J. (2010). The Fixation and Processing of the Iconic Gestures That Accompany Talk. *Journal of Language and Social Psychology*, 29(2), 194-213. doi: Doi 10.1177/0261927x09359589

- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, 7(11), 1190-1192. doi: Doi 10.1038/Nn1333
- 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.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-Related Potentials, Lexical Decision and Semantic Priming. *Electroencephalography and Clinical Neurophysiology*, 60(4), 343-355.
- Berger, H. (1929). Über das Elektroenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87, 527-570.
- Birbaumer, N., & Schmidt, R. F. (1990). *Biologische Psychologie*. Berlin: Springer-Verl.
- Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, 8(6), 631-642.
- Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.5.01). Amsterdam: Institute of Phonetic Sciences. Retrieved from <http://www.praat.org/>
- Bornkessel, I. (2002). *The argument dependency model a neurocognitive approach to incremental interpretation*. Leipzig: Max-Planck-Institute of Cognitive Neuroscience.
- Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience*, 21(1), 300-304.
- Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell & P. Smith (Eds.), *Recent Advances in the Psychology of Language* (pp. 347-360). New York: Plenum Press.
- Butterworth, B., & Hadar, U. (1989). Gesture, Speech, and Computational Stages - a Reply. *Psychological Review*, 96(1), 168-174.
- Cappa, S. F., Perani, D., Schnur, T., Tettamanti, M., & Fazio, F. (1998). The effects of semantic category and knowledge type on lexical-semantic access: a PET study. *Neuroimage*, 8, 350-359.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris*, 98(1-3), 191-205. doi: S0928-4257(04)00080-4 [pii] 10.1016/j.jphysparis.2004.03.018
- 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), 1-34.

- Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, 37(6), 871-887. doi: DOI 10.1016/j.pragma.2004.10.016
- 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.
- Coles, M. G. H., & Rugg, M. D. (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* (pp. 1-23). Oxford: Oxford University Press.
- 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.
- Condillac, E. V. de (1756). *An Essay on the Origin of Human Knowledge* (T. Nugent, Trans.). (Facsimile Reproduction by R. G. Weyant, Ed., 1971, New York: Scholar's Facsimiles and Reprints).
- Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The Effects of Processing Requirements on Neurophysiological Responses to Spoken Sentences. *Brain and Language*, 39(2), 302-318.
- Cook, S. W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, 7(2), 211-232.
- Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, 113(1), 98-104. doi: DOI 10.1016/j.cognition.2009.06.006
- Cotton, S., & Grosjean, F. (1984). The Gating Paradigm - a Comparison of Successive and Individual Presentation Formats. *Perception & Psychophysics*, 35(1), 41-48.
- Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57(4), 483-501. doi: DOI 10.1016/j.jml.2007.01.001
- De Jorio, A. (2000). *Gesture in Naples and Gesture in Classical Antiquity* (A. Kendon, Trans.). Bloomington: Indiana University Press. (Original work published in 1832).
- de Ruiter, J. P. (1998). *Gesture and Speech Production*. Ph.D. thesis. Nijmegen: Max Planck Institute for Psycholinguistics.

- de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.
- Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, 30(11), 3509-3526. doi: 10.1002/hbm.20774
- Diderot, D. (1916). Letter on deaf mutes. In M. Jourdain (Ed. & Trans.). Diderot's Early Philosophical Works (pp. 158-225), Chicago: Open Court Publishing Co. (Original work published in 1751).
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9(6), 719-721.
- Doehring, D. G. (1976). *Acquisition of rapid reading responses*. Washington: Society for Research in Child Development.
- Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERPs. In E. Callaway, P. Teuting & S. Koslow (Eds.), *Event-Related Brain Potentials in Man* (pp. 349-441). New York: Academic Press.
- Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10(20), R731-R735.
- Efron, D. (1941). *Gesture and Environment*. Morningside Heights, NY: King's Crown Press.
- Ekman, P., & Friesen, W. (1969). The repertoire of non-verbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1, 49-98.
- Emmorey, K., & Corina, D. (1990). Lexical Recognition in Sign Language - Effects of Phonetic Structure and Morphology. *Perceptual and Motor Skills*, 71(3), 1227-1252.
- Engle, R. A. (2000). *Toward a Theory of Multimodal Communication: Combining Speech, Gestures, Diagrams, and Demonstratives in Instructional Explanations*. Ph.D. thesis, Stanford University.
- Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, 18(2), 185-205. doi: Doi 10.1080/09541440540000158
- Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, 23(2), 153-171.

- Fischler, I., & Bloom, P. A. (1980). Rapid Processing of the Meaning of Sentences. *Memory & Cognition*, 8(3), 216-225.
- Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *American Journal of Psychology*, 111(1), 43-62.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78-84.
- Friederici, A. D. (2004). Event-related brain potential studies in language. *Curr Neurol Neurosci Rep*, 4(6), 466-470.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Brain Res Cogn Brain Res*, 1(3), 183-192. doi: 0926-6410(93)90026-2 [pii]
- Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *Neuroreport*, 12(15), 3391-3394.
- Fu, K., Johnston, T., Shah, A., Arnold, L., Smiley, J., Hackett, T., et al. (2003). Auditory cortical neurons respond to somatosensory input. *Journal of Neuroscience*, 23, 7510-7515.
- Fujii, T., Fukatsu, R., Watabe, S. I., Ohnuma, A., Teramura, K., Kimura, I., et al. (1990). Auditory sound agnosia without aphasia following aright temporal lobe lesion. *Cortex*, 26(2), 263-268.
- Gall, S., Kerschreiter, R., & Mojzisch, A. (2002). *Handbuch Biopsychologie und Neurowissenschaften ein Wörterbuch mit Fragenkatalog zur Prüfungsvo*. Bern: Huber.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12(5-6), 613-656.
- Goldin-Meadow, S. (2007). Pointing sets the stage for learning language - and creating language. *Child Development*, 78(3), 741-745.
- Grant, K. W., Tufts, J. B., & Greenberg, S. (2007). Integration efficiency for speech perception within and across sensory modalities by normal-hearing and hearing-impaired individuals. *Journal of the Acoustical Society of America*, 121(2), 1164-1176. doi: Doi 10.1121/1.2405859
- Grant, K. W., van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. *Speech Communication*, 44(1-4), 43-53. doi: DOI 10.1016/j.specom.2004.06.004



- Gredebäck, G., Melinder, A., & Daum, M. (2010). The development and neural basis of pointing comprehension. *Soc Neurosci*, 1-10. doi: 919286413 [pii] 10.1080/17470910903523327
- Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Human Brain Mapping*, 30(10), 3309-3324. doi: 10.1002/hbm.20753
- Greenhouse, S. W., & Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika*, 24(2), 95-112.
- Grosjean, F. (1980). Spoken Word Recognition Processes and the Gating Paradigm. *Perception & Psychophysics*, 28(4), 267-283.
- Grosjean, F. (1996). Gating. *Language and Cognitive Processes*, 11(6), 597-604.
- Grosjean, F., & Hirt, C. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. *Language and Cognitive Processes*, 11(1-2), 107-134.
- Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, 14(1), 53-82. doi: <http://dx.doi.org/10.1075/pc.14.1.05gul>
- Gullberg, M., & Kita, S. (2009). Attention to Speech-Accompanying Gestures: Eye Movements and Information Uptake. *Journal of Nonverbal Behavior*, 33(4), 251-277. doi: DOI 10.1007/s10919-009-0073-2
- 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.
- Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2010). The Role of Synchrony and Ambiguity in Speech-Gesture Integration during Comprehension. *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn.2010.21462
- Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, 115(1-2), 147-172.
- Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture - Implications for gesture classification and function. *Journal of Language and Social Psychology*, 23(2), 204-214. doi: Doi 10.1177/0261927x04263825

- Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage*, *15*, S18-S29.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*, 416-423.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441. doi: DOI 10.1126/science.1095455
- Handy, T. C. (2005). *Event-Related Potentials: A methods handbook*. Cambridge, Mass.: MIT Press.
- Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen der Physik und Chemie*(89), 211-233, 354-377.
- Hirata, Y., & Kelly, S. D. (2010). Effects of Lips and Hands on Auditory Learning of Second-Language Speech Sounds. *Journal of Speech Language and Hearing Research*, *53*(2), 298-310. doi: Doi 10.1044/1092-4388(2009/08-0243)
- Holcomb, P. J. (1993). Semantic Priming and Stimulus Degradation - Implications for the Role of the N400 in Language Processing. *Psychophysiology*, *30*(1), 47-61.
- Holcomb, P. J., & Mcpherson, W. B. (1994). Event-Related Brain Potentials Reflect Semantic Priming in an Object Decision Task. *Brain and Cognition*, *24*(2), 259-276.
- Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision - a comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4).
- Holle, H. (2007). *The comprehension of co-speech iconic gestures: behavioral, electrophysical and neuroimaging studies*. Leipzig: Max-Planck-Institute for Human Cognitive and Brain Sciences.
- Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175-1192. doi: 10.1162/jocn.2007.19.7.1175
- Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage*, *39*(4), 2010-2024. doi: S1053-8119(07)00989-5 [pii] 10.1016/j.neuroimage.2007.10.055
- Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage*, *49*(1), 875-884. doi: S1053-8119(09)00970-7 [pii] 10.1016/j.neuroimage.2009.08.058

