THE UNIVERSITY OF HULL

An exploration of the integration of speech with co-speech gesture with Non-Invasive Brain Stimulation

Thesis submitted for the Degree of Doctor of Philosophy

In Psychology at the University of Hull

Wanying Zhao

September 2017

Acknowledgements

I am not a good talker, I also don’t know how to express my feelings. Maybe it’s due to the family culture I grew up in, from a child, I was told to ‘work hard and talk less, the facts will speak by themselves’. But still, I would like to take this opportunity to thank everyone who has helped and supported me in the last three years, from the bottom of my heart.

First of all, I would like to thank Dr. Henning Holle, my first supervisor, who has led and guided me in this psychological research. Thank you for enduring my ‘Chinese-English’ accent, thank you for answering my smallest questions, thank you for being patient and tolerant with my stubbornness. In my father’s words ‘I could never have found a better supervisor with such patience’.

I would like to thank my second supervisor, Dr. Kevin Riggs. Thank you for all the advice and comments you provided, you are of my father’s age, and I always feel so close to you. In addition, to my ‘third supervisor’, Dr. Igor Schindler, thank you for all the useful feedback on conducting and planning experiments.

Secondly, I would like to thank Mrs. Kanthilatha Hettigoda Gamage, my best friend, roommate, colleague, and moral supporter. Thank you for keeping me when I was homeless, thank you for being still with me despite all my laziness and occasionally being ‘out of my mind’. I feel so lucky to have you in my life.

Thirdly, I would like to thank Hull-China Scholarship for my tuition waiver, and the Chinese Scholarship Council for providing me with living expenses. Without all this help, I could not have afforded.

Fourthly, I would like to express my apologies to my family, for not being with them in need. I thank all their friends, relatives, acquaintances and even strangers for their help, support and company.

Lastly, I would like to thank everyone I have met; the colleagues in the Psychology department, the friends I have in the UK, and the neighbour who smiles at me every morning (I don't even know her name). Thank you all for making this journey so enjoyable and lovely, to make this totally strange country not so 'strange and beyond reach'.

The world is a better place because of people's kindness. If I ever become a better person, it is only because of you all.

Wanying Zhao

09/2017

Abstract

The current PhD project focuses on the integration of gesture with their co-occurring speech with the use of non-invasive brain stimulation. The project investigated ‘where’ and ‘when’ gesture-speech integration takes place. Building on the paradigm of Kelly et al., (2010) which provides a reaction time index of automatic gesture-speech integration, it was tested whether left middle temporal gyrus (pMTG) as well as left Inferior frontal gyrus (LIFG) are causally involved in gesture-speech integration. A follow-up study investigated the time window for this integration of gesture and speech in pMTG. This study found that gesture has a priming effect on the semantic retrieval of speech. This effect only manifested itself after gesture had been clearly understood and before the semantic analysis of speech. Based on the common coding hypothesis, this finding was interpreted in terms of gesture and speech originating from a common coding system, with both LIFG and pMTG as its neural underpining, enabling bi-directional influences between both domains.

Contents

Figures

Tables

Chapter 1. General introduction

Language has a standard of form and structure and conveys meaning in a linear-segmented way. Words are composed of morphemes with a strict and systematic linear sequence, sentences are composed of words with a layered structure, and discourse is composed of sentences governed by the syntax of a language. The structure unfolds along the single dimension of time—morphemes, words, sentences, discourse. The process of comprehension of language involves mapping sound onto meaning. The meaning of language in nature is an arbitrary system referring to objects and concepts. For the comprehension of language, not only the linguistic information but also the context and world knowledge should be taken into consideration. Furthermore, because of the multisensory system of human beings, any extra-linguistic information that co-occurs with language may also have an influence on the comprehension of language.

In daily communication, language comprehension can be enhanced by the access to some other sources of information, for example, visual information like facial expressions and lip movements, especially in noisy environments or when the hearing is impaired (Grant et al., 1998; Grant et al., 2000; Sekiyama et al., 2003; Summerfield, 1992). These behavioural advantages are explained as reflecting the superiority of bi-modal information over unimodal information (Frens et al., 1995; Perrott et al., 1990). This thesis focuses on one source of information that has often been found to be accompanying speech: hand gestures. The co-occurring of speech and gesture happens not only during face-to-face communication but also when the communicators cannot see each other (Bavelas et al., 1992; Krauss et al., 1995). An ongoing question concerns the role those hand gestures play in conversation.

As extra-linguistic information, gestures have often been observed to accompany language, for example, an inverted V-hand in a wiggling manner while saying “he walked across the street”, or an upward hand movement in a climbing manner when saying “he climbs up the wall”. Speakers convey information in both gesture and speech, while listeners are sensitive to information available in both modalities (Goldin-Meadow et al., 1999; Kelly et al., 1998). This thesis focuses on how the information conveyed through the two different modalities interacts to form a ‘single unified system’, as has been hypothesized by some researchers (Hagoort, 2004, 2005; Ozyurek et al., 2007). To investigate the relation of information conveyed in both gesture and speech, how information is conveyed through gesture will be considered first, before moving on to investigate the relationship between gesture and speech.

1.1. Gestures

Gestures are defined as spontaneous hand movements that are produced during the act of speech (McNeill, 1992). There are two parameters for gesture. One is the form of the hand which includes aspects like the shape of the hand, shape of trajectory etc; the other is the motion of the hand, which includes aspects like palm and finger orientation, gesture space, space where motion is articulated and so on. Together the two parameters make up a gesture to take on a symbolic function. The form of the hand can serve as a symbol of a character that the co-occurring language is referring to and the motion as a symbol of the character in motion. For example, consider the gesture of an index figure rolling from left to right, together with the words, ‘he rolls over the bridge’. In this example, the index figure is a symbol of ‘he’ and the rolling of the figure from left to right is a symbol of ‘the rolling of the character’. Gestures are free from standards of form. In the last case, one index finger was used as a symbol of the character, one might also use two figures, or a fist as the character symbol. To symbolize the ‘roll over the bridge’ motion, it can makes two rolls, three rolls or draw an arc in the air. Gestures always occur with speech and represent aspects of the same representation described in speech.

1.1.1. Definition of gesture

When people talks, there are always hand and arm movements, but not all of them can be classified as gestures. There are two criteria that need to be fulfilled in the definition of a hand movement as a gesture.

First of all, gestures are hand movements that are part of communication. Gestures are produced as part of an intentional communicative act that is constructed at the moment of speaking (McNeill, 1992). This makes gestures different from other forms of nonverbal behaviour such as adaptors or emblems that can also accompany speech. Adaptors are adaptive hand movements that are maintained by habit; examples of adaptors would be smoothing of the hair, pushing up of the glasses now and then, holding or rubbing of the chin and so on. Such adaptive hand movements are performed with little awareness and have no intent of communication; thus, they cannot be classified as gestures. Emblems, at the other end of the awareness spectrum, are hand movements that are held to standards of form, for example, the ‘thumbs up’, the ‘okay’, the ‘shush’. In emblems, the change of handedness or the shape of the hand will lead to a different movement with a different meaning. The meaning of emblems does not depend on speech, they can act independently and can be understood with or without the accompanying speech (Goldin-Meadow, 2003).

Secondly, gestures do not have a codified system. They are free to take on forms that speech cannot assume and are consequently free to reveal meanings that speech cannot accommodate. The continuum by McNeill (1992) is useful in distinguishing gestures from other hand movements. According to this continuum, there is an ordering of gestures:

Figure 1.1 Gestural continuum by McNeill (1992)

From left to right: (1) the obligatory presence of speech declines, (2) the presence of language-like properties increases, and (3) idiosyncratic gestures are replaced by socially regulated signs

‘Sign languages’, such as American Sign Language (ASL) are fully-fledged linguistic systems in which hand signs are used to communicate with and by the deaf. Sign language, to the deaf, is like spoken language to hearing people. Sign languages have standards of well-formedness, are conventionalized, symbolic and are produced in hierarchical combinations (McNeill, 1992).

‘Emblems’ as discussed above, must meet standards of well-formedness. Like words, they hold a conventionally structured code and can be understood by members of the community in the absence of context of explanation (McNeill, 1992).

‘Pantomimes’ are enactions or demonstrations of an action without using an object. Therefore pantomimes can ‘stand on their own’ in conveying information, as speech is not obligatory. McNeill (1992) stated that successive pantomimes can create sequence-like demonstrations.

‘Language-like gestures’ have similar form and appearance to gesticulation; the difference is that language-like gestures are grammatically integrated into the utterance. For example, in the sentence, ‘The parents were all right, but the kids were [gesture]’, instead of an adjective, a language-like gesture is used to fill the grammatical slot.

Gestures refer here to the left end of the spectrum: ‘gesticulations’ are idiosyncratic spontaneous hand movements without holding conventional codes like those in language. According to McNeill (1992), in gesticulation, the meaning cannot be combined.

Thus, there is a clear definition for gestures. They are spontaneous hand and arm movements that are directly related to the speech they accompany and are produced with communicative intention.

1.1.2. Gesture types

There are five gesture classification schemes that have been proposed (see Table 1-1 for detail). As can be seen from the table that even though the sub-categorization of gestures varies, all schemes classify the same movements as gestures. In that sense, all these schemes are interconvertible. This thesis will use the scheme offered by McNeill (1992), which classifies gesture movements into four major categories based on the meaning and function that gesture possesses.

Krauss, Chen, and Gottesman (2000)	McNeill (1992)	Ekman and Friesen (1969)	Freedman and Hoffman (1967)	Efron (1941)
Lexical gestures	Iconic gestures	Kinetograph gestures Pictograph gestures	Literal-reproductive gestures	Physiographic gestures Kinetographic gestures
	Metaphoric gestures	Ideograph gestures Underline gestures Spatial movement gestures	Concretization minor and major qualifying	Ideographic gestures
Deictic gestures	Deictic gestures	Deictic gestures		Deictic gestures
Motor gestures	Beat gestures	Baton gestures Rhythmic gestures	Punctuating	Baton gestures

Table 1-1 Overview of gesture classification schemes (cited from McNeill, 1992; Goldin-Meadow, 2003)

1.1.2.1. Iconic gestures

A gesture can be classified as an iconic gesture if it fulfils two conditions: first of all, it should represent images concretely and transparently, for example, an arcing-upward motion produced while saying ‘I have to go upstairs to find my slippers’, or a tracing of a circle in the air with the index finger while saying, 'It is a round ornament'. The image that iconic gesture represents can be body movements, movements of objects (without holding that object, thus different from object manipulation), people in space as well as the shapes of objects or people.

Secondly, the meaning of an iconic gesture has to be generated online on the basis of gesture form and the co-speech context in which the gesture is observed. For example, a rotating gesture made with a pointing hand could be accompanied by the speech, 'She does lovely pirouettes'. The gesture represents a ballerina's movements. However, the same gesture could refer to a hand twisting off a jar lid accompanied by, 'Which direction shall I turn this?' This illustrates that iconic gestures do not have conventional or unambiguous meanings in the absence of speech (Feyereisen et al., 1988; Krauss et al., 1991).

1.1.2.2. Metaphoric gestures

Metaphoric gestures can represent abstract images like concepts, knowledge, language itself, the genre of the narrative and so on (McNeill, 1992), for example, the drop of the right hand to illustrate the abstract feature of the sentence, 'The man gets down to business '. Metaphorics are similar to iconic gesture in that they both represent images. While iconic gesture depicts some concrete event or object by creating a homology to aspects of the event/object, metaphoric gestures present abstract images that are invisible, or depict a concrete metaphor for a concept by creating a visual and kinetic image that is similar to that concept.

1.1.2.3. Deictic gestures

Deictic gestures are pointing movements that are used to indicate objects, people, or locations (McNeill, 1992), for example, a pointing to a hat while saying, 'This is our new fashion'. Deictic gesture can also indicate object or event that is not in the real world. In narratives and conversations, a deictic gesture can be made meaningful by selecting a part of the gesture space and attaching a referential value to that space. An example would be pointing to the space between self and interlocutor while saying, 'Where did you come from?'

1.1.2.4. Beats

Beats are short, quick hand movements that move along with the rhythmical contour of speech. A typical beat is a simple flick of the hand or fingers up and down, or back and forth in the space of the periphery of the gesture space. Beats tend to have the same form, regardless of the content. Beats do not present a discernible meaning; their value is their semiotic value to index the word of phrase they accompany as being significant (McNeill, 1992).

1.1.3. Gesture phases

According to McNeill (1992), gesture phases are different based on whether the gesture depicts imagery or not. Iconics and metaphorics are imagistic and generally consist of three phases: a preparation phase, a stroke phase, and a retraction phase (McNeill, 1992). Figure 1.2 provides an example of the three phases with the gesture 'break'. In the preparation phase, the hand and arm move away from their rest position to a position in gesture space where the stroke begins. As can be seen in the first row, that in illustrating the preparation phase of the 'break' gesture, the hands held into fist down are rise to centre front. In the stroke phase, the meaning of the gesture is expressed. It is performed in the central gesture space, and it is this phase that is synchronized with the linguistic segments with which it is co-expressive. An illustration of the gesture 'break' in the stroke phase would be the two fists moving 90 degrees outside and then back into the face down position (see second row of Figure 1.2). The retraction phase is the return of the hand from the stroke to the rest phase; in the 'break' example, it is where the fists move from the centre front back into the rest position (see third row of Figure

1.2). For non-imagistic gestures like beats, there are only two movement phases (see beats in 1.1.2 gesture types).

Figure 1.2 An illustration of the gesture phases of the 'break' gesture.

The first row shows the preparation phase, the second the stroke phase and the third the retraction phase

1.1.4. Meaning conveyed in gesture

This thesis focuses on one type of imagery gestures--the iconic gesture, and their relationship with speech. In the following part of this thesis, whenever referred to gesture, it means iconic gesture. From the earlier description, it can be learned that gesture occurs together with speech and conveys meaning relevant to the speech (Goldin-Meadow&Sandhofer, 1999; Kelly&Church, 1998). Gestures are particularly adept at describing concrete information like spatial or motor contents (Beattie et al., 1999; Krauss, 1998). Gestures can also convey additional information not present in the accompanying speech (Church et al., 1986). Gestures do not have a standard of form like speech, the meaning is dependent on the accompanying speech. Therefore iconic gestures have the potential to offer a different view of the speaker's thought in addition to speech, as has been described by Goldin-Meadow (2003) that iconic gesture provides a ‘window on the mind’.

Compared to language, which has standards of form and conveys meaning in a linear segmented way, gestures have some unique properties of conveying meaning (McNeill, 1992):

First of all, the meaning conveyed in an iconic gesture is global and synthetic. Gesture makes sense as a whole, and the meaning of each kinetic of the gesture is determined by the whole. The meaning of the whole is not a linear-segmented combination of separate meaningful parts, it is a syncretization of different meaningful segments (McNeill, 1992). For example, the meaning of the sentence, ‘He rolled over the bridge’ can be expressed by a single arc of the finger drawing in the air.

Secondly, gestures do not combine to form larger, hierarchically structured gestures. Each gesture is a complete expression of meaning, depicts the same aspect of the underlying thought or memory as that described in accompanying speech (McNeill, 1992).

Thirdly, gestures do not have standards of form. Speakers display the gesture based on their own understanding and experience. Therefore, even for the same context, different speakers may gesture in a slightly different way, incorporating a core meaning but adding details based on their inner thought.

1.1.5. Summary

In summary, gestures are hand movements that are produced as an intentional part of the verbal language. Types of gesture are categorized based on the meaning they possess. Iconics are defined as gestures that depict concrete aspects of the images, while metaphorical gestures depict abstract aspects of the images. Deictics are used for indication, and beats are used to highlight the accompanying speech. Gestures are free of form, yet generally, there are three phases for each gesture. Meanings are conveyed in gestures in a global and synthetic way. The meaning of iconic gesture has to be generated online on the basis of gesture form and the co-speech context in which the gesture is observed. As has been pointed out by Goldin-Meadow (2003), gestures provide a 'window on the mind' to describe the speaker's inner thought in addition to speech.

1.2. The relationship between gesture and speech

Speakers gesture in face-to-face communication, but also when they are on the phone where they cannot see the person they are talking to (Bavelas et al., 1992; Krauss et al., 1995). It has been found that speakers gesture even when they are blind and they themselves cannot see the gesture (Iverson et al., 1998). The meaning conveyed in gesture is global, synthetic and needs to be analysed as a whole, while the meaning conveyed in speech is linear, segmented and unfolding over time. This raises the question, what functions gestures serve. How is the speaker's inner thought described in the two different modalities? The following part will review evidence about the function of gestures for both speakers and listeners, about the relationships between the information conveyed in the two modalities, before moving on to explain how gesture and speech interact with each other.

1.2.1. Gesture function

1.2.1.1. Gesture functions for speaker

With the use of gesture, a speaker is able to produce a speech description more informatively and fluently. For example, in a study carried out by Rime et al. (1984), after subjects were impeded from making the principal movements they normally performed during a conversation, the content analysis revealed a significant decrease in the vividness of imagery[1] during movement restriction. In a study carried out by Rauscher et al. (1996), they found out that compared to when the speakers could not gesture, there tended to be fewer filled pauses in speech when they were free to gesture. They also found out that the spatial content of speech was less fluent when the speakers could not gesture, but the nonspatial content was not affected by gesture condition (Rauscher et al., 1996). In terms of speech content and fluency, the effects of preventing speakers from gesturing resembled those of increasing the difficulty of lexical access, either by trying to use as many obscure words as possible (obscure-speech condition) or by avoiding using words that contained a specified letter {constrained-speech condition). Based on these results, Rauscher and colleagues concluded that their results illustrated that co-speech gesture can facilitate access to the mental lexicon.

Gestures can also provide formats of new thoughts that have not been developed enough to be coded in speech (Goldin-Meadow, 1999). For example, Crowder (1996) found the gestural foreshadowing effect in science lessons, children tended to use gestures to precede the ideas they themselves eventually articulated in speech. By offering a route to try out and leave an expression of new ideas, gestures can facilitate the change of one's repertoire (Goldin-Meadow, 1999).

1.2.1.2. Gesture functions for listener

Gestures can communicate information about speakers' inner thoughts that can also be comprehended by listeners. There are plenty of studies showing that compared to a speech only condition, listeners have a better comprehension when they are presented with both speech and gesture (Hostetter et al., 2011; Kelly, 2001; Valenzeno et al., 2003). This is especially the case when gestures contain supplementary information with co-occurring speech. Gestures can facilitate comprehension of the spoken message

[1] The degree of speech imagery was quantified using a computer program (Hogenraad et al., 1981), in which the imagery valence of the verbal material was compared within a dictionary.

when they convey the same information, while impeding comprehension when they convey conflicting information (Goldin-Meadow, 1999). For example, in a study by Broaders et al. (2010), they found that spontaneous gestures containing mismatched information could lead witnesses to report incorrect details. In another study carried out by Goldin-Meadow, &Sandhofer (1999), gestures that conveyed different information from the speech were observed to hinder the listener's ability to identify information in speech.

When the information is hard to comprehend, as in the case of non-native speakers or someone with weak verbal abilities, listeners tend to attend to the accompanying gestures for additional cues. For example, Sueyoshi et al. (2005) demonstrated that the presence of gestures improved the comprehension of bilinguals who were of low proficiency in their second language, while for those of high proficiency, there was no such improved gesture effect. McNeill et al. (2000) asked both preschool children and kindergarten children to select blocks under either a no gesture condition, gesture reinforced speech condition or gesture conflict speech condition. The speech message was complex for preschool children but not for kindergarten children. They found that for preschool children, the reinforcing gesture condition facilitated speech comprehension, and the conflicting gesture condition hindered comprehension. For kindergarten children, however, there was no such facilitation or hindrance effect among the three gesture conditions.

1.2.2. Theoretical views on the relationship between gesture and speech

The relationship between gesture and speech has been investigated for quite some time. Some accounts deny a communicative impact of gesture, with the thought that gestures are just random hand movements to dissipate tension during the action of speech (T. Dittmann et al., 1969) or the opinion expressed by Krauss et al. (1991) that hand gestures are not as communicative as speech, the relationship between gesture and speech is comparatively unreliable. Despite all of these disagreements, the evidence that gestures formulate the production of speech for speakers and serve as a way of conveying information for listeners reduces the possibility of the gesture being just random.

McNeill et al. (1994) artificially created gesture-speech mismatches to investigate the relationship of gesture and speech during comprehension. In the study, participants were presented with a videotape of someone telling a story with gesture-speech mismatches in three conditions (manner mismatches, perspective mismatches, anaphor mismatches) and asked to retell the story. The results showed that all of the three gesture-speech mismatches in the videotape produced a detectable effect on the participants' retellings. In addition, the effect of the mismatches was evident in gesture, speech, as well as in both gesture and speech. Based on these findings, McNeill concluded that participants were sensitive to information from both gesture and speech. Information was coded independent of modality, participants tried to form a coherent understanding of the information and coherent information was stored in memory without specific indexing of the modality.

