Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

Zygimante Dukauskaite

Neural Correlates of Speech and Gesture Integration: A Literature Review

Master's thesis in English Linguistics and Language Acquisition
Supervisor: Giosuè Baggio

May 2024

NTNU
Norwegian University of
Science and Technology

Zygimante Dukauskaite

Neural Correlates of Speech and Gesture Integration: A Literature Review

Master’s thesis in English Linguistics and Language Acquisition
Supervisor: Giosuè Baggio
May 2024

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

Abstract

Human communication extends beyond verbal exchanges, encompassing a variety of non-verbal cues, such as gestures, that play a crucial role in conveying information and emotions. Despite the recognized importance of both verbal and non-verbal communication, the neural mechanisms underlying their integration remain less explored. Therefore, the present thesis aims to bridge this gap by investigating the neural networks that facilitate this integration, enhancing the understanding of how these two modes of communication influence each other and contribute to the evolution of language.

The thesis is structured around three main objectives: (1) identifying brain regions and networks involved in the processing of speech and gestures, (2) exploring the impact of different types of co-speech gestures on neural mechanisms, and (3) examining what the gestures and vocalizations of non-human primates can reveal about the evolutionary origins of human language.

Through the analysis of various fMRI and EEG/ERP studies, significant neural activations were observed in IFG, STG, MTG, and Broca’s area during the processing of co-speech gestures. Furthermore, it was found that different types of co-speech gestures uniquely engage distinct brain regions. As for the third objective, the thesis focused on studies on non-human primates, which revealed that both gestures and vocalizations are integral to the evolutionary trajectory of language, supporting the theory of multimodal integration.

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Giosuè Baggio, who helped me throughout the writing process, suggested great ideas and increased my interest in a topic that was initially unfamiliar to me.

Besides my supervisor, I would like to thank my colleagues, neighbours, and friends who not only endured my frequent complaints but also supported me greatly throughout this journey. I feel truly fortunate to be surrounded by so many incredible people, both in Lithuania and in Norway, who believe in me more than I believe in myself. Special thanks to my dear friend Auguste, whose consistent presence and insightful feedback were invaluable.

Taip pat noriu padėkoti savo šeimai, ypač mamai. Šie metai nebuvo patys lengviausi dėl įvairiausių iššūkių su kuriais turėjau susidurti, bet tu visada mane palaikei, palaikai ir žinau, kad palaikysi. Kiekvieną kartą kai jaučiausi blogai ar neužtikrintai, visada žinojai ką tokiu metu pasakyti kas man padėjo judėti tolyn ir siekti savo tikslų.

Noriu padėkoti ir savo vaikinui Dovydui, kuris turėjo atlaikyti mano kartais sunkų charakterį ir dažnus breakdownus rašymo procese. Ačiū mano katei Žibai, esu labai dėkinga, kad atsiradai mano gyvenime ir atnešei tiek daug džiaugsmo.

Reflecting on the past two years at NTNU, it was indeed an interesting journey filled with learning and growth. Despite the rollercoaster of experiences, this time has profoundly changed me for the better, and I will always cherish these memories.

Dulcius ex asperis

Table of Contents

List of Figures ix

List of Tables x

List of Abbreviations xi

1. Introduction 1
   - 1.1. The Present Study: Research Questions 2
   - 1.2. Overview of the Thesis 3
2. Gestures 4
   - 2.1. Definition of Gestures 4
   - 2.2. Classification of Gestures 4
   - 2.3. Gestures of Great Apes 7
3. Gestures and Speech 9
   - 3.1. Competing Theories 9
   - 3.2. The Systematic Relationship between Gestures and Speech 10
   - 3.3. Gestures and Language Acquisition 11
   - 3.4. Gestures and Vocalizations in Non-Human Primates 13
4. Brain 17
   - 4.1. Overview of Brain Anatomy 17
     - 4.1.1. Broca's Area and the Inferior Frontal Gyrus 17
     - 4.1.2. The Auditory Cortex and Planum Temporale 18
     - 4.1.3. The Superior and Middle Temporal Gyrus 19
     - 4.1.4. The Inferior Parietal Lobule 20
5. Methodological Approaches in Studying Neural Integration 22
   - 5.1. Functional Magnetic Resonance Imaging (fMRI) 22
   - 5.2. Electroencephalography/Event-Related Potential (EEG/ERP) 23
6. Neural Bases of Speech and Gesture 28
   - 6.1. Neural Processing of Co-Speech Gestures 28
     - 6.1.1. fMRI Studies 28
     - 6.1.2. EEG/ERP Studies 36
     - 6.1.3. Conclusion and Discussion 43
   - 6.2. Neural Processing of Different Types of Gestures 45

6.2.1. Conclusion and Discussion ... 52

7. The Evolutionary Perspective ... 54

7.1. Language Origin Theories ... 54

7.1.1. The Gesture-First Theory ... 54

7.1.2. The Vocal-First Theory ... 59

7.1.3. The Multimodal Theory ... 64

7.1.4. Conclusion and Discussion ... 68

8. Conclusion ... 71

8.1. Limitations and Suggestions for Further Research ... 73

9. List of References ... 75

Appendix: Searching Strategies ... 87

List of Figures

Figure 4.1 Neuroanatomy of language processing (Friederici, 2017, p. 6) *17*
Figure 4.2 Anatomical representation of the planum temporale and primary auditory cortex (Murphy et al., 2017, p. 25) *19*
Figure 4.3 Anatomical representation of the inferior parietal lobule (Dary et al., 2023, p. 2948) *20*
Figure 5.1 Detection of neural response to a stimulus with fMRI BOLD signal (Arthurs & Boniface, 2002, p. 28) *23*
Figure 5.2 Electrode placement based on the international 10-20 system (Rojas et al., 2018, p. 2) *24*
Figure 5.3 Representation of frequency bands of EEG signal (Oon et al., 2018, p. 2) *25*
Figure 5.4 Representation of ERP acquisition (Key, 2016, p. 142) *25*
Figure 6.1 Brain activation patterns in response to concrete vs. abstract speech gestures (Nagels et al., 2013, p. 6) *29*
Figure 6.2 Hemispheric differences in IFG activation during gesture processing (Dick et al., 2009, p. 3518) *30*
Figure 6.3 Activation of Broca's area sub-regions under different conditions (Skipper et al., 2007, p. 273) *32*
Figure 6.4 Mean activation levels for gesture and speech integration across four conditions (Willems et al., 2007, p. 2327) *33*
Figure 6.5 Illustration of significantly activated brain regions (Holle et al., 2010, p. 881) *35*
Figure 6.6 ERP responses to gesture-speech conditions (Kelly et al., 2004, p. 256) *37*
Figure 6.7 Experiment 1: ERP responses to congruous and incongruous gestures according to median decision times (Wu & Coulson, 2005, p. 660). *38*
Figure 6.8 Experiment 2: Grand averaged ERP responses to the onset of highly and moderately congruent gestures (Wu & Coulson, 2005, p. 663). *39*
Figure 6.9 Comparative ERP responses to gesture-speech congruency conditions (Özyürek et al., 2007, p. 611) *40*
Figure 6.10 ERP waveforms illustrating neural responses to congruent and incongruent gestures (Cornejo et al., 2009, p. 47) *42*
Figure 6.11 Brain activation patterns for different types of co-speech gestures (Steines et al., 2021, p. 5) *46*
Figure 6.12 Brain activation patterns in four experimental conditions (Straube et al., 2011, p. 529) *47*
Figure 6.13 Brain activation in response to dominant gestures vs. grooming (Holle et al., 2008, p. 2018) *49*
Figure 6.14 Brain activation in response to subordinate gestures vs. grooming (Holle et al., 2008, p. 2019) *50*

Figure 6.15 Neural activity related to processing speech with and without beat gestures (Hubbard et al., 2009, p. 1032) *51*
Figure 7.1 Comparative analysis of contextual correlation for gestural and facial/vocal signals in apes (Pollick & de Waal, 2007, p. 8187) *55*
Figure 7.2 Hand preference patterns in reaching task (a) and manual gesture task (b) (Meunier et al., 2013, p. 184) *58*
Figure 7.3 Call response patterns of titi monkeys to predatory threats (Cäsar et al., 2013, p. 2) *60*
Figure 7.4 Distribution of inter-call durations in bonobo vocal exchanges (Levréro et al., 2019, p. 2) *63*
Figure 7.5 Rhythmic dynamics of neonatal lip-smacks (Morrill et al., 2012, p. 562) *65*
Figure 7.6 Rhythmic dynamics of juvenile lip-smacks (Morrill et al., 2012, p. 562) *65*
Figure 7.7 Rhythmic dynamics of adult lip-smacks (Morrill et al., 2012, p. 561) *66*
Figure 7.8 Scatterplots of eye-blink measures in primates by body size and group size (Tada et al., 2013, p. 7) *67*
Figure 7.9 Comparison of mean blink rates between diurnal and nocturnal primates (Tada et al., 2013, p. 5) *68*

List of Tables

Table 7.1 Order of signal combinations produced by adult male and female chimpanzees (Crockford & Boesch, 2005, p. 410) *62*
Table 7.2 Socio-demographic influences on vocal sharing in bonobos (Levréro et al., 2019, p. 4) *63*

List of Abbreviations

AC – Auditory Cortex

BA – Brodmann Area

BOLD – Blood Oxygen Level Dependent

CSG – Co-Speech Gestures

EEG – Electroencephalography

ERP – Event-Related Potential

fMRI – Functional Magnetic Resonance Imaging

IFG – Inferior Frontal Gyrus

IFGOp – Inferior Frontal Gyrus, Opercular part

IFGTr – Inferior Frontal Gyrus, Triangular part

IPL – Inferior Parietal Lobule

LIFG – Left Inferior Frontal Gyrus

LPC – Late Positive Component

MR – Magnetic Resonance

MTG – Middle Temporal Gyrus

NSL – Norwegian Sign Language

PAC – Primary Auditory Cortex

POp – Pars Opercularis

PT – Planum Temporale

PTr – Pars Triangularis

SNR – Signal-to-Noise Ratio

STG – Superior Temporal Gyrus

STS – Superior Temporal Sulcus

1. Introduction

Communication guides people through the social world, facilitating the exchange of knowledge and the expression of ideas and emotions. In face-to-face interaction, information is multimodal as it is carried out not just through words but also through non-verbal cues like gestures. Gestures, encompassing various hand and body movements, are essential in the communication process, providing cues beyond what words can express (Hans & Hans, 2015). This implies that even though a lot of information is conveyed through speech, gestures can, for instance, provide additional insights into the actual emotions or intentions behind the speaker's words.

The interest in gestures as a research field gained significant momentum with Goldin-Meadow & Mylander's (1983) study on children's gestures. Following their influential work, an increasing number of researchers started to take a keen interest in exploring and understanding gestures in a broader sense, revealing that gestures are more than mere expressions; they also provide insights into cognitive processes, affecting learning and language acquisition (Krahmer & Swerts, 2007; Gluhareva & Prieto, 2017; Goldin-Meadow, 2018; Andrä et al., 2020). Building on this foundation, recent research has shifted its focus to investigate the connection between gestures and speech at a neural level, revealing the specific neural networks within the brain that process both speech and gestures (Steines et al., 2021; Hubbard et al., 2009; Holle et al., 2008). Methodologies used in neuroimaging studies, such as functional Magnetic Resonance Imaging (fMRI) and Electroencephalography/Event-Related Potential (EEG/ERP), have significantly deepened the understanding of how speech and gestures are intricately linked and processed in the brain (Willems et al., 2009; Holle et al., 2008; Green et al., 2009; Özyürek et al., 2007; Kelly et al., 2004). These studies collectively demonstrate that speech-gesture integration engages a complex network in the brain, highlighting the pivotal role of multimodal communication in human interaction.

Despite these advancements, a notable gap still remains in fully understanding how these two modes of communication interact at a neural level. Therefore, this literature review aims to synthesize and critically analyse the existing research on the neural processing of speech and gestures. By exploring the findings of various studies, the study seeks to shed light on how the brain processes these forms of communication, exploring their potential evolutionary implications.

1.1. The Present Study: Research Questions

To meet the objective of the study and to provide a clear description of the relationship between speech and gestures at a neural level, the following three questions will be answered:

1. What are the key brain regions and neural networks involved in the processing and integration of speech and gesture?

The first question seeks to identify the neural networks that underlie the simultaneous processing of speech and gestures, with an emphasis on understanding the shared and distinct brain regions involved. Through a comprehensive review of various studies employing fMRI and EEG/ERP methodologies, the present study aims to identify the specific areas of the brain involved in the multimodal integration of verbal and non-verbal communication.

2. How do different types of gestures engage distinct neural mechanisms in relation to speech?

The second question aims to uncover the variations in neural processing related to different types of gestures when presented in conjunction with speech. Specifically, through the analysis of multiple fMRI studies, the thesis aims to determine whether various co-speech gestures elicit distinct brain activation patterns. The goal is to elucidate how these gestures enhance language comprehension and communication process. By identifying specific neural activations for each type of gesture, the study aims to provide insights into the underlying mechanisms that support multimodal communication.

3. Can the findings from existing studies shed light on the potential evolutionary origins of gesture and speech integration in the brain?

The last question investigates the broader evolutionary implications of speech-gesture integration, taking into account studies on non-human primates to discern whether language might have originated from a gestural, vocal, or multimodal foundation.

1.2. Overview of the Thesis

The current thesis consists of eight chapters. Chapter 2 focuses on the existing research on gestures. Specifically, it delves into the various definitions scholars have proposed and outlines the diverse classification systems of gestures, as well as their functions and characteristics in both human and non-human primate communication. Chapter 3 delves into the relationship between gestures and speech, examining how they interact and complement each other. It discusses the competing theories, the systematic relationship between these two modes of communication, and their role in language acquisition and non-human primate communication. Chapter 4 discusses the brain regions involved in processing speech and gestures, detailing their specific functions and how they support multimodal communication. Chapter 5 outlines the methodological approaches used to study the neural integration of speech and gestures, focusing on the application and insights gained from non-invasive brain imaging techniques: fMRI and EEG/ERP. Chapter 6 delves into the neural interplay between speech and co-speech gestures, drawing on evidence from various fMRI and EEG/ERP studies. It examines the specific brain regions involved in processing speech and gestures and discusses how these regions exhibit distinct responses to different types of gestures. Chapter 7 examines the evolutionary origins of language by analysing different theories and their supporting evidence, focusing on the roles of vocal and gestural communication. More specifically, it discusses various studies of non-human primates to illustrate how their communicative systems might reflect the evolutionary underpinnings of human language. Lastly, chapter 8 provides a comprehensive summary of the findings and a conclusion based on the research questions raised in the study. It concludes with acknowledging thesis limitations and providing some suggestions for future research.

2. Gestures

This chapter focuses on gestures, exploring the diverse scholarly perspectives on their definition and examining the various classification systems that highlight their functions and characteristics in both humans and non-human primates.

2.1. Definition of Gestures

Over the years, various studies have examined the role of gestures (McNeill, 1992; Ekman & Friesen, 1969; Kendon, 1988). However, despite many scholars conducting thorough research on the nature of gestures, there is still no widely agreed-upon understanding or definition of what they truly mean. In a broad sense, gestures can be described as a form of non-verbal communication involving movements of specific parts of the body, such as "the head, hands, arms, and legs […] to convey specific messages" (Hişmanoğlu & Hişmanoğlu, 2008, p. 168). These 'specific messages' revolve around cultural differences and societal norms; what is defined as gestures in one culture may be differently interpreted in another (Hişmanoğlu & Hişmanoğlu, 2008, p. 168). This description, therefore, raises many other questions, which demonstrate the complexity of gestures. To provide a more precise definition or view on gestures, McNeill (1992, p. 1) described them as movements of the arms and hands closely related to speech, emphasizing that they are not just supplements to language but rather an integral part of communication. Furthermore, Hostetter & Alibali (2008, p. 495) provided yet another view of gestures; they described them as a form of embodied cognition, viewing them as a means through which language and cognition are physically embodied. These varying definitions show that gestures are a complex and sophisticated phenomenon that can have multiple roles, depending on the perspective from which they are viewed. Despite the varying views on gestures, the current study adopts McNeill's (1992) perspective, perceiving gestures as hand movements that complement and enhance speech.

2.2. Classification of Gestures

Not only do scholars hold varying views in terms of the definition of gestures, but they also propose diverse classification systems.

According to Ekman & Friesen (1969, pp. 62-92), gestures fall into five types: emblems, illustrators, affect displays, regulators, and adaptors.

1. Emblems – are body movements that can convey specific phrases, words, or meanings with no additional context (Ekman & Friesen, 1969, pp. 63-65). To be more precise, emblems carry explicit meanings and can substitute for verbal communication, especially under challenging auditory environments (Ekman & Friesen, 1969, p. 64). For example, moving the head up and down can indicate agreement or affirmation.
2. Illustrators – are closely tied to spoken words and serve to visually represent or highlight a specific word or idea (Ekman & Friesen, 1969, p. 68). For instance, using fingers to count or emphasize numerical information during speech.
3. Affect displays – express emotional states, primarily through facial expressions, but can also include body postures and movements (Ekman & Friesen, 1969, pp. 70-71). For example, crying indicates sadness.
4. Regulators – play a crucial role in guiding and managing interactions during communication. These involve eye movements, head nods, and body language that indicate turn-taking (Ekman & Friesen, 1969, p. 82). For example, raising a hand while someone else is speaking can indicate the desire to speak without interrupting verbally.
5. Adaptors – unconscious movements, specifically involving unintentional touching of oneself, others, or objects during interaction. They serve as a means for individuals to satisfy personal needs or to adapt to various circumstances, especially in response to stress or anxiety (Ekman & Friesen, 1969, pp. 84-92). For example, biting nails can signify that a person is distressed or nervous.

Later, another classification system for gestures was introduced by Kendon (1988) that served as the foundation for the subsequent work by McNeill (1992) as he expanded and refined this categorization, leading to the development of "Kendon's Continuum" (McNeill, 1992, p. 37). In this continuum, gestures are divided into five types based on their relationship to speech and their linguistic properties (McNeill, 1992, p. 37). The five types are defined as follows:

1. Gesticulations – spontaneous hand or arm movements closely connected to speech, typically accompanying verbal communication (McNeill, 1992, p. 37). The gestures analysed in the current study fall under the category of gesticulations. For example, a speaker might stretch their arms wide to emphasize the size while saying: ‘I have a big apple’.
2. Language-like gestures – resemble gesticulations in that they accompany speech, but they can also serve a grammatical function or replace certain parts of the speech, such as adjectives or adverbs (McNeill, 1992, p. 37). For example, saying: “the parents were all right, but the kids were [gesture]” (McNeill, 1992, p. 37).
3. Pantomimes – depict objects or actions, but they do not necessarily require speech for context (McNeill, 1992, p. 37). For instance, mimicking the act of drinking with the hands.
4. Emblems – socially recognized gestures that carry specific meanings (McNeill, 1992, pp. 37-38). For example, the ‘thumbs up’ gesture indicates ‘good’ or ‘yes’ in many countries.
5. Sign Language – differs from previously described gesture types in a way that it is a complete language system encompassing all the elements and rules that define any spoken language (McNeill, 1992, p. 38). For example, Norwegian Sign Language (NSL).

Despite the various existing classifications of gestures, the present study adopts McNeill’s (1992) framework for categorizing gestures based on their functions and characteristics. The four categories of gestures include iconic, metaphoric, deictic, and beat gestures (McNeill, 1992, pp. 12-18).

1. Iconic gestures – mimic the physical characteristics of the object or action being described (McNeill, 1992, pp. 12-13). It can include shaping the hands to represent some object (e.g., a ball, an apple, etc.)
2. Metaphoric gestures – represent abstract concepts or ideas (McNeill, 1992, p. 14). For instance, spreading hands apart from each other to indicate a huge idea.
3. Deictic gestures – indicate locations in space, often using pointing gestures (McNeill, 1992, p. 18). For instance, if a person asks for directions, one can show with the hand that he/she should go straight and turn to the left.

4. Beat gestures – simple, fast movements of the hands reflecting the rhythm of speech or emphasizing important aspects of it (McNeill, 1992, p. 15). For instance, waving a hand briefly up and down with stressed words.

The main reason for adopting McNeill's (1992) categorization in the present thesis revolves around its specific focus on hand gestures that are deeply linked to spoken language. As mentioned in the 2.1 Definition of Gestures section, McNeill (1992) considers gestures more than supplementary to speech as they can also enhance verbal communication. This perspective aligns with the objectives raised in the study, which seek to explore how gestures and speech influence each other and how this multimodal information is processed in the brain.

2.3. Gestures of Great Apes

Similar to humans, non-human primates, more specifically bonobos, chimpanzees, gorillas, and orangutans, use a wide range of gestures to communicate with others (Pollick & de Waal, 2007; Nishida et al., 2010; Pika et al., 2003; Call & Tomasello, 2007). These gestures are not only used across a variety of contexts but are also employed with notable flexibility, allowing a single gesture to convey different meanings depending on the situation (Graham & Hobaiter, 2019, p. 372). Just as humans use non-verbal cues to complement and enhance verbal communication, non-human primates rely on gestures to convey various messages and emotions. It can include signals for initiating play, demonstrating submission or dominance, requesting food, or even as part of mating rituals (Graham et al., 2018; Hobaiter & Byrne, 2014). This diversity in meanings shows that primates' gestures are not just instinctive reactions but rather deliberate actions used to achieve specific goals or, in general, to meet their individual and social needs.

Research has demonstrated significant variability in the gestural repertoires among different great ape species, reflecting adaptations to their specific social and ecological environments (Graham et al., 2022; Pika et al., 2003; Pollick & de Waal, 2007). More precisely, great apes residing in different environments may develop distinct sets of gestures, adapting their communicative behaviors to suit their surroundings. This adaptation thus leads to a great diversity in the gestural repertoires used by apes. Consequently, gestures have been broadly categorized into three types: tactile, auditory, and visual, each employed under specific conditions to achieve particular aims (Call & Tomasello, 2007, p. 21). These gestures are defined as follows:

1. Tactile gestures – involve direct physical contact, such as pushing or grooming (Call & Tomasello, 2007, pp. 21-24).
2. Auditory gestures – encompass actions that produce sound, such as hand clapping or foot stamping, to capture attention and signal alarm (Call & Tomasello, 2007, p. 24).
3. Visual gestures – involve movements that are primarily intended to be seen, such as arm waves or display behaviours. They can convey multiple messages, from signalling readiness for play to asserting dominance or submission within the social hierarchy (Call & Tomasello, 2007, p. 21).