- Holler, J., & Beattie, G. (2002). A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica*, 142(1-4), 31-69.
- Holler, J., & Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process? *Semiotica*, 146(1-4), 81-116.
- 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.
- Holler, J., Shovelton, H., & Beattie, G. (2009). Do Iconic Hand Gestures Really Contribute to the Communication of Semantic Information in a Face-to-Face Context? *Journal of Nonverbal Behavior*, 33(2), 73-88. doi: DOI 10.1007/s10919-008-0063-9
- Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, 26(1), 4-27. doi: Doi 10.1177/0261927x06296428
- Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*, 24(2), 267-289. doi: Doi 10.1080/01690960802095545  
Pii 793501367
- Hommel, B., Müsseler, J., Aschersleben, & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849-878.
- Hoskin, J., & Herman, R. (2001). The communication, speech and gesture of a group of hearing-impaired children. *International Journal of Language & Communication Disorders*, 36(2), 206-209.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495-514. doi: Doi 10.3758/Pbr.15.3.495
- Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22(3), 313-336. doi: Doi 10.1080/01690960600632812
- Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, 30, 1028-1037.

- Huettel, S. A., Song, A. W., & McCarthy, G. (2008). *Functional magnetic resonance imaging*. Sunderland, Mass.: Sinauer.
- Iacoboni, M. (1999). Cortical Mechanisms of Human Imitation. *Science*, 286 (5449), 2526-2528.
- Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396(6708), 228-228.
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367-371.
- Jansen, E., & Povel, D. J. (2004). The processing of chords in tonal melodic sequences. *Journal of New Music Research*, 33(1), 31-48.
- Jasper, H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*(10), 371-375.
- Jescheniak, J. D. (2002). *Sprachproduktion: Der Zugriff auf das lexikale Gedächtnis beim Sprechen*. Göttingen: Hogrefe.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science*. New York, NY: McGraw-Hill Publ.
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40(4), 577-592.
- Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating Speech and Iconic Gestures in a Stroop-like Task: Evidence for Automatic Processing. *Journal of Cognitive Neuroscience*, 22(4), 683-694.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253-260. doi: Doi 10.1016/S0093-934x(03)00335-3
- Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, 24(2), 313-334. doi: Doi 10.1080/01690960802365567 Pii 905457157
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2), 260-267. doi: 0956797609357327 [pii] 10.1177/0956797609357327
- Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, 101(3), 222-233. doi: DOI 10.1016/j.bandl.2006.07.008

- 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. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge Univ. Press.
- Kidd, E., & Holler, J. (2009). Children's use of gesture to resolve lexical ambiguity. *Developmental Science*, 12(6), 903-913. doi: DOI 10.1111/j.1467-7687.2009.00830.x
- Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., Green, A. (2009). Neural interaction of speech and gesture: Differential activations of metaphorical co-verbal gestures. *Neuropsychologia*, 47, 169-179.
- Kita, S. (1990). *The Temporal Relationship between Gesture and Speech: A Study of Japanese-English Bilinguals*. Unpublished MA Thesis, University of Chicago.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162-185). Cambridge: Cambridge University Press.
- Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, 24(5), 761-775. doi: Doi 10.1080/01690960802327971 Pii 902557407
- 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. doi: Pii S0749-596x(02)00505-3
- Kita, S., & Shao, Z. (2010). *Time course of speech-gesture integration in comprehension: Insights from gating experiments*. Paper presented at the 4th conference of the international society for gesture studies, Frankfurt/Oder.
- Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phase in signs and co-speech gestures, and their transcriptions by human coders. *Lecture Notes In Computer Science*(1371), 23-35.
- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396-414. doi: DOI 10.1016/j.jml.2007.06.005

- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7(2), 54-60.
- Krauss, R. M., Chen, Y. S., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and Gesture* (pp. 261-284). Cambridge: Cambridge University 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 and Social Psychology*, 61(5), 743-754.
- Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences - Brain Potentials Reflect Semantic Incongruity. *Science*, 207(4427), 203-205.
- Kutas, M., & Hillyard, S. A. (1984). Brain Potentials during Reading Reflect Word Expectancy and Semantic Association. *Nature*, 307(5947), 161-163.
- Lalanne, C., & Lorenceau, J. (2004). Crossmodal integration for perception and action. *Journal of Physiology-Paris*, 98(1-3), 265-279. doi: DOI 10.1016/j.jphysparis.2004.06.001
- 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. doi: 10.1007/s00221-005-2370-2
- Lausberg, H., & Kita, S. (2003). The content of the message influences the hand choice in co-speech gestures and in gesturing without speaking. *Brain and Language*, 86(1), 57-69. doi: Doi 10.1016/S0093-934x(02)00534-5
- Lausberg, H., & Sloetjes, H. (2008). Gesture coding with the NGCS - ELAN system. In A. J. Spink, M. R. Ballintijn, N. D. Rogers, F. Grieco, L. W. S. Loijens, L. P. J. J. Noldus, G. Smit & P. H. Zimmermann (Eds.), *Proceedings of Measuring Behavior 2008, 6th International Conference on Methods and Techniques in Behavioral Research* (pp. 176-177). Maastricht: Noldus.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1-22.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.

- Levelt, W. J. M., Richardson, G., & Laheij, W. (1985). Pointing and Voicing in Deictic Expressions. *Journal of Memory and Language*, 24(2), 133-164.
- Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, 12(2), 264-271. doi: DOI 10.1111/j.1467-7687.2008.00758.x
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108(3), 732-739. doi: DOI 10.1016/j.cognition.2008.06.013
- Luck, S. J. (2005). *An introduction to the event-related potential te*. Cambridge, Mass.: MIT Press.
- Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage*, 21(2), 725-732. doi: DOI 10.1016/j.neuroimage.2003.09.049
- Macedonia, M., Muller, K., & Friederici, A. D. (2010). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*. doi: 10.1002/hbm.21084
- Mallery, G. (1881). *Sign Language among North American Indians compared with that among Other Peoples and Deaf-Mutes*. (Photomechanic reprint of the Smithsonian Report ed., 1972, The Hague: Mouton)
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1-2), 71-102. doi: 0010-0277(87)90005-9 [pii]
- Martin, C., Vu, H., Kellas, G., & Metcalfe, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 52(4), 813-839.
- McNeill, D. (1985). So You Think Gestures Are Nonverbal. *Psychological Review*, 92(3), 350-371.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press; US.
- McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press; US.
- McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, 27(3), 223-237.

- McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, 36(1), 53-65.
- Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, 22(4), 473-500. doi: Doi 10.1080/01690960600696916
- Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4(2), 119-141.
- Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures Evidence from a comparative analysis of human-human and human-machine interactions. *Gesture*, 9(1), 97-126. doi: DOI 10.1075/gest.9.1.04mol
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res*, 14(1), 115-28.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word Familiarity Predicts Temporal Asynchrony of Hand Gestures and Speech. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18(3), 615-622.
- Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, 117(3), 411-424.
- Morsella, E., & Krauss, R. M. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture-speech theories. *Journal of Psycholinguistic Research*, 34(4), 415-427. doi: DOI 10.1007/s10936-005-6141-9
- Nair, D. G. (2005). About being BOLD. *Brain Res Brain Res Rev*, 50(2), 229-243. doi: S0165-0173(05)00111-6 [pii] 10.1016/j.brainresrev.2005.07.001
- Nieuwenhuys, R. (1994). The Neocortex - an Overview of Its Evolutionary Development, Structural Organization and Synaptology. *Anatomy and Embryology*, 190(4), 307-337.
- Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? . In D. McNeill (Ed.), *Language and gesture* (pp. 186-198). Cambridge: Cambridge University Press.
- Noppeney, U. (2010, September). Audiovisual integration within the cortical hierarchy: Neural mechanisms and functional relevance. Talk given at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.
- Noppeney, U., Josephs, O., Hocking, J., Price, C.J., & Friston, K. J. (2008). The Effect of Prior Visual Information on Recognition of Speech and Sounds. *Cerebral Cortex*, 18, 598-609.