Kelly et al. (2010b) presented participants with action primes (e.g. someone chopping vegetables) and bimodal speech and targets where one part of the target (speech or gesture) was related to the prime. Kelly et al. (2010b) found that gestures influenced speech comprehension, and speech influenced gesture comprehension in a comparable way in terms of N400[2] latency and amplitude. Moreover, this bi-directional influence of gesture and speech still existed when participants were told to focus only on the speech information. From these results, Kelly et al. proposed the 'integrated system hypothesis' that during comprehension the information from both gestures and speech interacts

[2] A negative-going deflection of ERPs waveform between 300 and 550ms post-stimulus, see part 1.2.3.3 for detail.

mutually with each other in an automatic, obligatory way. According to this view, the relation between gesture and speech is like 'two sides of the same coin'.

From the above illustrations, it can be seen that speech and gesture are bi-directionally influenced by each other, as speech influences how people gesture and gesture influences what is produced in speech. The bi-directional influence of gesture and speech appears to take place automatically in that there is no need for extra effort or intention, people just cannot help paying attention to information from both modalities.

1.2.3. Bi-directional influence between gesture and speech

Green et al. (2009) stress three points for the bi-directional influence of gesture and speech: first of all, gesture interacts with speech and does not act as an 'auxiliary support'. Second, this interaction takes place at a semantic level and differs from the relationship between lip movement and speech, which is based on form matching. Third, this interaction is implicit, as gesture and speech are automatically processed in the brain. To have a clearer look at the relationship between gesture and corresponding speech, these three points will be developed in detail below.

1.2.3.1. Gesture and speech interact with each other

The interaction of gesture and speech involves two levels. First of all, speech and gesture are temporally aligned with each other. The onset of gesture usually precedes the onset of the relevant speech segment by less than a second (Morrelsamuels et al., 1992). Secondly, there exists a bidirectional influence of gesture and speech in terms of external form. In a study carried out by Bernardis et al. (2006), they found that compared to unimodal conditions, multimodal voice spectra (especially formant 2, F2)[3] were enhanced by gestures, whereas multimodal gesture parameters (like gesture time, maximal height,

[3] Formants are distinctive frequency components of the acoustic signal produced by speech. There are three types of formants in phonetics, formant 1 (F1) with the lowest frequency, formant 2 (F2) the second, and formant 3 (F3) with the highest. Formants can be calculated using the PRAAT software (Bernardis et al., 2006).

hand oscillation amplitude, maximal peak velocity, the number of velocity peaks) were reduced by words. They concluded that spoken words and gestures are coded as a single signal by a unique communication system. Last but not least, the bidirectional influence of speech and gesture also appears at the semantic level. In a study carried out by Kita et al. (2003), participants were presented with an American animated cartoon video and asked to tell the story to a listener who did not see the video in three languages: English, Turkish and Japanese. Their narrations were videotaped and the gestures and languages that were produced in the narratives elicited from the same stimulus were analysed. The findings showed that gestures used to express the same motion events were influenced simultaneously by how features of motion events were expressed in each language. Participants also tended to use gestures to express motion events if the spatial information in the stimulus could not be verbalized. Based on the results, Kita et al. (2003) concluded that gestures are generated from spatial-motoric processes that interact on-line with the speech production process, speech affects what people produce in gesture, and gesture, in turn, influences what people produce in speech. In another study by Kelly et al. (1999), participants were presented with video segments of someone pointing at objects, and were asked to identify the referred- to object. Results showed that participants identified the intended object more often in the speech-gesture condition than in the gesture only condition, demonstrating that the referent of the pointing gestures is determined by both the gestures and the speech accompanying them. Moreover, Kelly et al. (2010b) presented participants with action primes (e.g. someone chopping vegetables) and bimodal speech and targets, and asked them to identify if any part of the target (speech or gesture) was related to the prime. The results showed that participants related primes to targets more quickly and accurately when they contained congruent information (speech: ‘chop’, gesture: ‘chop’) than when they contained incongruent information (speech: ‘chop’, gesture: ‘twist’). More importantly, the results did not show an interaction between incongruence and target modality. The authors concluded that gestures influenced speech comprehension, and speech influenced gesture comprehension in a comparable way.

1.2.3.2. Gesture and speech interaction can be implicit

There are several studies suggesting that the interaction between gesture and speech is obligatory in the sense that people cannot help but pay attention to information in both modalities during language comprehension. For example, in Kelly et al. (2009), participants performed a Stroop-like task in which they watched videos of a man and a woman gesturing and speaking about common actions. The video differed as to whether the gender of the gesture and speaker was the same or different and whether the content of the gesture and speech was congruent or incongruent. The task was simply to identify whether the man or the woman produced the spoken portion of the videos while accuracy rate, reaction times (RTs) and Event-related potentials (ERPs) time-locked to the spoken targets were recorded. Results yielded an N400 effect for incongruent speech-gesture pairs compared with the congruent ones. Additionally, RTs were slower for incongruent versus congruent gesture-speech stimuli. This suggests that although not relevant to the task, participants paid attention to the semantic relationship between the speech and the gesture. Also, Kelly et al. (2010a) observed that even though the task did not include instructions to respond to gestural information contained in the videos, participants' accuracy scores demonstrated sensitivity to gesture by becoming decreasingly accurate as semantic incongruity with speech increased.

While speech and gesture integrate automatically, the interaction can also be modulated by some controlled processes. For example, Holle et al. (2007) showed that the presence of non-meaningful gestures (e.g., grooming behaviors) modulates the extent to which meaningful gestures set up semantic expectations for speech later in a sentence. Obermeier et al. (2011) found that when gestures did not temporally overlap with the homonyms and when participants were not explicitly asked to pay attention to gestures, speech- gesture integration did not occur. Kelly et al. (2007) demonstrated that if the two modalities are perceived as not intentionally coupled (i.e. gesture and speech being produced by two different persons) the integration is less strong (difference between incongruent speech-gesture pairs in terms of N400 effect) compared with when they are perceived as being intentionally coupled (i.e. gesture and speech being produced by the same person).

1.2.3.3. Gesture and speech interact on semantic level

Kutas et al. (1980) discovered a negative-going deflection of the event-related potentials (ERPs) waveform between 300 and 550ms post-stimulus (the so-called N400 effect) with an enhanced amplitude if the word of a sentence is semantically incongruent (I take my coffee with cream and dog) compared to when it is congruent (I take my coffee with cream and sugar). Subsequent research has shown that the N400 effect can be generated by more or less subtle differences in the semantic fit between the meaning of a word and its context, where the context can be a single word, a sentence or a discourse (Hagoort et al., 2007). The N400 effect has been used as a dependent variable in psycholinguistic experiments to examine whether there is a semantic or conceptual processing induced by the stimulus event (Kutas et al., 2000).

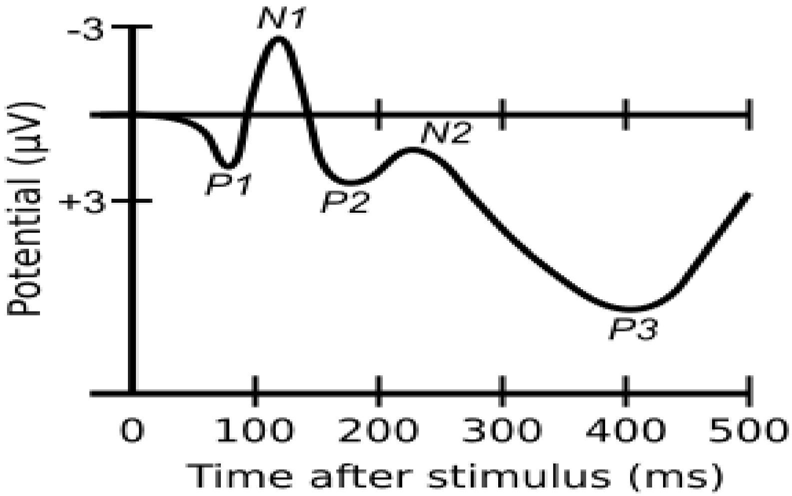


Figure 1.3 A waveform showing several ERP components (image from Wikipedia) https://en.wikipedia.org/wiki/Event-related_potential

The N400 effect has also been used to investigate the effect of iconic gestures on speech. Kelly et al. (2004) presented participants with gestures (primes) corresponding to a property of an object (size or shape of that object) preceded by spoken words (targets). When the information contained in gesture and words was incongruent, they observed an early P1/P2 effect[4], followed by an N400 effect. Kelly and his colleagues argued that the gesture primes influenced word comprehension, first at the level of "sensory/phonological" processing and later at the level of semantic processing. Also, researchers found an N400 effect in cases when incongruent iconic gestures produced in the course of conversation were compared with congruent ones (Wu et al., 2005). An N400 effect was also observed to a word later in the sentence if the meaning of that later word did not match with the meaning indicated by the gesture earlier in the sentence (Holle&Gunter, 2007).

In a study carried out by Ozyurek et al. (2007) ERPs were measured while subjects listened to spoken sentences with a critical verb (e.g., knock), which was companied by an iconic co-speech gesture (i.e., knock). The semantic fit of speech (i.e., a critical verb)

[4] P1 and P2 are ERPs waveforms. P1 is a waveform component that peaks at about 100ms post-stimulus, and P2 peaks at about 200ms.

and/or gesture in relation to the preceding part of the sentence (global integration) as well as the semantic relation between the temporally overlapping gesture and speech (local integration) were manipulated. This resulted in four conditions: a correct condition [gesture (G) +, language (L) +]; a language mismatch condition [G + L -]; a gesture mismatch condition [G - L +]; and a double mismatch condition [G - L -]. Results showed that language, gesture, and double mismatch conditions modulated the N400 in a similar way, in terms of N400 latency and amplitude. Ozyurek, et al. claimed that this illustrated that verb and gesture are integrated into parallel into the semantic context. This suggests that language comprehension involves the incorporation of information into a 'single unification space' (Hagoort, 2004, 2005; Ozyurek et al., 2007) with overlapping neuronal sources.

1.2.4. Time window for the interaction of Iconic gesture and speech

From above it was learned that even though gestures convey information in a global-synthetic way, while speech conveys information in a linear segmented way, there is a bi-directional influence between the two. Speech influences how people gesture and gesture influences what is produced in speech in an automatic, obligatory way (Kelly et al., 2010a). Yet gesture does not have independent meaning and needs to synchronize with speech in order to interact with speech (Habets et al., 2011; Krauss et al., 1991; Morrelsamuels et al., 1991; Obermeier et al., 2011). This raises the question of the critical time window for the interaction of gesture and speech.

Obermeier et al. (2011) explored the asynchrony of gesture and speech by replacing complete gesture with only meaningful gesture fragments. The original stimuli were two 2-sentence utterances in which there was either a dominant or a subordinate word that acted as target. Sentences were uttered simultaneously with the performance of a gesture. Gesture fragments were determined using a gating study, in which Obermeier and colleagues presented participants with gesture videos of various length that started at the onset of the preparation phase with 40ms increment of each of the video. Based on the gestural information presented in the gesture videos, participants had to decide whether the homonym that was presented 500ms ahead of the onset of gesture referred to dominant meaning or subordinate meaning. The disambiguation point (DP) was located when participants gave 10 consecutive correct answers. Meaningful gesture fragments with minimal length that had a communicative impact were then decided as the period from the onset up to the DP. Using a similar method, Obermeier et al., conducted another gating study to locate the identification point (IP) of the homonym as the earliest point in time at which participants could identify the homonym. In one experiment, Obermeier et al., replaced the original gesture in the sentence with gesture fragment in such a way that the temporal alignment of the onset of the gesture fragment with the homonym was identical to the alignment of the original gestures. Results showed that gesture fragments were able to disambiguate the speech only when the participants were explicitly asked to pay attention to the information in both the speech and the gesture fragments. There was no such disambiguating effect of gesture fragments on speech when participants were asked to do a shallow task in which they were not required to actively combine both streams to solve the task. In another experiment, Obermeier et al. manipulated the synchrony between gesture and speech by moving the offset of the gesture fragments to the same time point as the onset of the homonym identification point. Results showed that gesture fragments were able to disambiguate the speech even in the shallow condition. Based on these results, Obermeier et al. concluded that when gesture fragments and speech are in synchrony, there is an automatic integration between them.

To further investigate the time window for the synchrony of gesture and speech to be integrated, Obermeier et al. (2015) manipulated the temporal alignment of the homonym and the gesture into four conditions: the offset of the gesture fragments was either -600ms, -200ms ahead of, synchronous with (0ms) or +120ms behind the onset of the IP of the homonym. ERPs were recorded to both the homonym and the target word. Results showed significantly triggered N400 effect for both homonym and the target word in the -200ms condition, the 0ms condition, and the +120ms condition. For the -600ms condition, there were only significant ERP effects for the target word. Obermeier et al. (2015) concluded from the results that significantly triggered ERPs to the homonym reflected a local integration of gesture and speech, which happened at least in a time span from -200ms to +120ms. The significantly triggered ERPs to the target word reflected a global integration of gesture with the sentence, which may have taken place at multiple positions.

Habets et al. (2011) investigated the degree of synchrony with the manipulation of both the synchrony of gesture and speech and the semantic congruency between gesture and speech. In their study, there were three degrees of synchrony of gesture and speech: the SOA 0 condition in which the onset of gesture that started from the stroke phase and speech were presented simultaneously; the SOA 160 condition the SOA 360 condition in which the onset of the stroke phase of gesture was presented 160ms and 360ms respectively ahead of the onset of speech. Habet et al., (2011) compared the semantic congruency N400 effect in the three asynchrony conditions. They hypothesized that the semantic difference between the three asynchrony conditions in terms of N400 effect would reflect the different integration processes of gesture and speech. Results showed a greater negative N400 component in the semantic incongruent condition compared to the semantic congruent condition in only the SOA 0 condition and the SOA 160 condition, while in the SOA 360 condition, there was no such a significant N400 effect. Based on these results, Habets et al. concluded that there is a certain time span for the integration of gesture and speech, in this case, the time span should be within 360 ms's difference between the onset of the stroke phase of gesture and speech. It should be pointed out that Habets et al. also conducted a gating study which showed that participants can have a steady interpretation of the gesture at about 360ms after the onset of the gesture. Therefore in the SOA 360 condition, the interpretation of the gesture will no longer be influenced by the speech, resulting in no significant difference between the match and mismatch of gesture with speech, as explained by Habets et al.

In summary, all of the above studies investigated the time span for the integration of gesture with speech. Obermeier et al., (2011, 2015) used the disambiguation point (DP) to replace the full gesture with the hypothesis that from the onset of the gesture to the DP is the minimal fragment of gesture for which meaning can be disambiguated by the participants. They also used the identification point (IP) as the point from which participants can have a clear understanding of the meaning of the homonym. They found there is a direct integration of gesture with speech when the time span from the offset of the DP of the gesture to the onset of the IP of the speech is within -200ms to +120ms. Habets et al. (2011) used words instead of sentences and found the integration of gesture and speech when the onset of the stroke phase of gesture precedes the onset of the word by 0-160ms.

1.2.5. Summary

In summary, this part discussed the relationship between gesture and speech. Even though gestures convey information in a global-synthetic way, while speech conveys information in a linear segmented way, there is a bi-directional influence between them. First of all, in production, the product of speech will influence not only the external form of the gesture (Bernardis&Gentilucci, 2006), but also the semantic information conveyed in gesture (Kelly et al., 1999; Kita&Ozyurek, 2003), and vice versa, the produce of gesture influences the production of speech (Bernardis&Gentilucci, 2006), as well as the information conveyed in speech (Kita&Ozyurek, 2003). Secondly, in comprehension, an N400 effect will be triggered by the semantic fit between gesture and speech (Kelly et al., 2004) as well as the semantic fit between either gesture and speech with the context (Holle&Gunter, 2007; Ozyurek et al., 2007; Wu&Coulson, 2005), in a similar way in terms of N400 latency and amplitude (Ozyurek et al., 2007). Last but not least, the bi-directional influence between gesture and speech can be implicit (Holle&Gunter, 2007; Kelly et al., 2010a; Kelly et al., 2007; Obermeier et al., 2011).

Since gesture does not have independent meaning, the information a gesture conveys is dependent on the accompanying speech. The synchronization of gesture and speech has been investigated by several researchers (Habets et al., 2011; Obermeier&Gunter, 2015; Obermeier et al., 2011). Obermeier and colleagues defined the synchronization of gesture and speech as at least within -200ms to +120ms between the DP of the gesture and the IP of the speech. They also found the disambiguation function of the gesture to the global sentence. In the study of Habets et al., (2011), they examined the synchronization of gesture with word, with a definition of the synchronization as the time span between the onset of the stroke phase of gesture and the onset of the speech and found time span is at least within 160ms. Even though they have similar findings, their definition of synchronization is different.

Based on the bi-directional influence of gesture and speech as well as the synchronization of the two, some researchers assumed that gesture and speech are not separate communication systems as have been assumed by some researchers (Hadar et al., 1999; Hadar et al., 1998; Levelt et al., 1985). Instead there is a common origin in which a synthesis of opposite modes of thought – global-synthetic imagery information temporally extended to the linear-segmented temporally verbalization (McNeill, 1992). According to this view, utterances and thoughts realized in them are both imagery and language. Speakers use gestures as if they are the functional equivalents of lexical units in spoken language, alternating them with spoken elements within a sentence

(Slamacazacu, 1993). From a functional point of view, gestures can be regarded as ‘part of language’ (Kendon, 1997) to integrate with speech with overlapping neural architecture.

1.3. The neural architecture underlying gesture-speech integration

The evidence reviewed above suggests that gesture and speech are integrated with each other at the semantic level with overlapping neural architecture. As two different modalities (gesture: visual sense; speech: auditory sense), this integration may first involve the integration of auditory and visual modality, followed by the integration of semantic information from both gesture and speech. Therefore, the neural substrates underlying this integration should respond not only to isolated gesture (visual stimuli) or speech (auditory stimuli) but also to the bimodal integration, which should be no longer a linear combination of the two unimodal stimuli (Calvert et al., 2004).

Two traditional language areas, the posterior portion of the left superior temporal sulcus and adjacent superior temporal gyrus (pSTS/STG, Wernicke’s area) have been consistently implicated in the integration of speech-related audiovisual (lip movements) information (Callan et al., 2003; Calvert et al., 2000; Sekiyama et al., 2003). The left inferior frontal gyrus (LIFG, Broca’s area) has also been found to be activated in human action observation (Arbib, 2005). Researchers using either mismatch manipulations (Willems et al., 2007) or disambiguation paradigms (Holle et al., 2008) to manipulate semantic integration load between gesture and speech also reported activation of the pSTS/STG (Holle&Gunter, 2007; Holle et al., 2010; Straube et al., 2011) and LIFG (Willems et al., 2007).

1.3.1. Semantic integration in LIFG

In the study by Willems et al. (2007), participants were presented with iconic gestures that were either congruent (verb: write, gesture: write) or incongruent (verb: write, gesture:

hit) with speech, and found a significant greater activity in left IFG in incongruent conditions compared with congruent ones. Green et al. (2009) found a greater activation in IFG when iconic gestures were unrelated to the accompanying speech as compared to iconic gestures that were related to speech. Dick et al. (2012) also found stronger activation of LIFG when iconic gestures added information to nonspecific language (gesture contains disambiguating information with respect to speech), compared with when they conveyed the same information in a more specific language context (gesture contains redundant information with respect to speech).

Straube et al. (2011) compared the neural integration of metaphoric gestures (MP) and speech with that of iconic gestures (IC) and speech, with the implication that in MP gestures the 'integration load' is high compared with that of IC gestures. IC gestures refer to the concrete content of sentences (for example, the drop of the right hand to illustrate the concrete feature of the sentence, "The man goes down the hill"), while MP gestures illustrate abstract information (for example, the drop of the right hand to illustrate the abstract feature of the sentence, "The man gets down to business"). Results showed that IFG was activated during both MP and IC conditions. A similar result was also found by Willems et al. (2009) who investigated the brain area that underlines the integration of IC coverbal gestures and pantomimes. Pantomimes are enactions or demonstrations of an action without using an object, therefore pantomimes can 'stand on their own' in conveying information, while IC gestures need language to be meaningfully interpreted, therefore the 'integration load' is supposed to be high in IC compared to the pantomimes. Willems found the involvement of LIFG in the integration of both pantomimes and IC.