These categories of gestures demonstrate the complex and sophisticated communication strategies employed by non-human primates, revealing notable similarities to human language. In regard to the present thesis, discussing and analysing these gestures provides insights into potential evolutionary pathways that could have led to the development of human language. These aspects will be further discussed in subsequent sections of the study.

3. Gestures and Speech

This chapter delves into the relationship between gestures and speech by exploring contrasting theories on how these two modes of communication may be interconnected or independent. The section then examines their systematic relationship by discussing how gestures and speech interact and complement each other. Furthermore, it investigates the impact of different types of gestures in language learning, highlighting their significance in the development of communication skills. Finally, the section discusses non-human primates' communication, focusing on their vocalizations and gestural usage.

3.1. Competing Theories

The relationship between gestures and speech has sparked considerable debate among scholars, with some saying that gestures do not rely on speech production but instead have their own origin (Butterworth & Beattie, 1978; Hadar et al., 1998), while others consider gestures and speech as interconnected elements of a single communication system (McNeill, 1992). This disagreement led to the formation of two main theories: the Independence Hypothesis and the Interface Hypothesis (Özyürek, 2002, p. 500).

According to the Independence Hypothesis, gestures and speech are viewed as fundamentally separate systems of communication (Butterworth & Beattie, 1978; Hadar et al., 1998; Krauss & Hadar, 1999). Particularly, gestures stem from "spatial representations in working memory", and their creation and use are not directly tied to the verbal aspects of speech production (Özyürek, 2002, p. 500). This means that while a person is speaking, his/her gestures are not just mirroring the words but rather have their own separate origin and purpose in the communicative process. In other words, gestures do not simply accompany speech but rather are a form of communication that originates and operates independently.

In contrast, the Interface Hypothesis suggests that gestures and speech are deeply interconnected and emerge from the same cognitive processes (McNeill, 1992; Kita, 2000). To be more precise, this view supports the idea that there is a cognitive system in which gestures and speech are coactivated, and each can influence the other in the process of communication (Özyürek, 2002, p. 500). It means that when, for instance, a person wants to say something, he/she concurrently expresses this idea not only through speech but also through corresponding physical actions, which in this case are gestures. This view is supported by studies demonstrating how gestures can influence and facilitate language learning, memory retrieval, and cognitive processes (Andrä et al., 2020; Morett, 2014; Gluhareva & Prieto, 2017). Moreover, various neuropsychological studies have provided compelling evidence that gestures and speech activate the same neural networks in the brain (Holle et al., 2008; Steines et al., 2021; Hubbard et al., 2009), suggesting a deep neurological integration between the two. Therefore, this evidence shows that even though these two modes of communication are different in how they are expressed, one visually and the other audibly, they both come from the same cognitive system to provide a cohesive and comprehensive means of communication.

Overall, this section discussed two contrasting theories on the relationship between gestures and speech. While the Independence Hypothesis suggests they are separate systems, the Interface Hypothesis posits deep integration. The present study supports the view of the Interface Hypothesis, suggesting that gestures and speech are not separate but rather integrated and co-dependent processes, and aims to delve further into this connection at a neural level.

3.2. The Systematic Relationship between Gestures and Speech

Following the Interface Hypothesis, which posits a deep connection between speech and gestures, it becomes clear that these two modes of communication are not just related but are intricately linked. Research indicates that this relationship unfolds on three levels: semantic overlap, temporal synchronization, and complementary communication (Willems et al., 2007, pp. 2322-2323). These levels are described as follows:

1. Semantic overlap – speech and gestures often carry overlapping meanings, providing a dual modality of expression that enriches the communicative intent (Willems et al., 2007, p. 2322). For instance, while describing an action verbally, the accompanying hand gestures can visually represent the action, creating a more vivid depiction for the listener.

2. Temporal synchronization – signifies precise coordination in timing between speech and gestures that can be divided into three phases: *stroke*, *preparation*, and *retraction* (Kendon, 1980, p. 212). The *stroke* is the central component, representing the peak of the gestural movement and carrying the core of its semantic meaning (Kendon, 1980, p. 212). The *stroke* can be preceded by the *preparation* phase (when hands are moving to the position of the *stroke*) or followed by the *retraction* phase (when hands are moving away from the *stroke* either to the resting position or back to another *stroke*) (Kendon, 1980, p. 212). The *preparation* phase often starts before the related speech, setting up the context or complementing the upcoming verbal message (Kendon, 1980, p. 212).
3. Complementary communication – gestures can provide information that complements speech; to be more precise, gestures may emphasize certain aspects of spoken language, convey additional details, or even express something that was not explicitly stated (Willems et al., 2007, pp. 2322-2323). For instance, a person might say, ‘He is that taller than me’ while holding a hand above their head to visually demonstrate the height difference. In this way, a person provides precise additional information without directly specifying the exact number of centimeters in a speech.

In essence, these three levels (semantic overlap, temporal synchronization, and complementary communication) not only demonstrate the collaborative role of speech and gestures in the process of communication but also underscore their fundamental importance in cognitive processing and social interaction.

3.3. Gestures and Language Acquisition

Numerous researches have highlighted the critical role of gestures in facilitating language acquisition. In fact, it has been discovered that all types of gestures have a specific impact on the learning process, with studies demonstrating that the incorporation of gestures can be effective in enhancing vocabulary acquisition and communication skills (Andrä et al., 2020). Notably, children who use more gestures in early childhood often develop larger vocabularies, better grammar skills, and stronger language abilities later in life compared to children who use fewer gestures (Rowe et al., 2008). This is thought to be because gestures provide additional contextual information that can help children to make connections between words and their meanings, as well as to infer the meanings of unfamiliar words and concepts (Broaders et al., 2007, pp. 547-548). Therefore, encouraging gestural communication as a part of children's education is an essential and effective way to enhance their linguistic and cognitive development.

However, the impact of different types of gestures on language acquisition can vary. Studies have demonstrated that using iconic gestures can aid children in learning different word classes (Mertens & Rohlfing, 2021, p. 3). To be more precise, when children see an iconic gesture that corresponds to a word they are learning, it can improve their comprehension of that specific word. For instance, if a child sees someone make an eating motion with their hands while saying the word 'eat,' the accompanying gesture representing this action visually aids in an easier understanding of what the verb 'eat' means. Additionally, iconic gestures attract children's attention toward specific aspects of a scene or object, which can enhance their understanding of unfamiliar verbs (Mertens & Rohlfing, 2021, p. 3). For instance, if a parent dances and says, 'I love dancing bachata', the child immediately pays close attention to the action. This focused attention helps the child to more easily remember the meaning of the verb 'dancing' and associate 'bachata' with this specific movement.

Other studies have demonstrated that only deictic gestures play an especially essential role in language development as they tend to serve the purpose of guiding the recipient's focus toward a particular referent (Ramos-Cabo et al., 2019, p. 2). For instance, by pointing to objects and locations, children can draw the attention of their parents and communicate about their needs and interests. Studies on the use of metaphoric gestures showed that these gestures can add additional semantic meaning to speech and can help to clarify or emphasize a particular point (Bernard et al., 2015, p. 4017). For instance, a speaker might use a metaphoric gesture to indicate 'the unity of the company'. To be more precise, the speaker might hold their hands together in a clasping motion to symbolize the coming together of different parts to form a whole, which would add a visual context to the speech and help to convey the idea more effectively to the listener.

Furthermore, studies have indicated that incorporating beat gestures into language acquisition can be especially effective in helping learners acquire new vocabulary and improve their communication skills (Gluhareva & Prieto, 2017; Kushch et al., 2018; Morett & Fraundorf, 2019). Specifically, studies have demonstrated that beat gestures can enhance learners' prosodic skills (Gluhareva & Prieto, 2017), aid in the recall of words (Morett, 2014), predict narrative abilities (Vilà-Giménez et al., 2021), and significantly impact perceived prominence of specific words (Krahmer & Swerts, 2007). This evidence shows that even though beat gestures do not directly indicate or represent specific objects or ideas like other types of gestures, they still significantly contribute to language acquisition.

In essence, research highlights the diverse impact of gestures on language acquisition. Iconic, metaphoric, deictic, and beat gestures each contribute uniquely to language comprehension, vocabulary development, and communication skills. This evidence not only validates the importance of gestures in language learning but also highlights their integral role in cognitive development.

3.4. Gestures and Vocalizations in Non-Human Primates

Non-human primates, similar to humans, use gestures and vocalizations to communicate with each other. Research demonstrates that these interactions are complex and multifaceted, sharing a resemblance to human communication (Crockford & Boesch, 2005; Coye et al., 2016; Pollick & de Waal, 2007; Prieur et al., 2020; Taglialatela et al., 2011). Specifically, studies have shown that similar to human interactions, primates exhibit a sophisticated level of intent in their communicative behaviours (Fröhlich et al., 2019; Crockford et al., 2015; Maille et al., 2012; Graham & Hobaiter, 2019). Intentional communication in primates is evidenced by gestures that are not just spontaneous expressions but are instead directed with purpose and deliberation toward achieving specific social goals. These gestures can be characterized by four criteria: directedness, attentional stance, response waiting, and persistence or elaboration (Graham & Hobaiter, 2019, pp. 373-374).

1. Directedness – emphasizes that gestures in primate communication are not random but are directed with precision toward specific recipients (Graham & Hobaiter, 2019, p. 373). For instance, a chimpanzee can scratch its chest in front of a specific peer to indicate that they wish to be groomed by them.
2. Attentional stance – involves the signaller being attuned to whether the recipient is paying attention. To be more precise, gestures are made within the recipient's visual field, often accompanied by eye contact to ensure readiness for communication (Graham &

Hobaiter, 2019, p. 373). For example, starting to gesture only when the signaller sees that the recipient is paying attention.

3. Response waiting – indicates that the signaller must wait for the response from the recipient (Graham & Hobaiter, 2019, p. 373). For example, pausing after a gesture to observe the recipient.

4. Persistence or elaboration – is demonstrated by the signallers' continued efforts to communicate until the desired outcome is achieved (Graham & Hobaiter, 2019, p. 373). For instance, pointing to some food multiple times until the recipient becomes alerted.

The study by Maille et al. (2012) further supports this evidence by showing how red-capped mangabeys adjusted their gestural communication in response to human attention. In their experiment, monkeys were presented with a food item while the experimenter exhibited five different attentional states: eyes open, eyes distracted (looking away), eyes closed, head turned away, and body turned away. The findings revealed that red-capped mangabeys adjusted their behaviors in response to the attentiveness of the experimenter; more specifically, they gestured more frequently and promptly when the experimenter's body and head were oriented towards them, but showed no significant response to variations in eye state (open, distracted, and closed). By modifying their gestures based on the experimenter's body and head orientation, the mangabeys demonstrated what could be considered a form of intentional communication aimed at enhancing the effectiveness of their gestural request for food.

In contrast, the study by Bourjade et al. (2014) examined olive baboons and found that these primates do consider the state of human eyes when adjusting their communicative gestures. This study specifically investigated whether baboons could differentiate between an experimenter's eyes being open versus closed and whether they would modify their communicative strategies based on the experimenter's overall visual attention. In the experiment, researchers tested the reactions of baboons under four conditions: the experimenter's eyes open, eyes closed, back turned, and the experimenter absent ('Out'). The results indicated that olive baboons significantly increased their use of visual gestures when the experimenter's eyes were open compared to other conditions: when the eyes were closed, when the experimenter's back was turned, or when the experimenter was absent. Furthermore, when the experimenter was not visually attentive, baboons increased their production of auditory gestures, such as banging on surfaces, to draw visual attention. These findings suggest that, unlike red-capped mangabeys observed in Maille et al.'s (2012) study, olive baboons respond to detailed visual cues, such as eye movements, and can adjust their communicative methods to increase the chances that their signals are being noticed.

Other studies have revealed that primates can intentionally inform others not only through gestures but also through vocal signals (Fröhlich et al., 2019, p. 1814). For instance, the study by Crockford et al. (2015) demonstrated that wild chimpanzees use vocalizations intentionally to alert others about external threats. In their experiment, researchers exposed chimpanzees to two types of vocalizations: 'alert hoos' (indicating threat) and 'rest hoos' (indicating resting periods) and observed their reactions. Their findings revealed that the chimpanzees responded differently to two types of vocal calls. Specifically, when exposed to 'alert hoos', chimpanzees demonstrated a significant increase in search behavior and attention toward the source of the sound. This behavior included more frequent scanning of the area, indicating that they were alert and possibly looking for the threat the vocalization typically signifies. In contrast, the 'rest hoos' did not elicit the same level of alertness or search behavior, suggesting that the chimpanzees were able to discern the calls and tell the contextual difference: one signaling a potential threat and the other being a non-alerting communication.

Beyond being intentional, the vocal and gestural communication of non-human primates is also highly referential (Fröhlich et al., 2019, pp. 1814-1816). To be more precise, referential signals in primate communication play an essential part in their social interactions and survival, signaling the presence of predators, food sources, or social partners (Cäsar et al., 2013; Seyfarth & Cheney, 2003; Hobaiter & Byrne, 2014; Pika & Mitani, 2006). The study by Pika and Mitani (2006) demonstrated compelling evidence of wild chimpanzees' ability to use gestures referentially, similar to humans' use of pointing to direct attention. Specifically, their observations revealed that chimpanzees employ 'directed scratching' to communicate their grooming preference to others. More precisely, chimpanzees aimed to scratch a specific part of their body to indicate where they wished to be groomed by their grooming partner, suggestive of the ability to use gestures with a communicative intention and referential meaning.

Moreover, Slocombe & Zuberbühler's (2005) study revealed that chimpanzees use referential vocal signals, specifically 'rough grunts', to convey information about food to other peers. In their experiment, researchers exposed a chimpanzee to prerecorded 'rough grunts' that were distinctively associated with discovering high-value (bread) and low-value (apples) foods. With this approach, researchers aimed to determine whether the chimpanzee could differentiate foods based solely on these specific vocalizations without relying on visual or olfactory (smelling) cues. The findings demonstrated that when the chimpanzee heard the grunt associated with bread, it exhibited increased search behavior, indicating anticipation of finding this preferred food. In contrast, the grunt associated with apples did not evoke the same response. These results suggest that primates use their vocalizations referentially, particularly to refer to specific objects or events in their surroundings.

Furthermore, research indicates that primates' gestures exhibit a high level of iconicity (visually resemble their referents), allowing them to convey specific messages about their desires, surroundings, or emotional states (Douglas & Moscovice, 2015; Genty & Zuberbühler, 2014). In the study by Douglas & Moscovice (2015), researchers aimed to investigate the communicative functions of two specific gestures: 'foot-pointing' and 'hip shimmy' (rapid lateral movements of the hips) among wild female bonobos, focusing on how these gestures could represent actions or objects related to their social interactions, particularly genito-genital rubbing. Their findings demonstrated that bonobos' use of the 'hip shimmy' gesture was highly iconic, as it directly mimicked the physical act involved in genito-genital rubbing. These results provide significant evidence that non-human primates' gestural communication is not only referential but also highly iconic.

In essence, various studies demonstrate the complex nature of non-human primates' gestural and vocal communication. Specifically, it is evident that the communicative behaviors of primates are not mere instinctual reactions but rather intricate processes that are intentional, referential, and even include iconicity. Through the use of gestures and vocalizations, non-human primates can convey detailed information about their needs, desires, and social relations, which are similar to human communication.

4. Brain

This chapter explores aspects of functional brain anatomy, specifically focusing on areas integral to the processing of speech and gestures. Understanding the roles of these brain areas is essential for comprehension of the neurological interconnections between these two modes of communication.

4.1. Overview of Brain Anatomy

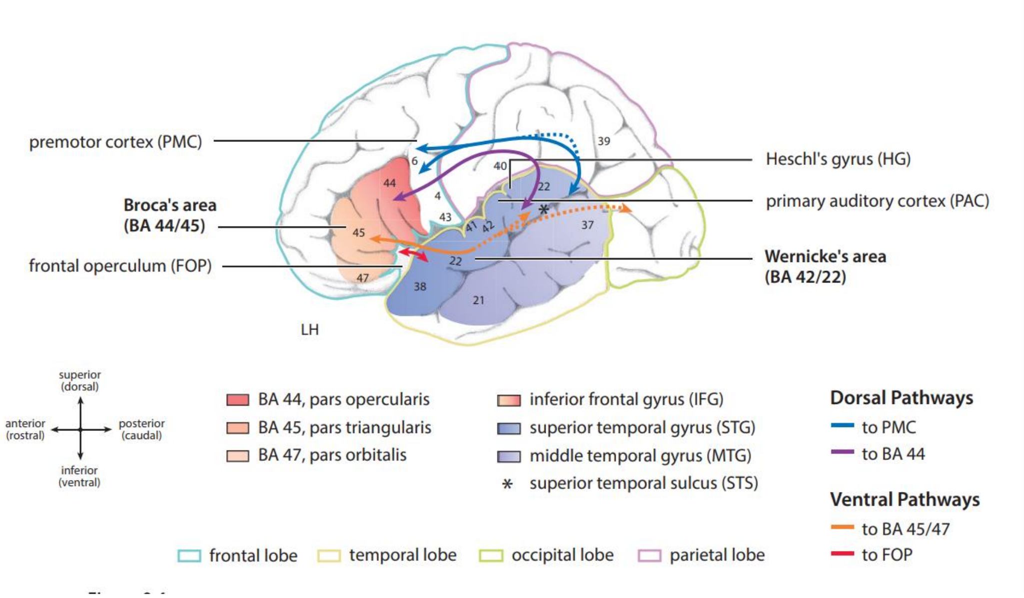


Figure 4.1 Neuroanatomy of language processing (Friederici, 2017, p. 6)

The figure above illustrates a simplified representation of the human brain, highlighting the specific regions involved in language processing. Specifically, it marks IFG, STG, MTG, STS, PAC, and Broca’s area, which will be further discussed in the following subsections. Additionally, the illustration outlines four major lobes of the brain: frontal, temporal, occipital, and parietal, providing a clearer visualization of the location of these areas.

4.1.1. Broca’s Area and the Inferior Frontal Gyrus

One of the most important brain regions involved in speech processing is Broca's area (see Figure 4.1). This region, located in the left hemisphere of the brain within the Inferior Frontal Gyrus (IFG) (Skipper et al., 2007, p. 261), plays an essential role in the formation of words and the construction of sentences (Brown & Yuan, 2018, p. 1214). To be more precise, during speech, Broca's area becomes active, facilitating the articulation and expression of ideas. However, recent research shows that Broca's area has a broader range of functions than previously known, encompassing both speech and gestures (Gentilucci et al., 2006; Skipper et al., 2007; Rizzolatti & Arbib, 1998), indicating its multimodal nature.

Expanding beyond Broca's area, the Inferior Frontal Gyrus, or IFG, is a broader region with a diverse set of language-related functions (see Figure 4.1). While it shares the role of language production with Broca's area, the IFG also supports language comprehension, processing the meanings of words and sentences (Ishkhanyan et al., 2020). Moreover, research indicates that the IFG is involved not only in speech processing but also in gestural communication (Willems et al., 2007; Steines et al., 2021; Straube et al., 2011; Kircher et al., 2009; Nagels et al., 2013). Specifically, the left IFG shows increased activity when there is a semantic mismatch between the context of a sentence and the accompanying gesture or word (Willems et al., 2007).

4.1.2. The Auditory Cortex and Planum Temporale

While Broca's area and IFG help to form words and sentences, the Auditory Cortex (AC) is primarily responsible for processing auditory information (Näätänen et al., 2001). This brain area is located within the temporal sulcus, primarily in the medial part of Heschl's gyrus (Formisano et al., 2003) (see Figure 4.1, Figure 4.2). It receives input from the ears and is responsible for deciphering various sounds, such as pitch, loudness, and timbre (Pontevendra, 2022). This region is crucial not only for the basic perception of sound but also for more complex auditory functions, such as interpretation of music and speech (Pontevendra, 2022). Furthermore, recent findings suggest that the Auditory Cortex, specifically the Bilateral Non-primary Auditory Cortex, also plays a role in integrating auditory information with visual cues (gestures), revealing its multimodal nature (Hubbard et al., 2009).

Adjacent to the Auditory Cortex lies the Planum Temporale (PT), a region that is deeply involved in the processing of language and speech (Stigler & McDougle, 2013, pp. 254-255) (see Figure 4.2). Often associated with Wernicke's area, the PT plays a pivotal role in the comprehension of spoken and written language, as well as in the complex task of lexical processing, which includes the recognition and understanding of words and their meanings (Stigler & McDougle, 2013, pp. 254-255). Moreover, recent studies have elucidated the PT's role in integrating auditory and visual information, highlighting its capacity to synchronize the prosodic elements of speech with corresponding rhythmic gestures (Hubbard et al., 2009).

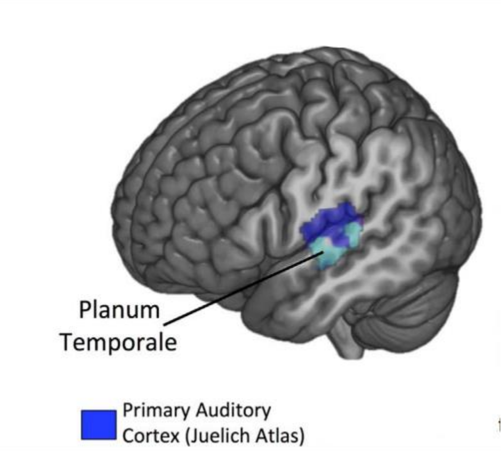


Figure 4.2 Anatomical representation of the planum temporale and primary auditory cortex (Murphy et al., 2017, p. 25)

4.1.3. The Superior and Middle Temporal Gyrus

Another important brain region is the Superior Temporal Gyrus (STG), which is primarily involved in processing auditory information and understanding language (Howard et al., 2000, p. 79) (see Figure 4.1). More specifically, STG plays an essential role in decoding the complex sounds of speech by identifying phonetic elements such as vowels and consonants and adjusting them based on context (Bhaya-Grossman & Chang, 2022). Furthermore, it has been discovered that STG also plays a role in non-verbal communication by matching speech sounds with gesture movements (Marstaller & Burianová, 2014).

The Middle Temporal Gyrus (MTG), on the other hand, is involved in language processing and semantic cognition (see Figure 4.1) (Briggs et al., 2021). Therefore, it plays a significant role in how the brain processes and comprehends complex language constructs. The MTG is integral to the understanding of not just the literal meanings of the words but also their contextual and conceptual relationships (making sense of metaphors and understanding abstract ideas conveyed through language) (Proverbio et al., 2009, p. 2). In addition, it has been shown that MTG is especially responsive to the semantic interaction between gestures and speech, underscoring its role in multimodal integration at the semantic level (Willems et al., 2009; Kircher et al., 2009).