- Nunez, P. L. (1981). *Electrical fields of the brain*. New York: Oxford University Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.
- 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.
- Perfetti, C. A., Goldman, S. R., & Hogaboam, T. W. (1979). Reading Skill and the Identification of Words in Discourse Context. *Memory & Cognition*, 7(4), 273-282.
- Ping, R., & Goldin-Meadow, S. (2010). Gesturing Saves Cognitive Resources When Talking About Nonpresent Objects. *Cognitive Science*, 34(4), 602-619. doi: DOI 10.1111/j.1551-6709.2010.01102.x
- 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.
- Prinz, W. (2005). An ideomotor approach to imitation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science* (Vol. 1, pp. 141-156). Cambridge, MA: MIT Press.
- Pulvermüller, F., Hauk, O., Nikulin, V. N., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21, 793-797.
- Quintilian, M. F. (1922). *The Institutio Oratoria of Quintilian* (H. E. Butler, Trans.). New York: G. P. Putnam and Sons.
- Rauschecker, G. H., Guard, D. C., Phan, M. L., & Su, T. K. (2000). Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *Journal of Neurophysiology*, 83, 2723-2739.
- Rauscher, F. H., Krauss, R. M., & Chen, Y. S. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7(4), 226-231.
- Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *Neuroreport*, 15(3), 409-412. doi: DOI 10.1097/01.wnr.0000113531.32218.0d

- Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebral Cortex*, *15*, 1261-1269.
- Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research*, *5*(1), 54-62.
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cerebral Cortex*, *17*(5), 1147-1153. doi: DOI 10.1093/cercor/bhl024
- Rowe, M. L., & Goldin-Meadow, S. (2009a). Differences in Early Gesture Explain SES Disparities in Child Vocabulary Size at School Entry. *Science*, *323*(5916), 951-953. doi: DOI 10.1126/science.1167025
- Rowe, M. L., & Goldin-Meadow, S. (2009b). Early gesture selectively predicts later language learning. *Developmental Science*, *12*(1), 182-187. doi: DOI 10.1111/j.1467-7687.2008.00764.x
- Rueschemeyer, S.-A., Lindemann, O., van Rooij, D., van Dam, W., & Bekkering, H. (2010). Effects of intentional Motor Actions on Embodied Language Processing. *Experimental Psychology*, *57*(4), 260-266.
- Rugg, M. D., & Coles, M. G. H. (1995). *Electrophysiology of mind event-related brain potentials and cognition*. Oxford, UK: Oxford Univ. Press.
- Sadoski, M., & Paivio, A. (2001). *Imagery and text. A dual coding theory of reading and writing*. Mahwah, NJ [u.a.]: Erlbaum.
- Salasoo, A., & Pisoni, D. B. (1985). Interaction of Knowledge Sources in Spoken Word Identification. *Journal of Memory and Language*, *24*(2), 210-231.
- Sassenberg, U., & van der Meer, E. (2010). Do We Really Gesture More When It Is More Difficult? *Cognitive Science*, *34*(4), 643-664. doi: DOI 10.1111/j.1551-6709.2010.01101.x
- Schegloff, E. A. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 266-298). Cambridge: Cambridge University Press.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and Automatic Human Information-Processing .1. Detection, Search, and Attention. *Psychological Review*, *84*(1), 1-66.
- Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology*, *15*, 454-458.

- Seliger, G. M., Lefever, F., Lukas, R., Chen, J., Schwartz, S., Codeghini, L., et al. (1991). Word deafness in head injury: Implications from coma assessment and rehabilitation. *Brain Injury*, 5, 53-56.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* (27), 379-423; 623-656.
- Sharbrough, F., Chatrian, G., Lesser, R., Lüders, H., Nuwer, M., & Picton, T. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8(200-202).
- 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.
- Simpson, G. B. (1981). Meaning Dominance and Semantic Context in the Processing of Lexical Ambiguity. *Journal of Verbal Learning and 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.
- Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, 40(1), 160-164.
- Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *Neuroreport*, 18(4), 347-350.
- Straube, B., Green, A., & Kircher, T. (2010). *The processing of iconic and metaphoric co-verbal gesture: Common and unique integration processes*. Paper presented at the 4th conference of the international society for gesture studies, Frankfurt/Oder.
- Straube, B., Green, A., Weis, S., Chatterjee, A., & Kircher, T. (2009). Memory effects of speech and gesture binding: Cortical and hippocampal activation in relation to subsequent memory performance. *Journal of Cognitive Neuroscience*, 21(4), 821-836.
- Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, 26(2), 212-215.

- Swinney, D. A. (1979). Lexical Access during Sentence Comprehension - (Re)Consideration of Context Effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645-659.
- Taylor, W. L. (1953). "Cloze" procedure: A new tool for measuring readability. *Journalism Quarterly* (30), 415.
- Thierry, G., & Price, C. J. (2006). Dissociating verbal and nonverbal conceptual processing in the human brain. *Journal of Cognitive Neuroscience*, 18(6), 1018-1028.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78(3), 705-722.
- Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and phases: The dynamics of speech-hand communication. *Ecological Psychology*, 20(1), 32-64. doi: Doi 10.1080/10407410701766643
- Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: a parallel independent model. *Psychonomic Bulletin & Review*, 7(1), 49-82.
- Tylor, E. B. (1865). *Researches into the Early History of Mankind and the Development of Civilization*. London. John Murray.
- Valbonesi, L., R., A., McNeill, D., Quek, F., Duncan, S. D., McCullough, K.-E., et al. (2001). *Temporal correlation of speech and gestural focal points*. Paper presented at the Congress of the International Society for Gesture Studies, Austin.
- van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, 17(4), 962-974. doi: bhl007 [pii] 10.1093/cercor/bhl007
- van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage*, 36(4), 1345-1360. doi: S1053-8119(07)00272-8 [pii] 10.1016/j.neuroimage.2007.03.065
- van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657-671.
- Van Petten, C., & 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 Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607. doi: DOI 10.1016/j.neuropsychologia.2006.01.001

- Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, 181(1), 173-181. doi: DOI 10.1007/s00221-007-0918-z
- Vatakis, A., & Spence, C. (2006a). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, 1111, 134-142. doi: DOI 10.1016/j.brainres.2006.05.078
- Vatakis, A., & Spence, C. (2006b). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters*, 393(1), 40-44. doi: DOI 10.1016/j.neulet.2005.09.032
- Vu, H., Kellas, G., & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory & Cognition*, 26(5), 979-1001.
- Wagner, S. (2002). *Verbales Arbeitsgedächtnis und die Verarbeitung*. Leipzig: Max-Planck-Institute of Cognitive Neuroscience.
- Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, 50(4), 395-407. doi: DOI 10.1016/j.jml.2004.01.002
- Walley, A. C. (1988). Spoken Word Recognition by Young-Children and Adults. *Cognitive Development*, 3(2), 137-165.
- 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.
- 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.
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107, 829-854.
- Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36(5), 559-570.
- Werner, S., & Noppeney, U. (2010). Distinct Functional Contributions of Primary Sensory Association Areas to Audiovisual Integration in Object Categorization. *The Journal of Neuroscience*, 30(7), 2662-2675.

- Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, 114(4), 591-600.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, 17(10), 2322-2333. doi: DOI 10.1093/cercor/bhl141
- Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage*, 47(4), 1992-2004. doi: DOI 10.1016/j.neuroimage.2009.05.066
- Woodall, W. G., & Burgoon, J. K. (1981). The Effects of Nonverbal Synchrony on Message Comprehension and Persuasiveness. *Journal of Nonverbal Behavior*, 5(4), 207-223.
- Wu, Y. C. (2005). Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures. *CRL Newsletters*, 17(2), 3-15.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654-667. doi: DOI 10.1111/j.1469-8986.2005.00356.x
- Wu, Y. C., & Coulson, S. (2007a). How iconic gestures enhance communication: An ERP study. *Brain and Language*, 101(3), 234-245. doi: DOI 10.1016/j.bandl.2006.12.003
- Wu, Y. C., & Coulson, S. (2007b). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, 14(1), 57-63.
- Wu, Y. C., & Coulson, S. (2010). Gestures modulate speech processing early in utterances. *Neuroreport*, 21(7), 522-526. doi: Doi 10.1097/Wnr.0b013e32833904bb
- Wundt, W. (1973). *The Language of Gestures* (J. S. Thayer, C. M. Greenleaf, & M. D. Silberman, Trans.). The Hague: Mouton. (Original work published in 1900).
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, 152(2), 198-210. doi: DOI 10.1007/s00221-003-1536-z