1.3.2. Semantic integration in pSTS/MTG

In a study carried out by Holle et al. (2008), participants watched videos in which sentences containing a homonym (e.g. she touched the mouse) that had both a more frequent dominant meaning (e.g. the animal), as well as a less frequent subordinate meaning (e.g. computer device). The homonym was accompanied by three gestural conditions: a meaningless gesture condition; a gesture supported the less frequent subordinate meaning condition, and a gesture supported the frequent dominant meaning condition. Holle et al. found that compared to the meaningless gesture condition, there is a significant activation of the left STS, but no LIFG, in both the dominant and the subordinate types of gesture condition. Additionally, Holle et al. (2010) provided evidence for the involvement of pSTS/STG in gesture-speech integration with the use of both bimodal enhancement property and inverse effectiveness property. Holle et al. hypothesised that the brain area that underpins the integration of gesture and speech should show a significantly greater activation in gesture-speech bimodal condition compared to gesture unimodal condition. Besides, such an advantage of bimodal over unimodal should survive in both the good signal to noise ratio (SNR) of speech condition as well as in the moderate SNR of speech condition, with a greater neural response in the moderate SNR condition compared to the good condition. With the above assumptions, Holle et al. created five experimental conditions: gesture-speech good SNR condition; gesture-speech moderate SNR condition; speech alone good SNR condition; speech alone moderate SNR condition; gesture alone condition. Results showed a main effect of both the property of bimodal enhancement and inverse effectiveness in pSTS/STG. Holle et al. concluded from the above two experiments that pSTS/STG plays a key role in the involvement of gesture in the facilitation of comprehension of speech.

Straube et al. (2011) found the left posterior middle temporal gyrus (lMTG) to be activated during the IC condition. Additionally, Willems et al. (2009) found left and right pSTS/MTG were involved in the semantic integration of speech and pantomimes. Dick et al. (2012) also found activation of MTGp when iconic gestures disambiguated the speech compared to when they conveyed the same information as speech.

1.3.3. The role of LIFG and pSTS/MTG in semantic integration

Based on the above evidence, it may be safe to say that two particular areas (IFG and MTG/STS) may be associated with this gesture-speech integration. Some researchers also pointed out that there may be a possible division of labour between LIFG and pSTS/MTG in the integration of iconic gestures and speech. The pSTS/MTG may mainly be involved in the integration of highly overlearned, strongly associated material (Hein et al., 2007; Naumer et al., 2009) where there is a relatively stable common object representation (Willems et al., 2009). In contrast, LIFG may integrate unrelated or incongruent gesture-speech combinations where the bimodal processing load is high (Straube et al., 2011) and where there is an online construction of a new and unified representation of the input streams (Willems et al., 2009). Some authors suggest that LIFG and pSTS/MTG work together to integrate multimodal information (Willems et al., 2009), with a modulatory role of LIFG to select/control/unify semantic information, and a more integrative role for pSTS/MTG (perceptual matching or activation of a common stable representation) (Willems et al., 2009). Finally, Fuhrmann Alpert et al. (2008) found pSTS/MTG precedes activation in LIFG in time by providing participants with simultaneous sounds (auditory stimuli) and images (visual stimuli) using a passive paradigm.

1.3.4. Summary

From the studies summarized above, it can be seen that some researchers found mere activation of IFG in gesture-speech integration (Green et al., 2009; Willems et al., 2007; Willems et al., 2009), while some researchers reported only the involvement of pMTG in gesture-speech integration (Holle&Gunter, 2007; Holle et al., 2010). There are also studies that found the activation of both IFG and pMTG during the integration of gesture and speech (Dick et al., 2012; Straube et al., 2011). To date, no consensus has been reached as to which of these areas is causally, and not merely epiphenomenally, involved in this merging process. Meanwhile, IFG and pMTG are anatomically connected

(Friederici, 2009) which can produce correlated patterns of activation between these regions (Whitney et al., 2011), which makes it possible for both areas to be activated during gesture-speech integration, with only one area crucially involved, while the activation of the other is merely a consequence of the strong anatomical connection. FMRI studies are in this respect limited with respect to the degree to which they allow inferences of causality. Therefore, this thesis used non-invasive brain stimulation (NIBS) to provide causal, direct evidence for the involvement of brain area(s) in gesture-speech integration.

1.4. Non-invasive brain stimulation (NIBS)

Non-invasive brain stimulation (NIBS) is a way to study cognitive functions. NIBS can induce transient changes in brain activity of the brain area being stimulated. If there are changes in behaviour following stimulation of that particular brain area, it leads to the conclusion that this area is critically involved in the behaviour being observed. In contrast, if no changes are observed, the assumption is that the stimulated area is not critical for that process. In this way, NIBS can establish a causal relationship between activity in a given cortical area and the ongoing motor, perceptual or cognitive processes (Hallett, 2000; Walsh et al., 2000), thus avoiding the drawback of correlative approaches used for functional imaging techniques (Miniussi et al., 2013). There are two main types of NIBS methods, transcranial magnetic stimulation (TMS) and transcranial electrical stimulation (tES). Each affects neuronal states through different mechanisms and thus may result in different neuronal effects.

1.4.1. Transcranial magnetic stimulation (TMS)

TMS relies upon the properties of electromagnetic induction. Specifically when a high voltage current is passed through a coil, held to the scalp over a cortical region of interest (ROI), a rapidly changing magnetic field is generated. That magnetic field induces electrical current in the underlying cortical tissue that interferes with the normal neural activity pattern. Thus creates a transient 'virtual lesion' by adding noise to the normal firing pattern, which results in behavioural changes, and thereby provides exactly the information on causal relations between brain and behaviour that cannot be provided by imaging techniques (Pascual-Leone et al., 2000). TMS parameters include stimulation intensity, frequency, and duration. In general, the stimulation intensity should be strong enough to produce a motor evoked potential (MEP), which is varied among different populations. The choices of other parameters (stimulation frequency and duration) depends on the purpose of the experiment, with different physiological or behavioural effects for different forms of stimulation.

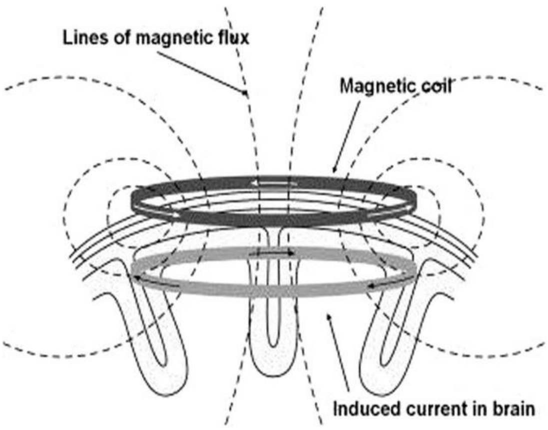


Figure 1.4 Illustration of the mechanism of TMS stimulation (image from Hallett, 2000)

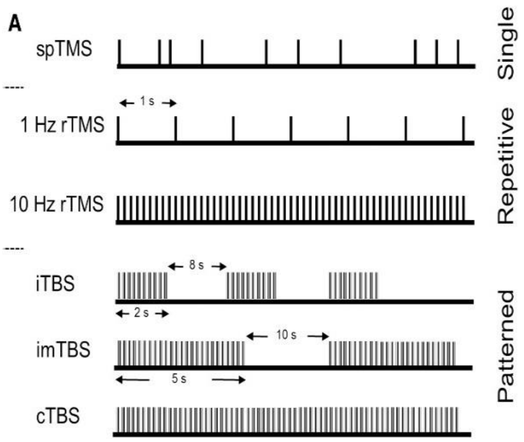


Figure 1.5 Illustration of the various types of TMS stimulation (image from Parkin et al., 2015)

1.4.1.1. Single Pulse TMS

This procedure involves the application of single pulses of TMS during task performance. Each pulse lasts approximately 1 ms, and has only a transient effect on normal cortical functioning. Due to its high temporal resolution (a few tens of milliseconds), single pulse TMS is normally used to establish at what point in time a specific brain area becomes involved in a cognitive process – often in conjunction with a highly sensitive measure of behaviour, such as the recording of motor-evoked potentials (Pascualleone et al., 1994). The interval between pulses in single-pulse TMS should be long enough to prevent consecutive-pulse interactions, a typical pulse interval is about 6-7s. As such, single pulses of TMS are sufficient to elicit a measurable response.

1.4.1.2. Low-frequency Repetitive (r) TMS

Offline low-frequency TMS protocols involve the application of a series of low frequency (≤1Hz) single pulses with an intensity set above motor threshold for several minutes (usually 10-25 minutes) before the experimental task. This type of protocol has been shown to decrease cortical excitability of the stimulated area for several minutes thereafter (Kobayashi et al., 2003). The effect of this type of protocol is that interference in cortical functioning carries over into the task, causing a performance deficit that lasts until excitability recovers to normal levels (approximately half the duration of the stimulation train, for example, 5 - 10 minutes following a 10-minute stimulation period)(Robertson et al., 2003). This offline paradigm of low-frequency TMS is useful in that it allows participants to be free of TMS during task performance, and thus avoid disruption from the side-effects of TMS stimulation such as the discomfort, noise, muscle twitches etc. (Rafal, 2001). Moreover, it allows experimenters to more realistically simulate the effects of brain lesions in the laboratory (Rafal, 2001).

1.4.1.3. High-frequency Repetitive (r) TMS

This procedure is similar in that TMS is also applied during task performance and at stimulator intensities above motor threshold. However, rather than using single pulses, a series (or "train") of pulses is delivered in rapid succession to a cortical region, hence repetitive TMS. A typical high frequency (10 Hz) rTMS procedure might involve the application of five pulses over a period of 500 ms at stimulus onset so that normal cortical functioning is interfered with for 500 milliseconds at a time. This method can produce behavioural effects that are stable enough to be measurable using reaction times (RT) and/or accuracy measures. High frequency rTMS protocols of this kind are typically used to establish whether a brain area of interest is essential for a cognitive process but not at which point in time.

1.4.1.4. Theta-burst TMS (TBS)

Theta-burst TMS (TBS) is similar to the low frequency rTMS, with the exception that TBS is applied at high frequency over a short period of time (e.g., three trains of 50-Hz pulses applied with an interval of 200ms) (Huang et al., 2005). There are two types of TBS protocols, continuous TBS (cTBS) and intermittent TBS (iTBS). cTBS involves applying the burst trains for 20-40s, at an intensity of 80% active motor threshold (aMT) and has an inhibitory effect on the excitability of the motor cortex (the aftereffect for 20s of cTBS is about 20 min, and up to 1 h for 40s of cTBS; Huang et al., 2005). iTBS, in which the burst trains are applied for 2s with an interval of 8s and then repeated for 190s, on the contrary, has a facilitatory effect over the motor cortex (Huang et al., 2005). The differences between low-frequency TMS and TBS are, for our purposes, related to experimental efficiency. Low-frequency TMS requires a 10-minute stimulation period to produce approximately five minutes' interference. TBS, on the other hand, requires only seconds of stimulation to produce an interference lasting up to 30 minutes (Nyffeler et al., 2006). Depending on the anatomical region of interest (ROI) it is sometimes preferable to use TBS when stimulation is causing discomfort, as it is finished more quickly.

1.4.2. Transcranial electrical stimulation (tES)

In transcranial electrical stimulation (tES), a weak direct current (typically used 1 – 2 mA) driven by a battery simulator with two or more surface electrodes is applied (Nitsche et al., 2000). At least one of the electrodes is applied to the scalp to induce current flow through the brain. This current flow will change the intrinsic neuronal excitability and induce a subthreshold polarisation of cortical neurons, in this way influencing the subject’s responses to an afferent signal. It is said that, there are three goals in the usage of tES in the cognitive field: (1) to provide a causal relationship between the brain area and the behaviour; (2) to understand the physiological mechanisms of the causal relationship and (3) to inhibit the excitability or facilitate the excitability of a given brain area (Parkin et al., 2015). There are three different variations of tES based on the type of electrical current used, namely, transcranial direct current stimulation (tDCS) ,transcranial alternating current stimulation (tACS) and transcranial random noise stimulation (tRNS).

1.4.2.1. Transcranial direct-current stimulation (tDCS)

There are two electrodes in tDCS, with current flow delivered under the electrode from the positive charge (anodal stimulation) to the negative charge (cathodal stimulation). tDCS can introduce a uniform steady state extracellular electric field in neural tissue (Bikson et al., 2004), with anodal stimulation over the motor cortex eliciting a larger motor evoked potential (MEP) with the application of TMS, cathodal stimulation over the motor cortex results in a reduced MEP. The principle behind tDCS is that tDCS can induce membrane depolarization with anodal stimulation, and membrane hyperpolarisation with cathodal stimulation. Depolarisation (anodal stimulation) will increase cortical excitability in the stimulated brain area and thus facilitate relevant behaviour, and hyperpolarisation (cathodal stimulation) will decrease cortical excitability, thus inhibiting the relevant behaviour. The position of the electrode is commonly located through the international 10-20 electrode placement system (Oostenveld et al., 2001). tDCS can be used online or offline. The after-effects elicited by tDCS dependent on the stimulation duration and/or intensity. For example, Nitsche, &Paulus (2000) showed that 9min of cathodal stimulation and 13 min of anodal stimulation at an intensity of 1mA can modulate the excitability of motor cortex for up to 1 hour.

1.4.2.2. Transcranial random noise stimulation (tRNS)

tRNS is a relatively new technique of transcranial electrical stimulation (tES) which involves the application of a random electrical oscillation spectrum over the cortex, thus stimulating the two brain areas at the same time in a similar way. There are three frequency ranges of tRNS stimulation: the entire spectrum (from 0.1 to 640 Hz), the low band spectrum (from 0.1 to 100Hz), and the high band spectrum (from 100 to 640 Hz) (Terney et al., 2008). The frequency band spectrum influences neuronal excitability as shown by Terney et al. (2008) who found that 10min tRNS stimulation over the motor cortex can induce an after-effect of up to 60 min, through both the measurement of motor-evoked potentials (MEPs) that resulted from single-pulse TMS stimulation and the behavioural measurement of implicit learning. Besides, high frequency stimulation increased excitability and low frequency stimulation decreased the excitability, according to Treney et al.(2008). Moreover, the neuroplastic effect of high-frequency tRNS is said to be analogous with clear advantages over the anodal tDCS stimulation. Fertonani et al. (2011) showed that the behavioural performance that is enhanced by high-frequency tRNS stimulation is larger than that caused by anodal tDCS stimulation if applied over the visual cortex.

1.4.2.3. *Transcranial alternating current stimulation (tACS)*

Unlike tDCS, which has an anodal electrode and a cathodal electrode, in tACS in half of one cycle of the alternating current oscillation, one electrode will serve as anodal and the other one as cathodal, while in the other half, the pattern will reverse, the former anodal electrode becomes cathodal and the former cathodal electrode becomes anodal (Woods et al., 2016). In this pattern, all stimulated brain areas are modulated in a similar way. tACS works by the principle of intrinsic resonance, which means that the coincidence of the stimulation frequency with the existing brain oscillation of a particular cognitive process will lead to facilitation of that cognitive process (Zaehle et al., 2010). Thus, by amplifying or interfering with the oscillatory pattern of a specific brain activity, tACS can build up the causal relationship between the brain and the corresponding behaviour. There is no linearity between the effects of tACS and the intensity of the stimulation. In a study carried out by Moliadze et al. (2012), they stimulated primary motor cortex with tACS at 140Hz, and found that low stimulation intensity of 0.4mA resulted in the decrease of the amplitudes of MEPs in response to TMS stimulation, high intensity of 1mA resulted in increase of the MEPs amplitude, whereas intermediate intensity of 0.6mA or 0.8mA showed no significant effect in terms of the amplitude of MEPs.

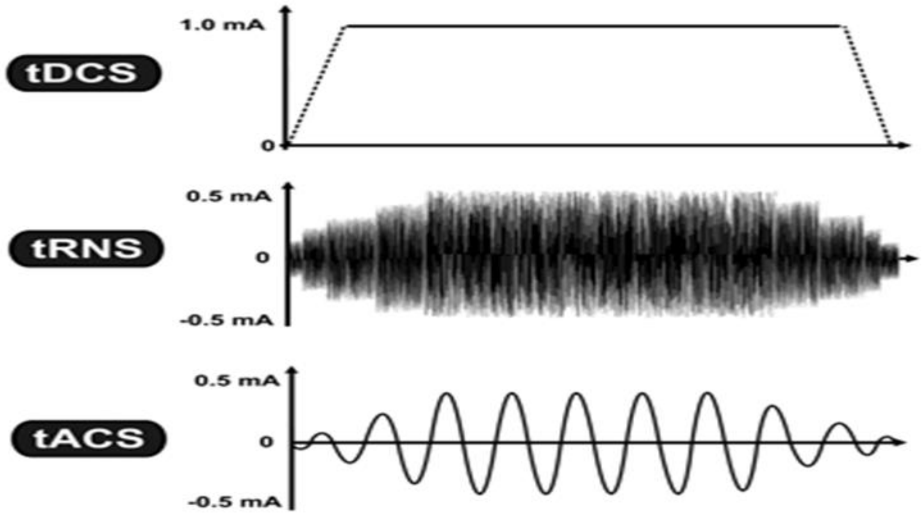


Figure 1.6 Three types of Transcranial Electrical Stimulation (image from Wikipedia)

https://en.wikipedia.org/wiki/Transcranial_direct-current_stimulation#cite_note-28.

1.4.3. Comparison of TMS and tES

1.4.3.1. Comparison of the two protocols from a theoretical perspective

TMS and tES are two types of non-invasive brain stimulation that follow different mechanisms. TMS can cause direct above-threshold depolarization, thus leading to trans-synaptic depolarization or hyperpolarisation of the targeted neurons depending on the stimulation protocol. In contrast, tES is too weak to evoke action potentials, but can modulate the responsiveness of neurons by inducing a polarisation/depolarization of neurons at a subthreshold level (Miniussi et al., 2013).

1.4.3.2. Comparison of the two protocols from a stimulation perspective

With a figure of eight coil, TMS stimulation can be very focal, and thus can effectively induce changes in the neuronal activity. In the meanwhile, stimulation of TMS can cause some unpleasant feelings, especially for the online phase-synchronize stimulation, which may cause some distraction from the behavioral task.

What's more, even though low-frequency TMS is supposed to decrease cortical activity, while high-frequency TMS is supposed to increase the cortical activity, this does not necessarily imply that applying these protocols to a task-relevant area will always result in an increased behavioral performance (e.g. reaction time). Luber et al. (2014) reported cases where TMS resulted in facilitated, rather than the usually expected inhibited cognitive performance.

Unlike the figure of eight coil used in TMS, in tES stimulation, the electrodes are quite large (usually 2cm*2cm). Therefore, tES induces a more wide-spread change in excitability which is less focal and effective compared to TMS. Yet, in tES, participants rarely experience sensations related to the treatment, may just be a slight itching sensation (Nitsche et al., 2008), which reduces the distraction from the experimental task, and makes it easier to conduct a sham condition.

There are two electrodes in tES stimulation, which makes it possible to both inhibit the excitability of a given brain area with cathodal stimulation and facilitate the excitability with anodal stimulation (Parkin et al., 2015). Thus, makes it possible to relate the caused behavioural changes to the change of brain excitabilities.

1.4.4. Summary

In summary, there are two types of non-invasive brain stimulation, transcranial magnetic stimulation and transcranial electrical stimulation. Whereas TMS stimulation can be very focal and effective, the muscular side effect caused by TMS might cause a distraction from the experimental task, besides, the behavioural performance caused by increased and/or decreased brain activity induced by TMS is quite task-dependent. On the contrary, tES induces a more wide-spread change in excitability that is less focal and effective, but the sensation caused by tES is almost ignorable, moreover, tES can provide the changes of behavioural performance by both increased and decreased brain excitability. Despite the side effects of NIBS such as headache, twitching, scalp discomfort, and tinnitus, both of them are safe modalities, especially when safety guidelines are followed (Krishnan et al., 2015), and can effectively induce changes in neuronal activity. This thesis will makes use of TMS to directly and focally interfere with the ongoing activity to investigate the integration of gesture and speech. Meanwhile, tDCS would be used to induce both increased and decreased excitability of IFG and pMTG to provide further evidence for the involvement of the two brain area in gesture-speech integration, as well as tRNS to stimulate the two brain areas synchronously to investigate whether the two anatomically connected brain areas compensate for each other.