In essence, STG and MTG work in conjunction with the previously mentioned Broca's area and IFG to ensure that communication is coherent and efficient.

4.1.4. The Inferior Parietal Lobule

The Inferior Parietal Lobule (IPL) is a region in the human brain known for its role in integrating sensory information across various modalities (see Figure 4.3). Specifically, it is essential for spatial sense, navigation, and a range of cognitive processes (Caspers et al., 2013; Numssen et al., 2021). Recent research further extends the understanding of the IPL's functionality by highlighting its involvement in gestural communication (Holle et al., 2008). To be more precise, IPL shows activation during the process of co-speech gestures, indicating its role in multimodal communication (Holle et al., 2008).

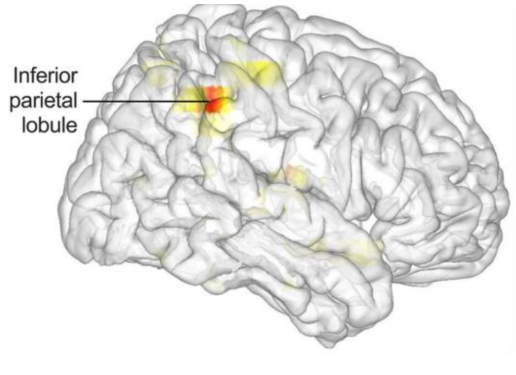


Figure 4.3 Anatomical representation of the inferior parietal lobule (Dary et al., 2023, p. 2948)

To conclude, this section delved into the critical brain regions playing an essential role in both speech and gesture processing. These regions, including Broca's area, IFG, AC, PT, STG, MTG, and IPL, not only serve language-related functions but also demonstrate their multimodal nature by processing gestures. This neural interaction forms the foundation for a better comprehension of the complex relationship between speech and gestures, which will be further explored in subsequent sections.

5. Methodological Approaches in Studying Neural Integration

This chapter explores various methods used in neuroscience research that provide insight into how the brain integrates and processes gestures and speech. It discusses the crucial role and limitations of advanced brain imaging techniques, particularly Functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG/ERP). These methods are essential for providing in-depth insights into brain functions, enabling researchers to observe and analyse neural activity.

5.1. Functional Magnetic Resonance Imaging (fMRI)

One of the most commonly used neuroimaging techniques to investigate brain function is Functional Magnetic Resonance Imaging (fMRI). It is a non-invasive technique that allows researchers to observe brain activity as it responds to various stimuli (Hay et al., 2022; Arthurs & Boniface, 2002). This method relies on Magnetic Resonance (MR) signals, which are responsive to alternations in blood flow, including the Blood Oxygenation Level Dependent (BOLD) signal (Buxton, 2013, p. 2). To be more precise, when a specific brain area becomes more active due to some stimulus, it requires more oxygen to support its increased activity, leading to stronger blood flow to that area. Magnetic resonance (MR) detects these changes in blood oxygenation levels (BOLD), which thereby allows researchers to pinpoint activated brain regions (Buxton, 2013) (see Figure 5.1). This method is especially useful for researchers investigating how the brain processes gestures and speech as it provides a non-invasive yet detailed view of brain activity. It helps to map out the specific brain regions involved in this process and enhances the understanding of how different areas of the brain collaborate during various cognitive tasks.

Even though fMRI provides valuable insights into brain functions, it also has its limitations. Its high spatial resolution allows for precise localization of brain activity (Detre & Wang, 2002, p. 621), which makes it exceptionally useful for identifying specific brain regions involved in processing both verbal and non-verbal communication. However, it exhibits lower temporal resolution, limiting its ability to capture the precise timing of these neural processes (Detre & Wang, 2002, p. 621).

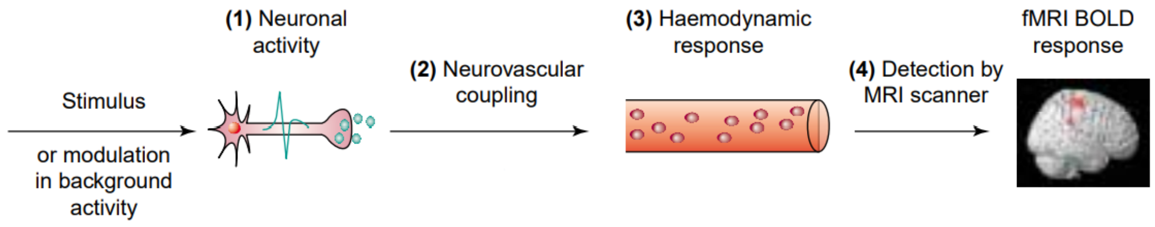


Figure 5.1 Detection of neural response to a stimulus with fMRI BOLD signal (Arthurs & Boniface, 2002, p. 28)

Figure 5.1 above depicts the underlying principles of fMRI in a stepwise manner. The process begins with a stimulus that activates neurons in a specific brain area (1). The active neurons then signal nearby blood vessels, indicating an increased demand for oxygen (neurovascular coupling) (2). In response to the signals from the active neurons, the blood flow increases to supply the necessary oxygen to the area (3). Finally, the MRI scanner detects these changes using BOLD contrast (4), depicted in highlighted areas in ‘fMRI BOLD response’, indicating regions of increased neural activity.

5.2. Electroencephalography/Event-Related Potential (EEG/ERP)

Electroencephalography (EEG), on the other hand, offers a different approach to neuroscientific research by capturing the brain’s electrical activity through electrodes positioned on the scalp. Specifically, this technique involves placing small metal electrodes on the surface of the head to detect tiny electrical charges generated by brain cell activity (Fleury et al., 2023, p. 3). These metal electrodes are strategically placed at specific locations corresponding to different brain regions, thus allowing for precise monitoring of brain activity (Subha et al., 2010, p. 195) (see Figure 5.2).

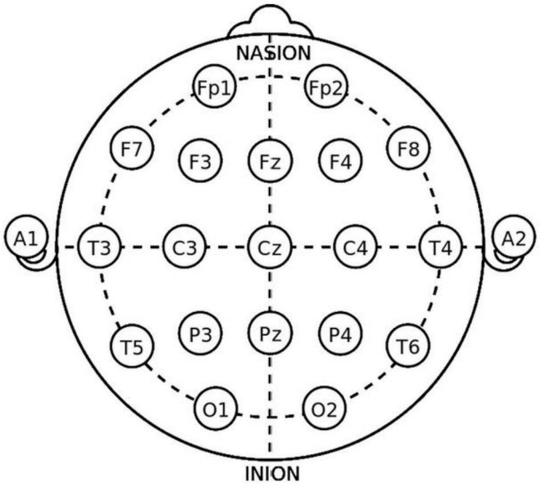


Figure 5.2 Electrode placement based on the international 10-20 system (Rojas et al., 2018, p. 2)

The raw EEG data obtained from electrodes is a combination of rhythmic waves known as neural oscillations (Morales & Bowers, 2022, p. 1). These waves are divided into five frequency bands: delta, theta, alpha, beta, and gamma, each varying frequency, amplitude, and shape (see Figure 5.3) (Campisi & La Rocca, 2014, pp. 783-784). These EEG frequency bands can help deduce the general mental states of individuals, such as alertness or relaxation (Coulson, 2007, pp. 402-403). However, isolating precise brain responses to specific stimuli from ongoing EEG data is challenging due to various types of noise; therefore, to navigate this complexity, researchers frequently use Event-Related Potentials (ERP) to extract specific and meaningful information (Coulson, 2007, p. 403). To be more precise, ERP is derived from the outgoing EEG data by focusing on the brain's electrical response that is temporally aligned with a specific stimulus (Sur & Sinha, 2009, p. 70). This process allows researchers to isolate and study specific neural responses while reducing the impact of external noise (see Figure 5.4).

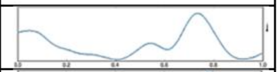
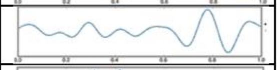
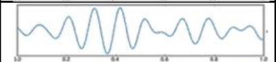
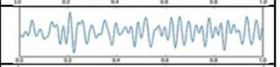
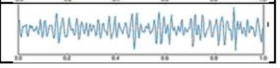
Waves	Frequency bands (Hz)	Behaviour Trait	Signal Waveform
Delta	0.3 – 4	Deep sleep	
Theta	4 – 8	Deep Meditation	
Alpha	8 – 13	Eyes closed, awake	
Beta	13 – 30	Eyes opened, thinking	
Gamma	30 and above	Unifying consciousness	

Figure 5.3 Representation of frequency bands of EEG signal (Oon et al., 2018, p. 2)

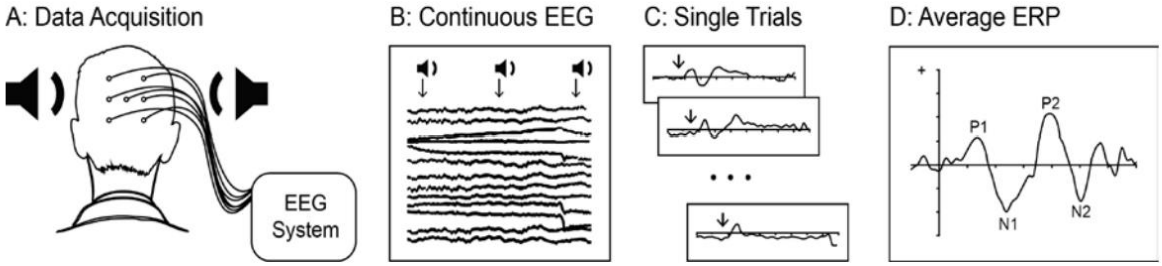


Figure 5.4 Representation of ERP acquisition (Key, 2016, p. 142)

Figure 5.4 illustrates the process of capturing and analyzing EEG and ERP data. Panel A demonstrates the initial data acquisition phase, where electrodes are placed on the scalp to record the brain's electrical activity triggered by external stimuli. Panel B depicts a segment of EEG data characterized by its rhythmic waveform, representing the brain's electrical activity captured by the electrodes over a period. In Panel C, individual EEG trials are extracted from the continuous data, with each trial corresponding to the brain's electrical response to a specific stimulus. These EEG responses are then averaged, meaning that researchers remove the noises and other external influences in order to get a clear representation of the brain's activation. Lastly, Panel D depicts the final ERP waveform, which includes different ERP components: P1, N1, P2, and N2, reflecting brain responses to the stimulus.

Furthermore, it is important to mention that, similar to fMRI, the EEG/ERP method has advantages and limitations. It can provide a direct measure of neuronal activity with very high temporal resolution (Detre & Wang, 2002, p. 621), more specifically, with milliseconds precision (Nunez & Srinivasan, 2009, p. 3). This exceptional temporal resolution makes the EEG/ERP method especially beneficial for researchers investigating the dynamic changes in brain activity during cognitive tasks such as language processing and gesture recognition. For instance, it can help researchers to reveal the precise timing of brain responses to specific stimuli, which provides a crucial insight into neural processing. However, compared to fMRI, its spatial resolution is minimal, meaning it is less effective in pinpointing the exact locations of brain activity (Detre & Wang, 2002, p. 621).

After describing the general methodology of EEG and ERP, it is crucial to delve into the specific ERP component that is particularly relevant to the present study. The N400 component, first discovered by Kutas & Hillyard (1980), is one of the key ERP components found in studies investigating the relationship between gestures and speech at a neural level (Kelly et al., 2004; Wu & Coulson, 2005). This component is a negative-going wave that peaks approximately 400 milliseconds after a specific stimulus, such as a word or a gesture (Key et al., 2005, p. 197). It is specifically associated with processing and reacting to semantic incongruities in language, responding more strongly to incongruous words within a given sentence context (Key et al., 2005, p. 197). However, numerous studies employing EEG/ERP methodology showed that the N400's influence is not confined only to linguistic elements; it also encompasses gestures (Kelly et al., 2004; Wu & Coulson, 2005; Cornejo et al., 2009). Specifically, these studies revealed that the N400 amplitude is greater for words or gestures that do not semantically align with their context than those that do, indicating that the brain is identifying a semantic inconsistency. This characteristic makes the N400 a valuable marker for investigating how the brain processes semantic information during multimodal communication. A more detailed discussion of these studies will be provided in the 6.1.2 EEG/ERP Studies section.

To conclude, fMRI and EEG/ERP are two distinct non-invasive neuroimaging techniques, each playing a crucial role in investigating how the brain processes speech and gestures. fMRI offers an exceptional spatial resolution, which allows researchers to identify the precise location within the brain where speech and gesture processes occur. This capability makes fMRI an excellent tool for exploring 'where' these cognitive processes occur and for understanding the brain's functional connectivity involved in multimodal communication. In contrast, EEG/ERP provides profound temporal resolution, capturing the brain's electrical activity with milliseconds precision. Specifically, EEG records overall brain wave patterns, while ERP focuses on the brain's direct responses to specific events or stimuli, such as hearing a word or seeing a gesture. This precision makes EEG/ERP particularly suited for examining 'when' the brain responds to speech and gesture. The N400 component of ERP, in particular, provides critical insights into semantic processing, revealing how the brain integrates verbal and non-verbal information. Overall, while each method has its limitations: fMRI's lower temporal resolution and EEG/ERP's lower spatial resolution, the combined use of these methods can provide a comprehensive view of how the brain processes speech and gestures, both spatially and temporally.

6. Neural Bases of Speech and Gesture

This chapter of the thesis delves into empirical studies that explore the neural processing of speech and gestures, employing methodologies such as functional Magnetic Resonance Imaging (fMRI) and Electroencephalography/Event-Related Potentials (EEG/ERP). The primary focus here is to elucidate how the brain processes co-speech gestures and to determine the impact of various types of co-speech gestures on different brain regions. Specifically, this chapter identifies and discusses specific brain areas activated during the observation and interpretation of gestural communication and investigates how different types of gestures activate distinct neural networks.

6.1. Neural Processing of Co-Speech Gestures

Human communication is a multifaceted process that extends beyond mere spoken words. Co-speech gestures (CSG) accompanying verbal communication, play a pivotal role in enhancing the richness and effectiveness of human interaction (Mertens & Rohlfing, 2021; Ramos-Cabo et al., 2019; Bernard et al., 2015; Gluhareva & Prieto, 2017; Kushch et al., 2018). As discussed in chapter 3, these gestures play an essential role in language acquisition by reinforcing and clarifying spoken words. Specifically, co-speech gestures support vocabulary development and enhance communication skills, thereby deepening understanding and improving retention of spoken content. Several neuroimaging studies using advanced non-invasive techniques such as fMRI and EEG/ERP have shed light on the complex gesture-speech relationship. Specifically, during the observation of co-speech gestures, the brain exhibits increased activity in regions traditionally associated with the auditory and semantic processing of language, such as STG, MTG, and IFG (Hubbard et al., 2009; Willems et al., 2009; Holle et al., 2008; Green et al., 2009; Dick et al., 2009; Kircher et al., 2009).

6.1.1. fMRI Studies

Various functional MRI studies have shown that the inferior frontal gyrus (IFG), in particular, plays an integral role in the interpretation of co-speech gestures, especially those that are complex or abstract (Steines et al., 2021; Nagels et al., 2013; Straube et al., 2011). Specifically, research indicates that the left IFG exhibits increased activity when speech is accompanied by abstract (metaphoric) gestures, as opposed to more literal (iconic) gestures (Steines et al., 2021; Straube et al., 2011; Kircher et al., 2009). This heightened response is also observed when gestures accompany speech, compared to when speech is presented alone (Dick et al., 2009; Straube et al., 2011), as well as in the cases of semantic mismatch between gestures and speech (Willems et al., 2007), highlighting IFG's role in navigating semantic processing and integration.

The study by Nagels et al. (2013) corroborates and further extends these findings by showing that the left IFG's activity increases when gestures accompany abstract speech. In their study, participants were exposed to short video clips of an actor performing gestures that either depicted spatial information or shape information, accompanied by sentences that were either concrete or abstract. The study's findings revealed that abstract utterances, regardless of whether they were accompanied by space or shape-related gestures, activated the left IFG (see Figure 6.1).

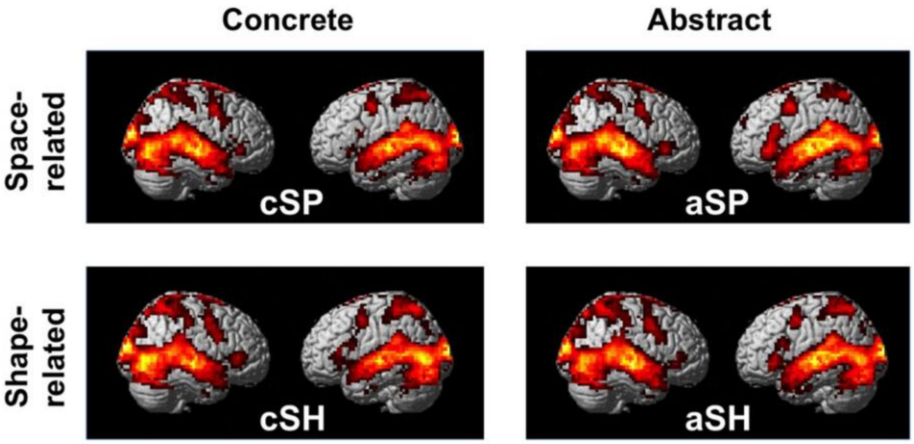


Figure 6.1 Brain activation patterns in response to concrete vs. abstract speech gestures (Nagels et al., 2013, p. 6)

Expanding the role of the IFG, the study by Dick et al. (2009) explored the neural impact of co-speech gestures focusing on various conditions: gestures that were either semantically related or unrelated to the spoken content, as well as conditions with no hand movements or visual cues. The findings revealed that not only the left but also the right IFG was critically involved in processing co-speech gestures. Specifically, it was found that the right IFG showed greater activation when gestures did not semantically match the spoken content, suggesting its role in resolving semantic discrepancies and enhancing the integration of conflicting verbal and non-verbal cues. Conversely, the left IFG showed more general activation in conditions where speech was accompanied by gestures (see Figure 6.2).

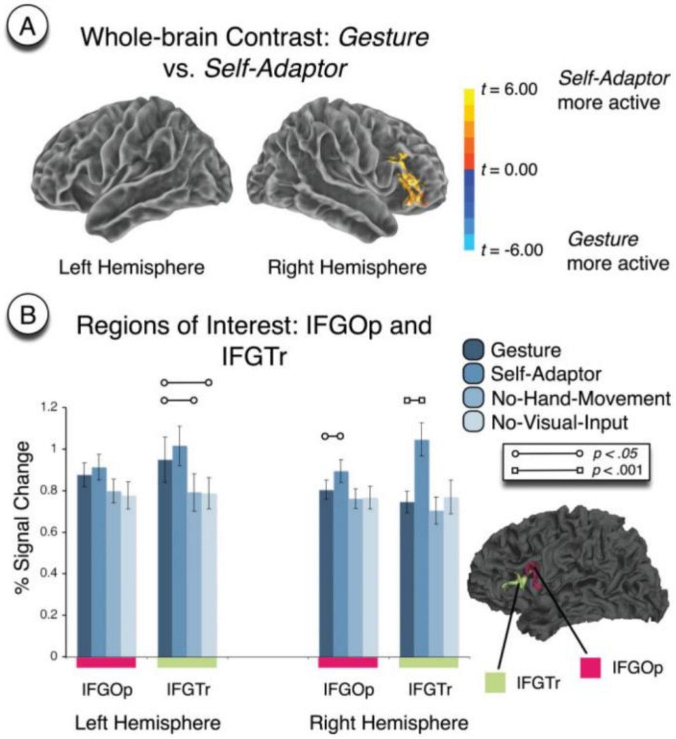


Figure 6.2 Hemispheric differences in IFG activation during gesture processing (Dick et al., 2009, p. 3518)

Figure 6.2 above illustrates the patterns of activation in the IFG during the processing of co-speech gestures. In Panel A, it can be clearly seen that the right hemisphere elicited increased activation as opposed to the left, particularly when participants were exposed to ‘Self-Adaptor’ hand movements. This suggests that the right IFG is notably engaged when processing gestures that do not semantically align with the spoken content. In Panel B, a comparison of activation in two sub-regions of the IFG can be seen: the opercular (IFGOp) and the triangular (IFGTr) in both the left and right hemispheres. Notably, the right IFGTr appeared particularly sensitive to gestures that lacked semantic congruence with speech, suggesting its crucial role in sorting out inconsistencies between what is said and shown. In contrast, the left IFG exhibited a more consistent activation pattern across various conditions tested, which may indicate a less variable response to the semantic relationship between gestures and speech. Collectively, these results highlight the distinct functions of the left and right hemispheres of the IFG. Specifically, the left IFG tends to facilitate the coherent flow of communication by combining congruent verbal and non-verbal cues, thus enhancing language comprehension. In contrast, the right IFG plays an essential role in managing semantic conflicts, ensuring that communication remains coherent even when confronted with potentially misleading gestural information.

However, other neuroimaging studies have demonstrated contradictory results regarding the activation of the left IFG. For instance, the study conducted by Kircher et al. (2009) aimed to explore how metaphoric gestures combined with abstract speech affect neural activity. The researchers focused on various conditions: gestures accompanied by speech, isolated speech, and isolated gestures. Their results revealed stronger activations in the left IFG during the audio-visual condition (using both gestures and speech) than in speech-alone or gesture-alone conditions. In contrast, the study by Wilson et al. (2008) focused on identifying brain regions involved in processing narrative speech with or without accompanying gestures. Their findings, unlike those found in the study by Kircher et al. (2009), did not show specific activations in the left IFG when participants were exposed to audio-visual conditions. These contrasting results demonstrate that the neural processing of speech and gestures is a complex, context-dependent process that may be influenced by methodological approaches such as the specific experimental setup, the type of gestures used in the study, and the nature of the accompanying speech.

In addition to these findings, other fMRI studies revealed that Broca's area, traditionally associated with the processing of spoken and written language, is also engaged in gestural communication (Häberling et al., 2016; Skipper et al., 2007; Brown & Yuan, 2018). Specifically, research shows that meaningful co-speech gestures have the capacity to reduce the cognitive workload on Broca's area during spoken language processing (Skipper et al., 2007). In the study by Skipper et al. (2007), participants listened to stories accompanied by various hand movements: meaningful speech-associated gestures, irrelevant self-grooming gestures, and no hand movements. Their findings demonstrated that Broca's area and other cortical areas interacted least when the spoken language aligned with meaningful speech-associated gestures. In contrast, these areas showed stronger activation when the spoken language was accompanied by self-grooming gestures (self-adaptors) or in the absence of hand movements (see Figure 6.3). These findings imply that meaningful gestures can actually reduce the workload on Broca's area during spoken language processing. To be more precise, gestures that semantically align with speech may add extra context or clarity, thus diminishing the need for intensive semantic retrieval or selection in Broca’s area.