## List of Figures

Figure 1.1:	The temporal structure of an iconic gesture .....	13
Figure 1.2:	Typical time course of an iconic gesture .....	14
Figure 2.1:	Layers of the human cortex .....	40
Figure 2.2:	Schematic illustration of the current flow after excitation in a pyramidal cell .....	41
Figure 2.3:	Electrode setup used in Experiment 2 .....	43
Figure 2.4:	ERP averaging procedure .....	43
Figure 2.5:	Example for an N400 effect .....	45
Figure 2.6:	Meta-analysis of fMRI studies on gesture-speech integration .....	58
Figure 2.7:	Timing of the original full-length gesture-speech material .....	60
Figure 3.1:	Schematic illustration of the gesture phases for the homonym <i>Kamm</i> .....	67
Figure 3.2:	Splicing procedure .....	70
Figure 3.3:	Temporal sequence of a typical trial of the gating .....	65
Figure 3.4:	Distribution of the disambiguation points of all gestures .....	76
Figure 3.5:	Illustration of the video stimuli construction .....	78
Figure 4.1:	Timing of gesture fragments and speech in Experiments 1 and 2 .....	86
Figure 4.2:	Regions of Interest (ROIs) used for data analysis .....	90
Figure 4.3:	ERPs as found in Experiment 1 .....	91
Figure 4.4:	Schematic illustration of Experiment 2 .....	96
Figure 4.5:	ERPs as found in Experiment 2 .....	97
Figure 5.1:	Timing of gesture fragments and speech in Experiment 3 .....	103
Figure 5.2:	ERPs as found in Experiment 3 .....	105
Figure 5.3:	The timing between gesture fragments and speech in Experiment 4 .....	108
Figure 5.4:	ERP results of Experiment 4 .....	112

*List of Figures*

171

Figure 6.1:	ERPs at the homonym as found in Experiment 5 .....	122
Figure 6.2:	ERPs at the target words as found in Experiment 5 .....	123
Figure 7.1:	The architecture of the Feature Integration Model .....	136



**List of Tables**

Table 3.1:	Stimulus examples .....	71
Table 3.2:	Stimulus properties of the full-length gestures .....	72
Table 3.3:	Stimulus properties of the gesture fragments .....	79
Table 4.1:	Stimulus examples for Experiments 1 to 5 .....	88

# Appendix

## Appendix A: Sentence material used in Experiments 1 to 5

### *Training sentences*

Item	Introduction	Dominant Target	Introduction	Subordinate Target
Abkuerzung	Regine wusste sich zu helfen.	*	Regine wusste sich zu helfen.	Sie benutzte die Abkürzung, weil vor dem Weg ein Hindernis stand.
Dichtung	Nicholas war offensichtlich beschäftigt.	Er stellte die Dichtung fertig, die das Rohr verbinden sollte.	Nicholas war offensichtlich beschäftigt	*
Leitung	Nichola traf eine Entscheidung.	Sie übernahm die Leitung, weil die Firma in einer Krise steckte.	Nichola traf eine Entscheidung.	*
Pass	Björn hatte so etwas noch nie gesehen.	Er betrachtete den Pass, der für die Grenze hervorgeholt wurde	Björn hatte so etwas noch nie gesehen.	*

\* Note, that the training sentences were only constructed for one meaning of the homonyms. In cases, in which the second meaning was needed (gating, cloze procedure), additional stimulus material was constructed.

*Stimulus sentences*

Item	Introduction	Dominant	Introduction	Subordinate
Absatz	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.
Anbau	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.
Auflauf	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.
Aufsatz	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.
Aussprache	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.
Ball	Alle waren von Sandra beeindruckt.	Sie beherrschte den Ball, was sich im Spiel beim	Alle waren von Sandra beeindruckt.	Sie beherrschte den Ball, was sich im Tanz mit dem Bräutigam

		Aufschlag zeigte.		deutlich zeigte.
Bogen	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.
Boxer	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.
Brause	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.
Bremse	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.
Eingang	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.
Fahne	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.

Fassung	Thomas musste den Job zu Ende bringen.	Er arbeitete an der Fassung, die für den Artikel vorgesehen war.	Thomas musste den Job zu Ende bringen.	Er arbeitete an der Fassung, die für die Lampe vorgesehen war.
Feder	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.
Fliege	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.
Flügel	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.
Futter	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 fertig gestellt werden sollte.
Gang	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.
Kamm	Marcos Entscheidung war eindeutig.	Er bevorzugte den Kamm, weil der Scheitel sich so leichter in Form	Marcos Entscheidung war eindeutig.	Er bevorzugte den Kamm, weil der Berg sich hier von seiner schönsten

		bringen ließ.		Seite zeigte.
Kapelle	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.
Kater	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.
Kundschaft	Hannes war bekannt für seine guten Manieren.	Er erwartete die Kundschaft, weil der Laden neu eröffnet wurde.	Hannes war bekannt für seine guten Manieren.	Er erwartete die Kundschaft, weil die Nachricht alles verändern könnte.
Linse	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.
Lösung	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.
Magazin	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.

Maus	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.
Messe	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.
Note	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.
Orden	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.
Ordner	Maren war aufgeregt.	Sie berührte den Ordner, als das Stadion betreten wurde.	Maren war aufgeregt.	Sie berührte den Ordner, als das Papier heraus fiel.
Pflaster	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.
Probe	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.



Quelle	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.
Rock	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.
Rolle	Susanne war zufrieden.	Sie bekam die Rolle, weil der Schauspieler so gut zu ihr passte.	Susanne war zufrieden.	Sie bekam die Rolle, weil der Schneider das Garn nicht mehr benötigte.
Schale	Meike war sehr vorsichtig.	Sie entfernte die Schale, weil beim Kristall kleinste Erschütterungen zum Bruch führen konnten.	Meike war sehr vorsichtig.	Sie entfernte die Schale, weil beim Apfel einige Stellen dreckig waren.
Schicht	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.
Schimmel	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.

Schloss	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 der Schlüssel stecken blieb und abbrach.
Spalte	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.
Spitze	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.
Stamm	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.
Stärke	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.
Strauss	Beate war	Sie freute sich über den Strauss,	Beate war	Sie freute sich über den Strauss,

	begeistert.	weil der Vogel so schnell rennen konnte.	begeistert.	weil die Blumen so schön dufteten.
Ton	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.
Wanze	Martha machte eine unangenehme Entdeckung.	Sie sah die Wanze, weil der Agent das Telefon nicht zugeschraubt hatte.	Martha machte eine unangenehme Entdeckung.	Sie sah die Wanze, weil der Käfer direkt darauf zugelaufen hatte.
Zeche	Ulrich übernahm Verantwortung.	Er gab alles für die Zeche, weil die Kneipe so teuer war.	Ulrich übernahm Verantwortung.	Er gab alles für die Zeche, weil das Bergwerk sein Lebensinhalt war.
Zirkel	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.

---

---

## **BIBLIOGRAPHIC DETAILS**

Obermeier, Christian

Exploring the significance of task, timing and background noise on gesture-speech integration

Fakultät für Biowissenschaften, Pharmazie und Psychologie

Universität Leipzig

*Dissertation*

182 pages, 237 references, 26 figures, 4 tables

---

In everyday face-to-face conversation, speakers not only use speech to transfer information but also rely on co-speech gestures. To date, a steadily increasing number of studies have shown that such co-speech gestures can indeed influence speech comprehension. Little, however, is known about the nature of gesture-speech integration. The present dissertation explored the significance of task, timing and background noise (speech quality) for the integration of gesture and speech. For this purpose, a disambiguation paradigm was used, in which participants were presented with short video clips of an actress uttering sentences like “She was impressed by the BALL, because the GAME / DANCE...”. The ambiguous noun, which was always a homonym (BALL), was accompanied by a dynamic iconic gesture fragment containing the minimal necessary amount of information to disambiguate the noun. This amount was determined by a context guided gating. In a series of 5 event-related potentials (ERP) experiments, both the direct integration of a gesture fragment with the ambiguous noun as well as the effect of this process on the sentence level (at the target word, e.g. GAME) were analyzed. Experiments 1 & 2 were set out to clarify the impact of task on gesture-speech integration, whereas Experiments 3 & 4 probed the role of timing between gesture and speech for their integration. Finally, Experiment 5 addressed how speech quality affects the impact of gesture on speech comprehension.

The combined ERP results suggest that the temporal alignment between gesture and the corresponding speech is crucial for a rather “automatic”, obligatory integration of both

information streams. Based on the results of Experiments 2-4, a time window ranging from -200 ms (auditory input lags visual input) to +120 ms (visual input lags auditory input) was determined for the local, immediate integration of gesture and corresponding speech unit. The findings of Experiments 1 and 5, however, suggest that integration of

gesture and speech is also possible beyond the temporal window of integration. In this case, however, the integration is no longer “automatic” and effortful gesture-related memory processes are necessary to be able to combine the gesture fragment and speech context in such a way that the homonym is disambiguated correctly. This “non-automatic” type of gesture-speech integration, can either be triggered externally via task / instruction (Experiment 1) or internally by the addressee her- / himself as in the case of an impaired speech signal (e.g. when embedded in noise, see Experiment 5).

In the General Discussion, the implications of the present findings for a potential model of gesture-speech integration are discussed in the light of previous gesture and multisensory research. Finally, a blueprint for such a model (*Feature Integration Model*) is provided that combines the present results with ideas from both linguistic and multisensory research within the framework of a 3-Stage-Model (Perceptual analysis, Feature extraction and Semantic integration).