1.5. Overview of the present thesis

1.5.1. Summary of background and research aims

As extra-linguistic information that has often been observed to accompany language, gesture is proposed to be regarded as 'part of language' (Kendon, 1997) to integrate with speech (Ozyurek et al., 2007). The information from gesture and speech is incorporated into a 'single unification space' (Hagoort, 2004, 2005; Ozyurek et al., 2007) with overlapping neuronal sources. This thesis used Non-invasive brain stimulation (NIBS) to investigate 'where' and 'when' the integration of gesture and speech takes place.

Even though IFG and pMTG have consistently been found to be involved in the integration of gesture and speech, there is no consensus regarding which brain area is causally involved. Some studies found the involvement of IFG (Green et al., 2009; Willems et al., 2007; Willems et al., 2009), some found the the activation of pMTG (Holle&Gunter, 2007; Holle et al., 2008; Holle et al., 2010) and others observed activation of both IFG and pMTG (Dick et al., 2012; Dick et al., 2014; Straube et al., 2011). In Chapter 3, it was aimed to provide causal evidence for the involvement of IFG and pMTG in the gesture-speech integration with the use of transcranial magnetic stimulation. It was also aimed to provide further evidence for the involvement of IFG and pMTG in gesture-speech integration with the use of transcranial electrical stimulation in Chapter 5.

Previous studies that investigated the time window for the integration of gesture and speech do not have a unified manipulation of the synchronization of gesture and speech. In the studies of Obermeier et al., (2011, 2015), they manipulated the synchronization between the offset of the DP of the gesture and the IP of the speech, while Habets et al., (2010) manipulated the synchronization between the onset of the stroke phase of gesture and the onset of the word. Moreover, given that both the closeness in time and the similarities in semantics will have an influence on whether two sources of information will be perceived as one unified event, it is believed that the question of the integration of gesture and speech cannot be addressed by simply manipulating the synchronization between the two. In Chapter 4, the 400ms after the onset of speech was split into 10 time windows and applied double-pulse TMS stimulation over each time window to provide clear evidence for the time window associated with the integration of gesture and speech.

1.5.2. Overview of present thesis

There are three purposes in this present thesis. 1. To provide causal evidence for the involvement of IFG and/or pMTG in the gesture-speech integration; 2. To investigate the time point for the integration of gesture and speech in pMTG; 3. To investigate the effect of decreased or increased excitability of IFG and/or pMTG on gesture-speech integration. There are three chapters of experiments in this thesis, corresponding to each of the questions.

1.5.2.1. Overview of Chapter 2

For all our experiments, the experimental paradigm of Kelly et al. (2010a) was used, in which they presented participants with videos of a man and a woman gesturing and speaking about common actions. The videos differed as to whether the gender of the gesture and speaker was the same or different and whether the content of the gesture and speech was congruent or incongruent. The reaction time of identifying the gender of the voice the subject heard was recorded. Kelly et al. found that even when participants were not asked to pay attention to the content of the gesture and speech, their reaction times were significantly longer when the content contained in gesture and speech was incongruent compared to when it was congruent. Kelly et al. interpreted this result as an illustration of an 'automatic integration' of gesture and speech.

Chapter 2 introduced the general methodology that was used for all the experiments. First of all, three pre-tests were conducted to create and validate our experimental stimuli. In pre-test 1, the relationship between gesture and speech was rated on a 5 point rating scale, to verify that the semantically congruent or incongruent combinations of gesture and speech were indeed perceived as such. In pre-test 2, stimuli was validated by replicating the results of Kelly et al. (2010a) with the hypothesis that even when participants are not asked to pay attention to gesture, the information contained in gesture still has an influence on the reaction time to judge the gender of the speaker, with significantly longer reaction time in gesture-speech incongruent condition compared to the congruent condition. In pre-test 3, participants were asked to use one verb to describe the gesture presented in the video, in order to make sure that the stimulus set could be characterized as containing iconic gestures.

A quantitative meta-analysis of several published functional magnetic resonance imaging (fMRI) studies on iconic-speech integration was also conducted to find the consistently activated brain areas, which were then used as our stimulation sites.

1.5.2.2. Overview of Chapter 3

In chapter 3, transcranial magnetic stimulation (TMS) was used to test for a potential causal role of IFG and/or pSTS/MTG in gesture-speech integration. LIFG and pMTG were stimulated with the assumption that if either area is critically involved in gesture-speech integration, TMS would disrupt that integration, which would be reflected in terms of the reaction time to judge the gender of the speaker.

There are two experiments in chapter 3. In Experiment 1, offline theta-burst TMS stimulation on both IFG and pMTG was applied to directly interfere with any ongoing activities. Participants were asked to judge the gender of the speaker by pressing the button as quickly and accurately as possible. It was hypothesised that even though the participants were not asked to pay attention to the semantic relationship of the gesture and speech, the semantic relationship would still have an influence on their reaction time (RT) for judging the gender of the speaker, with RTs in the gesture-speech semantic incongruent condition significantly larger than the congruent condition. In Experiment 2, online TMS stimulation on pMTG was applied, in which participants performed the experimental task under concurrent pMTG stimulation. It was intended to replicate the pMTG results from Experiment 1, and thus provide more evidence for the involvement of pMTG in gesture-speech integration, as well as validate the TBS stimulation.

1.5.2.3. Overview of Chapter 4

Chapter 4 aimed to find the time window for the integration between gesture and speech. The 400ms after the onset of the speech was divided into 10 time windows (TW): 0/40ms, 40/80ms, 80/120ms, 120/160ms, 160/200ms, 200/240ms, 240/280ms, 280/320ms, 320/360ms, 360/400ms and applied double pulse TMS stimulation to pMTG during each time window. It was assumed that if any time window is critically involved in gesture-speech integration, TMS would disrupt that integration, which would be reflected in terms of the reaction time to judge the gender of the speaker.

There are two experiments in Chapter 4. In Experiment 3, the gestures were presented 200ms ahead of speech, and double pulse TMS stimulation was applied over each of the ten time windows. It was hypothesized that double pulse TMS stimulation would interrupt any integration activity during a specific time window, resulting in a significantly reduced semantic congruency effect in the pMTG condition, relative to control site stimulation. The subsequent Experiment 4 determined the gating point for each gesture, with a definition that the gating point is the earliest point in time at which a clear understanding of that gesture emerges. Then reanalysed the data obtained in Experiment 3, but taking the individual gesture gating points into account. This involved equalizing the effect of different gating points for the gestures, so that Time Window 1 always corresponded to the earliest point in time at which a gesture interpretation became possible.

1.5.2.4. Overview of Chapter 5

In chapter 5, transcranial electrical stimulation (tES) was used to investigate the effect of selectively increasing or decreasing the excitability of pMTG and IFG on gesture-speech integration.

There are three experiments in chapter 5. In Experiments 5 & 6, transcranial direct current stimulation (tDCS) was applied over either IFG (Experiment 5) or pMTG (Experiment 6). The hypotheses were motivated by the observation that anodal tDCS stimulation increases cortical excitability, whereas cathodal tDCS stimulation decreases cortical excitability. Therefore, it was expected a significantly reduced semantic congruency effect in the cathodal condition compared to the sham condition and a significantly increased semantic congruency effect in the anodal condition as compared to the sham condition.

The effect of simultaneously changing excitabilityty in both areas at the same time was also being observed. Such a study could shed light on whether IFG and pMTG can compensate for each other in the integration of gesture and speech, or whether they make independent contributions.

Therefore, in Experiment 7, transcranial random noise stimulation (tRNS) was applied to both IFG and pMTG synchronously. It was hypothesized that low frequency tRNS over both IFG and pMTG would decrease the excitability of the two brain areas in a similar way, while high frequency would increase the excitability of the two brain areas. If both areas form a functional network, with one area compensating for the other, then it should be expected to only see an effect in Experiment 7 (where excitability of both areas is changed at the same time), but not Experiments 5 and 6, where only one area's excitability is changed at a time..

Chapter 2. General methods

2.1. Participants

The participants in all experiments were native speakers, and were classified as right-handed according to the Edinburgh Handedness questionnaire (Oldfield, 1971). They also had normal or corrected-to-normal vision and be screened for TMS exclusion criteria using a medical questionnaire. All the experiments were approved by the Ethics Committee of the Department of Psychology, University of Hull. Participants were remunerated at a rate of £8 per hour for taking part in the experiment. All participants had to sign the consent form prior to taking part in this study.

2.2. Stimuli

The stimuli in all experiments were based on a list of 44 common action verbs that could easily be expressed by a gesture without using an actual object (e.g., e.g., typing on a keyboard, throwing a ball, zipping up a coat). The action verbs were selected based on previous studies on co-speech gestures (Dick et al., 2014; Drijvers et al., 2017; Kelly et al., 2010a). Each action was performed by two native speakers of English (1 male, 1 female) with only the torso visible, while simultaneously naming the corresponding action verb, yielding 88 video items. Stimuli were edited subsequently so that each gesture started with the stroke phase. The spoken words were subsequently re-recorded using the same speakers, who were also gesturing while saying the word. The re-recorded words were then digitally inserted 200ms after the onset of each gesture stroke.

By digitally inserting sounds into the relevant videos, two experimental manipulations were created. First, gender congruency was manipulated. In order to achieve this, half of all the videos included a gender congruent voice (e.g. a man doing a gesture combined with a male voice; or a woman doing a gesture combined with a female voice), whereas the other half of the videos included a gender incongruent voice (e.g. a man doing a gesture with a female voice; or a woman doing a gesture with a male voice).

Second, semantic congruency was also manipulated. To achieve this, half of all the videos included a semantically congruent speech token (e.g. a man or a woman doing a 'cut' gesture while saying 'cut'), while the other half of the videos contained a semantically incongruent speech token (e.g. a man doing a 'cut' gesture while saying 'stir').

Thus, there are two factors in the experimental design, semantic congruency (semantic congruent vs semantic incongruent), and gender congruency (gender congruent vs gender incongruent) (see Fig.2).

2.2.1. Pre-test1: semantic congruency rating of gesture-speech stimuli

To verify that the semantically congruent or incongruent combinations of gesture and speech were indeed perceived as such, a separate set of participants (n=21) rated the relationship between gesture and the speech on a 5 point rating scale (1 means "no relation" and 5 means "very strong relation"). Based on the rating results, two pairs of stimuli were excluded from the stimulus set for ambiguous relationships and two further pairs of stimuli were moved from the main stimulus set to the practice set. The 40 remaining stimuli were used for further pre-tests. The mean rating for the remaining set of congruent pairs was 4.71 (SD = 0.32), and the mean rating for the incongruent pairs was 1.28 (SD = 0.29). Paired sample t-test showed that there is a significant difference between the congruent pairs and the incongruent pairs (t (79) = 98.8, p=.000). Thus, it was verified that the semantic congruent pairs were been perceived as have a very strong gesture-speech relation (4.71 out of 5), while the semantic incongruent pairs were been perceived as have no gesture-speech relation (1.28 out of 1). The semantic congruent and the semantic incongruent sit on the opposite side of the semantic congruency of gesture and speech with a significant difference.

2.2.2. Pre-test 2: validation of stimuli by behavioural replication of Kelly et al., 2009

Next, it was aimed to see whether the behavioural finding of Kelly et al. (2010a) could replicate with the stimulus set. Since each video-sound pair existed in four different versions, corresponding to the four conditions (see Figure 2.2 for an example), the total stimulus set consisted of 320 items. The video started at the stroke phase of the gesture, the speech onset occurred 200ms after the onset of the video. Reaction time was recorded relative to the onset of speech, the video stopped at the moment when the participant pressed the button. During the variable inter-trial interval (ITI) of 1.5 to 2.5 sec, a fixation cross was presented on the centre of the screen. All 320 items were presented in five blocks using Presentation software (www.neurobs.com) in a pseudo random order. Participants were allowed to have a break after each block and were told to press any button if they were ready to continue.

37 Participants (7 male, 30 female, age 18-45, mean age 21.57, SD=5.87) took part in Pre-test 2. They were informed that the gender they saw on the screen and the gender of the voice they heard might be different, or might be the same. They were asked to indicate the gender of the voice they heard by pressing one of two buttons on the keyboard (key assignment was counterbalanced across participants). During the experiment, they were asked to look at the screen but only pay attention to the gender of the voice they heard, and press the buttons as quickly and accurately as possible. Their reaction time and the button being pressed were recorded. Only the correct responses were used for analysis of RTs. Participants made very few errors on the task (overall accuracy > 99%), therefore accuracy data were not statistically analysed.

Following a 2.5% trimmed mean procedure (removal of the lowest 2.5% and the highest 2.5% for every participant), it was found that participants were slower to judge the gender of the speaker when speech and gesture were semantically incongruent (M=643.09ms, SD= 108.34) relative to when they were semantically congruent (M=632.92ms, SD=103.84). In addition, gender congruency negatively affected reaction time, with gender incongruent trials eliciting slower reaction times (M=650.68ms, SD= 102.81) than gender congruent trials (M=625.68ms, SD= 109.50).

	Semantic congruent		Semantic incongruent	
	Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
Reaction time (SE)	619(18)	648(17)	633(18)	654(17)

Table 1-2 Mean reaction times in ms(SEM) for Pre-test 2

A 2 (Semantic Congruency) by 2 (Gender Congruency) repeated measures ANOVA on the RT data revealed a significant main effect of gender congruency, F (1, 36) = 63.45, p<.001. There was also a significant main effect of Semantic Congruency, F (1, 36) = 15.12, p<.001, but no significant interaction of semantic congruency by gender congruency, F (1, 36) = 1.92, p=.18.

Thus, in Pre-test2, the experimental paradigm has been validated by replicating the results of Kelly et al. (2010a), who used a similar experimental setup and task. Along with a main effect of gender congruency, Kelly et al. also found that semantic congruency between gesture and speech has a significant influence on participants' reaction time in judging of the gender of the voice. Kelly et al. (2010a) concluded that gesture and speech are automatically integrated with each other, and that people cannot help but pay attention to information conveyed in both modalities. Thus, the magnitude of the reaction time effect for semantic congruency (RT (semantic incongruent) –RT (semantic congruent)) can be taken as a behavioural index of gesture-speech integration.

To maximize the statistical power of the brain stimulation experiments, the results of Pre-test 2 was used to select those item pairs that produced the strongest semantic congruency effect. To this end, those four stimulus pairs were excluded from the stimulus set that did not show a semantic congruency effect in the expected direction (RT (semantic incongruent) –RT (semantic congruent)). Thus, the final stimulus set consisted of 32 gestures (16 pairs, see Appendix A). Since each gesture was performed either by a male or a female actor, as semantically congruent or incongruent, and also as gender congruent or incongruent, the total stimulus set consisted of 256 videos (32*2*2*2).

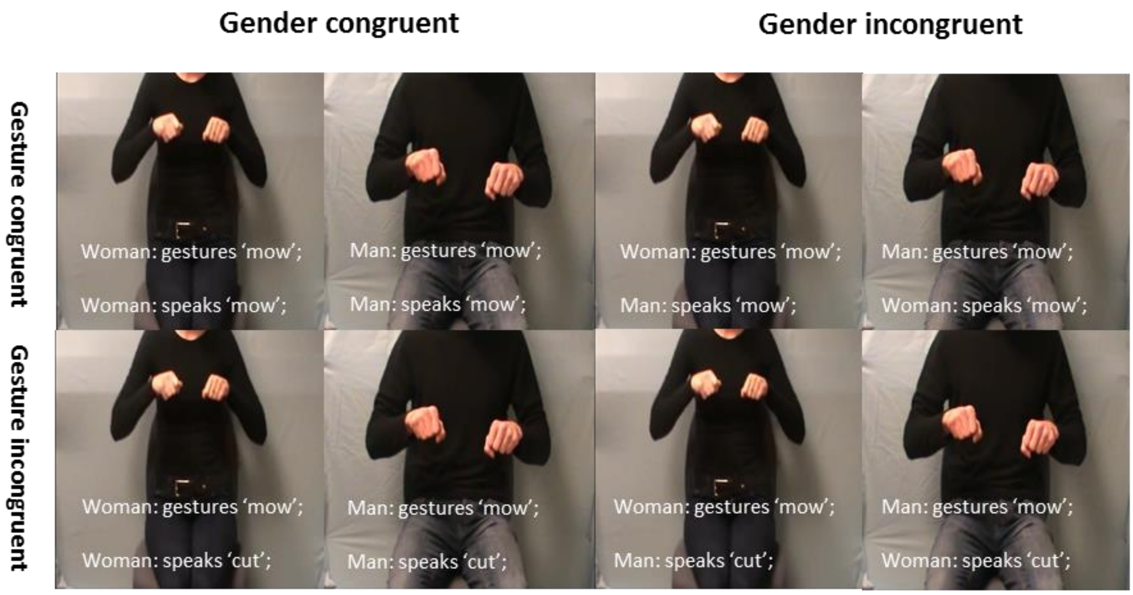


Figure 1.7 Still frame examples of the semantic congruency and gender congruency

2.2.3. Pre-test 3: naming study of gesture stimuli

Gestures differ with respect to the meaning and function that gesture possesses, According to McNeill (1992), there are four major categories of gestures (see the 'gesture types' in Chapter 1, page 9 for detail). This thesis focused on iconic gestures, which were defined by McNeill as spontaneous hand actions that represent body movements, movements of objects without holding the object, whose meaning has to be generated online on the basis of gesture form and the co-speech context in which the gesture is observed. To guarantee the hand actions used in the experiments were iconics, Pre-test 3 was conducted.

Another separate set of participants (42 native speakers, 20 male, 22 female) were asked to use one verb to describe the gesture presented in the video. All the items were presented without speech. For each gesture, the percentage of participants that provided the correct label was then calculated. The overall mean nameability index of the final stimulus set of 32 gestures was 49%. This indicates that as a whole, the stimulus set is best characterized as containing iconic gestures, which are characterized by a certain ambiguity when presented in the absence of speech (Drijvers&Ozyurek, 2017; Hadar et al., 2004).

2.3. Brain stimulation coordinates

A coordinate-based quantitative meta-analysis of several published functional magnetic resonance imaging (fMRI) studies using Ginger Ale 2.3 (www.brain.org) was performed. The activation likelihood estimation (Eickhoff et al., 2009) was based on iconic-speech integration (e.g. multimodal>unimodal) from 7 studies using healthy participants(Dick et al., 2012; Green et al., 2009; Holle et al., 2008; Holle et al., 2010; Straube et al., 2011; Willems et al., 2007; Willems et al., 2009). Two sites corresponding to Montreal Neurological Institute (MNI) coordinates were identified as consistently activated across studies, the left IFG (-62, 16, 22) and the left pMTG (-50, -56, 10), and were used as the two stimulation sites in my study. The Vertex was used as a control site.

To enable an image-guided TMS navigation, high resolution (1 x 1 x 0.6 mm) T1 weighted anatomical MRI scans of each participant were acquired at Hull Royal Infirmary using a GE medical systems scanner with a field strength of 3 Tesla. MNI coordinates of the target areas were defined as regions of interest (ROIs) using Marsbar (marsbar.sourceforge.net) and SPM12 (see Fig. 2.3). These ROIs were then backprojected from MNI space into each participant's native brain space, using SPM12's inverse transformation function. Subject-specific ROIs were then imported into BrainVoyager and superimposed on the surface reconstruction of the two hemispheres and defined as targets during neuronavigation. This ensured precise stimulation of each target region in each participant.

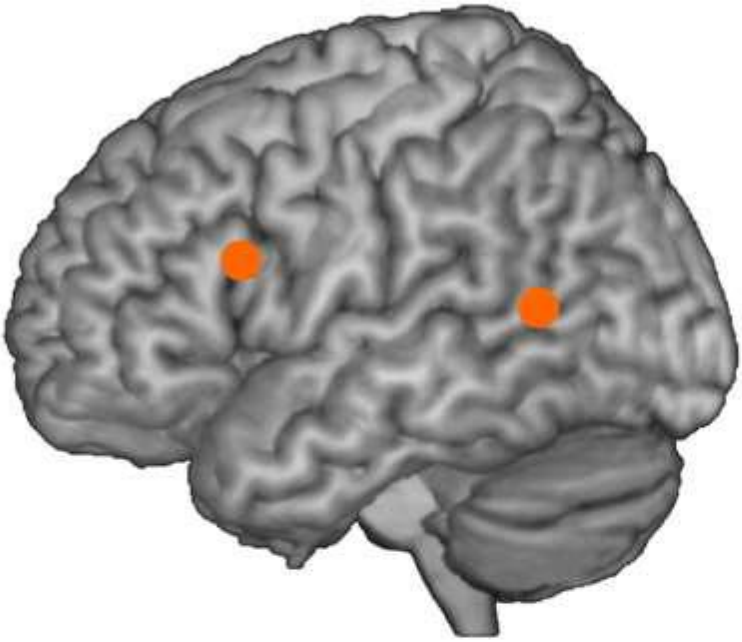


Figure 1.8 Overview of stimulation sites in MNI space.