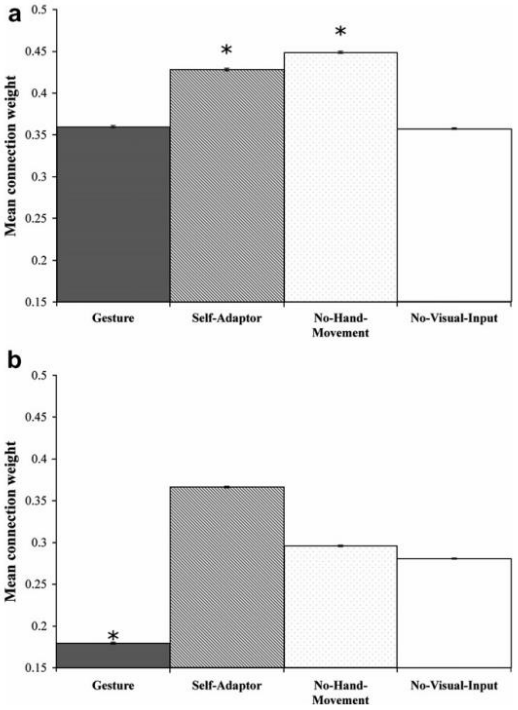


Figure 6.3 Activation of Broca’s area sub-regions under different conditions (Skipper et al., 2007, p. 273)

In Figure 6.3, graphs (a) and (b) illustrate two Broca’s area sub-regions (pars opercularis, POp (a); pars triangularis (PTr) (b). The ‘Gesture’ condition, represented by the lowest bars in both graphs, indicates a decrease of activation in Broca’s area when meaningful co-speech gestures were present. This contrasts with ‘Self-Adaptor’ and ‘No-Hand-Movement’ conditions, which exhibit higher activation.

In addition, the study by Willems et al. (2007) provides further evidence supporting the involvement of Broca’s area and LIFG in gestural communication. In their study, researchers explored how the semantic information from speech and gestures is integrated into the brain. During the experiment, participants were exposed to spoken sentences accompanied by co-speech gestures, with conditions designed to vary the semantic congruence between speech and gestures. To be more precise, the experiment was structured to test four distinct conditions: a correct match (where both gesture and language supported each other), a language mismatch (where the spoken language did not align with the gesture), a gesture mismatch (where the gesture did not match the verbal context), and a double mismatch (where both channels were incongruent). The study's findings revealed that the left inferior frontal cortex (BA 45/47), which includes Broca's area, showed increased activation in conditions where the speech and gestures were semantically mismatched (see Figure 6.4). These results suggest that Broca's area and LIFG are not merely involved in speech production but are also crucial for the semantic unification of language and action.

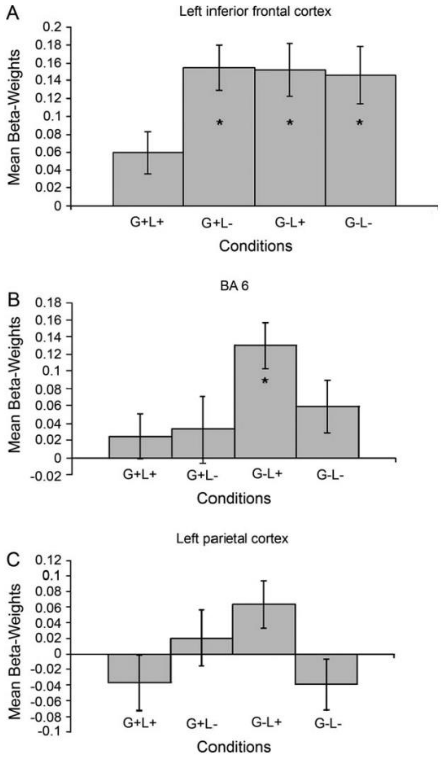


Figure 6.4 Mean activation levels for gesture and speech integration across four conditions (Willems et al., 2007, p. 2327)

Figure 6.4 illustrates the mean beta-weights reflecting neural activation within three brain regions across four experimental conditions. For the purpose of the present thesis, only Panel A of the figure will be discussed, which depicts activations in the left inferior frontal cortex (BA 45/47). Panel A shows that, compared to the baseline condition (G+L+), activation levels significantly increased when there was a mismatch in either gesture (G-L+), language (G+L-), or both (G-L-), as indicated by the asterisks. These increased responses in BA 45/47 suggest that

Broca’s area and LIFG respond more strongly to incongruities in speech and gestures than semantically congruent information. These findings further support previous studies on the IFG and Broca’s area by showing that these two brain areas play a crucial role in processing and integrating gestures and speech.

In addition to IFG and Broca’s area, fMRI research consistently identifies other brain regions, such as the STG and MTG, as crucial in the neural processing of co-speech gestures (Dick et al., 2009; Straube et al., 2011; Wilson et al., 2008; Willems et al., 2009). Specifically, studies have shown that the STG exhibits heightened activity when speech is accompanied by gestures as opposed to conditions where speech or gestures occur in isolation (Hubbard et al., 2009; Straube et al., 2011; Wilson et al., 2008).

Moreover, the study by Holle et al. (2010) revealed that the STG not only demonstrates a particularly strong response to combined audio and visual information (i.e., speech accompanied by gestures) but also is particularly sensitive to such integration in challenging (noisy) auditory environments. In their study, participants were shown videos with gesture-supported sentences and their unimodal components at varying levels of audio clarity (signal-to-noise-ratios, SNR) to explore how the brain processes co-speech gestures (iconic) under different listening conditions. The results demonstrated activation within the pSTS/STG regions in both hemispheres and notably increased activation in the left pSTS/STG when listening conditions were challenging (see Figure 6.5). These findings suggest that the STG plays a crucial role in multimodal communication. Specifically, it combines information from both speech and gestures, thereby improving speech comprehension. As this study demonstrated, this integration is particularly useful in challenging auditory environments, where gestures provide additional context that complements spoken words, making communication clearer.

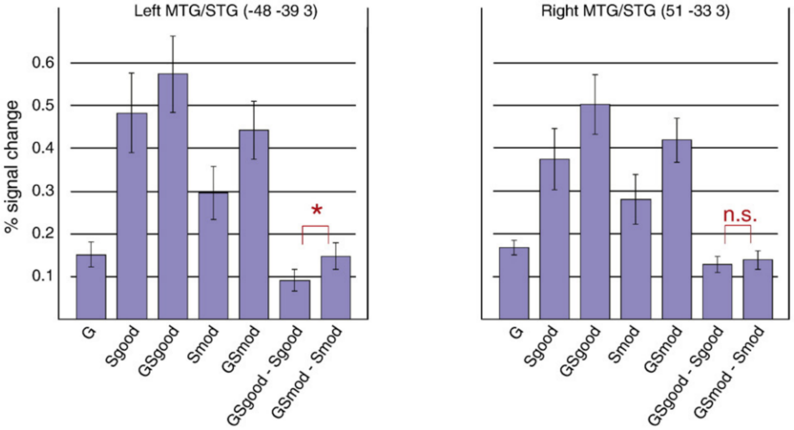


Figure 6.5 Illustration of significantly activated brain regions (Holle et al., 2010, p. 881)

Figure 6.5 above demonstrates the responsiveness of the left and right MTG/STG to various stimulus conditions: gestures-only (G), good SNR speech (Sgood), gesture combined with good SNR speech (GSgood), moderate SNR speech (Smod), and gesture combined with moderate SNR speech (GSmod). It can be clearly seen that the left MTG/STG exhibited stronger activation when gestures accompanied speech in a moderate SNR condition than in a good SNR condition, as indicated by the red asterisk. These findings demonstrate that gestures play an essential role in enhancing speech comprehension under challenging auditory conditions. Conversely, the right MTG/STG did not show a significant difference, indicating no particular enhancement from gestures in either SNR condition.

Furthermore, research indicates that the left MTG exhibits increased activation during the processing of co-speech gestures that either metaphorically align with the spoken language (Kircher et al., 2009) or that visually convey details not explicitly mentioned by speech (Dick et al., 2014). Supporting these findings, Green et al. (2009) aimed to explore how the brain processes the combination of speech and gestures, particularly focusing on pairs where the gestures visually represent or reinforce the meaning of spoken words (iconic gestures). The study also aimed to determine whether the brain shows a similar response to gestures that are unrelated to the spoken content. In their experiment, participants were exposed to short video clips of an actor speaking in two different languages, German (familiar) and Russian (unfamiliar), accompanied by gestures that were either directly related to the speech (iconic) or unrelated. Their results showed that the left MTG was significantly activated in conditions where gestures were both related and unrelated to the spoken language. However, it is important to mention that while there was overlapping activation in the MTG between two types of gestures, each type also activated distinct regions: iconic gestures primarily activated occipital areas, whereas unrelated gestures activated bilateral temporal and parietal regions. In essence, the findings from Green et al.'s (2009) study show that MTG plays a crucial role in multimodal communication by processing and integrating both congruent and incongruent gestures.

To conclude, this section discussed multiple fMRI studies and further supported the evidence that gestures and speech are intricately linked to each other. Specifically, it identified four critical brain regions, more precisely, IFG, Broca's area, STG, and MTG, that are not only involved in language processing but also play a critical role in multimodal gesture-speech integration. Even though each of the brain regions discussed here has its own specific roles in language processing, mainly focusing more on semantic or auditory processing of speech, all of them show increased activation during integration of multimodal input. These findings highlight the intricate nature of human cognition and emphasize the importance of considering both verbal and non-verbal cues in understanding the mechanisms underlying language comprehension and production. A more detailed discussion of the results provided in this section will be discussed in 6.1.3 Conclusion and Discussion section.

6.1.2. EEG/ERP Studies

In addition to fMRI studies, EEG/ERP methods have also been used to explore how the brain processes gestures that accompany speech (Özyürek et al., 2007; Kelly et al., 2004, 2007; Habets et al., 2011). Specifically, these studies have shown that observing co-speech gestures can influence brain activity, particularly affecting the N400 component. The N400 component, known for its role in processing meaning and semantic integration, was initially identified by Kutas & Hillyard (1980). As it was more thoroughly discussed in section 5.2 Electroencephalography/Event-Related Potential (EEG/ERP), this component manifests as a negative deflection in ERP's when an individual is presented with semantically incongruent stimuli, such as words that are out of context within a sentence (Kutas & Hillyard, 1980, p. 204).

Subsequent studies, such as the study by Kelly et al. (2004), have extended the understanding of the N400 component, demonstrating that it is not only responsive to linguistic stimuli but also modulated by gestures when they are synchronized with speech. In their study, participants were exposed to audio-visual clips where an actor produced words accompanied by gestures in four conditions: complementary (where gestures provided additional information related to the speech), matching (where gestures directly corresponded with the spoken content), mismatching (where gestures conflicted with the spoken content), and with no gestures. The findings showed that when participants viewed gestures incongruent with spoken language, a more pronounced N400 response was detected, indicating semantic incongruity (see Figure 6.6).

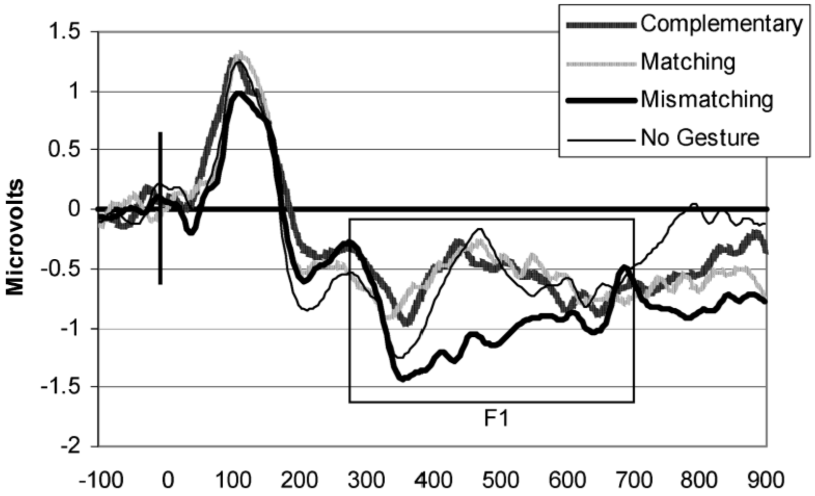


Figure 6.6 ERP responses to gesture-speech conditions (Kelly et al., 2004, p. 256)

The figure above illustrates a distinct pattern of neural responses to each of the four conditions, with the ‘Mismatching’ condition exhibiting a pronounced negative deflection in the ERP waveform (N400), indicating greater semantic processing demands when gestures conflict with verbal content. The N400 deflection was observed from approximately 324 to 648 milliseconds after the stimulus presentation. In contrast, the ‘Matching’ condition exhibited a smaller N400 response, indicative of less semantic conflict. Interestingly, the ‘Complementary’ and ‘No Gesture’ conditions did not show a significant deviation from the ‘Matching’ condition within the N400 time frame, suggesting that the semantic integration process may not be significantly impeded in these conditions.

The study by Wu & Coulson (2005) further expanded on these findings by examining how the brain processes the semantic content of iconic gestures in conjunction with visual stimuli across two experiments. In the initial experiment, participants were shown cartoon clips followed by videos of a person making hand gestures. The task required participants to discern whether the gestures were congruent (matching) or incongruent (not matching) with the cartoon narrative. The results revealed that incongruent gestures elicited a stronger N400-like brain response, which was referred to as the gesture N450 (see Figure 6.7). In the subsequent experiment, the same visual stimuli were used, but the focus shifted to probe words displayed after the gestures. Participants assessed the relatedness of those words to the prior gesture-cartoon combination. The findings showed that the congruency between gestures and the preceding cartoons had a tangible impact on the semantic processing of the subsequent linguistic information (see Figure 6.8).

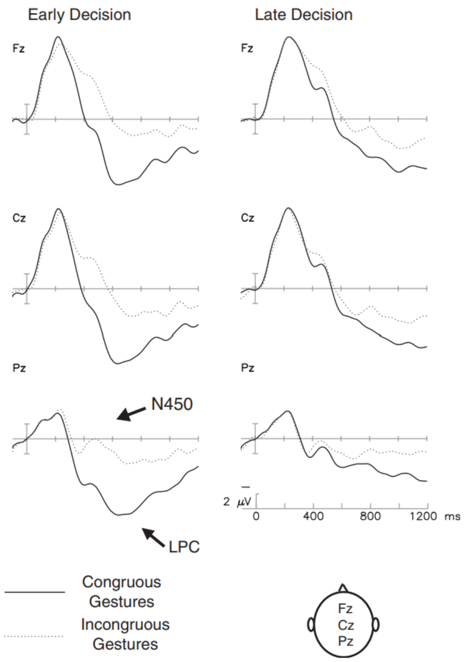


Figure 6.7 Experiment 1: ERP responses to congruous and incongruous gestures according to median decision times (Wu & Coulson, 2005, p. 660).

The figure above depicts ERP responses to matching (congruous) and mismatching (incongruous) gestures during early and late decision trials. It is evident that for incongruous gestures, the brain exhibited a generally stronger N450 response, indicating the brain's increased cognitive effort during speech and gesture processing. In contrast, when the gestures matched the cartoon, the brain's response was less pronounced, suggesting more efficient integration. Conversely, congruent gestures evoked a more pronounced Late Positive Component (LPC), reflecting the ease of processing congruent visual information.

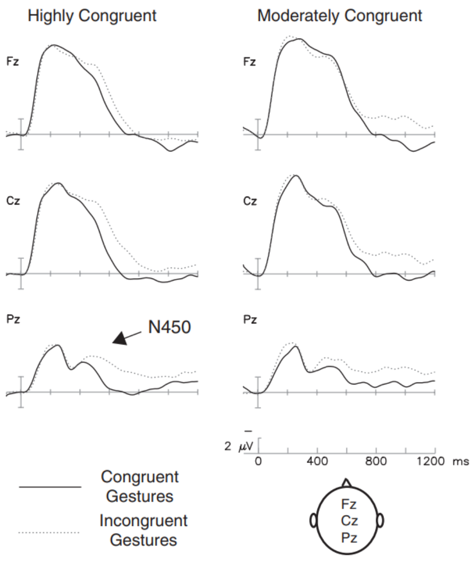


Figure 6.8 Experiment 2: Grand averaged ERP responses to the onset of highly and moderately congruent gestures (Wu & Coulson, 2005, p. 663).

The figure above illustrates results comparable to the first experiment, demonstrating an increased brain response to incongruent gestures (the N450 component) as opposed to congruent gestures. However, in contrast to the previous findings, the LPC was not detected, which may reflect the different cognitive demands of the tasks between the two experiments. While the first experiment required an explicit judgment of congruency, leading to a pronounced LPC, the second experiment did not require such explicit processing, which may explain the lack of an LPC. This difference highlights the sensitivity of ERP components to task demands and the specific cognitive operations engaged by participants during the experiments.

Furthermore, the study by Özyürek et al. (2007) discovered that the brain integrates gestural information with spoken language into the context of the preceding sentence. In other words, researchers found evidence that when people observe someone making gestures while speaking, the gestures are not just seen as separate or additional visual information. Instead, the gestures are processed by the brain as part of the overall meaning of the sentence being spoken. In their experiment, participants were shown video clips in which an actor uttered sentences while performing gestures. Each sentence featured a critical verb aligned with one of the four conditions: the matching or correct condition (the gesture directly matched the semantic content of a verb), gesture mismatch (the gesture conflicted with the verb), language mismatch (the spoken verb did not fit the overall sentence context, while gesture either matched the verb or was neutral), and double mismatch (both the gesture and the verb were incongruent with the sentence context). The study observed that all conditions where there was a mismatch (in gesture, speech, or both) elicited a larger N400 response (see Figure 6.9). These findings suggest that the brain processes semantic information from gestures in a manner similar to spoken language, supporting the notion that understanding spoken language involves integrating multimodal information.

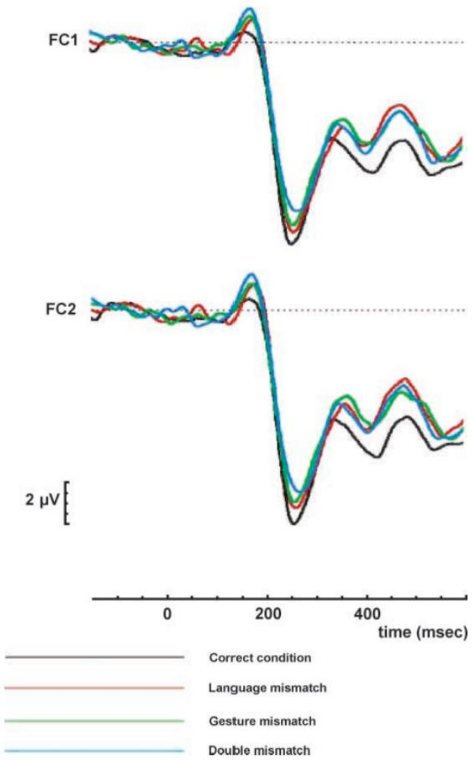


Figure 6.9 Comparative ERP responses to gesture-speech congruency conditions (Özyürek et al., 2007, p. 611)

Figure 6.9 illustrates the grand-average waveforms of ERPs, recorded from electrodes FC1 and FC2, in response to four conditions. These waveforms reveal distinct neural responses between the correct and mismatched conditions. Specifically, in ‘Language’, ‘Gesture’, and ‘Double mismatch' conditions, the ERPs show significantly greater negative deflections than in 'Correct condition', particularly around 400 milliseconds post-stimulus onset. This pronounced N400 response suggests increased cognitive processing effort as the brain attempts to resolve semantic discrepancies between the spoken word and the accompanying gestures.

In addition, the study by Cornejo et al. (2009) expands upon earlier research by shifting the focus from iconic gestures, as discussed in previous studies, to metaphoric gestures. In their study, researchers aimed to investigate how the brain processes metaphorical language when accompanied by gestures that either matched (congruent) or did not match (incongruent) the spoken words. Their experiment involved participants watching videos in which an actor produced metaphorical sentences alongside gestures designed to either align with or contradict the metaphor's abstract meaning. The results demonstrated distinct neural responses based on the congruence of the gestures with the metaphorical language. Specifically, incongruent gestures were found to elicit a larger N400 component, particularly in the 350-650 milliseconds time frame, compared to congruent gestures. This increase in N400 amplitude suggests that the brain faces greater semantic processing difficulties when the gestures conflict with the spoken language. In addition to the N400, the study also found that incongruous gestures led to a positive wave in posterior regions at a later stage (LPC) between 650-900 milliseconds time frame, which indicates that the brain further reanalyzes and tries to integrate the conflicting information (see Figure 6.10).

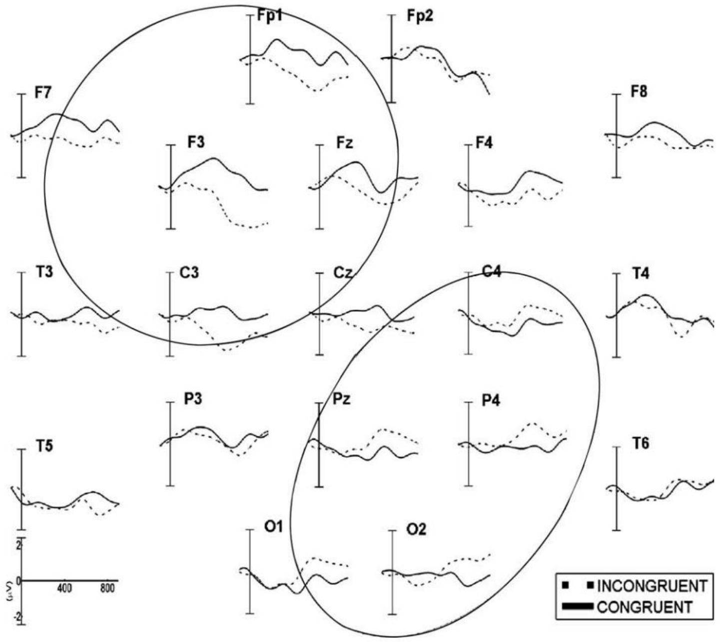


Figure 6.10 ERP waveforms illustrating neural responses to congruent and incongruent gestures (Cornejo et al., 2009, p. 47)

Figure 6.10 above depicts the brain's electrical activity, or waveforms, captured while participants viewed hand gestures that either matched or conflicted with the metaphorical language they heard. The researchers observed significant differences between congruent and incongruent gestures in the left-frontal and right-posterior areas, which are circled in the figure. In the left-frontal region, the circled area shows a larger negative deflection, N400 component, for incongruent gestures compared to congruent ones. This can be seen by incongruent waveforms being lower than the congruent. Conversely, in the right-posterior region, the circled area depicts a positive deflection, Late Positive Component (LPC), which is more pronounced for incongruous gestures than congruous ones. These patterns demonstrate how the brain processes the congruence of gesture and speech differentially, as indicated by the localized electrical activity corresponding to each condition. Specifically, the results regarding the N400 component reflect the brain's greater processing load when there is a mismatch between gestures and speech. This suggests that incongruent gestures create semantic conflict, which the brain tries to resolve, leading to a larger N400 deflection. The positive deflection, or LPC, indicates that, even after initial semantic processing, the brain continues to evaluate and integrate the gesture and speech information, working to resolve the mismatch and make sense of the overall context.