## Summary

### Introduction

In everyday face-to-face conversation, speakers not only use speech to transfer information but also rely on facial expressions, body posture and gestures. In particular, co-speech gestures are playing an important role in daily communication. This category of spontaneous hand movements consists of different sub-types such as beats, emblems, deictic, metaphoric and iconic gestures. Iconic gestures are distinguished by their “close formal relationship to the semantic content of speech” (McNeill, 1992, p.12) and are the most thoroughly studied gesture type. A steadily increasing number of studies has shown that such iconic gestures are not only closely linked to the content of the accompanying speech but that they also have an effect on speech comprehension (Event-related potential (ERP) studies: e.g. Kelly, Kravitz, & Hopkins, 2004; Wu & Coulson, 2005; Holle & Gunter, 2007; Kelly, Ward, Creigh, & Bartolotti, 2007; Özyürek, Willems, Kita, & Hagoort, 2007; functional magnetic resonance imaging (fMRI) studies: e.g. Willems, Özyürek, & Hagoort, 2007; Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, 2008; Holle, Obleser, Rüschemeyer, & Gunter, 2010; Green, et al., 2009). Whereas such experiments suggest that a listener can extract additional information from iconic gestures and use that information linguistically, little is known so far about the factors that impact gesture-speech integration. From a theoretical perspective, however, this is a very important aspect which has already attracted scientific interest early on (cf. Wundt, 1921/1973). To date, there has been no systematic, experimental approach that tried to shed light on this issue. Yet, previous multisensory and gesture research gives at least some indications that, for instance, task, timing and background noise could play a role in gesture-speech integration (task: van Atteveldt, Formisano, Goebel, & Blomert, 2007; timing: Dixon & Spitz, 1980; McNeill, 1992, 2005; van Atteveldt, Formisano, Blomert, & Goebel, 2007; van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2006; background noise: Rogers, 1978). There is little doubt, that identifying and clarifying the impact of such factors presents a condition sine qua non en route to a full-fledged cognitive model of gesture comprehension.

The aim of the present dissertation was to identify the significance of task, timing, and background noise on the integration of gestural information in sentence comprehension in a

series of 5 experiments. Because timing is a critical issue here, event related potentials (ERPs) of the electroencephalogram (EEG) were used as the dependent measure as they provide an excellent temporal resolution.

## **Experiments**

As in the experiments by Holle and Gunter (2007), a disambiguation paradigm was used. Participants were presented with sentences containing an unbalanced homonym (e.g. Ball / English: *ball*), which was disambiguated downstream in the sentence by a target word, which was either related to the dominant, more frequent meaning of the homonym (dominant target: Spiel / *game*) or related to the subordinate, less frequent meaning (subordinate target: Tanz / *dance*). In contrast to Holle and Gunter (2007), the homonym was not accompanied by a full-length iconic gesture which depicted either the dominant or subordinate meaning, but by a gesture fragment containing the minimal necessary information to either cue the dominant or subordinate meaning of a homonym (i.e. the disambiguation point [DP] of a gesture). The gesture fragments were determined with the use of a gating study. Note, that due to stimulus construction, the gesture fragments ended 1000 ms prior to the point in time where the corresponding homonym was identified (the homonym identification point [IP]). I.e. gesture and speech were asynchronous.

The use of gesture fragments has several advantages over the use of full-length gestures. Full-length gestures share an extensive temporal overlap with the critical speech material including homonym and target word. Hence, it is impossible to distinguish whether the effect at the target word is caused by the integration of gesture and target word or is a consequence of the local integration of gesture and homonym. Additionally, there is considerable variability within the full-length gestures at which point in time they become meaningful. In contrast, the last frame of the gesture fragments always contains the minimal necessary information to disambiguate the corresponding homonym, thus reducing the semantic variability in the gesture material. The use of gesture fragments, therefore, allows for a more precise investigation of timing between

gesture and speech. The major advantage, however, is that the use of gesture fragments offers the unique possibility to investigate the direct integration of gesture fragment and homonym separately from the delayed disambiguating effect at the target word.

Experiments 1 & 2 were set out to clarify the impact of task on gesture-speech integration, whereas Experiments 3 & 4 probed the role of timing between gesture and speech for their integration. Finally, Experiment 5 addressed how speech quality affects the impact of gesture on speech comprehension.

## **Results & Discussion**

Experiment 1 was carried out to determine whether addressees are able to use gesture fragment information at all, using an explicit congruency judgment task that required participants to integrate both streams of information to solve the task. Note, that due to the stimulus construction procedure, the gesture fragments used in Experiment 1 (as well as Experiments 2 and 5) ended almost 1000 ms prior to the homonym IP. If participants made use of the minimal gestural information, ERP effects for integration and disambiguation should be observed.

The ERPs triggered by the homonym IPs revealed a direct influence of gesture during the processing of the ambiguous word. Subordinate gesture fragments elicited a more negative deflection compared to dominant gesture fragments, indicating that the integration of subordinate gesture fragments with the homonym is more effortful. The ERP data at the target words showed that the gesture fragments were not only integrated with speech, but were also used to disambiguate the homonym. When a target word was incongruent with the meaning of the preceding gesture-homonym combination, a larger N400 was elicited as compared to when this meaning was congruent. The target word effect is similar to the findings by Holle & Gunter (2007) who used full-length gestures instead of fragments. Thus, the results of Experiment 1 revealed that participants were able to use the gesture fragments for disambiguation both at local homonym as well as global sentence level. In contrast to Experiment 1, no significant ERP effects were observed in Experiment 2 which featured a much shallower task that did not require the integration of gesture and speech information. One possible interpretation is that gesture-speech integration is task-dependent and thus not automatic according to the two-process

theories of information processing (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The data of Experiment 3, however, suggests a different reason for the null finding of Experiment 2.



In Experiment 3, the identical task as in Experiment 2 was used. The crucial difference was that the gesture fragments and homonyms were synchronized (to the homonym IP) this time in order to test whether the same “idea unit” (McNeill, 1992, p.27) must occur simultaneously in both gesture and speech to allow proper, i.e. obligatory, integration (semantic synchrony rule, McNeill, 1992). This synchrony manipulation led to a robust negativity for the subordinate gestures at the homonym as well as to a significant N400 effect at subordinate target words. Thus, the information uptake from gesture seemed to be rather obligatory when both streams of information were synchronous or shared some temporal overlap, which was not the case in Experiments 1 and 2. This finding supports McNeill’s semantic synchrony rule (1992) in which he proposes that the information from gesture and speech is automatically and involuntarily integrated when both streams of information are accessed simultaneously. Findings from other studies on gesture production suggest, however, that the timing between gesture and speech varies considerably (e.g. Morrel-Samuels & Krauss, 1992). It has been shown for other types of multimodal integration processes, that the exact synchrony is not decisive for the integration as the human perceptual system is able to compensate a certain degree of asynchrony (e.g. between visual and auditory speech information: Dixon & Spitz, 1980; van Wassenhove, et al., 2007). This so-called temporal window of integration varies with regard to stimulus complexity, i.e. it is larger for more complex stimuli (see Vatakis, et al., 2007). For example, for audiovisual speech onset, asynchronies within a range of -200 ms (visual lead) to +100ms (auditory lead) can be compensated. Experiment 4 explored whether there was also some kind of temporal window of integration for gesture and speech. For this purpose, the homonym IP was either prior (+120 ms), synchronous with (0 ms, identical to Experiment 3) or lagging behind the end of the gesture fragment (-600 ms / -200 ms). The ERPs triggered to the homonym IP showed integration of gesture fragment and speech only in the -200ms, 0ms and +120 ms conditions. In these conditions, the subordinate gesture fragments elicited a larger N400 than the dominant gesture fragments. The disambiguating influence of the gesture-homonym combination at the target word was unaffected by the timing manipulation. For all timing conditions, subordinate target words preceded by an incongruent gesture-speech context elicited a larger N400 than those

preceded by a congruent context. Thus, based on the present results, the temporal window for the integration of gesture fragment and speech seems to be somewhere between -200 ms (audio lag) and +120 ms (visual lag).

Using the same setup as Experiment 2, Experiment 5 addressed the question whether gesture-speech integration in a noisy environment differs from the processing of both streams in silence. It is known, for instance, that communicators especially make use of gestural information when the speech signal quality is bad (e.g. Rogers, 1978). Such a situation was reproduced by embedding speech in multi-speaker babble noise. We hypothesized that the information uptake is different in the sense that it might be more automatic in noise than in silence. In the silence condition, the null result of Experiment 2 was replicated. In contrast, the ERPs in the babble noise condition showed both clear effects of gesture-homonym integration and disambiguation of the subordinate target word. Thus, participants must have adapted their gesture processing strategy from “no integration due to asynchrony” to ‘integration to compensate speech difficulties’.