IFG (-62, 16, 22) and pMTG (-50, -56, 10). Vertex was used as control site.

2.4. Data analysis

In all subsequently reported experiments, the experimental paradigm of Kelly et al. (2010a) was used by presenting participants with videos of a man and a woman gesturing and speaking at the same time. Participants were instructed to look at the screen but only pay attention to the gender of the voice they heard, and press the relevant button as quickly and accurately as possible. Their reaction time was recorded and analysed as the primary outcome variable. All experiments were conducted with one type of non-invasive brain stimulation, either online (the rTMS used in Chapter 3) or offline (the TMS used in Chapter 3 and the tES used in Chapter 5). In the online stimulation studies, participants performed the experimental task under concurrent rTMS stimulation; in the offline studies, participants first received a period of brain stimulation (either to an area of interest or of a control site), followed by completion of the experimental task. The experimental factor of the stimulation site was realized between sessions in all offline studies, but within sessions in the online stimulation study.

To eliminate the influence of outliers, after excluding incorrectly answered trials, outliers were determined as those scores that were located in the extreme 5% on either end of the Z normalized distribution of reaction times. This is equivalent to removing scores above and below 1.65 SD of a subject's mean RT. It was interested in three effects: The first analysis of each study was a repeated ANOVA, typically consisting of the factors Semantic Congruency (congruent, incongruent), Gender Congruency (same different) and a factor Stimulation site (e.g, vertex, pMTG). Based on Kelly et al. (2010a) and the results of the pre-tests, longer reaction times under control site stimulation when gesture and speech were semantically incongruent were expected. The critical test was whether the magnitude of this semantic congruency effect interacted with the factor Stimulation Site. The ANOVA was then followed up by paired t-tests, where appropriate, to see whether the magnitude of the semantic congruency effect (RT (semantic incongruent) – RT (semantic congruent)) differed between the stimulation condition and the Vertex condition. To see whether brain stimulation selectively affects the semantic congruency effect, the gender congruency effect between the stimulation condition and the Vertex condition was additionally compared with the use of a paired sample t-test.

Chapter 3. The involvement of IFG and pMTG in gesture-speech integration[5]

3.1. Introduction

Although spoken communication is often considered a purely auditory-vocal process, research in the past decade has provided unequivocal evidence that speech is fundamentally multimodal (Ozyurek, 2014). Across all spoken languages, speakers additionally take advantage of the visual-manual modality during communication in the form of hand gestures. For example, speakers may spontaneously use their hands to outline patterns (e.g., when describing the layout of their house) or re-enact actions (making a strumming movement while saying 'He played the instrument'). This happens not only during face-to-face communication, but also when the communicators cannot see each other (Bavelas et al., 1992; Krauss et al., 1995). There is now also convincing evidence that listeners pick up the additional information provided by gestures (Gunther et al., 2015; Kelly et al., 2010a; Wu et al., 2007). For example, a study carried out by Kelly et al. (1999) presented participants with the speech "I stayed up all night" together with the gesture pantomiming drinking, and asked the participants to write down what they heard. Participants' responses convey information from both the gesture and the speech with the sentence "I stayed up all night drinking". They may even not been able to tell whether a particular piece of information originated in the speech or the gesture channel if asked later (Alibali et al., 1997), suggesting that the visual information originating from gesture is integrated with the auditory information from speech into a single coherent semantic representation (Ozyurek, 2014). A contentious question has been where in the brain the merging of gesture and speech information occurs. Since the information extracted from each modality is qualitatively different (linear, segmented information with arbitrary form-meaning mapping in the

[5] Please note that the work described in this chapter has been submitted for publication (Zhao, W., Riggs, K.J., Schindler, I., Holle, H..) Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. Under review. Journal of Neuroscience.

case of speech vs. holistic, parallel information with motivated form-meaning mapping in the case of gesture), answering this question promises to deepen our understanding of the cortical interface between linguistic and pragmatic information.

Some authors have argued for a critical role of the IFG (Willems et al., 2007; Willems et al., 2009) in the integration of information from both gesture and speech. For example, in a study carried out by Willems et al. (2007) participants were presented with iconic gestures that were either congruent (verb: write, gesture: write) or incongruent (verb: write, gesture: hit) with speech, and found a significant greater activity in left IFG in incongruent conditions compared with congruent ones. Green et al. (2009) also found a greater activation in IFG when iconic gestures were unrelated to the accompanying speech as compared to iconic gestures that related to speech.

Others suggest that pMTG is critically involved (Holle et al., 2008; Holle et al., 2010) in the integration of gesture and speech. Holle, &Gunter (2007) asked participants to watched videos in which sentences containing an ambiguous word (e.g. she touched the mouse) were accompanied by either a meaningless grooming movement, a gesture supporting the more frequent dominant meaning (e.g. animal) or a gesture supporting the less frequent subordinate meaning (e.g. computer device). They found that when compared to the grooming movement, both types of gesture (dominant and subordinate) activated left STS. Another study by Holle et al. (2010), provided evidence suggesting that the integration of iconic gestures and speech takes place in pSTS/STG.

There are also researchers who suggested that both areas might be causally involved in linking semantic information extracted from the two domains (Dick et al., 2014). For example, Dick et al. (2012) found stronger activation of both LIFG as well as the activation of pMTG when gestures contain disambiguating information with respect to speech compared with when they contain redundant information with respect to speech.

Straube et al. (2011) also found the left posterior middle temporal gyrus (MTG) been activated during the iconic gesture condition, with the involvement of the IFG restricted to paradigms that induce semantic conflict.

To date, no consensus has been reached as to which of these areas is causally, and not merely epiphenomenal, involved in this merging process. IFG and pMTG are anatomically well-connected (Friederici, 2009) which can produce correlated patterns of activation between these regions (Whitney et al., 2011). It is therefore also possible, for example, that activation of pMTG alone is crucial for gesture-speech integration, with IFG activation merely a consequence of its strong anatomical connection with pMTG (or vice versa). FMRI studies are in this respect limited with respect to the degree to which they allow inferences of causality.

In Experiments 1 and 2, we used transcranial magnetic stimulation (TMS), a method that is ideally suited to identify causal brain-behaviour relationships. This allowed us to disrupt activity in either left IFG or pMTG and observe its effect on gesture-speech integration. Using the Kelly et al. (2010a) task, we tested whether the magnitude of the semantic congruency effect is reduced when activity in these areas is perturbed using TMS, relative to control site stimulation.

In Experiment 1, we used 'off-line' theta-burst TMS stimulation over both IFG and pMTG to examine the role of the IFG and pMTG in gesture-speech integration. In the second experiment, we used online TMS stimulation over pMTG to directly interfere with any ongoing activities. Online TMS, in the form of repetitive transcranial magnetic stimulation, has the advantage that perturbation of cortical activity can be synchronized with the presentation of the experimental stimuli. This enables a more powerful statistical comparison of the effects of brain stimulation on gesture-speech integration compared to the 'off-line' theta-burst TMS stimulation.

3.2. Experiment 1

In Experiment 1, we explored whether disrupting activity in areas hypothesized to underlie gesture-speech integration (left IFG and/or left pMTG) leads to a reduction of the semantic congruency effect. In a within-subject design, participants underwent three sessions, where continuous theta burst stimulation (cTBS) was either applied to the left IFG, pMTG or a control site (vertex). The session order was counterbalanced across participants. After stimulation, which occurred at the beginning of each session, participants completed the reaction time task described above (section Pre-test 2, Chapter 2). Thus, the full experimental design was a 3 (Stimulation Site: IFG, pMTG, Vertex) x 2 (Gender Congruency) x 2 (Semantic Congruency) factorial design. During the statistical analysis, Greenhouse-Geisser correction was applied where necessary. We predicted an interaction between Stimulation Site and Semantic Congruency, in the form of a reduction of the semantic congruency effect following either IFG or pMTG stimulation, relative to control site stimulation. The factor of Gender Congruency was used as an additional control, to see whether brain stimulation specifically disrupts the processing of semantic (in)congruencies, or more generally interferes with task processing.

3.2.1. Method

3.2.1.1. Participants

Twenty participants took part in Experiment 1 having given written informed consent. Three participants were excluded from the analysis, one for not being able to follow instructions, and another two because of computer malfunction. The experimental protocol was approved by the Ethics Committee of the Department of Psychology. The final sample consisted of 17 participants (6 males and 11 females, age 20-42, mean age 25.06, SD= 5.87). All were native English speakers, and were classified as right-handed according to the Edinburgh Handedness form (LQ=73.51, SD=22.12), had normal or corrected-to-normal vision and were screened for TMS suitability using a medical questionnaire.

3.2.1.2. Apparatus and stimuli

The stimuli consisted of 256 videos, as described in Chapter 2. A Magstim Rapid² stimulator was used to generate repetitive magnetic pulses. A cTBS train of 804 pulses was used (268 bursts, each burst consisting of three pulses at 30 Hz, repeated at intervals of 100 ms). Stimulation intensity was set to 40% of maximum machine output and lasted for 40 s (Nyffeler et al., 2009; Nyffeler et al., 2008). The pulses were delivered with a standard 70mm figure-8 coil.

The centre of the cTBS coil was targeting either the left IFG (-62, 16, 22), the left pMTG (-50, -56, 10), (corresponded to MNI coordinates that was previously determined in the meta-analysis of FMRI studies), or Vertex. Each stimulation condition was realized in a separate session. The position of the coil was confirmed by the MRI based neuronavigation system, in this way the coil can targeting the stimulation point based on individual brain scan. Duration and intensity of the cTBS stimulation are both within participants' bearable limit and the neuropsychological application safety limit (Anand et al., 2002).

3.2.1.3. Procedure

There are three sessions with cTBS stimulating either pMTG, IFG or Vertex. Sessions were scheduled to be at least one week apart. In each session, participants were guided to sit in front of a computer and keyboard. After theta-burst stimulation, they were asked to complete the experimental task consisting of 256 gesture videos. Participants were asked to pay attention to the gender of the speaker and respond as quickly and accurately as possible. Responses were made using the left and right index finger, and key assignment was counterbalanced across participants.

3.2.2. Data analyses

All incorrect responses (302 out of the total number of 8160, 3.7% of trials) were excluded. To eliminate the influence of outliers, the lowest 2.5% and the highest 2.5% for every participant in each session were also excluded. We focus our analysis on the main effect of semantic congruency and its interactions with stimulate sites, as well as the gender congruency effect as a control effect.

We ran a 3 (stimulate site: Control condition, pMTG condition and IFG condition) * 2 (semantic congruency: semantic congruent vs semantic incongruent) *2 (gender congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs to see if Theta-burst TMS stimulation caused any significantly reduced magnitude of gesture-speech integration.

3.2.3. Results

The trimmed reaction time data were subjected to a repeated measures ANOVA with the factors Stimulation Site (IFG, pMTG, vertex), Gender Congruency (same, different) and Semantic Congruency (congruent, incongruent). The ANOVA yielded a significant main effect of Semantic Congruency ($F(1, 16) = 14.64$, $p = .001$), reflecting longer RTs or semantically incongruent trials ($M = 543$, $SE = 16.6$) than congruent trials ($M = 530$, $SE = 13.7$). Furthermore, a significant main effect of Gender Congruency ($F(1, 16) = 45.37$, $p < .001$) was observed, indicating that reaction times were longer when speech and gesture were produced by different genders ($M = 554$, $SE = 15.2$) than the same gender ($M = 518$, $SE = 15.4$). Crucially, there was a significant Semantic Congruency by Stimulation Site interaction ($F(1.944, 30.466) = 3.53$, $p = .042$), indicating that the magnitude of the semantic congruency effect was modulated depending on the brain area stimulated. No such modulation was observed for the gender congruency effect, as indicated by a non-significant Stimulation Site by Gender Congruency interaction ($F(1.944, 30.466) = 0.50$, $p = .60$). The full pattern of results is shown in Table 3.1

	Semantic congruent		Semantic incongruent	
	Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
IFG	519 (17)	549 (17)	526 (22)	562 (19)
pMTG	495 (11)	535 (13)	508 (13)	543 (13)
Vertex (control)	521 (17)	559 (19)	540 (20)	578 (22)

Table 3-1 Mean reaction time in ms (SEM) for Experiment 1

Simple effects analyses indicated that the size of the semantic congruency effect was significantly reduced (t (16) = 2.58, p = .020) when cTBS was applied to the left IFG (M = 19.3, SE = 9.10), relative to control site stimulation (M = 38.6, SE = 10.1) (see Fig. 3.2). A similar pattern was observed following stimulation of pMTG (M = 21.1, SE = 4.6), although this comparison did not reach full significance (t (16) = 2.05, p = .057). There was no evidence that stimulation of either pMTG or IFG modulated the size of the effect of gender congruency (all t < 0.77, all p > .451, see Fig. 3.3).

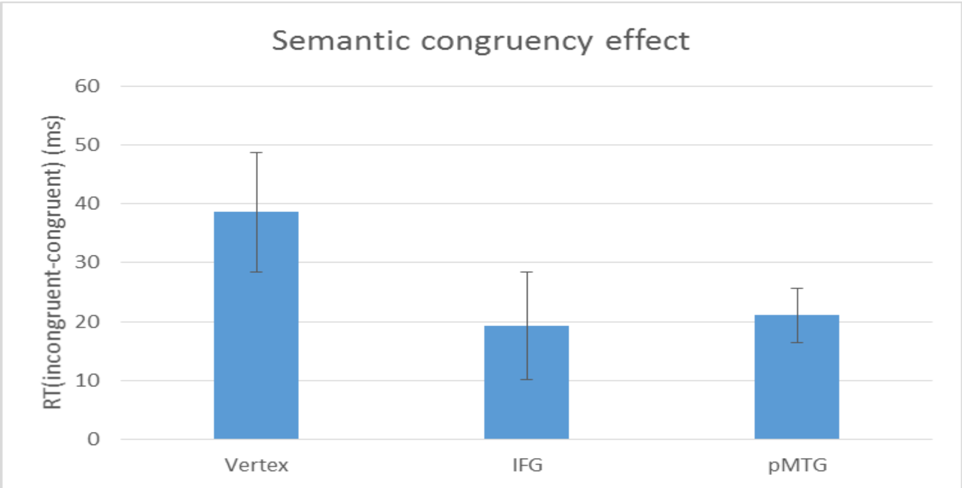


Figure 3.1 Magnitude of semantic congruency effect (ms) for Experiment 1. Error bars show M±SE.

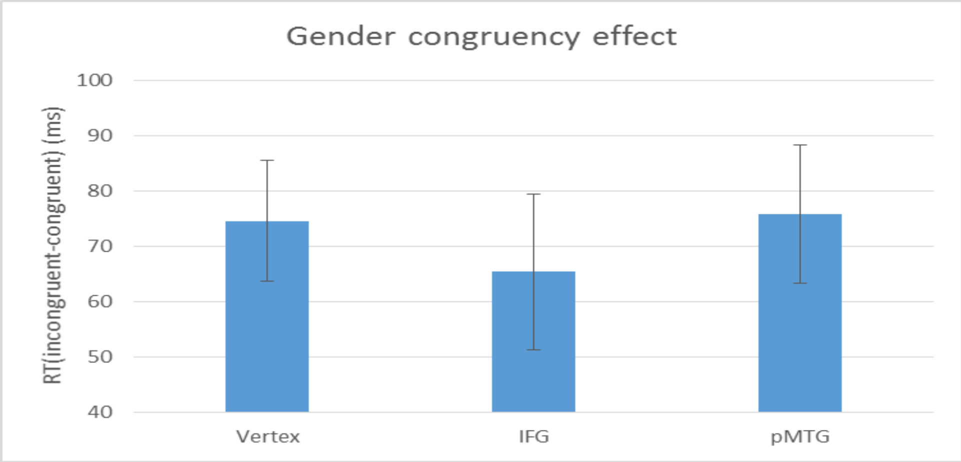


Figure 3.2 Magnitude of gender congruency effect (ms) for Experiment 1. Error bars show M±SE Discussion

Experiment 1 was designed to test whether IFG and/or pMTG is critically associated with gesture-speech integration with the use of Theta-burst TMS.

We can see from the results on the effect of semantic congruency that, compared to the control condition, cTBS stimulation of IFG caused a significantly reduced difference between the reaction time in the semantic incongruent condition and the semantic congruent condition. For the effect of gender congruency, there is no significant difference between the Control condition and the IFG condition, suggesting that cTBS stimulation of the left IFG specifically disrupts gesture-speech integration.

The results on the RTs in the Control condition (vertex stimulation) showed a significant difference between the semantic incongruent condition and the semantic congruent condition. That means that, even if participants were not asked to pay attention to the gesture information, the information contained in gesture still had an influence on participant's behaviour, suggested an automatic integration of gesture and speech, as reported by Kelly et al. (2010a).

In summary, we may conclude from Experiment 1 that IFG is causally involved in gesture-speech integration, whereas the pMTG has displayed only a tendency to be involved in the automatic integration of gesture with speech.

3.3. Experiment 2

In Experiment 1, the effect of brain stimulation to pMTG on the semantic congruency effect did not reach full significance (p = .057, see results section). To further investigate a possible role for pMTG, Experiment 2 was conducted using online TMS, as opposed to offline cTBS, to disrupt brain activity. Online TMS, in the form of repetitive transcranial magnetic stimulation (rTMS), has the advantage that perturbation of cortical activity can be synchronized with the presentation of the experimental stimuli. This enables a more powerful statistical comparison of the effects of brain stimulation on gesture-speech integration. Experiment 2 employed a 2 (Stimulation Site: pMTG, vertex) by 2 (Gender Congruency) by 2 (Semantic Congruency) factorial design. We predicted that pMTG would significantly reduce the semantic congruency effect, as indicated by a significant Stimulation Site by Semantic Congruency interaction. Furthermore, we hypothesized that rTMS of pMTG would specifically disrupt gesture speech integration, but not general task processing, as indicated by an absent interaction of Stimulation Site and Gender Congruency.

3.3.1. Method

3.3.1.1. Participants

Thirteen participants participated in Experiment 2. One participant was excluded for not following instructions. The final sample used for statistical analysis consisted of 5 males and 7 females (age range: 20-36, mean age: 24.08, SD= 4.36). All were English native speakers, and were classified as right-handed according to the Edinburgh Handedness form questionnaire (Oldfield, 1971), with a Mean Laterality Coefficient of 71.32 (SD=21.42). All other participant details were as described in Experiment 1.

3.3.1.2. Apparatus and stimuli

The same stimuli as before were used. rTMS was delivered using a 70mm figure-of-eight coil. Five pulses over a 500 ms period, with a frequency of 10HZ repetitive TMS (rTMS) were delivered at a fixed intensity of 45% of the maximal stimulator output were used in Experiment 2. These pulses were delivered 'on-line' during each video item. The first pulse coincided with the onset of the sound, and the last pulse was 400ms after the onset of the sound. Given the fact that the gesture started 200ms prior to sound, and previous research indicates that semantic gesture-speech integration takes place between 350-550ms after the onset of the gesture stroke (Ozyurek et al., 2007), we predicted that such stimulation over a relevant brain should impair the possible integration process. Duration and intensity of the rTMS stimulation are both within participants' bearable limit and the neuropsychological application safety limit (Anand&Hotson, 2002).

3.3.1.3. Procedure

Each participant completed four blocks of alternating pMTG (-50, -56, 10) or vertex stimulation in a single experimental session. Presentation order was counterbalanced across participants. Within each block, 64 trials were presented. The experimenter explained to the participant that they would be presented with a number of videos with rTMS stimulation on either pMTG or Vertex as a control site. The videos were presented using Presentation software (Version 17.2, www.neurobs.com). All other experimental details were as described in Experiment 1.

3.3.2. Data analyses

A total number of 106 trials were excluded because of incorrect responses or failing to respond within 2000ms (5.5% of trials). To eliminate the influence of outliers and prevent the possibility of a false positive, a 2.5% trimmed mean for every participant were also excluded. We focus our statistical analysis on the effect of semantic congruency (which is taken as an index of gesture-speech integration) and its interactions with the factor stimulation site (pMTG, control) to see whether the magnitude of gesture-speech integration is reduced when activity in pMTG is stimulated, relative to vertex stimulation. Besides, we used the effect of gender congruency (the magnitude of reaction time between gender incongruent condition and gender congruent condition) as a control effect, with the assumption that rTMS stimulation would only have an effect on semantic congruency effect.

Accordingly, we ran a 2 (stimulation site: pMTG and Control) * 2 (semantic congruency: semantic congruent vs semantic incongruent) * 2 (gender congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs.