To conclude, this section discussed and analyzed various EEG/ERP studies, which further reinforced the notion that the brain's language processing regions extend beyond verbal communication to encompass non-verbal cues such as gestures. Specifically, the observations of the ERP's N400 component, associated with semantic processing, demonstrated how the brain responds to semantic congruency versus incongruency in multimodal communication, providing insights into how the brain resolves potential semantic discrepancies, ensuring coherent and effective communication. The following subsection will delve more thoroughly into the findings of EEG/ERP studies.

6.1.3. Conclusion and Discussion

The research evidence presented from both fMRI and EEG/ERP studies highlights the brain's significant role in processing multimodal information. Specifically, findings from fMRI studies revealed that brain areas traditionally associated with language processing, such as the IFG, STG, MTG, and Broca's area, play critical roles in the interpretation and integration of co-speech gestures.

The findings regarding IFG elucidate its pivotal role in processing speech and gestures, with each hemisphere specializing in different aspects of this integration. Specifically, the left IFG was found to be more active when processing gestures that are abstract rather than literal, underscoring its importance in the cognitive processing that reinforces semantic integration and the comprehension of complex, multimodal inputs. The right IFG, on the other hand, was found to be critically engaged when there was a discrepancy between the semantic content of speech and accompanied gestures. Its responsiveness to incongruent speech and gestures highlights its role in conflict resolution within the communication process. These findings suggest that while the left IFG facilitates understanding by integrating cohesive verbal and non-verbal information, the right IFG ensures that communication remains effective even when faced with contradictory information.

Moreover, the research on Broca's area revealed that this part of the brain exhibits reduced activity when speech is accompanied by meaningful gestures, suggesting that gestures can, in fact, ease the cognitive demands typically placed on Broca's area during verbal communication. These findings align with various other studies suggesting that incorporating multiple modes (such as gestures and speech) can enhance the effectiveness of language processing. In addition, it was found that when there was a semantic mismatch between speech and gestures, a more increased activation was observed in Broca's area. This finding indicates that Broca's area is not only involved in language processing but also in analyzing semantic content from multiple sources.

Furthermore, the results regarding STG and MTG showed that both regions play an essential role in multimodal communication. The STG, in particular, shows increased activation not only when speech is accompanied by gestures but also under challenging auditory conditions. These findings suggest that the STG plays a broader role than only processing auditory information; it gathers information from both visual and verbal cues to enhance the communication process. On the other hand, MTG demonstrates a broader involvement in processing the semantic content of gestures that accompany speech. Its activation increases not only when gestures are metaphorically aligned with spoken language but also when they provide supplementary visual information that is not explicitly verbalized. This was particularly noted in Green et al.'s (2009) study, where the MTG showed activation to both related and unrelated gestures accompanying speech. Such findings suggest that the MTG is essential in bridging the gap between gestural input and the linguistic context, thereby enriching the overall comprehension and cognitive processing.

Complementing these findings, EEG/ERP studies have provided further insights into the neural processing of co-speech gestures. These studies employed a match-mismatch experimental paradigm, examining conditions where gestures were either congruent or incongruent with the spoken content to assess how this impacts semantic integration. A primary focus was mainly placed on the N400 component associated with processing meaning and semantic integration. The findings revealed consistent results, mainly that the N400 component elicited stronger responses in mismatching conditions (incongruent gestures, speech, or both), indicating increased semantic processing demands. In essence, the evidence from the studies demonstrates that the brain's language processing mechanisms encompass not only verbal but also non-verbal cues, such as gestures. The N400 component serves as a reliable indicator of semantic integration, responding similarly to incongruences in both speech and gestures.

6.2. Neural Processing of Different Types of Gestures

Understanding how the brain controls and processes gestures can provide significant information about how people communicate without speaking, which is an essential but often overlooked aspect of human interaction. Previous sections have identified key brain regions, such as IFG, Broca's area, MTG, and STG that play a crucial role in processing gestures and speech. Recent scientific research shows that these brain areas engage differently for each type of co-speech gesture (Steines et al., 2021; Hubbard et al., 2009).

Specifically, various studies indicate that metaphoric gestures increase activation in the left IFG (Straube et al., 2011; Nagels et al., 2013; Andric et al., 2013). The study by Steines et al. (2021) further supports this evidence by examining how the left and right IFG interact with three types of gestures: metaphoric, iconic, and unrelated. Employing functional magnetic resonance imaging (fMRI), the study observed brain activity in 74 participants who were shown videos of an actor speaking accompanied by one of the three gesture types. The findings revealed that the brain's response to metaphoric and unrelated gestures was markedly different compared to iconic gestures, particularly in the activation patterns within the bilateral IFG. To be more precise, the left IFG demonstrated heightened activity when processing metaphoric gestures, while the right IFG was more responsive to unrelated gestures (see Figure 6.11).

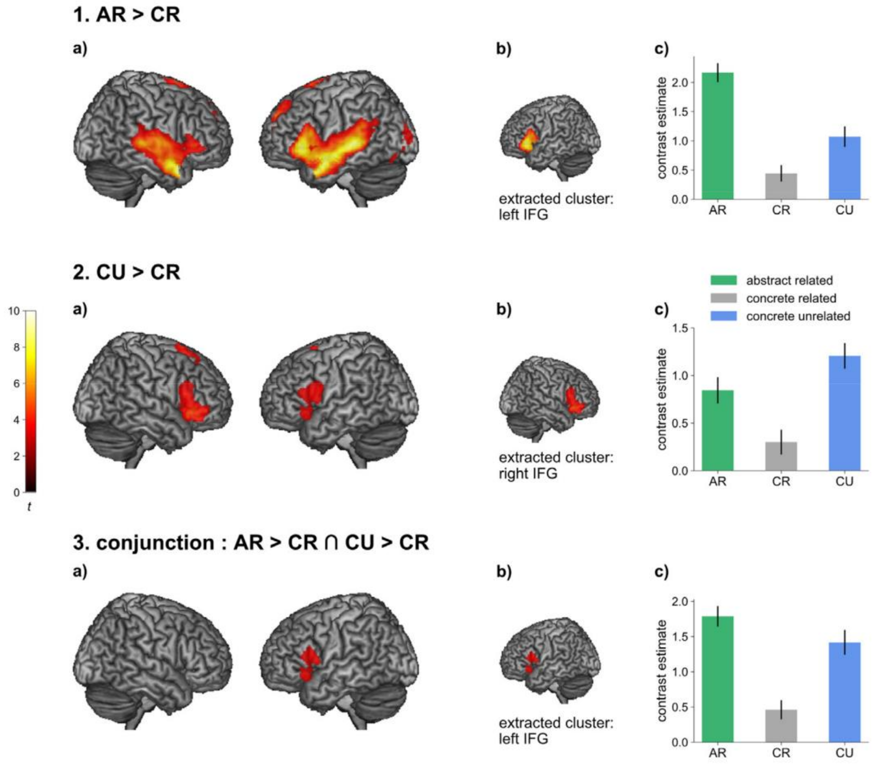


Figure 6.11 Brain activation patterns for different types of co-speech gestures (Steines et al., 2021, p. 5)

The first part of the figure (AR>CR) demonstrates heightened activity in the left IFG when abstract speech is paired with related metaphoric gestures as opposed to concrete speech with related iconic gestures, indicating this region's significant role in processing abstract tasks. The second part (CU>CR) demonstrates heightened activity in the right IFG for concrete speech with unrelated gestures compared to concrete speech with related iconic gestures. The last part (AR > CR ∩ CU > CR) depicts the overlapping activation in the left IFG for both abstract-related and concrete-unrelated gestures as compared to concrete-related gestures, suggesting a shared neural substrate involved in processing both types of complex gestures.

Similar results were also observed in the Straube et al.'s (2011) study. Specifically, in their study, researchers focused on metaphoric and iconic gestures to determine how the brain processes these types of co-speech gestures. During the experiment, participants were exposed to videos where an actor performed gestures that either directly represented the spoken content (iconic gestures) or symbolized abstract concepts related to the speech (metaphoric gestures). The study included four experimental conditions: iconic gestures with speech, metaphoric gestures with speech, gestures only, and speech only. The findings showed that each type of gesture engaged the brain differently. To be more precise, iconic gestures evoked a significant activation in the left posterior MTG and right STG, while metaphoric gestures activated both the posterior temporal regions and the left IFG (see Figure 6.12). The evidence regarding the left IFG further complements previous studies, indicating that IFG is deeply involved in semantic integration, particularly combining abstract concepts conveyed by metaphoric gestures.

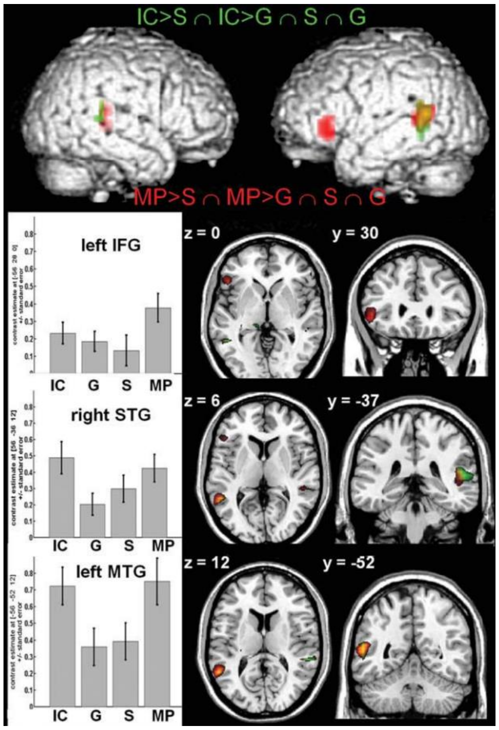


Figure 6.12 Brain activation patterns in four experimental conditions (Straube et al., 2011, p. 529)

Figure 6.12 above illustrates the activation of three brain regions: left IFG, right STG, and left MTG in four different conditions. It can be seen that metaphoric gestures elicited a significantly higher response in the left IFG compared to iconic gestures and isolated gesture conditions. However, iconic gestures evoked slightly higher activations in the right STG than metaphoric gestures. Additionally, both metaphoric and iconic gestures elicited an increased response in the left MTG. These findings indicate that even though both iconic and metaphoric gestures elicited activation in the left MTG, each gesture also tended to evoke distinct brain regions that perhaps reflect their distinct cognitive and communicative functions. To be more precise, as IFG is traditionally associated with processing the meanings of words and sentences, it becomes more active when processing metaphoric gestures that express abstract content. At the same time, STG is more associated with comprehending spoken words and language, which, therefore, becomes more active when processing iconic gestures that express concrete concepts.

Furthermore, the study conducted by Holle et al. (2008) corroborates the findings regarding iconic gestures by showing that these types of co-speech gestures elicit activation in the left posterior superior temporal sulcus (STS), the bilateral inferior parietal lobule, and the bilateral ventral precentral sulcus. During their experiment, participants watched video clips where sentences included words with ambiguous meanings. These sentences were paired with different types of hand movements: a non-meaningful grooming movement, a gesture supporting the dominant meaning of the word, or a gesture supporting the subordinate meaning of the word. The researchers found that meaningful iconic gestures (whether they emphasized the dominant or subordinate meaning of the word) activated the left posterior superior temporal sulcus (STS), the bilateral inferior parietal lobule, and the bilateral ventral precentral sulcus (see Figure 6.14).

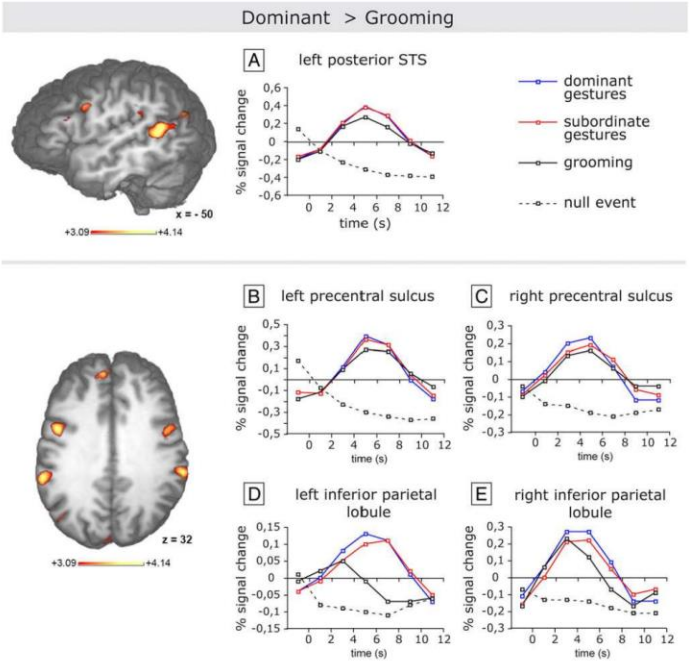


Figure 6.13 Brain activation in response to dominant gestures vs. grooming (Holle et al., 2008, p. 2018)

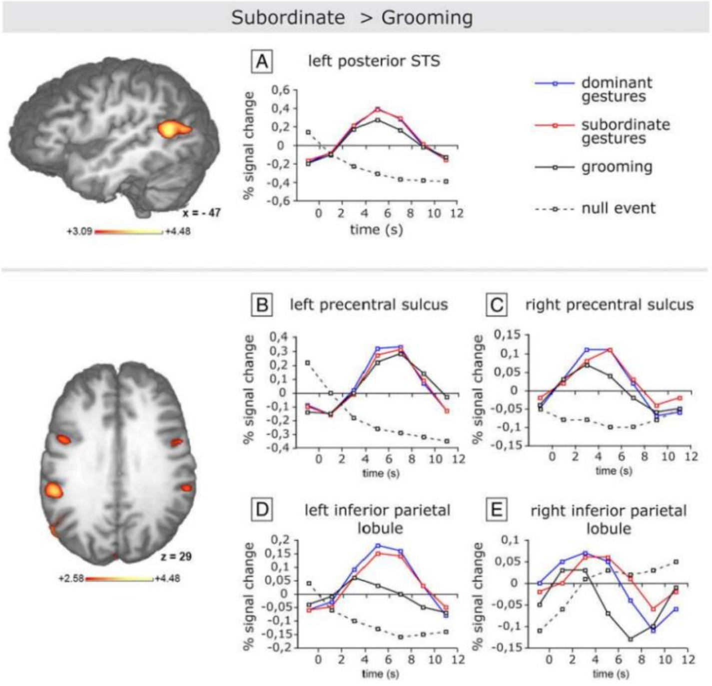


Figure 6.14 Brain activation in response to subordinate gestures vs. grooming (Holle et al., 2008, p. 2019)

Figure 6.13 and Figure 6.14 illustrate the brain's differential activation patterns when processing dominant and subordinate gestures in relation to non-meaningful grooming movements. Both figures reveal increased activity in the left posterior STS, the precentral sulcus bilaterally, and the inferior parietal lobule bilaterally when dominant and subordinate iconic gestures were used.

The study by Hubbard et al. (2009) focused on how beat gestures influence the brain's processing of spoken language. Using the fMRI method, researchers examined the neural responses to speech accompanied by beat gestures compared to speech presented alone or with non-meaningful hand movements. Their findings demonstrated significant neural activations extending beyond the traditional auditory processing areas. Specifically, when speech was paired with beat gestures, there was an increase in activity within the bilateral auditory cortex, suggesting that the presence of gestures enhances the auditory processing of speech. Furthermore, the left STG/S was particularly responsive to speech paired with beat gestures as opposed to nonsensical hand movements. Additionally, the right PT emerged as a critical area for the multisensory integration of speech and beat gestures. This region demonstrated increased activity for combined speech and gesture inputs, exceeding the activity observed when each modality was presented individually (see Figure 6.15 below). This finding demonstrates the PT's essential role in merging the rhythmic aspects of both speech and gesture.

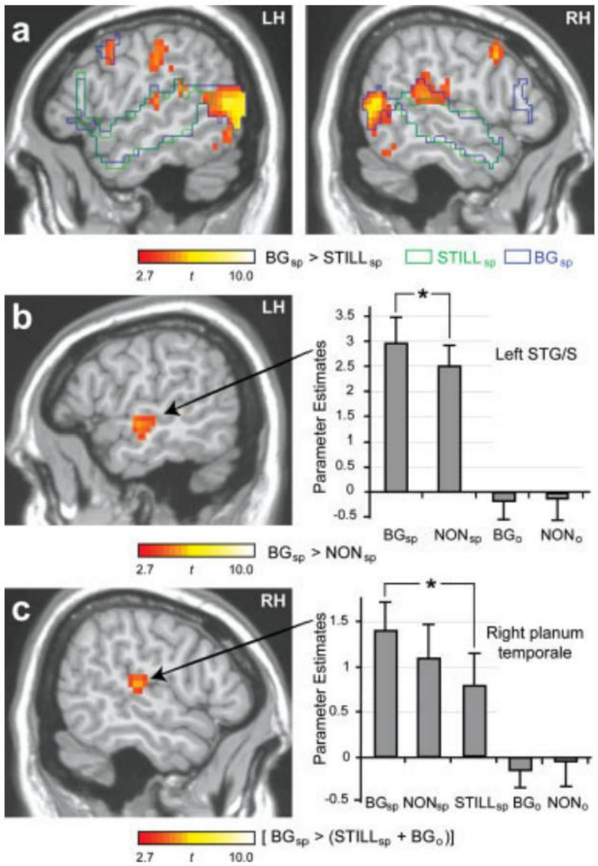


Figure 6.15 Neural activity related to processing speech with and without beat gestures (Hubbard et al., 2009, p. 1032)

In Figure 6.15 (a), increased neural activity can be seen in both the left and right hemispheres when speech is accompanied by beat gestures as opposed to when no gestures are used (still body). Figure 6.15 (b) specifically highlights that the left STG/S was especially responsive to beat gestures paired with speech, more so than when speech was accompanied by nonsensical hand movements. This finding suggests a significant role for the left STG/S in the semantic processing of speech and rhythmically relevant gestures. Figure 6.15 (c) depicts the right PT showing increased activity for the combination of speech with beat gestures, which is greater than the summed activation levels of speech and gestures when each is presented alone.

Furthermore, it is important to mention that despite an extensive search across available literature, no studies were found that specifically addressed how the brain processes deictic gestures. Studies in the field of gestures and speech have primarily focused on investigating other co-speech gestures, such as metaphoric, beats, and especially iconic gestures. One potential reason for the lack of focus on deictic gestures in neuroimaging research may be related to their perceived simplicity. To clarify, unlike iconic gestures, which visually depict objects or actions, deictic gestures involve simple pointing or indicating, which might lead researchers to prioritize more complex gestures.

In essence, this section focused on different types of co-speech gestures, more precisely, metaphoric, beat, and iconic, and examined whether the brain responds to these gestures in a similar or distinct manner. The analysis of multiple fMRI studies revealed that each gesture type uniquely engages distinct brain regions. This diversity in neural activation patterns demonstrates the specialized ways the brain processes different types of gestures, each contributing uniquely to human communication. A more detailed discussion of the results will be provided in the following 6.2.1 Conclusion and Discussion section.

6.2.1. Conclusion and Discussion

A vast number of studies conducted on the relationship between speech and gestures clearly demonstrate that these two modes of communication are intricately linked not only in the way they are expressed but also in the way the brain processes them. Studies focusing on the neural mechanisms underlying the processing of co-speech gestures have identified four critical brain regions: IFG, Broca's area, MTG, and STG, all of which were discussed in detail in previous sections. Interestingly, the evidence shows that different types of co-speech gestures tend to evoke responses in different brain regions.

To elaborate further, metaphoric gestures, which represent abstract meanings, tend to elicit a notable response in brain regions that are responsible for processing semantic information or, in general, help to comprehend spoken words and language. Particularly, findings demonstrated that during the processing of metaphoric gestures, the left IFG shows increased activation. This increased activation suggests that the left IFG is crucial in understanding metaphoric gestures, as it works in a greater capacity to comprehend what is being said and shown. The study by Straube et al. (2011) expanded on previous findings by showing that this type of co-speech gesture elicits a significant response not only in the left IFG but also in the left MTG. However, this response was evoked not only by metaphoric gestures but also by iconic ones. These findings demonstrate that although both types of gestures activate these brain regions, the left IFG, in particular, is more actively involved in the processing of metaphoric gestures.

Iconic gestures, on the other hand, as was mentioned, evoked a response in the left MTG; however, the most noticeable increase in the activation was observed in the right STG, in which this brain region reacted more deeply than to metaphoric gestures or isolated gesture and speech conditions. Furthermore, Holle et al.'s (2008) study demonstrated that brain areas involved in processing sensory, linguistic, and motor information, particularly STS, bilateral inferior parietal lobule, and the bilateral ventral precentral sulcus, show activation during the processing of iconic gestures. This evidence shows that gestures that visually mimic the physical characteristics of objects or actions tend to be more deeply processed in these brain areas that focus on combining sensory information with language comprehension and movement.

Another type of co-speech gestures, known as beat gestures, which mainly reflect the rhythm of the speech, predominantly show activation in brain areas traditionally associated with auditory processing, such as the bilateral auditory cortex, left STG/S, and the right PT. These findings further align with the view that each type of co-speech gesture tends to evoke brain regions that reflect their distinct cognitive and communicative functions more.

To conclude what has been stated so far, the neural processing of speech and gestures involves complex interaction of different brain regions. Each type of gesture tends to activate specific neural networks based on its meaning and function. This connection highlights the role of gestures in enhancing and clarifying spoken language, demonstrating that speech and gestures are deeply interconnected. However, as the relationship between these two modes of communication in the brain is complex and involves many factors, a more in-depth research is needed to fully comprehend the intricate mechanisms at play and how they shape the understanding and use of language and gestures.