## Summary

Taken together, the present dissertation provided a first insight into the complexity and flexibility of gesture-speech integration by showing the significance of task, timing and background noise for this process. In particular, the temporal alignment between gesture and the corresponding homonym proved to be crucial for a rather automatic integration of both information streams. Based on the present results, a time window ranging from -200 ms (auditory input lags visual input) to +120 ms (visual input lags auditory input) was determined for the local, immediate integration of gesture and corresponding speech unit. The present work also provides first evidence that integration of gesture and speech is possible beyond the temporal window of integration. This process, however, is no longer “automatic” but driven by additional executive, top-down (e.g. task, memory) influences as well as situational, bottom-up influences (e.g. noisy environment). The present findings have implications for a potential model of gesture-speech integration. A blueprint for such a model (*Feature Integration Model*), which combines the

present results with ideas from both linguistic and multisensory research within the framework of a 3-Stage-Model (Perceptual analysis, Feature extraction and Semantic integration) is presented.



## **Zusammenfassung**

### **Einleitung**

Wenn wir über menschliche Kommunikation sprechen, denken wir als erstes natürlich an Sprache - aber es gibt auch andere Möglichkeiten, unseren Gesprächspartnern Informationen zu vermitteln, z. B. Mimik, Körpersprache und sprachbegleitende Gesten. Unter Letzteren versteht man spontan produzierte, sprachrelatierte Handbewegungen, wie zum Beispiel Taktstock-Gesten, Embleme, Zeige-Gesten, metaphorische und ikonische Gesten. Ikonische Gesten sind dadurch gekennzeichnet, dass sie formal-inhaltlich eng an den semantischen Gehalt der Sprache angelehnt sind (McNeill, 1992). Sie gehören zu den bisher am besten erforschten Gesten. Eine stetig steigende Zahl an Studien hat gezeigt, dass ikonische Gesten nicht nur inhaltlich eng mit der dazugehörigen Sprache verbunden sind, sondern auch das Sprachverstehen beeinflussen können (EKP-Studien: z.B. Kelly, Kravitz & Hopkins., 2004; Wu & Coulson, 2005; Kelly, Ward, Creigh & Bartolotti, 2007; Özyürek, Willems, Kita & Hagoort, 2007; functional magnetic resonance imaging (fMRI) studies: e.g. Willems, Özyürek & Hagoort, 2007; Holle, Gunter, Rüschemeyer, Hennenlotter & Iacoboni, 2008; Holle, Obleser, Rüschemeyer & Gunter, 2010; Green et al., 2009). Die Ergebnisse der genannten Studien zeigen, dass ein Adressat zusätzlich Informationen aus ikonischen Gesten gewinnen und diese Informationen linguistisch nutzen kann. Allerdings ist bisher wenig darüber bekannt, auf welche Art diese Gesten-Sprach-Integration abläuft und wodurch die relevanten Verarbeitungsprozesse beeinflusst werden können. Aus theoretischer Sicht sind dies natürlich außerordentlich bedeutsame Aspekte. So verwundert es wenig, dass sie schon relativ früh in der Forschung Beachtung fanden (z.B. bei Wundt, 1921/1973). Bis dato gab es jedoch in der Gesten-Forschung keinen systematischen, experimentalpsychologischen Ansatz, mit dem versucht wurde, Gesten-Sprach-Integration und ihre Einflussfaktoren genauer zu charakterisieren. Einige Faktoren, die dabei eventuell eine wichtige Rolle spielen könnten, lassen sich aus der bereits vorhandenen Literatur zu Gesten bzw. multisensorischer Integration erahnen. So gibt es Befunde, die zeigen, dass Aufgabenstellung, zeitliche Anordnung von Geste und Sprache, sowie Hintergrundlärm bzw. Sprachqualität eine Rolle spielen könnten (Aufgabe: van Atteveldt, Formisano, Goebel & Blomert, 2007; zeitlicher Ablauf: Dixon & Spitz, 1980; McNeill, 1992, 2005; van Atteveldt, Formisano, Blomert & Goebel, 2007; van Wassenhove, Grant & Poeppel, 2007; Vatakis & Spence, 2006;

Hintergrundlärm: Rogers, 1978). Eben diese Identifikation von Einflussfaktoren auf die Gesten-Sprach-Integration, sowie die genauere Spezifizierung ihrer Auswirkung auf den Integrations-Prozess ist zweifellos eine wichtige Voraussetzung für die Erstellung eines Modells zur Gesten-Sprach-Integration.

Erklärtes Ziel der vorliegenden Dissertation war es, den Einfluss von Aufgabe, Synchronizität und Hintergrundlärm auf die Integration von Geste und Sprache beim Satzverstehen zu untersuchen. Zu diesem Zweck wurde eine Serie von fünf Experimenten durchgeführt. Da der genaue zeitliche Ablauf der Integration ein wichtiger Bestandteil der Fragestellungen dieser Dissertation war, wurden Ereigniskorrelierte Potentiale (EKPs) aufgrund ihrer exzellenten zeitlichen Auflösung als abhängiges Maß für alle Studien gewählt.

## **Experimente**

In Anlehnung an die Experimente von Holle und Gunter (2007), wurde ein Disambiguations-Paradigma zur Untersuchung der Fragestellungen gewählt. Dabei wurden den Versuchsteilnehmern Sätze präsentiert, die ein unbalanciertes Homonym enthielten (d.h. ein orthographisch und phonologisch identisches, mehrdeutiges Wort, wie z.B. Ball). Im weiteren Verlauf des Satzes wurde das Homonym durch ein Zielwort, das entweder in Bezug zur höher frequenten (dominanten) oder niedriger frequenten (subordinierten) Bedeutung des Homonyms stand, disambiguiert (dominante Bedeutung: Spiel, subordinierte Bedeutung: Tanz). Im Gegensatz zur Studie von Holle und Gunter (2007), die komplette ikonische Gesten simultan zu dem Homonym präsentierten, wurden in den vorliegenden Experimenten lediglich *Gesten-Fragmente* verwendet. Diese Fragmente beinhalteten die für einen Gesprächspartner minimal notwendige Information, um die richtige (also entweder dominante oder subordinierte) Bedeutung des entsprechenden Homonyms auszuwählen. Der Punkt, zu dem die notwendige Information in der Geste verfügbar ist, ist der sogenannte Disambiguations-Punkt (DP), der gleichzeitig auch dem Ende des Gesten-Fragments entspricht. Bestimmt wurden die Disambiguations-Punkte für die einzelnen Gesten mit Hilfe eines Gating-Experimentes. Hierbei ist es im Zusammenhang mit den Materialien wichtig zu erwähnen, dass die Gesten-Fragmente als Folge der Bestimmungsmethodik etwa 1000 ms vor dem Zeitpunkt in den Sätzen endeten, an

denen das zugehörige Homonym identifiziert wurde (dem sogenannten Homonym-Identifikationspunkt [IP]). Mit anderen Worten, die Gesten-Fragmente und zugehörigen Homonyme wurden asynchron präsentiert.

Für die Beantwortung der Fragestellungen dieser Doktorarbeit hatte die Verwendung von Gesten-Fragmenten gegenüber vollständigen Gesten diverse Vorteile. Die vollständigen Gesten überlappen zeitlich mit großen Teilen des kritischen Satzmaterials, inklusive Homonym und Zielwort. Dies erschwerte die Interpretation der EKP-Effekte am Zielwort, da es nicht eindeutig zu klären ist, ob diese Effekte durch die Integration von Geste und Zielwort entstanden sind oder durch die lokale Integration von Geste und Homonym. Darüber hinaus gibt es eine hohe Varianz innerhalb der kompletten Gesten in Bezug darauf, wann ein Adressat ihre Bedeutung erkennen kann. Im Gegensatz dazu enthält bei den Gesten-Fragmenten das letzte zu sehende Einzelbild immer die minimal notwendige Information, die man braucht um, das zugehörige Homonym zu disambiguieren. Damit ist genau bekannt, wann die Gesten-Fragmenten bedeutungsvoll werden. Die Benutzung von Gesten-Fragmenten als Stimuli erlaubt daher eine präzisere Untersuchung des zeitlichen Ablaufs von Gesten-Sprach-Integration im Satzkontext. Der größte Vorteil jedoch ist, dass die Verwendung von Gesten-Fragmenten es, ermöglicht die lokale, direkte Integration von Gesten-Fragment und Homonym getrennt von dem globaleren Disambiguations-Effekt am Zielwort zu untersuchen.

Das Ziel von Experimente 1 und 2 war, den Einfluss unterschiedlicher Aufgabentypen auf die Integration von Gesten und Sprache zu untersuchen. Dagegen wurde in den Experimenten 3 und 4 die Wichtigkeit des zeitlichen Zusammenspiels von Geste und Sprache getestet. Darüber hinaus sollte Experiment 5 klären, auf welche Weise die Qualität des Sprachsignals die Verarbeitung von Gesten beeinflusst.