3.3.3. Results

A 2 x 2 x 2 repeated measures ANOVA on the trimmed reaction time data (see Table 4) revealed a significant main effect of Gender Congruency ($F(1,11) = 41.36$, $p < .001$), with longer RTs to gestures and speech from actors of different genders ($M = 584.7$, $SE = 19.0$) than to actors of the same gender ($M= 549.6$, $SE = 19.3$). Crucially, there was a significant interaction of Semantic Congruency and Stimulation Site ($F (1, 11) = 9.01$, $p = .012$), indicating that the magnitude of the semantic congruency effect was modulated by rTMS (see Fig. 3.5). No such effect of brain stimulation on the gender congruency effect was observed (see Fig. 3.6), as indicated by a non-significant Stimulation Site by Gender Congruency interaction ($F (1, 11) = 1.51$, $p = .245$). All other main effects or interactions of the ANOVA were not significant (all $F < 3.00$, all $p > .111$).

	Semantically congruent		Semantically incongruent	
	Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
pMTG	557 (21)	594 (19)	549 (15)	592 (21)
Vertex (control)	537 (25)	568 (17)	556 (19)	584 (23)

Table 3-2 Mean reaction time in ms (SEM) for Experiment 2

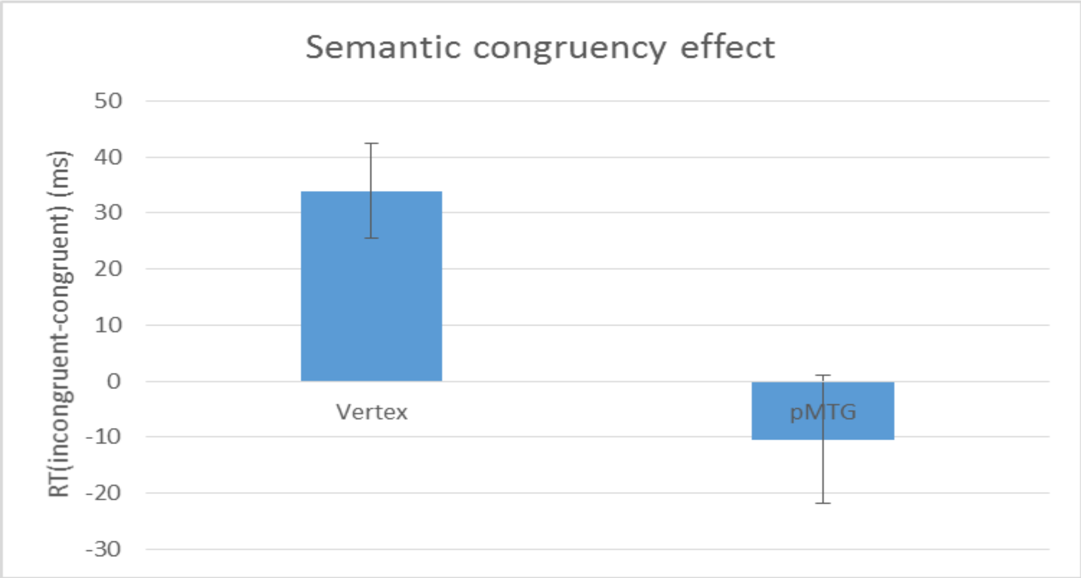


Figure 3.3 Magnitude of semantic congruency effect (ms) for Experiment 2. Error bars show M±SE.

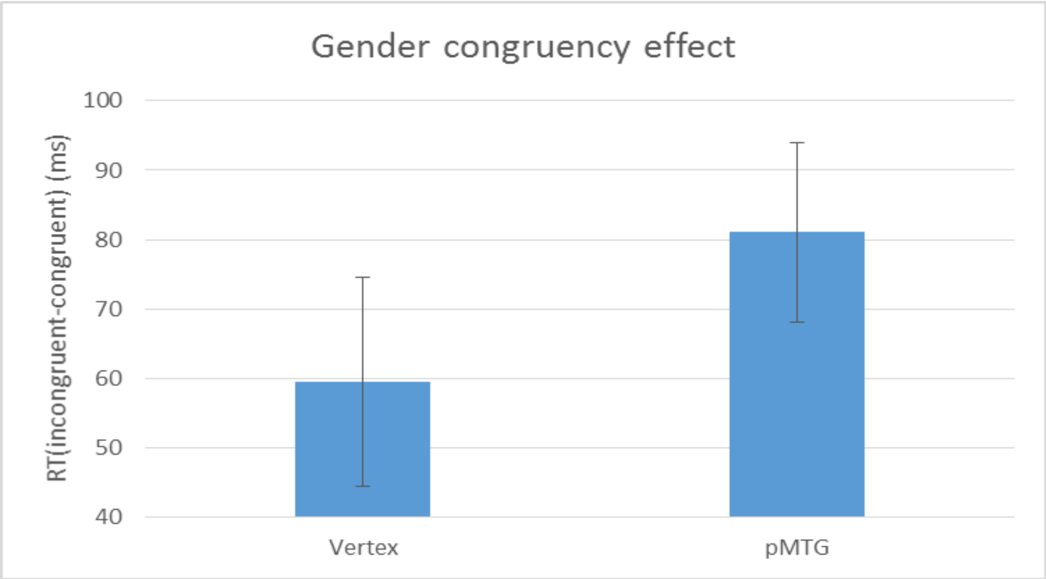


Figure 3.4 Magnitude of gender congruency effect (ms) for Experiment 2. Error bars show M±SE.

3.3.4. Discussion

The aim of Experiment 2 was to test whether pMTG is crucially involved in gesture-speech integration with the more direct and powerful rTMS stimulation.

We can see from the results on the effect of semantic congruency that rTMS stimulation of pMTG caused a significantly reduced semantic congruency effect as compared to the control condition. There was no significantly reduced gender congruency effect when rTMS stimulation over pMTG compared to the Vertex condition, suggesting that rTMS stimulation of pMTG specifically disrupts gesture-speech integration.

3.4. General discussion of Experiments 1 and 2

Taken together, the results of the two studies presented here provide clear evidence that both IFG and pMTG are involved in the merging of semantic information from iconic gestures and speech. When cortical excitability of these areas was decreased via TMS, a reduced RT cost was observed indicating a reduction in semantic integration. TMS of IFG and pMTG was found to specifically disrupt gesture speech integration, but not general task processing. By directly linking brain activity to behaviour, Experiments 1 and 2 demonstrate, for the first time, that both IFG and pMTG are causally involved in integrating information from the two domains of gesture and speech.

The results of Experiment 1 provided clear evidence that the left IFG is critically involved in the multimodal merging of information from the visual-manual (i.e., iconic gestures) and auditory-vocal (i.e. speech) modalities. Under control site stimulation, semantically incongruent combinations of gesture and speech elicited the typical reaction time costs, reflecting the automatic integration of information from the two modalities. The finding of a semantic congruency effect in the unperturbed brain (longer RTs for semantically incongruent gesture-speech units) replicates the results of the pre-test 2, as well as the finding by Kelly et al. (2010a). Crucially, these RT costs were significantly reduced when activity in the left IFG was disrupted, suggesting that this brain area is causally involved in the integration process. This finding is consistent with the idea that the functional role of the IFG during the comprehension of co-speech iconic gestures is the strategic recovery of context-appropriate semantic information, as explained below.

According to Whitney et al. (2011), semantic cognition involves (a) accessing information within the semantic store itself, and (b) executive mechanisms that direct semantic activation so that it is appropriate for the current context. In terms of its neural instantiation, it is often assumed that semantic cognition involves modulatory signals from the IFG acting upon temporal storage areas (Whitney et al., 2011). Since iconic gestures represent objects and actions by bearing only a partial resemblance to them (Wu et al., 2011), their comprehension may require strategic recovery of semantic activation, in order to come up with an interpretation of the observed gesture that is compatible with the accompanying speech context. For example, a gesture consisting of two closed fists moving forward from the body centre may initially only activate the general concept *push*, but needs additional strategic recovery of semantic activation, via modulatory signals from the IFG acting upon posterior temporal storage areas, to achieve an interpretation that is consistent with the accompanying speech unit *mow*.

Relative to congruent gesture-speech pairs, semantically incongruent combinations trigger an increased need for strategic recovery of semantic information, in an (eventually probably unsuccessful) attempt to resolve the semantic conflict between gesture and speech. Disrupting activity in the IFG interferes with this strategic recovery process, as reflected in the significantly decreased RT costs associated with the semantic congruency manipulation. The effect of cTBS on the IFG cannot be dismissed as a general disruption of cognitive processing, since stimulation of IFG specifically reduced the (task-irrelevant) semantic congruency effect, but not the (task-relevant) gender congruency effect.

IFG and pMTG most likely work together in integrating gesture with speech, with modulatory signals originating in the IFG acting upon temporal storage areas. The posterior temporal cortex, encompassing the middle temporal gyrus and adjacent superior temporal sulcus, has been suggested to be involved in accessing semantic information (Lau et al., 2008), either by serving as an interface to a widely distributed network of brain region representing semantic knowledge, or by accessing feature knowledge directly stored in pMTG. When presented in isolation, spoken words, lexicalized gestures, as well as less formalized iconic gestures all activate pMTG, which was interpreted that this area is indeed a hub for supramodal access of semantic information (Straube et al., 2012; Xua et al., 2009). Incongruent combinations of gesture and speech most likely place a higher semantic access load on this area than their congruent counterparts. From this perspective, our finding that rTMS of pMTG significantly reduces the size of the semantic congruency effect can be interpreted as an interference in the access of supramodal representations.

In conclusion, Experiments 1 and 2 provided clear evidence that IFG and pMTG are both critically involved in the integration of gestural and spoken information during comprehension. By linking cortical activity in these areas directly to observed behaviour, our study is the first to provide evidence that both areas are causally involved in this process.

Chapter 4. An investigation of the time window for the integration of gesture and speech using Transcranial Magnetic Stimulation

4.1. Introduction

As extra-linguistic information, gestures have often been observed to accompany language, for example, the inverted V of the fingers moved in a wiggling manner while saying, “He walked across the street”, or an upward climbing hand movement when saying, “He climbs up the wall”. Speakers convey information in both gesture and speech, while listeners are sensitive to information made available in both modalities (Goldin-Meadow&Sandhofer, 1999; Kelly&Church, 1998). This raises the question of what function gestures play in conversation. The traditional view is that gesture functions as a communicative device and that people have better comprehension when they are presented with speech accompanied with gesture (Hostetter&Alibali, 2011; Kelly, 2001; Valenzeno et al., 2003). Studies that shown gesture conveys not only relevant information (Goldin-Meadow&Sandhofer, 1999; Kelly&Church, 1998) but also additional information not present in the accompanying speech (Church&Goldinmeadow, 1986), proved the communicative function of gesture. However, the phenomenon that speakers gesture not only in face-to-face communication but also when they are on the phone (Bavelas et al., 1992; Krauss et al., 1995), or even in cases when they are blind and they themselves cannot see the gesture (Iverson&Goldin-Meadow, 1998) illustrates that gesture functions more than just for being communicative. Another proposal has been made by some researchers that gesture helps lexical retrieval of speech (Krauss et al., 2000; Pine et al., 2013; Rauscher et al., 1996; Rime et al., 1984; Wesp et al., 2001; Yap et al., 2010).

There are some theories regarding how gesture primes lexical retrieval. Rauscher et al. (1996) proposed that gesture primes lexical representation of the concept. Hadar et al. (1997) further explained that gesture is generated under the pre-lexical concept, the presentation of a motoric gesture will activate the equivalent lexical representation that is stored in memory, which will aid the following lexical selection. Krauss et al. (2000) proposed a lexical gesture process model to explain the facilitation effect of gesture on speech. According to Krauss, the motoric feature of the gesture will be picked up by a kinesthetic monitor, which will monitor the features of the source information represented in gesture, and then transfer the picked up information into the phonological encoder of speech, where it facilitates the lexical retrieval of word.

The facilitation effect of gesture on speech has been investigated by several studies. Kelly et al. (2010b) presented participants with action primes (e.g. someone chopping vegetables) and bimodal speech and targets and asked them to identify if any part of the target (speech or gesture) was related to the prime. The results showed that participants related primes to targets more quickly and accurately when they contained congruent information (speech: 'chop', gesture: 'chop') than when they contained incongruent information (speech: 'chop', gesture: 'twist'). Yap et al. (2011) presented participants with action primes followed by words (e.g. bird) or nonwords (e.g. flirp) and ask participants to respond as quickly and accurately as possible to whether the presented sound was a word or not. Results showed that participants responded faster in gesture-word related pairs (gesture: pair of hands flapping, word: bird) compared to gesture-word unrelated pairs (gesture: pair of hands drawing a square, word: bird). Pine et al. (2013) asked participants to name image of which the clarity increased gradually under three gestural conditions (gesture-image congruent condition, gesture-image incongruent condition, no gesture condition), results showed that participants named the objects significantly fastest in gesture-image congruent condition, and slowest in gesture-image incongruent condition.

There are two steps in the lexical retrieval of speech, one is during the phonological encoding phase, and the other one is during the semantic encoding phase. The studies reported above showed a facilitation effect of gesture on speech without distinguishing between the two phases. One may wonder of which stage the priming effect of gesture on the lexical retrieval of speech takes place.

In a study carried out by Kelly et al. (2004), they presented participants with gestures (primes) corresponding to a property of an object (size or shape of that object) preceded by spoken words (targets). When the information contained in gesture and words was incongruent, they observed an early P1/P2 sensory effect, followed by an N400 semantic effect. Kelly and his colleges argued that the gesture primes influenced word comprehension, first at the level of "sensory/phonological" processing and later at the level of semantic processing. However other electrophysiological studies that also found an N400 semantic effect in gesture-speech integration failed to find this early P1/P2 sensory effect. For example, Wu, &Coulson (2005) found the N400 effect in cases when incongruent iconic gestures produced in the course of conversation were compared with congruent ones (Wu&Coulson, 2005). Holle, &Gunter (2007) found the N400 effect to a word later in the sentence if the meaning of that later word did not match with the meaning indicated by the gesture earlier in the sentence. Ozyurek et al. (2007) manipulated the semantic fit of speech (i.e., a critical verb) and/or gesture in relation to the preceding part of the sentence (global integration) as well as the semantic relation between the temporally overlapping gesture and speech (local integration), and found that speech, gesture, and double mismatch conditions modulated the N400 in a similar way.

There are other studies that tried to explore the time phase for the effect of gesture on speech by manipulating the degree of the synchronization between the two. For example, in the study carried out by Habets et al. (2011), they investigated the degree of synchrony by manipulating the synchrony of gesture and speech into three conditions: the SOA 0 condition in which the onset of the stroke phase of gesture and speech were presented simultaneously; the SOA 160 condition and the SOA 360 condition, in which the onset of gesture was presented 160ms and 360ms respectively ahead of the onset of speech. Results showed a greater N400 component in the semantic incongruent condition compared to the semantic congruent condition only in the SOA 0 condition and the SOA 160 condition, while in the SOA 360 condition, there was no such significant N400 effect. Like most of the electrophysiological studies, Habets et al. also found merely the N400 effect, which shown that gesture primes speech in the semantic phase, when gesture was presented either synchronously or 160ms ahead of the speech.

We saw from the above studies that some studies found the facilitation of gesture solely on the semantic retrieval of speech (Habets et al., 2011; Holle&Gunter, 2007; Ozyurek et al., 2007; Wu&Coulson, 2005), while Kelly et al. (2004) found both the "sensory/phonological" processing and the semantic processing facilitation effect of gesture on speech. We concluded that there is no consensus at the moment. Moreover, we would also like to point out that the synchronization manipulation of gesture and speech in Habets's study cannot answer the question of the time phase of the lexical priming effect of gesture on speech. It appears that, in the face of two sources of information, there are two sets of factors which influence whether they will be perceived as one unified event or as two separate events. the low-level factors include the closeness of events in space or time (Calvert&Thesen, 2004), and the high-level factors include semantic similarities (Laurienti et al., 2003). In the study of Habets et al. (2011) they only considered the closeness of gesture and speech in time, without checking on which phase of speech the gesture had a priming effect.

In another two studies, Obermeier et al. (2011) made use of the disambiguation point (DP) of gesture and the identification point (IP) of speech to investigate this problem. Obermeier et al., (2011, 2015) replaced the full gesture with a minimal gesture fragment, with the hypothesis that from the onset of the gesture to the DP is the minimal fragment of gesture from which meaning can be disambiguated by the participants. They also used the identification point (IP) as the symbolization that from this point onwards participants can have a clear understanding of the meaning of the homonym. Obermeier, &Gunter (2015) manipulated the temporal alignment of the homonym and the gesture into four conditions: the offset of the gesture fragments was either -600ms, -200ms ahead of, synchronous with (0ms) or +120ms behind the onset of the IP of the homonym. Results showed significantly triggered N400 effect to both homonym and the target word in the -200ms condition, the 0ms condition, and the +120ms condition, which illustrated that when gesture was presented -200ms to +120ms in reference to IP of speech, it influenced the speech on the semantic phase of lexical retrieval.

From all the above studies, we conclude that there is no consensus regarding whether the priming effect of gesture on the lexical retrieval of speech takes place during the phonological phase or the semantic phase. Some researchers tend to choose the phonological phase (Hadar & Butterworth, 1997; Krauss et al., 2000; Rauscher et al., 1996), some find evidence only for the semantic phase (Holle&Gunter, 2007; Ozyurek et al., 2007; Wu&Coulson, 2005), while some others found an effect in both of the two phases (Kelly et al., 2004), and yet others neglected the discrimination between the two phases (Habets et al., 2011; Obermeier&Gunter, 2015; Obermeier et al., 2011).

In the present chapter, we used online TMS stimulation over pMTG to investigate the time phase for the lexical priming effect of gesture on speech. We split the 400ms after the onset of speech into 10 time windows with the assumption that TMS stimulation would interrupt the ongoing gesture-speech integration within a certain time window, thus providing direct evidence for the involvement of the time window in the gesture-speech priming effect.

4.2. Experiment 3: timing study

In Experiment 3, we intended to explore the time window for the priming effect of gesture on speech. We used the same stimuli as was validated in Chapter 2, in which gestures were presented from the stroke phase, and speech was 200ms after gesture. Then we split the 400ms after the onset of speech into 10 time windows and applied double pulse TMS stimulation over each of the time windows. We hypothesized that double pulse TMS stimulation would interrupt the ongoing effects of gesture on speech in a certain time window, resulting in a significantly reduced semantic congruency effect for the pMTG condition as has been show in Experiment 2, relative to vertex stimulation in that time window. As before, gender congruency was used as a control effect, with the assumption that double pulse TMS stimulation would selectively influence only the semantic congruency effect.

4.2.1. Method

4.2.1.1. Participants

Eight participants (4 male and 4 female, age 21-36, mean age 25.9, SD= 4.4) participated in Experiment 3. All were native English speakers and were right-handed according to the Edinburgh Handedness form (LQ=83.68, SD=18.33), had a normal or corrected-to-normal vision and were screened for TMS suitability using a medical questionnaire. Before the experiment, all signed consent forms approved by the Ethics Committee of the Department of Psychology, and were allowed for rTMS stimulation according to the TMS Subject Questionnaire.

4.2.1.2. Apparatus and stimuli

The same stimuli consisting of 256 videos that were validated in Chapter 2 were used. Double pulse TMS at an intensity of 45% of maximum stimulator output was delivered ‘online’ at each time window using a 70mm figure-of-eight coil. For example, in a trial where stimulation took place in time window 1, a participant would receive a pulse at time 0 (coinciding with the onset of speech) and a second pulse 40ms later (40ms after the onset of speech). These pulses were delivered during each video item. All other apparatus details were as described in Experiment 2.

4.2.1.3. Procedure

To investigate the time window for the priming effect of gesture on speech in pMTG (-50, -56, 10), we divided the 400ms after the onset of sound into 10 time windows (TW): 0/40ms, 40/80ms, 80/120ms, 120/160ms, 160/200ms, 200/240ms, 240/280ms, 280/320ms, 320/360ms, 360/400ms. To counterbalance all items effects across participants, we split all the gestures into 10 lists using a balanced Latin Square design.

Each participant completed one of the 10 lists. Each list consisted of 4 blocks with 64 trials each. In each block, either pMTG or the Control site were stimulated by the double pulse TMS. After the first session, participants were reinvited for a second session at least one week later, where they completed the whole list again with a counterbalanced order of blocks in the second session.