7. The Evolutionary Perspective

This chapter delves into the evolutionary origins of human language by addressing the main theories that propose different ways through which language may have evolved. Specifically, through the analysis of various studies on non-human primates, this chapter aims to discern whether human language originated from gestural, vocal, or multimodal communication.

7.1. Language Origin Theories

The quest to understand the origins of language has generated numerous theories and extensive scholarly debate over the years (Cäsar et al., 2013; Zuberbühler et al., 2009; Gentilucci & Corballis, 2006; Corballis, 2003). Although there is a broad consensus that language evolved from simpler forms of communication, the specific pathway of this evolution remains unclear. Some scholars believe that as humans primarily communicate with each other using speech, it would be a natural or plausible theory that human language evolved from vocalizations (Zuberbühler et al., 2009; Cäsar et al., 2013; Seyfarth & Cheney, 2003). However, others provide a contradictory view, saying that gestural communication observed in non-human primates exhibits more flexibility than their vocal calls, which thus suggests that language may have originated from gestures rather than vocalizations (Tomasello, 2007; Pollick & de Waal, 2007). Therefore, to explain the evolution of language, three major theories have been proposed: Gesture-First theory, Vocal-First theory, and the Multimodal theory (Prieur et al., 2020, p. 538).

7.1.1. The Gesture-First Theory

The gestural theory of language origin posits that early humans, or hominins, primarily relied on manual gestures and body language for communication, which subsequently laid the foundation for the development of spoken language (Corballis, 2003; Arbib, 2005; Arbib et al., 2008; Gentilucci & Corballis, 2006). This theory is supported by various research, indicating that in non-human primates, gestures provide a richer and more flexible means of communication compared to vocalizations (Call & Tomasello, 2007; Tomasello, 2007; Pollick & de Waal, 2007; Arbib et al., 2008).

The study by Pollick & de Waal (2007) exemplifies this by highlighting the adaptability and complexity of gestural communication in contrast to vocal calls. In their study, researchers examined the natural communication behaviors of bonobos and chimpanzees focusing on gestural flexibility and its potential implications for the evolutionary development of language. In their study, researchers identified 31 manual gestures and 18 facial/vocal signals used by bonobos and chimpanzees and analyzed the context and flexibility of these manual gestures compared to facial/vocal signals. The findings showed that bonobos and chimpanzees used facial and vocal signals similarly, however, their use of gestures was more flexible and adaptable (see Figure 7.1).

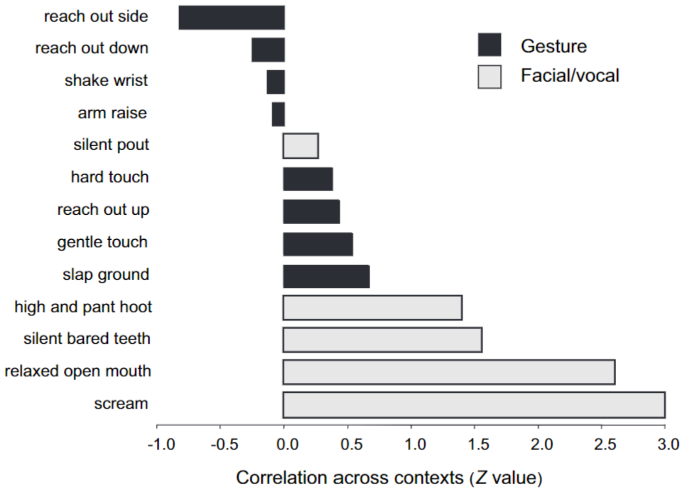


Figure 7.1 Comparative analysis of contextual correlation for gestural and facial/vocal signals in apes (Pollick & de Waal, 2007, p. 8187)

Figure 7.1 demonstrates the varying degrees of consistency in the use of gestural and facial/vocal signals by bonobos and chimpanzees across different behavioral contexts. It can be seen that both facial and vocal signals showed significantly high positive correlations. This is especially evident in their vocal signals, such as ‘scream,’ and facial signals, such as ‘relaxed open mouth.’ Interestingly, ‘silent pout’ exhibited lower correlations than other vocal signals. These results demonstrate that bonobos and chimpanzees tend to use their vocal/facial signals consistently in specific contexts. In contrast, gestural signals exhibited relatively low or even negative correlations, particularly ‘reach out side’, reflecting their more flexible use of gestures as opposed to their vocal/facial signals.

Additionally, the discovery of mirror neurons in monkeys which are homologues to Broca’s area in the human brain, lends further support to the gestural theory of language origins (Arbib, 2005; Rizzolatti & Arbib, 1998; Fabbri-Destro & Rizzolatti, 2008). These neurons, located within the ventral premotor cortex (F5) area, show activation when monkeys perform a grasping motion and also when they observe others performing the same action (Rizzolatti & Arbib, 1998, p. 188). Additionally, research indicates that mirror neurons are activated not only by actions but also by sounds associated with those actions, use of tools, and mouth actions observed in social partners (Prieur et al., 2020, p. 540). This evidence supports the gestural theory of language origins, suggesting that the foundational elements of linguistic communication might derive from the innate ability to decode and mimic observed behaviours (Rizzolatti & Arbib, 1998).

However, other researchers offer contradictory views of linking mirror neurons to the evolution of language (Hickok, 2009; Jacob & Jeannerod, 2005). For instance, Hickok (2009) challenges the idea by pointing out the lack of direct evidence from both monkeys and human studies that mirror neurons support action understanding in a way that is necessary for language evolution. To be more precise, he questions the broad or ‘popular’ claims surrounding mirror neurons; how can cognitive processes that allow interpreting and mimicking actions be attributed to mirror neurons if there is no sufficient or clear evidence. Similarly, Jacob & Jeannerod (2005) argue that social cognition, including language, requires an understanding of other’s goals, beliefs, and desires, which often extend beyond the direct observable actions and intentions that mirror neurons are proposed to encode. The authors contend that the complexities of linguistic communication and symbolic representation involve cognitive processes that cannot be solely explained by motor simulations provided by mirror neurons. For instance, mirror neurons are mainly concerned with mimicking actions, such as grabbing food. However, language and symbols are more complex as they can evolve meanings that extend beyond grasping actions. Additionally, researchers highlight the absence of direct evidence linking mirror neuron activity to the syntactic structures and symbolic nature of language. Therefore, attributing the evolution of language to mirror neurons overlooks the broader and more complex cognitive mechanisms at play, which could be misleading.

Moreover, another argument supporting the gestural theory of language origins revolves around similar cerebral lateralization (specialization of brain hemispheres) and its correlation with handedness (hand preference) found in both humans and non-human primates (Meunier et al., 2013; Knecht et al., 2000; Hopkins & Cantero, 2003; Hopkins et al., 2005). Specifically, the majority of people exhibit right-handedness, a phenomenon linked to the specialized roles of the brain's left hemisphere (Papadatou-Pastou, 2011). This hemispheric specialization includes not just motor control but also extends to the domain of language, suggesting a link between hand use and cognitive functions (Knecht et al., 2000).

The study conducted by Meunier et al. (2013) provides compelling evidence that the phenomena of hemispheric lateralization and handedness extend beyond humans, as it is also observed in non-human primates. In their study, researchers subjected Tonkean monkeys to two types of tasks: communicative (pointing, showing objects) and non-communicative (grasping objects) tasks. In the experiment, these objects were strategically placed in different positions around the macaques: to the left side of the subject (positions 1, 2, 3), to the right side (positions 5, 6, 7), and in the center (position 4). The findings revealed a significant preference for using the right hand during manual communicative tasks compared to non-communicative tasks (see Figure 7.2). Additionally, the authors claimed that this right-hand preference for communicative gestures was consistent with the patterns observed in human infants and baboons, suggesting a common evolutionary basis for hemispheric specialization in manual gestures and, potentially, in language.

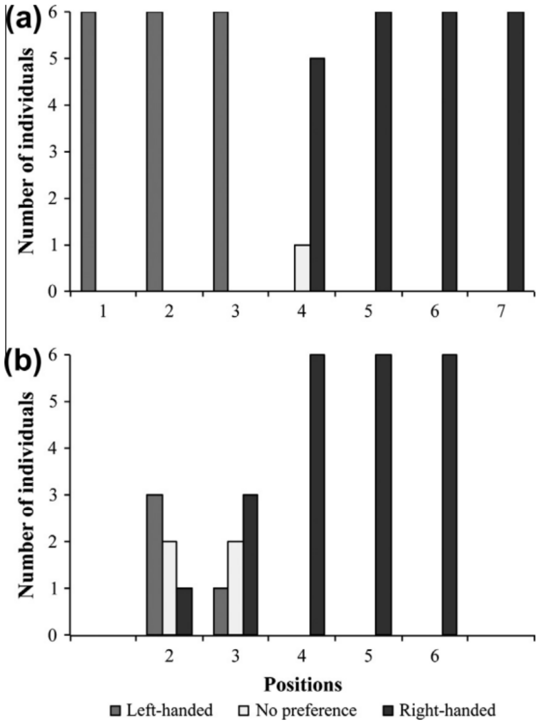


Figure 7.2 Hand preference patterns in reaching task (a) and manual gesture task (b) (Meunier et al., 2013, p. 184)

The figure above shows the distribution of hand preferences among Tonkean macaques during non-communicative (a) and communicative (b) tasks. In the non-communicative task (a), macaques showed a left-hand preference for object positions on their left side (1, 2, and 3) and a right-hand preference for positions on their right side (5, 6, and 7). In the central position (4), a right-hand preference was also observed, with only one individual showing no clear preference. For the communicative task (b), positions 1 and 7 were omitted due to their proximity to the subjects. In this task, a more varied pattern was observed: Position 2 showed a balance between 'Left-handed' and 'No preference'. In contrast, positions 3 through 6 demonstrated a stronger right-hand preference, particularly for central and right-side positions. These results highlight a distinct lateralization when the macaques engaged in communicative tasks, with some variation depending on the spatial position of objects.

Furthermore, the gestural theory is also supported by shared properties found in non-human primates' gestural communication and the foundational elements of human language. This includes gestures that are intentional, referential, involve turn-taking, comply with linguistic principles, and exhibit iconicity (Prieur et al., 2020, p. 541). These shared properties suggest that gestural communication could have provided the initial framework upon which verbal language was built, with gestures evolving to accompany and eventually give rise to spoken language.

In essence, the Gesture-First theory suggests that language evolved from manual gestures. This theory is supported by various studies demonstrating the significant adaptability and flexibility of primates' gestural communication. Further support comes from observed similarities in cerebral lateralization, handedness, and the functioning of the mirror neuron system across both humans and primates. Additionally, the presence of shared characteristics between non-verbal communication in primates and the foundational aspects of human language suggests that gestural communication might have been the foundation of the evolution of language.

7.1.2. The Vocal-First Theory

The vocal theory, on the other hand, posits that human language originated from vocal calls, suggesting that the complex languages in use today evolved from more basic vocalizations akin to those produced by primates (Zuberbühler et al., 2009). Specifically, research has shown that certain vocalizations among non-human primates, particularly alarm calls, carry specific meanings that communicate the nature and urgency of threats, such as the type and proximity of predators (Cäsar et al., 2013; Seyfarth & Cheney, 2003).

In the study by Cäsar et al. (2013), researchers explored the complexity and informativeness of alarm call systems in non-human primates, particularly focusing on the black-fronted titi monkeys. Researchers aimed to determine whether these primates' alarm calls could convey specific information about both the type of the predator and its location. During the experiment, titi monkeys were presented with two models of predators: an oncilla and a caracara, which were placed either on the ground or within the canopy (in the tree). The findings demonstrated that titi monkeys produced distinct alarm call sequences depending on the predator type and its location (see Figure 7.3).

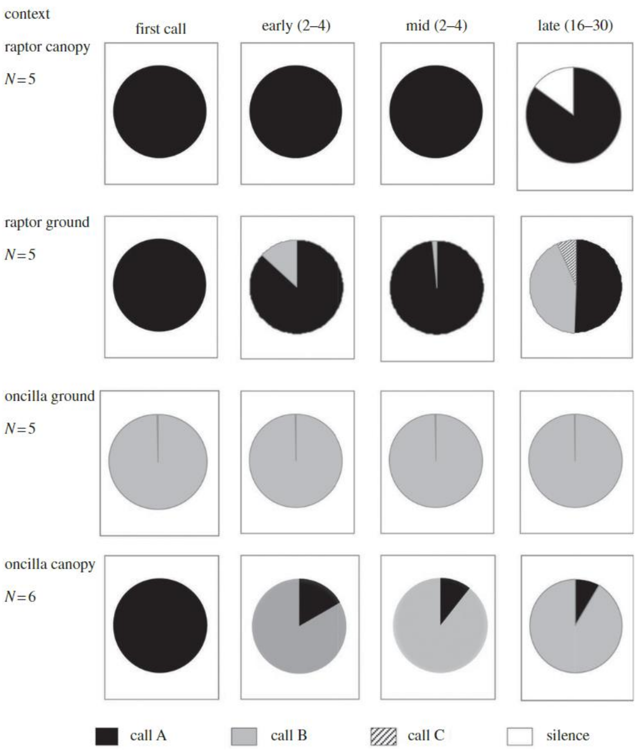
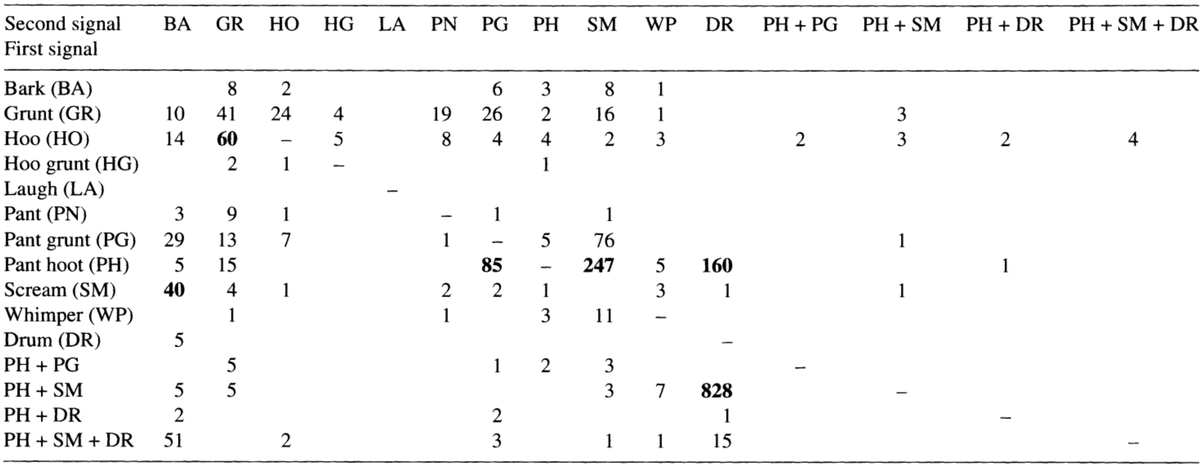


Figure 7.3 Call response patterns of titi monkeys to predatory threats (Cäsar et al., 2013, p. 2)

The figure above provides a clear illustration of how black-fronted titi monkeys adapted their call sequences based on both the type and location of a potential predator. Specifically, when a raptor was in the canopy, the monkey's responses were consistent; showing that only 'call A' was made across all stages with only brief moments of silence. In contrast, when a raptor was on the ground, the initial response was 'call A', but as the sequence progressed, they began to incorporate 'call B' and a small part of 'call C' into their alarms. The change from one type of call to another over time could indicate the monkey's ability to convey that the threat (usually spotted in the trees) changed their usual location to a different one. The response to oncilla, a ground predator, contrasted this pattern. More specifically, when oncilla was seen on the ground, monkeys predominantly used 'call B' from the start, suggesting an innate ability to distinguish between aerial and terrestrial threats and use calls accordingly. Interestingly, when an oncilla was spotted in the canopy, the call sequences started with a 'call A', indicating an aerial threat, however over time, their call changed to a mixture of 'call A' and 'call B', indicating that a predator that is usually seen on the ground was spotted in the tree. These varied call patterns demonstrate the monkeys' capacity to convey detailed information about the nature and location of potential dangers. This supports the vocal theory of language origins by showing that primate vocalizations can indeed carry complex, context-specific information similar to elements of human language.

Furthermore, certain foundational elements of grammar, similar to those in human language, have been observed in the vocal communication of primates, which further supports the vocal theory of language origins (Crockford & Boesch, 2005; Coye et al., 2016). Specifically, Crockford and Boesch (2005) conducted a pivotal study investigating the complexity and informational content of vocal communications among chimpanzees. In their study, researchers aimed to understand how chimpanzees use combinations of different call types to communicate and how these combinations compare to the use of single calls in terms of the information conveyed. The experiment involved observing two groups of habituated chimpanzees and documenting their vocalizations and associated contexts. The results showed that nearly half of all chimpanzee vocalizations occurred in combination with other vocalizations of different call types, indicating a complex vocal communication system. These call combinations were found to be context-specific, often conveying different or more detailed information than single calls alone. Moreover, the production of call combinations was non-random, with some combinations occurring more frequently or in a specific sequence, suggesting an element of 'lexical syntax' (see Table 7.1). However, the study notes that while these combinations reflect some structural consistency, they should not be directly equated with lexical syntax observed in human language as it requires more in-depth research.



Second signal First signal	BA	GR	HO	HG	LA	PN	PG	PH	SM	WP	DR	PH + PG	PH + SM	PH + DR	PH + SM + DR
Bark (BA)		8	2				6	3	8	1					
Grunt (GR)	10	41	24	4		19	26	2	16	1			3		
Hoo (HO)	14	**60**	–	5		8	4	4	2	3		2	3	2	4
Hoo grunt (HG)		2	1	–				1							
Laugh (LA)					–										
Pant (PN)	3	9	1			–	1		1						
Pant grunt (PG)	29	13	7			1	–	5	76				1		
Pant hoot (PH)	5	15					**85**	–	**247**	5	**160**			1	
Scream (SM)	**40**	4	1			2	2	1		3	1		1		
Whimper (WP)		1				1		3	11	–					
Drum (DR)	5										–				
PH + PG		5					1	2	3			–			
PH + SM	5	5							3	7	**828**		–		
PH + DR	2						2				1			–	
PH + SM + DR	51		2				3		1	1	15				–

Table 7.1 Order of signal combinations produced by adult male and female chimpanzees (Crockford & Boesch, 2005, p. 410)

Table 7.1 illustrates the structured patterns of vocal combinations in chimpanzee communications, demonstrating that these combinations were used systematically rather than randomly. Specifically, the table shows that ‘Pant hoot (PH)’ frequently initiated sequences, particularly leading a ‘Pant grunt (PG)’ or a ‘Scream (SM)’, with a notably high occurrence observed in the PH+SM combination. This pattern suggests a preferential and possibly syntactic ordering in these vocal interactions. However, other call pairings, such as ‘Hoo (HO)’ and ‘Grunt (GR)’, showed more variability in their sequencing, indicating a more flexible arrangement within these specific vocal combinations. In essence, this variability highlights the complexity of chimpanzee vocal communication, reflecting structured and systematic patterns of interaction.

Moreover, research reveals the existence of structured communication patterns within primate vocal interactions that are similar to conversational rules observed in human speech, such as turn-taking, call-overlap avoidance, and acoustic matching (Prieur et al., 2020, p. 539). The study by Levréro et al. (2019) further supports this evidence by demonstrating complex vocal exchanges of great apes. Specifically, in their study, researchers observed that bonobos avoid overlapping their calls and tend to respond quickly to each other; following specific rules that resemble the turn-taking seen in human communication (see Figure 7.4). Notably, bonobos were more likely to engage in these vocal exchanges with individuals they had stronger social bonds with, demonstrating a preference for communicating with specific group members. Additionally, the study revealed that the rate of vocal sharing was primarily affected by the age difference between bonobos rather than other factors like sex, kinship, or social and vocal affinity (see Table 7.2). This implies that age significantly influences vocal communication patterns among bonobos.

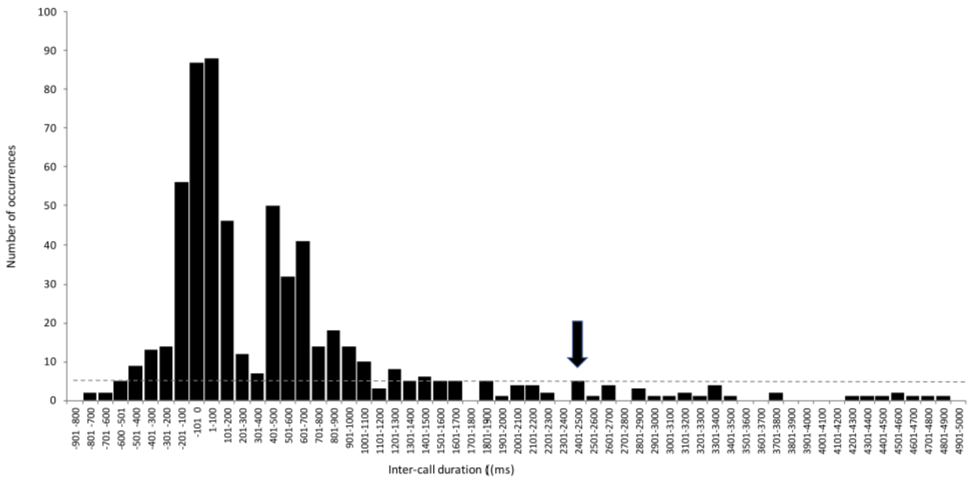
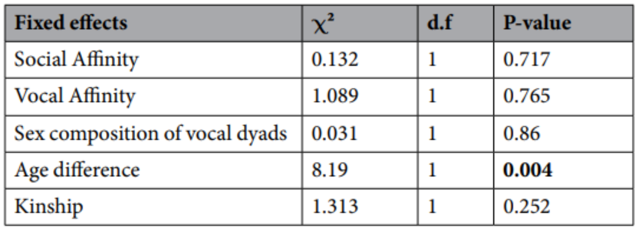


Figure 7.4 Distribution of inter-call durations in bonobo vocal exchanges (Levréro et al., 2019, p. 2)

The figure above illustrates the timing between vocalizations in bonobos' communication, highlighting three patterns of vocal interaction: successive, overlapped, and isolated. It can be seen that the majority of the distribution is characterized by short, positive inter-call durations, aka successive pattern, indicating a tendency for bonobos to wait briefly before responding to another's call. The second pattern emerges from the inter-call durations, representing instances where calls from different bonobos overlap. However, these occurrences were relatively infrequent, as the graph predominantly displays positive intervals, implying that overlapping is not the norm in bonobo vocal interactions and there is a general preference for sequential calling. The third pattern is distinguished by very short and slightly longer positive inter-call durations, indicating a more deliberate pace of communication in certain contexts. These intervals reflect instances where bonobos exhibit greater patience, possibly allowing for more contemplation or processing of the vocalizations heard before issuing a response.