## **Ergebnisse und Diskussion**

Experiment 1 diente der Beantwortung der Frage, ob die Adressaten einer kommunikativen Nachricht überhaupt die Information, die in den Gesten-Fragmenten enthalten war, zum besseren

Verständnis der Nachricht nutzen. Dabei hatten die Versuchsteilnehmer die Aufgabe, zu beurteilen, ob die Gesten-Fragmente und der Inhalt der gehörten Sprache kongruent waren oder nicht. Mit anderen Worten, die Probanden mussten Gesten-Fragmente und Sprache integrieren, um die Aufgabe zu lösen. Wie bereits oben erwähnt, endeten die Gesten-Fragmente etwa 1000 ms vor dem Homonym-IP. Das heißt, Geste und zugehöriges Wort waren asynchron. Falls die Versuchsteilnehmer die Gesten-Information nutzten, sollte sich dies anhand von EKP-Effekten für die lokale Integration und die Disambiguation zeigen. Tatsächlich zeigte sich für die EKPs, die auf dem Homonym-IP gemessen wurde, ein signifikanter Effekt. Gesten-Fragmente, die zur subordinierten Bedeutung des Homonyms in Bezug standen, lösten eine stärkere Negativierung aus als Fragmente, die zur dominanten Bedeutung related waren. Dieser Effekt zeigt, dass die Gesten-Fragmente von den Versuchspersonen mit dem Homonym integriert wurden. Darüber hinaus zeigte sich auch an der Position des Zielwortes ein signifikanter EKP-Effekt. Wenn die Bedeutung des Zielwortes inkongruent zu der durch die Integration von Gesten-Fragment und Homonym vorher determinierten Bedeutung des Homonyms war, löste dies eine stärkere N400 aus, als wenn die Bedeutung des Zielwortes kongruent zum vorhergehenden Kontext war. Dieser Effekt wurde in ähnlicher Form auch schon für vollständige Gesten gefunden (Holle & Gunter, 2007). Er zeigt nicht nur, dass die Gesten-Fragmente direkt mit dem Homonym integriert wurden, sondern auch, dass die in ihnen enthaltene Information zu Disambiguation des Homonyms genutzt wurde. Die Ergebnisse von Experiment 1 liefern somit einen Beleg dafür, dass die Gesten-Fragmente sowohl zur Disambiguation auf lokaler Homonym-Ebene als auch auf globaler Satz-Ebene hilfreich sind. Im Gegensatz zu Experiment 1, wurden in Experiment 2 keine signifikanten EKP-Effekte gefunden. Der einzige Unterschied zwischen beiden Experimenten war, dass die Aufgabe in Experiment 2 nur der Aufmerksamkeitskontrolle diente und keine Integration von Gesten-Fragmenten und Sprache erforderte. Eine mögliche Interpretation der unterschiedlichen Resultate in Experiment 1 und 2 wäre, dass die Integration von Gesten-Fragmenten und dem dazugehörigen Homonym aufgabenabhängig ist. Dies wiederum würde bedeuten, dass der Integrationsprozess von Gesten und Sprache nicht die Kriterien für einen „automatischen“ Prozess im Sinne der Zwei-Prozess-Theorie für Informationsverarbeitung erfüllen würde (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Allerdings lassen die Ergebnisse von Experiment 3 eine alternative und differenziertere Erklärung für den Null-Effekt in Experiment 2 als wahrscheinlicher erachten.

Die Aufgabe der Probanden blieb in Experiment 3 identisch wie zuvor in Experiment 2. Der entscheidende Unterschied zwischen beiden Experimenten war, dass nun in Experiment 3 die Gesten-Fragmente mit den Homonymen *synchron* dargeboten wurden. Genauer gesagt, wurden die Disambiguations-Punkte der Gesten zeitgleich mit den Identifikationspunkten der Homonyme präsentiert. Dies ermöglichte zu testen, ob tatsächlich die gleiche „Idee“ simultan in Geste und Sprache dargeboten werden muss, damit ein Adressat beide Information gut integrieren kann, d.h. dass es zu einer „automatischen“, obligatorischen Integration von Gesten und Sprache kommt (Semantische Synchronizitäts-Regel, McNeill, 1992). Die verwendete Synchronizitäts-Manipulation führte zu signifikanten EKP-Effekten sowohl auf Homonym- als auch auf Zielwort-Ebene. Ähnlich wie in Experiment 1, lösten die subordinierten Gesten-Fragmente eine stärkere Negativierung auf Homonym-Ebene aus als die dominanten Gesten-Fragmente. Ebenso wurde ein signifikant größere N400 gefunden, wenn die Bedeutung des Zielwortes inkongruent zum vorhergehenden, durch Gesten-Fragment und Homonym bestimmten Kontext war, als wenn sie kongruent war. Im Gegensatz zu Experiment 1 galt dies allerdings nur für Zielworte, die zur subordinierten Bedeutung des Homonyms related waren. Den Ergebnissen von Experiment 3 nach zu urteilen, scheint die Integration von Geste und Sprache obligatorisch zu sein, wenn die zugehörigen Informationen in beiden Kanälen simultan oder zumindest zeitlich überlappend dargeboten werden. Somit liefern die Resultate von Experiment 3 erste Anhaltspunkte für McNeills Semantische Synchronizitäts-Regel (1992), die besagt, dass Gesten und Sprache nur dann automatisch integriert werden können, wenn beide zeitgleich verarbeitet werden. Allerdings deuten Ergebnisse aus der Gesten-Produktions-Forschung darauf hin, dass die Semantische Synchronizitäts-Regel in ihrer strikten Form möglicherweise nicht zutreffend ist. So zeigte sich unter anderem, dass das zeitliche Zusammenspiel von Gesten und Sprache eine hohe Varianz aufweist (z.B. Morrel-Samuels & Krauss, 1992). Wenn man dazu noch Ergebnisse aus der Forschung zur multimodalen Integration betrachtet, so wird schnell deutlich, dass nicht exakte Synchronizität entscheidend für die Integration zweier Informationen ist, sondern, dass die menschliche Wahrnehmung die Kompensation eines bestimmten Maßes an Asynchronizität erlaubt (z.B. für visuelle und auditorische Sprachinformation: Dixon & Spitz, 1980; van Wassenhove et al., 2007). Dieses sogenannte Integrations-Zeitfenster bzw. dessen Größe hängt unter anderem von der Stimuluskomplexität ab: Es ist größer für komplexere Stimuli (siehe Vatakis et al., 2007). Für



audiovisuelle Sprachverarbeitung können, zum Beispiel, Asynchronizitäten von -200 ms (die visuelle Information kommt vor der auditorischen) bis + 100 ms (die auditorische Information kommt vor der visuellen) kompensiert werden. Experiment 4 ging daher der Frage nach, ob es auch ein Integrations-Zeitfenster für Gesten und Sprache gibt. Hierzu wurde der Homonym-IP in unterschiedlicher zeitlicher Relation zum Gesten-DP dargeboten. Der Homonym-IP wurde entweder vor dem Gesten-DP (+120 ms), zeitgleich (0 ms, identisch zu Experiment 3), oder danach (-600 ms / -200 ms) platziert. Die EKPs, gemessen am Homonym-IP, zeigten signifikante Effekte in der -200 ms, 0 ms und +120 ms Bedingung. In all diesen Bedingungen lösten die subordinierten Gesten-Fragmente eine höhere Negativierung aus als die dominanten Gesten-Fragmente. Mit anderen Worten, in den -200 ms, 0 ms und +120 ms Bedingungen fand eine Integration von Gesten-Fragment und Homonym statt. Im Gegensatz zu den Effekten am Homonym, war der am Zielwort gefundene Effekt unabhängig von der Synchronizitäts-Manipulation. In allen vier Bedingungen löste ein inkongruenter Gesten-Homonym-Kontext eine größere N400 am subordinierten Zielwort aus als ein kongruenter Kontext. Aufgrund der Ergebnisse von Experiment 4 kann also festgestellt werden, dass es auch für Gesten-Sprach-Integration ein kritisches Integrations-Zeitfenster gibt, welches, basierend auf den Ergebnissen, einen Zeitraum von -200 ms bis +120 ms umfasst.

Experiment 5 dagegen befasste sich mit der Frage, ob sich die Integration von Gesten und Sprache bei Hintergrundlärm von der Integration in Stille unterscheidet. Dazu wurden dieselben Stimuli und dieselbe Aufgabe wie in Experiment 2 verwendet. Aus der Gesten-Literatur ist bekannt, dass Adressaten vermehrt Gesten-Information nutzen, wenn die Qualität des Sprachsignal schlecht ist (z.B. Rogers, 1978). Diese Situation wurde in Experiment 5 dadurch simuliert, dass die Sätze in ein Stimmengewirr aus mehreren Sprechern eingebettet wurden, etwa vergleichbar mit einer Unterhaltung in einer lauten Bar. Es wurde angenommen, dass die Integration von Gesten und Sprache in einer lärmreichen Umgebung „automatischer“ von statten geht, als in Stille. In letzterer Bedingung wurde in Experiment 5 der Nullbefund aus Experiment 2 repliziert. Bei Hintergrundlärm dagegen fanden sich klare ERP-Effekte sowohl für die Integration von Gesten-Fragment und Homonym, als auch für die Integration des Zielwortes in den vorhergehenden Gesten-Sprach-Kontext. Das heißt, die Versuchspersonen haben ihre Gesten-Verarbeitung an die veränderten Umweltbedingungen angepasst („keine Integration

wegen der Asynchronizität“ wurde zu „Integration um die Sprachverständnisprobleme zu kompensieren“).