The experimenter explained to the participant that they would be presented with a number of videos with double-pulse TMS stimulation on either pMTG or Vertex site. The videos were presented using Presentation software (Version 17.2, www.neurobs.com). All other procedural details were as described in Experiment 2. Before the experiment proper, participants were asked to do 16 training trials to get used to the experimental procedure.

4.2.1.4. Data analyses

All the incorrect responses (331 out of the total number of 8192, 4.04% of trials) were excluded. To eliminate the influence of outliers, a 2.5% trimmed mean for every participant in each session was conducted. We focused our analysis on the effect of semantic congruency and its interactions with the factor Stimulation Site (pMTG, control) and the factor Time Window, in order to find out at which time window the magnitude of semantic congruency effect (the RT difference between semantic incongruent and semantic congruent conditions) is reduced when activity in pMTG is stimulated, relative to vertex stimulation. We also used the effect of gender congruency (the RT difference between gender incongruent and gender congruent conditions) as a control effect, with the assumption that double-pulse TMS stimulation would only have an effect on semantic congruency.

Accordingly, we ran a 2 (Stimulation Site: pMTG and Control) *10 (Time Window: 0/40ms, 40/80ms, 80/120ms, 120/160ms, 160/200ms, 200/240ms, 240/280ms, 280/320ms, 320/360ms, 360/400ms after the onset of the sound) *2 (Semantic Congruency: semantic congruent vs semantic incongruent) * 2 (Gender Congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs.

4.2.2. Results

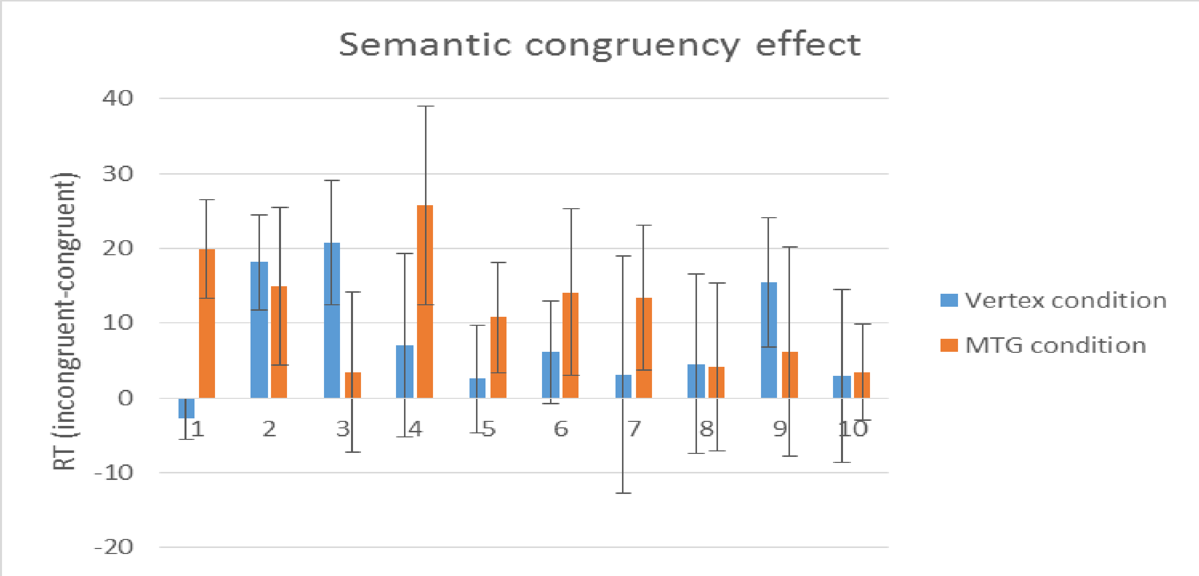
The ANOVA on the RTs revealed a significant main effect of Semantic Congruency (F (1, 7) = 6.93, p=.034), reflecting longer RTs of semantically incongruent trials (M = 517, SE = 23.9) than the congruent trials (M = 507, SE = 20.6). Furthermore, there was a significant main effect of Gender Congruency (F (1, 7) = 89.6, p=.000), with longer reaction times when speech and gesture were produced by different genders (M = 522, SE = 22.5) than when they were produced by the same gender (M = 502, SE = 21.9). The full pattern of results is shown in Table 4.1.

There was no significant main effect of Stimulation Site (F (1, 7) = 1.46, p=.27), neither was there a significant main effect of the Time Window (F (9, 63) = .23, p=.99). The ANOVA showed a significant interaction of Stimulation Site by Time Window (F (9, 63) = 2.98, p=.005). Pairwise comparisons showed that this significant interaction was located only in Time Window 3 (F (1, 7) = 10.30, p=.015). However, there was no significant interaction of Stimulation Site by Semantic Congruency (F (1, 7) = 1.13, p=.32), nor a significant interaction of Stimulation Site by Time Window by Semantic Congruency (F (9, 63) = 1.58, p=.14), illustrating that the magnitude of the semantic congruency effect was not modulated by double-pulse TMS stimulation over the ten time windows.

There was no modulation of the gender congruency effect, as shown by the ANOVA, revealing no significant interactions, either for the interaction of Time Window by Semantic Congruency ($F(9, 63) = .26$, $p=.98$), the interaction of Stimulation Site by Gender Congruency ($F(1, 7) = .73$, $p=.42$), the interaction of Time Window by Gender Congruency ($F(9, 63) = 1.48$, $p=.18$), the interaction of Semantic Congruency by Gender Congruency ($F(1, 7) = 4.47$, $p=.07$), or the three-way interaction of Stimulation Site by Semantic Congruency by Gender Congruency ($F(1, 7) = .71$, $p=.43$), the interaction of Stimulation Site by Time Window by Gender Congruency ($F(9, 63) = .92$, $p=.51$), the interaction of Time Window by Semantic Congruency by Gender Congruency ($F(9, 63) = .70$, $p=.71$). Neither was there a significant four-way interaction of Stimulation Site by Time Window by Semantic Congruency by Gender Congruency ($F(9, 63) = .85$, $p=.57$).

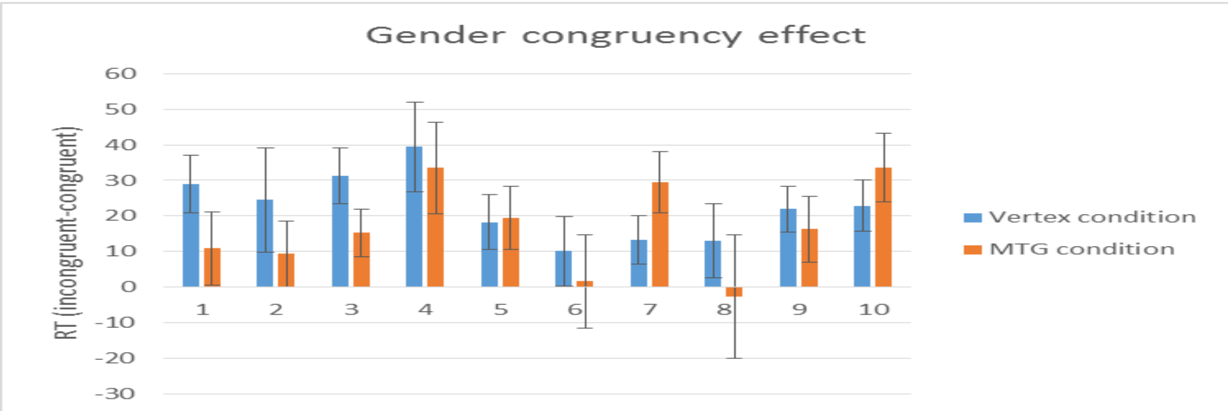
		Semantic congruent		Semantic incongruent	
		Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
Time window 1	pMTG	499(20)	507(16)	515(18)	529(24)
	Vertex	508(17)	531(18)	501(19)	535(18)
Time window 2	pMTG	502(19)	521(29)	527(31)	525(27)
	Vertex	495(21)	534(34)	526(28)	537(33)
Time window 3	pMTG	490(27)	502(26)	493(28)	508(29)
	Vertex	479(21)	520(24)	510(25)	532(33)
Time window 4	pMTG	485(25)	511(23)	503(30)	544(39)
	Vertex	483(23)	528(24)	495(27)	529(36)
Time window 5	pMTG	492(26)	520(33)	513(31)	522(27)
	Vertex	489(25)	530(34)	515(32)	508(24)
Time window 6	pMTG	507(30)	514(34)	528(41)	520(21)
	Vertex	487(28)	504(24)	501(31)	505(22)
Time window 7	pMTG	496(21)	527(26)	511(24)	540(28)
	Vertex	485(22)	513(20)	501(19)	502(19)
Time window 8	pMTG	516(23)	532(18)	539(28)	517(18)
	Vertex	507(14)	511(15)	504(15)	525(21)
Time window 9	pMTG	506(22)	524(22)	516(36)	527(25)
	Vertex	488(21)	516(20)	510(25)	525(25)
Time window 10	pMTG	498(32)	535(34)	503(29)	534(37)
	Vertex	491(30)	511(29)	492(27)	517(22)

Table 4-1 Mean reaction time in ms(SEM) for Experiment3.



Onset of speech(ms)	0-40	40-80	80-120	120-160	160-200	200-240	240-280	280-320	320-360	360-400
Onset of gesture(ms)	200-240	240-280	280-320	320-360	360-400	400-440	440-480	480-520	520-560	560-600

Figure 4.1 Semantic congruency effect in the 10 time windows in both the Control condition and the pMTG condition. Error bars show M±SE.



Onset of speech(ms)	0-40	40-80	80-120	120-160	160-200	200-240	240-280	280-320	320-360	360-400
Onset of gesture(ms)	200-240	240-280	280-320	320-360	360-400	400-440	440-480	480-520	520-560	560-600

Figure 4.2 Gender congruency effect in the 10 time windows in both the Control condition and the pMTG condition. Error bars show M±SE.

4.2.3. Discussion

In Experiment 3, by splitting the 400ms after the onset of speech into 10 time windows to investigate at which of the 10 time windows the priming effect of gesture on speech takes place, we stimulated each time window using double pulse TMS and found a significantly decreased RT in time window 3 (80-120ms after the onset of speech, and 280-320ms after the onset of gesture) when double pulse TMS was stimulated over pMTG compared to the same stimulation over the Vertex.

However, none of the interactions of Stimulation Site by Time Window by Semantic Congruency ($F(9, 63) = 1.58$, $p=.14$) or the interaction of Stimulation Site by Time Window by Gender Congruency ($F(9, 63) = .92$, $p=.51$) were significant. Therefore, we cannot tell whether the significantly decreased RT in the third time window when pMTG was stimulated with double pulse TMS compared to the Vertex was caused by a semantic factor, or a gender factor, or a combined effect of both factors.

4.3. Experiment 4: gating study

One limitation of Experiment 3 was that we did not take into account that some gestures may be processed faster than others. This raises the possibility that for some gestures, a representation may already have been available at the onset of the word (when the pulses of the first time windows were applied), whereas other more complex gestures may take more processing time before a meaning can be extracted. In Experiment 4, we aimed to take these temporal differences into account, by first determining the identification point (IP) of each gesture, followed by a re-analysis of the data from Experiment 3 that took these IPs into account.

4.3.1. Method

4.3.1.1. Participants

Thirty participants (15 male and 15 female, age 20-36, mean age 25.8, SD= 4.28) participated in Experiment 4. All of the participants were native English speakers. Before the experiment, all signed consent forms approved by the Ethics Committee of the Department of Psychology and were given one credit hour for their participation.

4.3.1.2. Procedure and stimuli

The 32 gestures that formed the basis of the stimulus set for Experiment 3 were used in Experiment 4. All gestures were presented without sound for the gating paradigm. Since each gesture was performed by either male or female without audio, 64 items were used in Experiment 4. Each item was presented in segments of increasing duration. The first segment had a length of 80ms, the second one was of 120ms' duration, and so on.

Participants were told that during the experiment, they would be presented with a number of videos of somebody performing an action without holding the actual object. For each of the actions, there were several videos of various durations. Participants were asked to look carefully and try to infer what was described in the action, by using a single action word to describe it. There were two conditions to move to the next action, the first condition was the end of the action, then it would move to the next action automatically; the second condition was when they were told to type 'next'. There were no time limits in Experiment 4 for making responses.

4.3.2. Data analyses

We defined the Identification point (IP) of a gesture as the first time the participant gave the last answer. For example, in Table 4.2, the participant responded 'open' for the first four times, but then changed to the answer 'tear' and 'tare', before he came to the answer 'open' again which had been provided for four continuous times. It was believed that the answer 'open' was a stable answer for the break gesture, so the participant was told to type 'next' which will move on to the next gesture in 520ms response time. In other words, when that participant was presented with the 'break' gesture, he thought the gesture represents 'open'. The first time that participant got the idea of the represent of the gesture as ‘open’ was in the 80ms response time. Therefore, the identification point of the ‘break’ gesture was defined to be 80ms.

Gesture	Response time	Response answer	Identification point
f-break.avi	80	open	x
f-break.avi	120	open	
f-break.avi	160	open	
f-break.avi	200	open	
f-break.avi	240	tear	
f-break.avi	280	tear	
f-break.avi	320	tare	
f-break.avi	360	open	
f-break.avi	400	open	
f-break.avi	440	open	
f-break.avi	480	open	
f-break.avi	520	next	

Table 4-2 An example of how Identification point of the 'f-break.avi' gesture was found

In Experiment 3, double-pulse TMS was applied to each of the 10 time windows. For the first time window, it was 0-40ms after the onset of speech and 200-240ms after the onset of gesture - the first TMS pulse was applied at the beginning of the time window (0ms after the onset of speech and 200ms after the onset of gesture) and the second at the end of the time window (40ms after the onset of speech and 240ms after the onset of gesture). Some gestures could be understood before 200ms when the onset of speech started (for example, the ‘break’ gesture), while some gestures could not (for example, the ‘vacuum’ gesture). To take these temporal differences into account, we shifted the time window according to the IP of the gesture. This then gave rise to the new modified IP Time Window. For example, in Fig. 4.3, the IP of the gesture ‘vacuum’ is 309ms, which locates in the third time window (80-120ms after the onset of speech). Therefore, we define the third time window as the first IP_TW for gesture ‘vacuum’, the fourth time window as the second IP_TW, the fifth as the third IP_TW, and so on. As has been shown in Fig. 4.3, after the shift of time window into IP_TW, there are only 8 IP_TWs left for gesture ‘vacuum’. So do some other gestures such as gesture ‘close’, there are only 9 IP_TWs left after the shifting. This will led to the fact that in the 8^{th}, 9^{th} and 10^{th} IP_TW, there are fewer gesture items than that in the other IP_TWs, or in the corresponding TWs. Yet we do not think that will make a difference to the semantic congruency effect we are interested in, given that in Experiment 3, we found a tendency of a reduced semantic congruency effect in the third TW, we hypothesised that in Experiment 4, after taken the identification point of each of the gesture into consideration, a significantly reduced semantic congruency effect would be found in the third IP_TW in the pMTG condition compared to the Vertex condition.

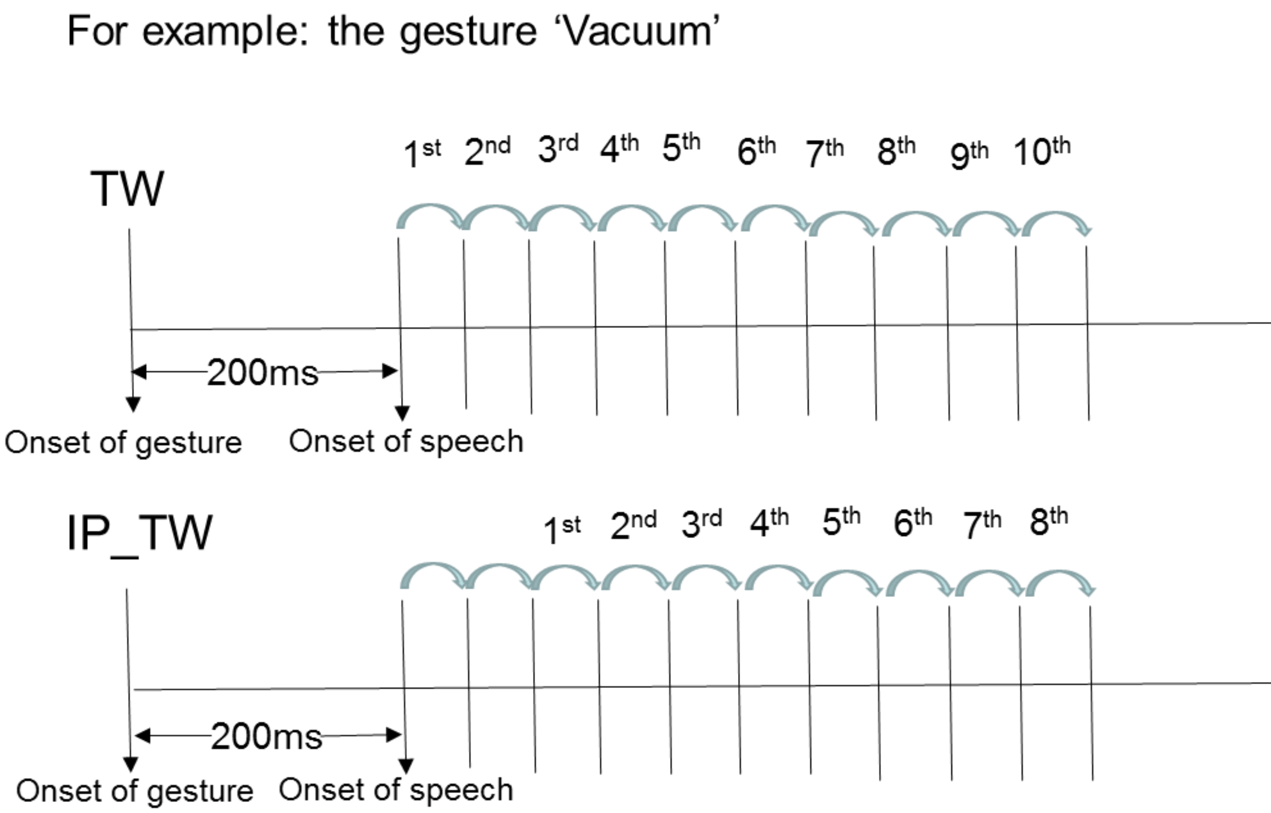


Figure 4.3 An example how the information from the gating study was used to shift the time windows for the re-analysis of the data of the TMS timing study.

Next, we replaced the time windows in Experiment 3 with IP_TWs and reanalysed the data. As we did in Experiment 3, we deleted all the incorrect responses and used a 2.5% trimmed mean to eliminate the influence of outliers. A 2 (Stimulation Site: pMTG and Control) *10 (IP_TW: 0/40ms, 40/80ms, 80/120ms, 120/160ms, 160/200ms, 200/240ms, 240/280ms, 280/320ms, 320/360ms, 360/400ms after the IP of the gesture) *2 (Semantic Congruency: semantic congruent vs semantic incongruent) * 2 (Gender Congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs was conducted.

4.3.3. Results

Identification points (IP) and IP_TWs and their relationship with TW, as well as the onset of speech for each of the 32 gestures (Mean_IP= 199.2ms, SE=9.96) are presented in Table 4.3. For example, the identification point of the 'break' gesture is 124ms, which locates in the period before the onset of speech, therefore, the first TW equals to the first

IP_TW, which is 0-40ms after the onset of speech. For the ‘close’ gesture, the identification point is 279ms, which locates in the second TW. Therefore, the second TW would be shifted into the first IP_TW, which is 40-80ms after the onset of speech.

Geature	Identification_point (IP) (ms)	Location of IP in terms of TW	Degree of shift of TW into equivalent IP_TW	Location of first IP_TW after onset of speech (ms)
break	124	<1	0[6]	0-40
close	279	2	1[7]	40-80
cut	139	<1	0	0-40
dial	193	<1	0	0-40
flip	240	1	0	0-40
hammer	215	1	0	0-40
iron	275	2	1	40-80
knock	104	<1	0	0-40
light	169	<1	0	0-40
open	145	<1	0	0-40
peel	259	2	1	40-80
pull	225	1	0	0-40
saw	276	2	1	40-80
scrub	253	2	1	40-80
sew	215	1	0	0-40

[6] 0 means no shift from TW into IP_TW, that all the IP_TWs equal to the TWs.