Fixed effects	χ^2	**d.f**	**P-value**
Social Affinity	0.132	1	0.717
Vocal Affinity	1.089	1	0.765
Sex composition of vocal dyads	0.031	1	0.86
Age difference	8.19	1	**0.004**
Kinship	1.313	1	0.252

Table 7.2 Socio-demographic influences on vocal sharing in bonobos (Levréro et al., 2019, p. 4)

The table above illustrates a statistical analysis of factors influencing vocal sharing among bonobos. It can be seen that the only variable with a significant effect was the age difference between callers (p=0.004), suggesting that vocal sharing is predominantly influenced by age disparity rather than other socio-demographic influences.

In conclusion, the Vocal-First theory suggests that the origins of human language are rooted in vocalizations rather than gestures. This theory is supported by various studies on non-human primate communication, demonstrating the complex use of vocalizations to convey immediate, context-specific information, such as the nature and location of potential threats. Furthermore, primates' vocalizations exhibit certain foundational elements of grammar and structured conversational patterns, such as turn-taking, overlapping avoidance, and acoustic matching, that mirror those observed in human language. This collective evidence suggests that the vocal communication system in non-human primates could have laid the foundation for the human language.

7.1.3. The Multimodal Theory

Other research, however, challenges the notion that language evolution must be attributed to a single mode of communication and states that language simultaneously evolved from both gestures and speech (Taglialatela et al., 2011; Gillespie et al., 2014). To be more precise, the Multimodal theory posits that human language developed through the integrated evolution of gestural, vocal, facial expressions, and eye signaling systems (Prieur et al., 2020, p. 541). This theory is supported by empirical research indicating that these various modes of communication are not isolated but deeply interconnected in both humans and non-human primates, highlighting their collective role in the evolution of communicative systems (Morrill et al., 2012; Tada et al., 2013; Emery, 2000).

The study by Morrill et al. (2012) contributes to this theory by demonstrating that lip-smacking by monkeys mirrors the progression from infant babbling to adult human speech, both converging on a similar rhythmic frequency. In their study, researchers recorded and analyzed lip-smacking behaviors by rhesus monkeys across three distinct age groups: neonatal (infant), juvenile, and adult monkeys, focusing on the rhythm and variability of these movements. The study's results showed that, much like human infants, younger monkeys produced slower and more variable lip-smacks, which became faster and more consistent as they aged (see Figure 7.5, Figure 7.6, and Figure 7.7). Additionally, the study compared lip-smacking with another rhythmic mouth movement – chewing, which revealed that unlike lip-smacking, the frequency and variability of chewing did not change significantly as the monkeys grew.

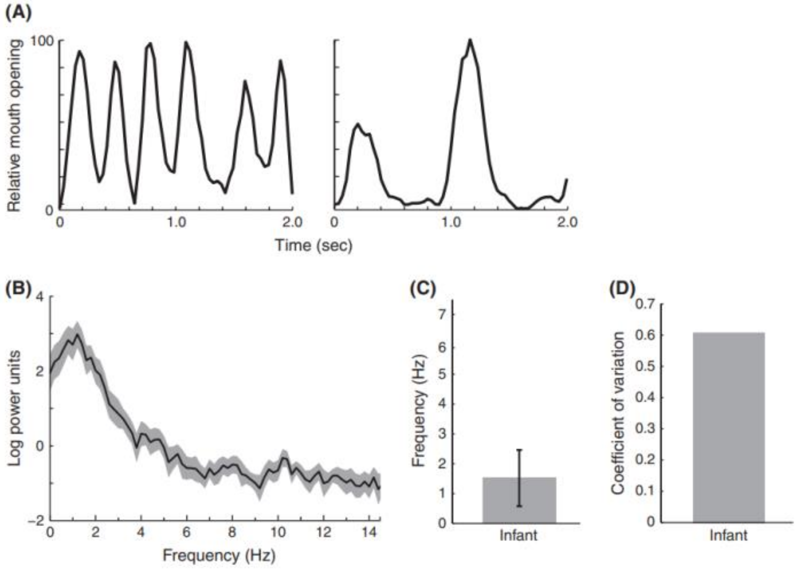


Figure 7.5 Rhythmic dynamics of neonatal lip-smacks (Morrill et al., 2012, p. 562)

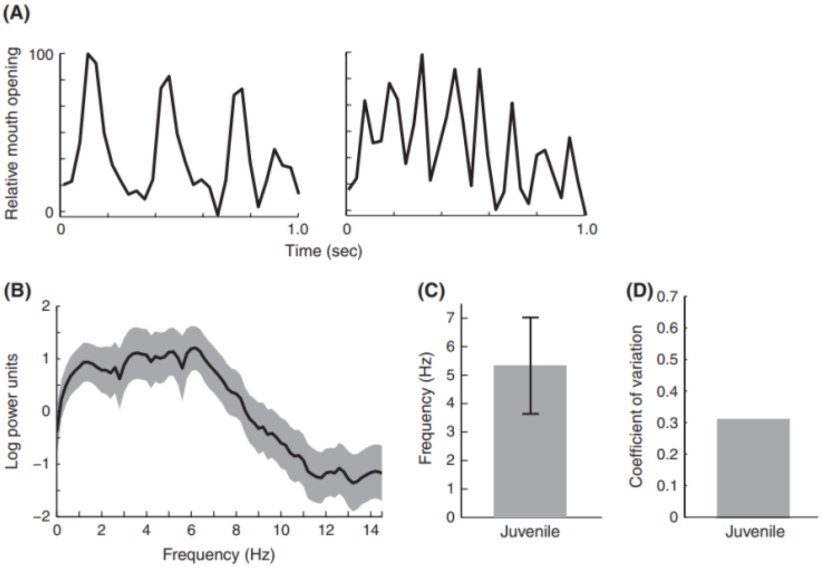


Figure 7.6 Rhythmic dynamics of juvenile lip-smacks (Morrill et al., 2012, p. 562)

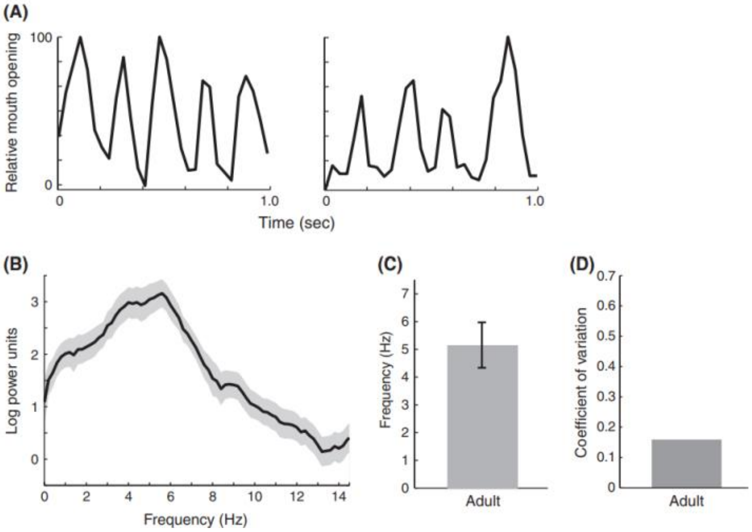


Figure 7.7 Rhythmic dynamics of adult lip-smacks (Morrill et al., 2012, p. 561)

Figures 7.5, 7.6, and 7.7 illustrate differences in the rate and consistency of lip-smacking across three developmental stages. Panel (A) in each stage shows the timing of mouth opening. It can be seen that infants open their mouths slower compared to juveniles and adults. Furthermore, the power spectrum (B) reveals the dominant lip-smacking frequencies, which are wide-ranging and less defined in neonates but narrow and more defined in juveniles and adults. Panel (C) shows the average frequency in all stages, confirming that neonates lip-smack at a slower rate, and this rate increases with age. Lastly, panel D illustrates consistency, with infants showing the most variation in their lip-smacking rhythm and adults the least, signifying a developmental change in their behavior.

Furthermore, research indicates that eye signaling (eye gazing or blinking) plays an essential role in both human and primates’ communication (Emery, 2000; Bard et al., 2005; Hömke et al., 2018). Tada et al. (2013) study supports this evidence by examining patterns and implications of eye-blink behaviors across a wide range of non-human primate species. In their experiment, researchers video-recorded 141 individual primates and analyzed these recordings to measure the blink rate, duration of blinks, and the proportion of blinks that occurred without accompanying eye or head movements (named ‘isolated blinks’). The findings revealed that blinking behaviors in primates varied widely among species and were influenced by body size and social dynamics. Specifically, results showed that larger primates blinked more frequently, had longer blink durations, and exhibited a higher percentage of isolated blinks compared to smaller primates.

Furthermore, the analysis revealed a significant relationship between blink rate and the size of the social group. To be more precise, primates living in larger groups tended to blink more frequently than those in smaller groups, hinting at the potential role of blinking in social communication and interaction (see Figure 7.8). In addition, the study showed that diurnal (active during daytime) primates blinked more often than nocturnal species, suggesting that environmental demands during their active hours may have an impact on their blinking patterns (see Figure 7.9).

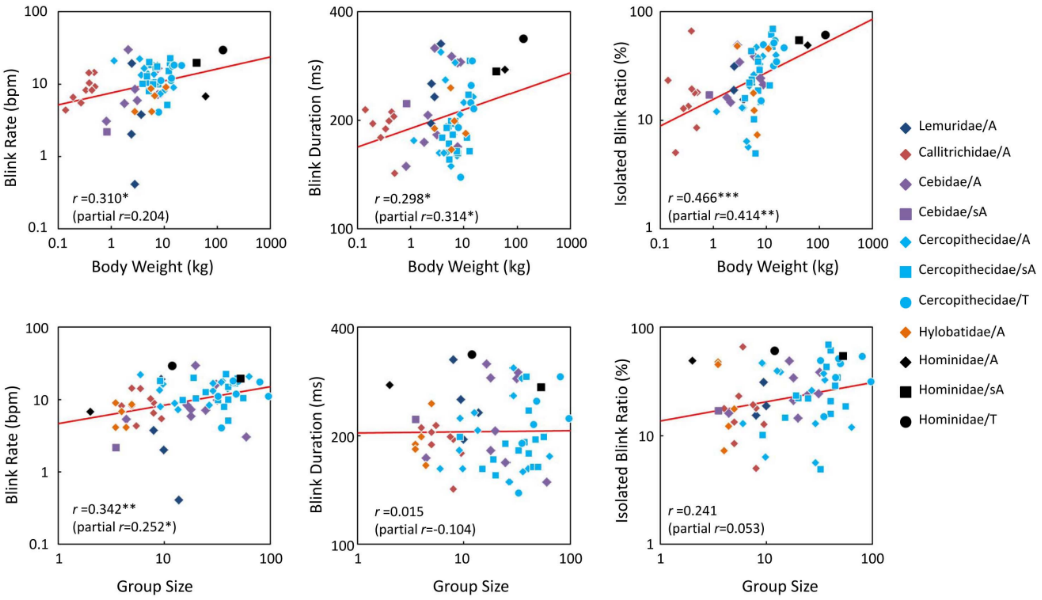


Figure 7.8 Scatterplots of eye-blink measures in primates by body size and group size (Tada et al., 2013, p. 7)

In Figure 7.8, the first row of scatterplots illustrates the relationship between the body weight of primates and three specific eye-blink measures: blink rate, blink duration, and isolated blink ratio. It can be clearly seen that as the body weight increases, so does the blink rate and blink duration, and a higher proportion of blinks occur without concurrent eye or head movements (isolated blinks). The second row depicts the same blinking measures against the size of the social group to which the primates belong. It can be seen that a positive correlation occurs between group size and blink rate, suggesting that primates in larger social groups tend to blink more frequently. In contrast, the relationship between group size and blink duration or isolated blink ratio is less pronounced or non-significant, as the lower correlation coefficients indicate.

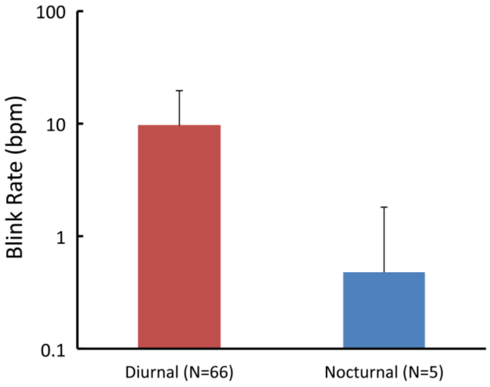


Figure 7.9 Comparison of mean blink rates between diurnal and nocturnal primates (Tada et al., 2013, p. 5)

In essence, the multimodal theory challenges the unimodal explanations for language evolution, such as the gesture-first and the vocal-first theories by proposing a more integrated approach to the development of human communication. Specifically, it posits that language evolved from the interconnected evolution of gestural, vocal, facial, and eye signaling systems. This theory is supported by various research demonstrating similarities between primate and human communication methods, such as the rhythm of lip-smacking or the use of eye-signaling in social interactions. Such evidence suggests that the foundations of language can be traced back to multimodal communicative strategies, prevalent not only in non-human primates but also inherent to human language.

7.1.4. Conclusion and Discussion

This section has explored the evolutionary origins of human language by examining three main predominant theories: the Gesture-First theory, the Vocal-First theory, and the Multimodal theory. Each of these theories offers unique insights into the complex pathways through which human language might have evolved, highlighting different aspects of both human and non-human primate communication.

The Gesture-First theory, as the name suggests, posits that human language developed from manual gestures, which were a primary form of communication among early hominins. This theory is supported by evidence from various studies on non-human primates, which demonstrate that their gestures are not only very flexible but also richly contextual. To be more precise, non-human primates use a vast amount of gestures which are intentional, referential, involve turn-taking, comply with linguistic principles, and exhibit iconicity, which shares a resemblance to human communication. Additionally, neurological evidence, particularly the discovery of mirror neurons, significantly bolsters this theory. Mirror neurons, which fire both when an individual carries out an action and when they observe the same action performed by others, were found in regions of the brain analogous to Broca's area in humans – a key area involved in language processing. These neurons allow primates to understand and replicate gestures, facilitating a pre-linguistic method of communicating intentions and meaning, much like the basic functions of language. However, despite substantial support for the theory of mirror neurons, considerable skepticism remains regarding their direct role in language evolution. Critics highlight the absence of direct evidence supporting the idea that mirror neurons are capable of supporting the complex cognitive processes required for language evolution. Additionally, anatomical and functional correlations between humans and non-human primates, such as cerebral lateralization and a preference for right-handedness, suggest that gestures could be linked to how language developed in the brain.

Vocal theory, on the other hand, suggests that language evolved not from gestures but rather from vocalizations. The evidence from studies on primates, including titi monkeys and chimpanzees, reveal that their vocalizations are not random but possess specific patterns and meanings, indicating a primitive form of 'syntax'. This complexity in primate calls, which varies according to the context and the nature of the threat, suggests an evolutionary pathway towards more complex language systems seen today. For instance, the vocal strategies observed in non-human primates involve combinations of calls that are context-specific and carry detailed information, which parallels the syntactic structures in human speech. Such structured vocal patterns, found in the orderly sequences of chimpanzee calls, imply an inherent capacity for grammatical construction that could predate human language. Furthermore, the vocal interactions among bonobos, characterized by turn-taking and a preference for responding to familiar individuals, mirror the social aspects of human communication, indicating that the social use of language also has deep evolutionary roots.

Lastly, the Multimodal theory of language evolution presents a view that language developed not from a singular mode of communication but from a blending of several systems, including gestures, vocalizations, facial expressions, and eye signals. This theory posits that these modes of communication co-evolved, enhancing the complexity and effectiveness of interaction both in humans and non-human primates. This view is supported by research indicating that various communication strategies are deeply intertwined; supporting the notion that human language is an integrated system rather than a byproduct of one dominant communicative form. Studies like the one by Morrill et al. (2012) demonstrate how behaviors like lip-smacking in monkeys resemble infant babbling in humans, evolving in complexity and speech as they mature, which suggests that vocal rhythm might have played a critical role in the language development process. This idea is complemented by studies on eye signaling, which suggest that non-verbal cues like eye movements and blinking patterns are not just supplementary but pivotal in communication. They contribute to the regulation of social interactions and convey nuanced messages, which are essential to human communication.

In essence, while each theory presents compelling arguments and evidence, the intricate relationship between speech and gestures discussed in the previous chapters gives further support that language evolution was likely a multifaceted process involving an integrated development of various communication modes. The findings from behavioral and neurobiological studies of both human and non-human primates indicate that neither gestures alone nor vocalizations in isolation fully account for the complexities of human language. Instead, as the Multimodal theory proposes, the simultaneous evolution of gestures, vocalizations, facial expressions, and even eye signaling likely contribute to the sophisticated communicative systems that exist today. While the present thesis lends support to the Multimodal theory as a plausible explanation for the origins of human language, further research is required in order to more confidently delineate the pathways of language evolution.

8. Conclusion

The present thesis explored the intricate relationship between gestures and speech at a neural level, uncovering how these two modes of communication are not merely interconnected but are co-dependent systems that enhance human interaction. More specifically, the thesis focused on three main objectives, which aimed to deepen the understanding of how gestures and speech are processed and integrated by the brain and what this might indicate about the evolutionary origins of human language.

The first objective aimed to identify specific brain regions involved in processing speech and gestures. In order to achieve this goal, various fMRI and EEG/ERP studies were discussed and critically analysed. The evidence from the studies demonstrated that brain areas: Inferior Frontal Gyrus (IFG), the Superior Temporal Gyrus (STG), the Middle Temporal Gyrus (MTG), and Broca's area are not only crucial for language processing but also for interpreting and integrating co-speech gestures. Results regarding IFG showed that both parts of this area responded to semantic incongruence but in slightly different ways. To be more precise, the left IFG was mainly concerned with processing abstract information and integrating complex inputs for coherent comprehension. Meanwhile, the right IFG resolved semantic mismatches between speech and accompanying gestures, ensuring that communication remains clear and effective. Similar findings were observed in Broca's area as during semantic mismatch between gestures and speech, this area of the brain showed increased activation, which demonstrates its critical role in integrating and analyzing information from multiple sources. Additionally, when gestures semantically aligned with spoken content, Broca's area exhibited reduced activity. This evidence shows that gestures are essential to communication as they can ease the cognitive demands of language processing. Furthermore, STG was found to be exceptionally responsive to gesture-speech integration when auditory conditions were challenging. This shows that STG plays an integral part in multimodal communication as it gathers information from both speech and gestures to comprehend and enhance information. Lastly, MTG showed increased activation during semantic processing in situations when gestures metaphorically aligned with spoken language and in cases when they visually complemented verbal information. Moreover, the results from the EEG/ERP studies provided complementary insights, revealing that when gestures and speech were incongruent, the brain showed heightened semantic processing demands, as evidenced by increased activation of the N400 component. These findings from both fMRI and EEG/ERP studies show that the relationship between gestures and speech is complex and multifaceted, affecting how people gather and comprehend information.

The second objective of the present thesis aimed to elucidate how different types of co-speech gestures, namely metaphoric, iconic, beat, and deictic, influence neural activity. However, due to a lack of research done on deictic gestures, only three of them were discussed. The analysis of fMRI studies revealed that each type of gesture stimulated distinct brain regions. Particularly, gestures tended to evoke activations in those brain areas that corresponded to their cognitive functions. For instance, metaphoric gestures, which carry abstract meaning, elicited activations in both the left IFG and left MTG, brain areas primarily associated with comprehension and production of spoken language. However, the most noticeable response was found in the left IFG. This evidence shows that as metaphoric gestures are not so ‘straightforward’, they evoke responses in brain areas that are responsible for decoding given information. Iconic gestures, directly representing objects or actions, have been found to elicit activations in several brain regions: the right STG, the left MTG, the STS, the bilateral inferior parietal lobule, and the bilateral ventral precentral sulcus. These regions are responsible for various cognitive processes, including language, perception, and action, which align with the role of iconic gestures. Therefore, these findings further support the idea that specific gestures are processed by specific neural networks systematically rather than randomly. The results regarding beat gestures, reflecting the rhythm of the speech, were no exception. They tended to elicit an increased response in brain regions associated with auditory processing: the bilateral auditory cortex, left STG/S, and the right PT. In essence, even though the present thesis came to the conclusion that gestures tend to evoke responses in brain areas that seem to be the most reflective of their functions, it is difficult to claim it as a fact. More in-depth research and analysis are needed to determine this complex processing of gesture-speech integration.

The third objective aimed to investigate whether existing studies on non-human primates could shed light on the potential evolutionary origins of human language. More specifically, the study aimed to determine whether human language evolved from gestural, vocal, or multimodal forms of communication. The gestural theory revolves around the view that in the early stages of human evolution, the early ancestors, or hominins, primarily used manual gestures to communicate, eventually laying the foundation for the present language. This view is supported by multiple studies on non-human primates, showing that primates' gestures are highly flexible and context-dependent, displaying intentional, referential, iconic, and turn-taking qualities, resembling human communication. Additionally, neurological findings, such as discovering mirror neurons, cerebral lateralization, and handedness in both humans and non-human primates, provide further arguments for this theory. However, even though it is clear that gestures may have been crucial in language evolution, the role of vocalizations cannot be overlooked. Findings have shown that, similarly to gestures, primates' vocalizations are complex and context-dependent. More specifically, they can carry detailed information about the environment surrounding them. Even though their vocalizations are not as flexible as gestures, they can easily alert others about potential threats, indicate their preferences and desires, or, in general, have a meaningful conversation by involving turn-taking. Overall, these findings demonstrate that both gestures and vocalizations observed in non-human primates' communication resemble human interaction, which thus leads to multimodal theory. This theory revolves around the view that not only one but multiple forms of communication, such as gestural, vocal, facial, and eye signalling, have played a part in the origins of human language. This view is supported not only by research on non-human primates but also by various neuroimaging studies, which were discussed in the present thesis. Specifically, it is clear that human communication is inherently multimodal; people use gestures when they speak, even if they do not realize it. To claim that only gestures or speech is the foundation of the present language would be misleading. Therefore, the current thesis comes to the conclusion that as human communication is a sophisticated and complex process, it must have evolved in a complicated way involving multiple modes of communication.