### **Zusammenfassung**

Die vorliegende Dissertation gibt einen ersten Einblick in die Komplexität und Flexibilität der Integration von Gesten und Sprache. Die experimentellen Ergebnisse zeigen, dass sowohl die *Aufgabe*, die *Synchronizität* von Gesten und Sprache als auch die *Qualität* des Sprachsignals einen Einfluss auf die Integration haben. Dabei scheint die *zeitliche Anordnung* von Geste und Sprache besonders wichtig zu sein. Innerhalb eines Zeitfensters von -200 ms bis + 100 ms erscheint die lokale Integration von Gesten und Homonym mehr oder weniger automatisch vonstatten zu gehen. Doch auch außerhalb des genannten Zeitfensters ist eine Integration von Gesten- und Sprach-Information möglich, wie die Ergebnisse der Dissertation zeigen. Allerdings erfolgt dieser Prozess in diesem Fall nicht mehr automatisch, sondern muss entweder durch zusätzliche top-down Faktoren (wie Aufgabe und Gedächtnis) oder auch durch situationale, bottom-up Einflüsse (wie z.B. eine lärmige Umgebung) angeschoben / ausgelöst? werden.

Die Ergebnisse der vorliegenden Dissertation weisen einige Implikationen für die Konstruktion eines möglichen Gesten-Sprach-Integration Modells auf. Ein Entwurf für ein solches „*Feature Integration Model*“, welches die gegenwärtige Resultate mit linguistischen und multimodalen Forschungsergebnissen im Rahmen eines dreigestuften Modells (Perzeptuelle Analyse, Extraktion von Bestandteilen, Semantische Integration) integriert, wird diskutiert.

## **Curriculum Vitae**

Name	Christian Obermeier
Date of Birth	23.11.1976
Place of Birth	Kaufbeuren
Country	Germany

### **EDUCATION**

Since July 2007	PhD student at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
2001-2007	Diploma in Psychology, Catholic University of Eichstätt-Ingolstadt, Germany
1997-2000 Germany	Studies in dentistry, Ludwig-Maximilians University, Munich, Germany
1996	Abitur (A-Levels equivalent), Jakob-Brucker-Gymnasium, Kaufbeuren, Germany

### **PROFESSIONAL EXPERIENCE**

2005	Student Research Assistant, Department of Experimental Psychology, Catholic University of Eichstätt-Ingolstadt, Germany
2005	Internship at the Clinic for Psychiatry, Psychotherapy and Psychosomatic Medicine, Havellandklinik Nauen, Germany
2003 - 2005	Student Research Assistant, Department of Experimental Psychology, Catholic University of Eichstätt-Ingolstadt, Germany
2003 - 2004	Assistance in the German retranslation and revision of "Psychology and Life" by P.G. Zimbardo and R.J. Gerrig (Graf, R., Nagler, M., & Riecker, B. (Eds.). (2004). Psychologie. Munich: Pearson Education.)
2003	Internship at the Department of Experimental Psychology, Catholic University of Eichstätt-Ingolstadt, Germany

## **Verzeichnis der eigenen Publikationen**

Obermeier, C., Dolk, T., & Gunter, T.C. (in press). The benefit of gestures during communication: Evidence from hearing and hearing impaired individuals. *Cortex*.

Obermeier, C., Holle, H., & Gunter, T.C. (2010). What iconic gesture fragments reveal about gesture–speech integration: When synchrony is lost, memory can help. Uncorrected proof. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn.2010.21498  
<http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2010.21498>

## **Selbstständigkeitserklärung**

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

Leipzig, 25. Februar 2011

---

Christian Obermeier

## 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 elektro-physiologischer 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 Aufgabenwechselparadigmas 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 Objektpräsentation: 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 Benennungsalternativen 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 get 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  
*Aufgabeneinhibition 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 sensorimotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen*
- 79 Päivi Sivenon  
*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 ist? 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: Neurophysiologische 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*
- 95 Henning Holle  
*The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies*
- 96 Marcel Braß  
*Das inferior frontale Kreuzungsareal und seine Rolle bei der kognitiven Kontrolle unseres Verhaltens*
- 97 Anna S. Hasting  
*Syntax in a blink: Early and automatic processing of syntactic rules as revealed by event-related brain potentials*
- 98 Sebastian Jentschke  
*Neural Correlates of Processing Syntax in Music and Language – Influences of Development, Musical Training and Language Impairment*
- 99 Amelie Mahlstedt  
*The Acquisition of Case marking Information as a Cue to Argument Interpretation in German An Electrophysiological Investigation with Pre-school Children*
- 100 Nikolaus Steinbeis  
*Investigating the meaning of music using EEG and fMRI*
- 101 Tilmann A. Klein  
*Learning from errors: Genetic evidence for a central role of dopamine in human performance monitoring*
- 102 Franziska Maria Korb  
*Die funktionelle Spezialisierung des lateralen präfrontalen Cortex: Untersuchungen mittels funktioneller Magnetresonanztomographie*
- 103 Sonja Fleischhauer  
*Neuronale Verarbeitung emotionaler Prosodie und Syntax: die Rolle des verbalen Arbeitsgedächtnisses*
- 104 Friederike Sophie Haupt  
*The component mapping problem: An investigation of grammatical function reanalysis in differing experimental contexts using event-related brain potentials*
- 105 Jens Brauer  
*Functional development and structural maturation in the brain's neural network underlying language comprehension*
- 106 Philipp Kanske  
*Exploring executive attention in emotion: ERP and fMRI evidence*
- 107 Julia Grieser Painter  
*Music, meaning, and a semantic space for musical sounds*
- 108 Daniela Sammler  
*The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies*
- 109 Norbert Zmyj  
*Selective Imitation in One-Year-Olds: How a Model's Characteristics Influence Imitation*
- 110 Thomas Fritz  
*Emotion investigated with music of variable valence – neurophysiology and cultural influence*
- 111 Stefanie Regel  
*The comprehension of figurative language: Electrophysiological evidence on the processing of irony*

- 112 Miriam Beisert  
*Transformation Rules in Tool Use*
- 113 Veronika Krieghoff  
*Neural correlates of Intentional Actions*
- 114 Andreja Bubić  
*Violation of expectations in sequence processing*
- 115 Claudia Männel  
*Prosodic processing during language acquisition: Electrophysiological studies on intonational phrase processing*
- 116 Konstanze Albrecht  
*Brain correlates of cognitive processes underlying intertemporal choice for self and other*
- 117 Katrin Sakreida  
*Nicht-motorische Funktionen des prämotorischen Kortex: Patientenstudien und funktionelle Bildgebung*
- 118 Susann Wolff  
*The interplay of free word order and pro-drop in incremental sentence processing: Neurophysiological evidence from Japanese*
- 119 Tim Raettig  
*The Cortical Infrastructure of Language Processing: Evidence from Functional and Anatomical Neuroimaging*
- 120 Maria Golde  
*Premotor cortex contributions to abstract and action-related relational processing*
- 121 Daniel S. Margulies  
*Resting-State Functional Connectivity fMRI: A new approach for assessing functional neuroanatomy in humans with applications to neuroanatomical, developmental and clinical questions*
- 122 Franziska Süß  
*The interplay between attention and syntactic processes in the adult and developing brain: ERP evidences*
- 123 Stefan Bode  
*From stimuli to motor responses: Decoding rules and decision mechanisms in the human brain*
- 124 Christiane Diefenbach  
*Interactions between sentence comprehension and concurrent action: The role of movement effects and timing*
- 125 Moritz M. Daum  
*Mechanismen der frühkindlichen Entwicklung des Handlungsverständnisses*
- 126 Jürgen Dukart  
*Contribution of FDG-PET and MRI to improve Understanding, Detection and Differentiation of Dementia*
- 127 Kamal Kumar Choudhary  
*Incremental Argument Interpretation in a Split Ergative Language: Neurophysiological Evidence from Hindi*
- 128 Peggy Sparenberg  
*Filling the Gap: Temporal and Motor Aspects of the Mental Simulation of Occluded Actions*
- 129 Luming Wang  
*The Influence of Animacy and Context on Word Order Processing: Neurophysiological Evidence from Mandarin Chinese*
- 130 Barbara Ettrich  
*Beeinträchtigung frontomedianer Funktionen bei Schädel-Hirn-Trauma*
- 131 Sandra Dietrich  
*Coordination of Unimanual Continuous Movements with External Events*
- 132 R. Muralikrishnan  
*An Electrophysiological Investigation Of Tamil Dative-Subject Constructions*