8.1. Limitations and Suggestions for Further Research

Although the current research has provided valuable insights into the neural processing of speech and gestures, it is important to mention several limitations inherent in the study.

The present thesis primarily revolved around analysing and discussing various studies employing fMRI and EEG/ERP methodologies. While these studies have provided profound insights, it is crucial to acknowledge that their findings were often constrained by specific experimental conditions and settings. These controlled environments may not fully mirror the complexities of naturalistic communication, potentially limiting the validity of the results.

Moreover, the studies discussed in the present thesis mostly used a narrow demographic, primarily consisting of right-handed young adults. While this demographic choice may have facilitated controlled experimental conditions, it also raises concerns regarding the broader applicability of the findings across different age groups and populations.

Additionally, the methodologies employed in the studies under review inherently come with their own set of limitations, which were briefly described in Chapter 5. While fMRI offers excellent spatial resolution, it lacks the temporal precision necessary to capture the rapid dynamics of speech and gesture integration. On the other hand, EEG/ERP provides superior temporal resolution but suffers from poorer spatial localization. Thus, a comprehensive understanding/analysis of the neural processing of speech and gesture integration necessitates a multi-modal approach that mitigates the limitations of individual methodologies.

In light of these considerations, future research should address these limitations by incorporating more diverse participants, employing several neuroimaging techniques, and embracing ecologically valid experimental paradigms. Furthermore, as the gesture-speech relationship is complex and multifaceted, it requires a more in-depth analysis and research to fully comprehend how these two modes of communication interact. While previous research has explored metaphoric, iconic, and beat gestures, there is a notable gap in the research regarding deictic gestures. Therefore, future research should focus on all types of co-speech gestures in order to gain a more comprehensive understanding of how gestures and speech interact at a neural level.

9. List of References

1. Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning Foreign Language Vocabulary with Gestures and Pictures Enhances Vocabulary Memory for Several Months Post-Learning in Eight-Year-Old School Children. *Educational Psychology Review, 32*(4), 815-850.
2. Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia, 51*(8), 1619-1629.
3. Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28*(2), 105–124.
4. Arbib, M. A., Liebal, K., Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology, 49*(6), 1053–1076.
5. Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences, 25*(1), 27–31.
6. Bard, K. A., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Costall, A., & Matsuzawa, T. (2005). Group differences in the mutual gaze of chimpanzees (Pan troglodytes). *Developmental Psychology, 41*(4), 616-624.
7. Bernard, J. A., Millman, Z. B., & Mittal, V. A. (2015). Beat and Metaphoric Gestures are Differentially Associated with Regional Cerebellar and Cortical Volumes. *Human Brain Mapping, 36*, 4016-4030.
8. Bhaya-Grossman, I., & Chang, E. F. (2022). Speech computations of the human superior temporal gyrus. *Annual Review of Psychology, 73*, 79-102.
9. Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons, Papio anubis, adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour, 87*, 121-128.
10. Briggs, R. G., Tanglay, O., Dadario, N. B., Young, I. M., Fonseka, R. D., Hormovas, J., Dhanaraj, V., Lin, Y.-H., Kim, S. J., Bouvette, A., Chakraborty, A. R., Milligan, T. M., Abraham, C. J., Anderson, C. D., O'Donoghue, D. L., & Sughrue, M. E. (2021). The

Unique Fiber Anatomy of Middle Temporal Gyrus Default Mode Connectivity. *Operative Neurosurgery, 21*(1), E8-E14.

11. Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making Children Gesture Brings Out Implicit Knowledge and Leads to Learning. *Journal of Experimental Psychology: General, 136*(4), 539-550.
12. Brown, S., & Yuan, Y. (2018). Broca's area is jointly activated during speech and gesture production. *Neuroreport, 29*(14), 1214-1216.
13. Butterworth, B., & Beattie, G. (1978). Gesture and Silence as Indicators of Planning in Speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language* (pp. 347-360). New York: Plenum.
14. Buxton, R. B. (2013). The physics of functional magnetic resonance imaging (fMRI). *Reports on Process in Physics, 76*(9), Article 096601. https://doi.org/10.1088/0034-4885/76/9/096601
15. Call, J., & Tomasello, M. (2007). *The Gestural Communication of Apes and Monkeys.* Lawrence Erlbaum Associates.
16. Call, J., & Tomasello, M. (2007). The gestural repertoire of chimpanzees (Pan troglodytes). In J. Call & M. Tomasello (Eds.), *The Gestural Communication of Apes and Monkeys* (pp. 17-39). Lawrence Erlbaum Associates.
17. Campisi, P., & La Rocca, D. (2014). Brain waves for automatic biometric-based user recognition. *IEEE Transactions on Information Forensics and Security, 9*(5), 782-800.
18. Cäsar, C., Zuberbühler, K., Young, R. J., & Byrne, R. W. (2013). Titi monkey call sequences vary with predator location and type. *Biology Letters, 9*(5), Article 20130535. https://doi.org/10.1098/rsbl.2013.0535
19. Caspers, S., Schleicher, A., Bacha-Trams, M., Palomero-Gallagher, N., Amunts, K., & Zilles, K. (2013). Organization of the Human Inferior Parietal Lobule Based on Receptor Architectonics. *Cerebral Cortex, 23*(3), 615-628. https://doi.org/10.1093/cercor/bhs048
20. Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences, 26*(2), 199-260.
21. Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition, 70*(1), 42-52.

22. Coulson, S. (2007). Electrifying results: ERP data and cognitive linguistics. In M. González-Marquez, I. Mittelberg, S. Coulson, & M. J. Spivey (Eds.), *Methods in cognitive linguistics*, (pp. 400–423). John Benjamins Publishing Company.
23. Coye, C., Zuberbühler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female Diana monkeys. *Animal Behaviour, 115*, 97-105.
24. Crockford, C., & Boesch, C. (2005). Call Combinations in Wild Chimpanzees. *Behaviour, 142*(4), 397-421.
25. Crockford, C., Wittig, R. M., & Zuberbühler, K. (2015). An intentional vocalization draws others' attention: A playback experiment with wild chimpanzees. *Animal Cognition, 18*(3), 581–591.
26. Dary, Z., Lenggenhager, B., Lagarde, S., Medina Villalon, S., Bartolomei, F., & Lopez, C. (2023). Neural bases of the bodily self as revealed by electrical brain stimulation: A systematic review. *Human Brain Mapping, 44*(7), 2936-2959.
27. Detre, J. A., & Wang, J. (2002). Technical aspects and utility of fMRI using BOLD and ASL. *Clinical Neurophysiology, 113*(5), 621–634.
28. 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.
29. Dick, A. S., Mok, E. H., Raja Beharelle, A., Goldin-Meadow, S., & Small, S. L. (2014). Frontal and Temporal Contributions to Understanding the Iconic Co-Speech Gestures That Accompany Speech. *Human Brain Mapping, 35*(3), 900-917.
30. Douglas, P. H. & Moscovice, L. R. (2015). Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. *Scientific Reports, 5*, Article 13999. https://doi.org/10.1038/srep13999
31. Ekman, P., & Friesen, W. V. (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica, 1*(1), 49-98.
32. Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604.
33. Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology, 23*(3), 171-179.

34. Fleury, M., Figueiredo, P., Vourvopoulos, A., & Lécuyer, A. (2023). Two is better? Combining EEG and fMRI for BCI and neurofeedback: A systematic review. *Journal of Neural Engineering, 20*(5), Article 051003. https://doi.org/10.1088/1741-2552/ad06e1
35. Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., Goebel, R. (2003). Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. *Neuron, 40*(4), 859–869.
36. Friederici, A. D. (2017). *Language in Our Brain: The Origins of a Uniquely Human Capacity*. Cambridge: MIT Press.
37. Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. *Biological Reviews, 94*(5), 1809–1829. https://doi.org/10.1111/brv.12535
38. Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews, 30*(7), 949–960.
39. Gentilucci, M., Bernardis, P., Crisi, G., Dalla Volta, R. (2006). Repetitive Transcranial Magnetic Stimulation of Broca's Area Affects Verbal Responses to Gesture Observation. *Journal of Cognitive Neuroscience, 18*(7), 1059– 1074.
40. Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology, 24*(14), 1601–1605.
41. Gillespie-Lynch, K., Greenfield, P. M., Lyn, H., & Savage-Rumbaugh, S. (2014). Gestural and symbolic development among apes and humans: Support for a multimodal theory of language evolution. *Frontiers in Psychology, 5*, Article 1228. https://doi.org/10.3389/fpsyg.2014.01228
42. Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research, 21*(5), 609–631.
43. Goldin-Meadow, S. (2018). Taking a Hands-on Approach to Learning. *Policy Insights from the Behavioral and Brain Sciences, 5*(2), 163-170.
44. Goldin-Meadow, S., & Mylander, C. (1983). Gestural Communication in Deaf Children: Noneffect of Parental Input on Language Development. *Science*, *221*(4608), 372-374.

45. Graham, K. E., & Hobaiter, C. (2019). Gestural communication in the great apes. In J. C. Chloe (Ed.), *Encyclopedia of Animal Behavior* (2nd ed., Vol. 1, pp. 371-377). Elsevier Academic Press.
46. Graham, K. E., Badihi, G., Safryghin, A., Grund, C., & Hobaiter, C. (2022). A socio-ecological perspective on the gestural communication of great ape species, individuals, and social units. *Ethology Ecology & Evolution, 34*(3), 235-259.
47. Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS Biology, 16*(2), e2004825. https://doi.org/10.1371/journal.pbio.2004825
48. Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K. & Kircher, T. (2009). Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapping, 30*(10), 3309-3324.
49. Häberling, I. S., Corballis, P. M., & Corballis, M. C. (2016). Language, gesture, and handedness: Evidence for independent lateralized networks. *Cortex, 82*, 72-85.
50. Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845-1854.
51. Hadar, U., Wenkert-Olenik, D., Krauss, R., & Soroker, N. (1998). Gesture and the Processing of Speech: Neuropsychological Evidence. *Brain and Language, 62*(1), 107–126.
52. Hans, A., & Hans, E. (2015). Kinesics, Haptics and Proxemics: Aspects of Non-Verbal Communication. *IOSR Journal of Humanities and Social Science, 20*(2), 47-52.
53. Hay, L., Duffy, A. H. B., Gilbert, S. J., & Grealy, M. A. (2022). Functional magnetic resonance imaging (fMRI) in design studies: Methodological considerations, challenges, and recommendations. *Design Studies, 78*, Article 101078. https://doi.org/10.1016/j.destud.2021.101078
54. Hickok, G. (2009). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience, 21*(7), 1229-1243. https://doi.org/10.1162/jocn.2009.21189
55. Hişmanoğlu, M. & S. Hişmanoğlu. (2008). The use of body language in foreign language learning and teaching. *Journal of Social Sciences, 19*, 165-178.

56. Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology, 24*(14), 1596-1600.

57. Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage, 39*, 2010-2024.

58. 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.

59. Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS ONE, 13*(12), e0208030. https://doi.org/10.1371/journal.pone.0208030

60. Hopkins, W. D., & Cantero, M. (2003). From hand to mouth in the evolution of language: the influence of vocal behavior on lateralized hand use in manual gestures by chimpanzees (Pan Troglodytes). *Developmental Science, 6*(1), 55-61.

61. Hopkins, W. D., Russel, J., Freeman, H., Buehler, N., Reynolds, E., & Schapiro, S. J. (2005). The distribution and development of handedness for manual gestures in captive chimpanzees (Pan Troglodytes). *Psychological Science, 16*(6), 487-493.

62. Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495-514.

63. Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., & Brugge, J. F. (2000). Auditory Cortex on the Human Posterior Superior Temporal Gyrus. *The Journal of Comparative Neurology, 416*(1), 79–92.

64. 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*(3), 1028-1037. https://doi.org/10.1002/hbm.20565

65. Ishkhanyan, B., Lange, V. M., Boye, K., Mogensen, J., Karabanov, A., Hartwigsen, G., & Siebner, H. R. (2020). Anterior and Posterior Left Inferior Frontal Gyrus Contribute to the Implementation of Grammatical Determiners During Language Production. *Frontiers in Psychology*, *11*, Article 685. https://doi.org/10.3389/fpsyg.2020.00685

66. Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Sciences, 9*(1), 21-25.

67. Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253-260.
68. 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.
69. Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key. (Ed.), *The relationship of verbal and non-verbal communication* (pp. 207-227). Mouton.
70. Kendon, A. (1988). How gestures can become like words. In F. Poyatos (Ed.), *Cross-cultural perspectives in nonverbal communication* (pp. 131-141). Toronto: Hogrefe.
71. Key, A. P. (2016). Human auditory processing: Insights from cortical event-related potentials. *AIMS Neuroscience, 3*(2), 141-162.
72. Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: an ERP primer. *Developmental Neuropsychology, 27*(2), 183-215.
73. 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 metaphoric co-verbal gestures. *Neuropsychologia, 47*(1), 169-179.
74. Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162-185). Cambridge: Cambridge University Press.
75. Knecht, S., Deppe, M., Dräger, B., Bobe, L., Lohmann, H., Ringelstein, E. B., & Henningsen, H. (2000). Language lateralization in healthy right-handers. *Brain, 123*(1), 74-81.
76. Krahmer, E. J., & Swerts, M. G. J. (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.
77. Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign* (pp. 93-116). Oxford: Oxford University Press.
78. Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience, 33*(8), 992-1004.

79. Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

80. Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J. P., & Lemasson, A. (2019). Social bonding drives vocal exchanges in Bonobos. *Scientific Reports, 9*, Article 711. https://doi.org/10.1038/s41598-018-36024-9

81. Maille, A., Engelhart, L., Bourjade, M. & Blois-Heulin, C. (2012). To beg, or not to beg? That is the question: mangabeys modify their production of requesting gestures in response to human's attentional states. *PLoS One, 7*(7), Article e41197. https://doi.org/10.1371/journal.pone.0041197

82. Marstaller, L., & Burianová, H. (2014). The multisensory perception of co-speech gestures – A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics, 30*, 69-77. https://doi.org/10.1016/j.jneuroling.2014.04.003

83. McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: The University of Chicago Press.

84. Mertens, U. J., & Rohlfing, K. J. (2021). Progressive Reduction of Iconic Gestures Contributes to School-Aged Children's Increased Word Production. *Frontiers in Psychology, 12*, Article 651725. https://doi.org/10.3389/fpsyg.2021.651725

85. Meunier, H., Fizet, J., & Vauclair, J. (2013). Tonkean macaques communicate with their right hand. *Brain and Language, 126*(2), 181-187.

86. Morales, S., & Bowers, M. E. (2022). Time-frequency analysis methods and their application in developmental EEG data. *Developmental Cognitive Neuroscience, 54*, Article 101067. https://doi.org/10.1016/j.dcn.2022.101067

87. Morett, L. M. (2014). When Hands Speak Louder Than Words: The Role of Gesture in the Communication, Encoding, and Recall of Words in a Novel Second Language. *The Modern Language Journal, 98*(3), 834–853.

88. Morett, L. M., & Fraundorf, S. H. (2019). Listeners consider alternative speaker productions in discourse comprehension and memory: Evidence from beat gesture and pitch accenting. *Memory & Cognition, 47*(8), 1515-1530.

89. Morrill, R. J., Paukner, A., Ferrari, P. F., & Ghazanfar, A. A. (2012). Monkey lipsmacking develops like the human speech rhythm. *Developmental Science, 15*(4), 557–568.

90. Murphy, C., Rueschemeyer, S. A., Watson, D., Karapanagiotidis, T., Smallwood, J., & Jefferies, E. (2017). Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. *NeuroImage, 147*, 19–31.
91. Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, *24*(5), 283-288.
92. Nagels, A., Chatterjee, A., Kircher, T., Straube, B. (2013). The role of semantic abstractness and perceptual category in processing speech accompanied by gestures. *Frontiers in Behavioral Neuroscience, 7*, Article 181. https://doi.org/10.3389/fnbeh.2013.00181
93. Nishida, T., Zamma, K., Matsusaka, T., Inaba, A., & McGrew, W. C. (2010). *Chimpanzee behavior in the wild: An audio-visual encyclopedia*. Springer.
94. Numssen, O., Bzdok, D., & Hartwigsen, G. (2021). Functional specialization within the inferior parietal lobes across cognitive domains. *eLife, 10*, Article e63591. https://doi.org/10.7554/eLife.63591
95. Nunez, P. L., & Srinivasan, R. (2009). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press.
96. Oon, H. N., Saidatul, A., & Ibrahim, Z. (2018). Analysis on non-linear features of electroencephalogram (EEG) signal for neuromarketing application. In *2018 International Conference on Computational Approach in Smart Systems Design and Applications* (ICASSDA). Kuching, Malaysia. http://dx.doi.org/10.1109/ICASSDA.2018.8477618
97. Özyürek, A. (2002). Speech-gesture relationship across languages and in second language learners: Implications for spatial thinking and speaking. In B. Skarabela, S. Fish, & A. H. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development* (pp. 500-509). Somerville, MA: Cascadilla Press.
98. 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.
99. Papadatou-Pastou, M. (2011). Handedness and language lateralization: Why are we right-handed and left-brained? *Hellenic Journal of Psychology, 8*, 248-265.

100. Pika, S., & Mitani, J. C. (2006). Referential gestural communication in wild chimpanzees (Pan Troglodytes). *Current Biology, 16*(6), 191–192.

101. Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (Gorilla gorilla): Gestural repertoire, learning, and use. *American Journal of Primatology, 60*, 95–111.

102. Pollick, A. S, de Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences of the United States of America, 104*(19), 8184-8189.

103. Pontevendra, M. S. (2022). Function and Association of Auditory Cortex with Brain. *Bipolar Disorder: Open Access, 8*(4), Article 184.

104. Prieur, J., Barbu, S., Blois-Heulin, C., & Lemasson, A. (2020). The origins of gestures and language: History, current advances, and proposed theories. *Biological Reviews*, 95(3), 531-554. https://doi.org/10.1111/brv.12576

105. Proverbio, A. M., Crotti, N., Zani, A., & Adorni, R. (2009). The role of left and right hemispheres in the comprehension of idiomatic language: An electrical neuroimaging study. *BMC Neuroscience, 10*, Article 116. https://doi.org/10.1186/1471-2202-10-116

106. Ramos-Cabo, S., Vulchanov, V., & Vulchanova, M. (2019). Gesture and Language Trajectories in Early Development: An Overview From the Autism Spectrum Disorder Perspective. *Frontiers in Psychology, 10*, Article 1211. https://doi.org/10.3389/fpsyg.2019.01211

107. Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *TINS, 21*(5), 188–194.

108. Rojas, G. M., Alvarez, C., Montoya, C. E., de la Iglesia-Vayá, M., Cisternas, J. E., & Gálvez, M. (2018). Study of resting-state functional connectivity networks using EEG electrodes position as seed. *Frontiers in Neuroscience, 12*, Article 235. https://doi.org/10.3389/fnins.2018.00235

109. Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture’s role in predicting vocabulary development. *First Language, 28*(2), 182-199.

110. Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology, 54*(1), 145-173.

111. Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language, 101*(3), 260-277.

112. Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology, 15*(19), 1779–1784.

113. Steines, M., Nagels, A., Kircher, T., & Straube, B. (2021). The role of the left and right inferior frontal gyrus in processing metaphoric and unrelated co-speech gestures. *NeuroImage, 237*, Article 118182. https://doi.org/10.1016/j.neuroimage.2021.118182

114. Stigler, K. A., & McDougle, C. J. (2013). Structural and Functional MRI Studies of Autism Spectrum Disorders. In J. D. Buxbaum & P. R. Hof (Eds.), *The Neuroscience of Autism Spectrum Disorders* (pp. 251-266). Oxford, UK: Elsevier.

115. Straube, B., Green, A., Bromberger, B., & Kircher, T. (2011). The differentiation of iconic and metaphoric gestures: Common and unique integration processes. *Human Brain Mapping, 32*(4), 520–533.

116. Subha, D. P., Joseph, P. K., Acharya, U. R., & Lim, C. M. (2010). EEG signal analysis: A survey. *Journal of Medical Systems, 34*(2), 195–212. https://doi.org/10.1007/s10916-008-9231-z

117. Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal, 18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865

118. Tada, H., Omori, Y., Hirokawa, K., Ohira, H., & Tomonaga, M. (2013). Eye-blink behaviors in 71 species of primates. *PLoS ONE, 8*(5), e66018. https://doi.org/10.1371/journal.pone.0066018

119. Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PLoS ONE, 6*(4), e18852. https://doi.org/10.1371/journal.pone.0018852

120. Vilà-Giménez, I., Dowling, N., Demir-Lira, Ö. E., Prieto, P., & Goldin-Meadow, S. (2021). The predictive value of non-referential beat gestures: Early use in parent-child interactions predicts narrative abilities at 5 years of age. *Child Development, 92*(6), 2335-2355.

121. 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.

122. 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.

123. Wilson, S. M., Molnar-Szakacs, I., Iacoboni, M. (2008). Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension. *Cerebral Cortex, 18*(1), 230–242.

124. Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654-667.

125. Zuberbühler, K., Ouattara, K., Bitty, A., Lemasson, A., & Noë, R. (2009). The primate roots of human language: Primate vocal behavior and cognition in the wild. In F. d'Errico & J. M. Hombert (Eds.), *Becoming eloquent: Advances in the emergence of language, human cognition, and modern culture* (pp. 235-264). John Benjamins Publishing Company.

Appendix: Searching Strategies

To find published sources on the neural processing of speech and gestures, the following databases and providers were searched intermittently between October 2022 and May 2024.

1. NTNU – *NTNU Universitetsbiblioteket*
2. JSTOR – *JSTOR: Journal Storage*
3. ScienceOpen Inc. – *ScienceOpen*

Additionally, other search engines were used to find various studies and books, which are listed below.

1. Google Scholar
2. CiteSeerX
3. archive.org

Search terms

The numbered search terms below were used to find relevant studies. The search strategy involved using both individual key words and their combinations.

1. co-speech gestures; gestures accompanying speech; primates gestures; non-human primates gestures; classification; gesture types
2. gestures; speech; language acquisition; language acquisition; non-human primates vocalizations
3. brain functions; human brain; language processing areas; STG; MTG; IFG; Broca's area; PT; parietal lobule; brain lobes; neural processing; neural bases; neural networks
4. neuroimaging techniques; fMRI; EEG; ERP; EEG/ERP; ERP components; N400; negative deflection;
5. language origins; language evolution; evolutionary perspective; evolution theories; gesture theories; vocal theories; multimodal theories

NTNU

Norwegian University of
Science and Technology