[7] 1 means one shift from TW into IP_TW, that the second TW equals to the first IP_TW, the third TW equals to the second IP_TW, the fourth TW equals to the third IP_TW, and so on.

shake	231	1	0	0-40
sharpen	165	<1	0	0-40
spray	215	1	0	0-40
spread	223	1	0	0-40
steer	166	<1	0	0-40
stir	210	1	0	0-40
swipe	183	<1	0	0-40
throw	146	<1	0	0-40
turn	230	1	0	0-40
type	92	<1	0	0-40
vacuum	309	3	2[8]	80-120
wash	208	1	0	0-40
weigh	147	<1	0	0-40
whisk	201	1	0	0-40
wipe	141	<1	0	0-40
write	272	2	1	40-80
zip	125	<1	0	0-48

Table 4-3 The relation between the Identification point, IP_TW, TW, and the onset of speech for each gesture

The ANOVA on the RTs revealed a significant main effect of Gender Congruency (F (1, 7) = 51.7, p=.000), with longer reaction times when speech and gesture were produced

[8] 2 means two shifts from TW into IP_TW, that the third TW equals to the first IP_TW, the fourth TW equals to the second IP_TW, the fifth TW equals to the third IP_TW, and so on.

by different genders (M = 522, SE = 22.5) than by the same gender (M = 502, SE = 21.9). There was no significant main effect of Semantic Congruency (F (1, 7) = 3.21, p=.12), reflecting no significantly different RTs in semantically incongruent trials (M = 517, SE = 23.8) and congruent trials (M = 506, SE = 20.7). The full pattern of results is shown in Table 4.4.

Additionally, there was a significant three-way interaction of Stimulation Site by IP_TW by Semantic Congruency (F (9, 63) = 3.11, p=.004), which also survived the Greenhouse-Geisser comparison with a corrected p-value of .031. This illustrated that the magnitude of semantic congruency was modulated by different IP_TWs when double-pulse TMS stimulation was applied over these IP_TWs in different brain areas (pMTG or Vertex). Simple effect analysis showed that the significant difference in terms of Reaction times in the semantic incongruent condition compared to the semantic congruent condition existed only in the third IP_TW in the Vertex (control) condition (F (1, 7) = 11.81, p=.011), but with double-pulse TMS stimulated over pMTG, these significantly different RTs in the Semantic Congruency factor in third IP_TW disappeared (F (1, 7) = 1.31, p=.29). This illustrated that double-pulse TMS stimulation over pMTG significantly modulated the Semantic Congruency factor in the third IP_TW. Furthermore, a paired sample t-test showed that the magnitude of the semantic congruency effect (RT (semantic incongruent) – RT (semantic congruent)) in pMTG had been significantly reduced as compared to that in the Vertex, as shown with t (7) = 2.81, p=.026. However this significance may not survive a multiple comparison.

There was no modulation of the gender congruency effect, as the ANOVA failed to show any other significant interactions. This can be seen in the following results : the interaction of IP_TW by Semantic Congruency (F (9, 63) = .58, p=.81), the interaction of Stimulation Site by Gender Congruency (F (1, 7) = 1.50, p=.26), the interaction of IP_TW by Gender Congruency (F (9, 63) = 1.26, p=.28), the interaction of Semantic Congruency by Gender

Congruency (F (1, 7) = 4.92, p=.06), the three-way interaction of Stimulation Site by Semantic Congruency by Gender Congruency (F (1, 7) = .29, p=.61), the interaction of Stimulation Site by IP_TW by Gender Congruency (F (9, 63) = .85, p=.58), the interaction of IP_TW by Semantic Congruency by Gender Congruency (F (9, 63) = .79, p=.63). Neither was there a significant four way interaction of Stimulation Site by IP_TW by Semantic Congruency by Gender Congruency (F (9, 63) = 1.47, p=.18).

There was also no significant main effect of Stimulation Site (F (1, 7) = 1.65, p=.24), neither was there a significant main effect of IP_TW (F (9, 63) = .24, p=.99). The ANOVA showed a significant interaction of Stimulation Site by IP_TW (F (9, 63) = 2.59, p=.013), but this significance did not survive the Greenhouse-Geisser comparison with a p-value of .073. Nor was there a significant interaction of Stimulation Site by Semantic Congruency (F (1, 7) = .29, p=.61).

		Semantic congruent		Semantic incongruent	
		Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
IP_TW 1	pMTG	502(17)	517(18)	520(19)	515(25)
	Vertex	510(18)	531(22)	504(23)	525(20)
IP_TW 2	pMTG	490(26)	520(31)	520(25)	530(30)
	Vertex	487(22)	524(29)	512(27)	538(22)
IP_TW 3	pMTG	501(27)	499(25)	504(26)	517(24)
	Vertex	484(25)	522(27)	522(24)	560(39)
IP_TW 4	pMTG	476(29)	519(23)	501(37)	528(35)
	Vertex	503(26)	534(24)	492(29)	506(35)
IP_TW 5	pMTG	491(24)	505(32)	521(29)	524(26)
	Vertex	469(20)	522(29)	516(30)	504(19)
IP_TW 6	pMTG	504(26)	536(33)	508(36)	521(23)
	Vertex	491(22)	517(24)	502(31)	514(24)
IP_TW 7	pMTG	509(30)	525(22)	526(27)	524(22)
	Vertex	491(20)	500(23)	493(15)	506(17)
IP_TW 8	pMTG	519(25)	535(19)	546(32)	514(21)
	Vertex	517(19)	525(17)	498(19)	518(26)
IP_TW 9	pMTG	506(25)	533(29)	491(31)	515(31)
	Vertex	484(20)	518(26)	521(35)	520(24)
IP_TW 10	pMTG	503(32)	535(37)	509(33)	544(42)
	Vertex	491(30)	511(29)	470(32)	520(26)

Table 4-4 Mean reaction time in ms(SEM) for Experiment 4

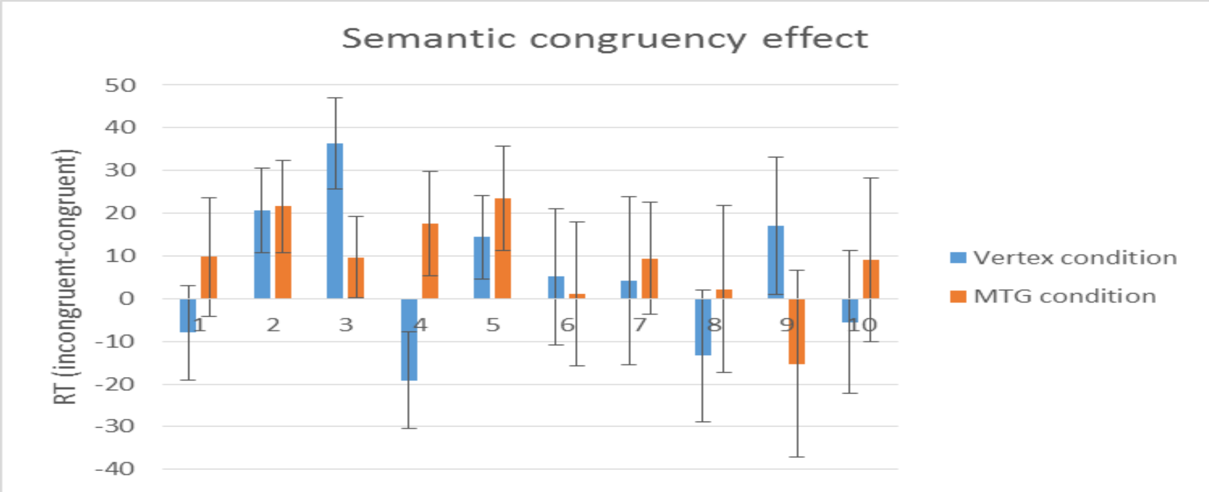


Figure 4.4 Semantic congruency effect in the 10 gating time windows in both the control condition and the pMTG condition. Error bars show M±SE.

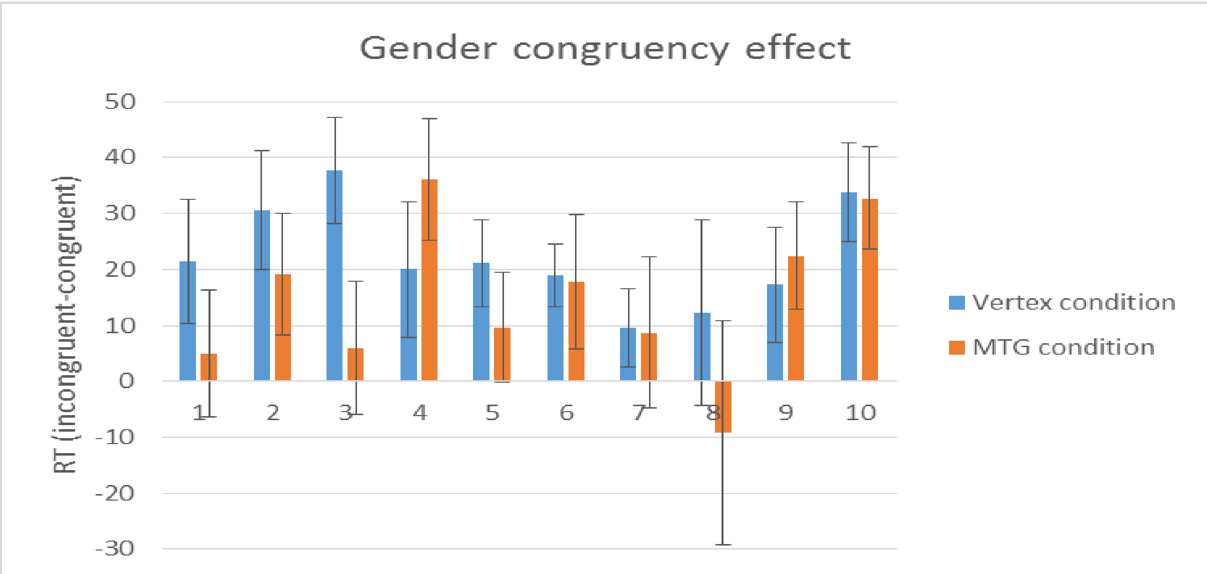


Figure 4.5 Gender congruency effect in the 10 gating time windows in both the Control condition and the pMTG condition. Error bars show M±SE.

4.3.4. Discussion

In Experiment 4, by shifting the time windows into IP TWs, all gestures were at the same level of comprehension upon TMS stimulation. Therefore, from the first IP_TW onwards, all gestures were clearly understood and co-occurred with speech.

Results show a significantly larger RT in the third IP_TW in the semantic incongruent condition compared to the congruent condition in the control condition (double-pulse TMS stimulation over Vertex). This replicates the results found by Kelly et al. (2010a) as well as our pre-test 2. This significant difference of RT in the third IP_TW disappeared when we applied double-pulse TMS over pMTG. Moreover, the semantic congruency effect, which was the magnitude of RT difference between the semantic incongruent condition and the semantic congruent condition, also showed a significant difference between the pMTG condition and the Vertex condition in the third IP_TW. We conclude from the above results that the automatic priming effect of gesture on speech took place in the third IP_TW, which is 80-120ms after the IP of gesture.

Given that the onset of speech is correspondent to TW (the first TW is defined as 0-40ms after the onset of speech) and the fact that there is a discrepancy between IP_TW and TW, in the third IP_TW, not all speech was presented exactly for 80-120ms. Yet even for the gesture 'vacuum', which had the largest shift from TW to IP_TW (the first IP_TW equals to the third TW), speech was presented at 160-200ms in the third IP_TW. According to Friederici (2002), this is still in the second phase (from 100-300ms after the onset of speech) which is to identify the word category, and before the processing of lexical semantic information that took place in the third phase, during 300-500ms of speech presentation. Therefore, we may conclude from this that the lexical priming effect of gesture on speech took place during the phonological phase and before the lexical analyse of speech.

There was no significant gender congruency effect in the third IP_TW, which illustrates that the TMS stimulation selectively interfered with only the semantic congruency effect.

4.4. General discussion

Overall, by applying double pulse TMS stimulation over pMTG during various time windows after the onset of speech, we observed a significantly reduced reaction time in the third time window (80-120 after the onset of speech and 280-320ms after the onset of gesture) when participants were asked to judge the gender of the speaker. After shifting the time window based on the identification point of gesture to ensure all gestures were at the same stage of comprehension, we found the significantly reduced reaction time was caused by the semantic congruency effect (task-irrelevant factor), rather than the gender congruency effect (task-relevant factor). Our results provide direct evidence that the automatic priming effect of gesture on speech takes place 80-120ms after a clear understanding of the gesture and before the semantic analysis of speech.

We explain our results in the following ways. First of all, In Experiment 4, when we applied double-pulse TMS over pMTG, we found a significantly reduced semantic congruency effect in the third IP_TW (80-120ms after the IP of gesture) compared to the Vertex condition. We hypothesized that these results showed the ‘priming’ effect of gesture on speech. According to Friederici (2002), the processing of lexical semantic information takes place 300-500ms after the onset of speech, therefore, we also hypothesized that the ‘priming’ effect of gesture on speech happens before speech’s semantic analysis phase. Our results support the hypothesis in the model proposed by Krauss et al., (2000) that gesture primes lexical retrieval of speech. To be specific, the motoric feature of the gesture is picked by the kinesthetic monitor, which will monitor the features of the source information represented in gesture, and then transfer the picked up information into the phonological encoder of speech, where it facilitates the lexical retrieval of word. In other words, the representation that gesture presents acts as the one for speech to retrieve lexical information that is expressed as speech (Hadar&Butterworth, 1997; Rauscher et al., 1996).

Secondly, our results showed that there was no interaction between the stimulation site and the semantic congruency effect in the third time window in Experiment 3, yet this interaction became significant when we shifted the time window to make sure all the gestures are at the same stage of comprehension in Experiment 4. We hypothesized that gesture needs to be comprehended to have a priming effect on speech, in which the information conveyed in gesture is encoded as a whole (McNeill, 1992). In other words, the representation of gesture should be established first, before it can act as an equivalent lexical formation, to further influence the lexical selection. The phenomenon that action is encoded based on its goal, rather than the individual movement of the action form, has been suggested by several researchers (Rizzolatti et al., 2009; Umilta et al., 2008). For example, in the study carried out by Umilta et al. (2008), they found the activation of the cortical motor neurons, in both cases when a monkey was asked to grasp with hands, to grasp with normal pliers (which required close hand), and also when the monkey was asked to grasp with reversed pliers (which required open hand).

Thirdly, it has previously been shown that there is a bi-directional influence between gesture and speech (Bernardis&Gentilucci, 2006), in a similar way in terms of N400 latency and amplitude (Ozyurek et al., 2007). Kelly et al. (2010b) proposed the 'integrated system hypothesis' to illustrate this bidirectional influence. According to Kelly et al. (2010b), the relation between gesture and speech is like 'two sides of the same coin', the information from both gestures and speech interact mutually with each other in an automatic, obligatory way. In our study, we have shown the lexical priming effect of gesture on speech. We further hypothesized that this priming effect is how gesture and speech 'integrate' with each other. To be more specific, on one hand, the representation of gesture acts as an equivalent source of analytic information for lexical selection of speech, as has been shown in our results. On the other hand, the representation of speech will also act as an equivalent source of spatial-motoric information for motoric retrieval of gesture. There is a cross-modal priming effect between gesture and speech (Rauscher et al., 1996; Krauss et al., 2000). Further studies can examine the priming effect of speech on the motoric retrieval of gesture.

Fourthly, we used the experimental paradigm of Kelly et al. (2010a) which provided an automatic paradigm to examine the relationship between gesture and speech. Therefore, our results showed that even if participants were not asked to pay attention to the information conveyed in gesture, the representation of gesture that had been learned by participants still had an effect on the lexical retrieval of speech that was co-occurring. Given that the bidirectional influence between gesture and speech can be implicit (Holle&Gunter, 2007; Kelly et al., 2010a; Kelly et al., 2007; Obermeier et al., 2011), we hypothesized that the cross-modal priming effect between gesture and speech took place in an automatic, obligatory way.

Last but not least, in this chapter we found a significant priming effect of gesture on lexical retrieval of speech during the phonological phase. These results suggest that this cross-modal priming effect provides an exploration of how gesture and speech integrate with each other. We have provided direct evidence for the involvement of both pMTG and IFG in the integration of gesture with speech in Chapter 3. Therefore, we hypothesized that both of the areas are involved in the cross-modal priming effect between gesture and speech. There is a division of labor between the two, in that pMTG is the brain area for the perceptual matching/activation of a common stable representation, and IFG is to select/control/unify semantic information (Willems et al., 2009). It would be of interest to examine the divided labor of pMTG and IFG in the priming effect of gesture on speech, to see whether in IFG the prime effect of gesture on lexical retrieval of speech still works in the phonological phase.

In conclusion, Chapter 4 provided direct evidence that there is a priming effect of gesture on the lexical retrieval of speech, which took place after a clear understanding of gesture and before the semantic analysis of speech.

Chapter 5. The effect of decreased or increased excitability of IFG and pMTG on gesture-speech integration

5.1. Introduction

The brain areas associated with the integration of gesture and speech have been investigated by various researchers. Some observed activation of the Left Inferior frontal gyrus (LIFG) (Green et al., 2009; Willems et al., 2007) in gesture-speech integration, whereas others reported evidence suggesting an involvement of the posterior temporal gyrus (pMTG) (Holle et al., 2008; Holle et al., 2010). There are also some researchers who observed activation of both IFG and pSTS/MTG during the integration of gesture and speech (Dick et al., 2012; Straube et al., 2011). In Chapter 3, we provided a closure to this controversial question with the use of transcranial magnetic stimulation (TMS). By applying theta-burst offline TMS stimulation and repetitive online TMS stimulation over IFG and/or pMTG, we found a significantly reduced semantic congruency effect in both the IFG condition and the pMTG condition compared to the control condition. We concluded that both two areas are involved in the semantic integration of gesture and speech. In Chapter 4, when we zoomed in to determine the point in time at which pMTG starts to make a functional contribution to the integration process, we found a significantly decreased semantic congruency effect in the third IP_TW (80 – 120ms after the identification point of gesture). Given that speech is still in the phonological processing phase in the third IP_TW, we concluded that the integration of gesture and speech takes place after a clear understanding of the gesture and before the semantic analysis of speech.

TMS has initially been considered to cause an effect of a ‘virtual lesion’ in the stimulated brain area (Pascual-Leone et al., 2000; Walsh et al., 1999), leading to impaired performance. However, there are also reports of facilitated behavioural performance following TMS in various perceptual and cognitive tasks (for a review see Vallar &

Bolognini, 2011), prompting researchers to reconsider the 'virtual lesion' interpretation of TMS. Walsh and Rushworth (1999) defined the effect of TMS as 'neural noise that the pulse adds random activity in the midst of organized activity in the cortical region'. The noise generation hypothesis has also been supported by various TMS studies (Rahnev et al., 2012; Ruzzoli et al., 2010; Schwarzkopf et al., 2011). For example, Ruzzoli et al. (2010) compared the neural activity induced by a visual motion direction discrimination task and that induced by TMS. In that study, there were five levels of motion coherence in a motion direction discrimination task (50, 55, 70, 75, and 90% accuracy). By plotting a logistic regression of the "proportion of rightward choices" against the percentage of motion coherence, the results showed that TMS induced a decrement in the slope of the curve. The authors concluded that the results indicated a generalized reduction in system sensitivity and thus provided evidence for the hypothesis that TMS operates by adding neural noise to the perceptual process. Based on the neural noise hypothesis of TMS, we can re-interpret the results obtained in Chapter 3, in that when we applied 5-pulse repetitive TMS stimulation over pMTG, we created a general noise background, which 'disturbed the excitability of the neurons for any other activity', therefore resulting in decreased semantic congruency effect in TMS condition (the pMTG condition) compared to the control condition (the Vertex condition).

In this chapter 5, we make use of transcranial direct current stimulation (tDCS) to investigate the effect of selectively increasing or decreasing the membrane excitability of pMTG and IFG, to provide further evidence on the involvement of the two brain areas in gesture-speech integration. By inducing a subthreshold polarization of cortical neurons, anodal tDCS can induce membrane depolarization with a decreased threshold, whereas cathodal tDCS can induce membrane hyperpolarization with an increased threshold (Nitsche et al., 2005). We hypothesized that the increased threshold induced by cathodal tDCS would 'disrupt' the neural response to the ongoing task, similar to the effects of theta-burst TMS, while the decreased threshold induced by anodal tDCS would reduce the background neural noise and help the emergence of the signal (Dockery et al., 2009). Since tDCS is a continuous stimulation procedure, with the effect of stimulation lasting for up to 60 minutes (Nitsche&Paulus, 2000), tDCS stimulation was applied in an offline fashion, with 20 minutes of tDCS followed by the experimental task.

One aim of Chapter 5 is to replicate the results from Chapter 3 using a different form of brain stimulation. We applied anodal and cathodal tDCS stimulation over both IFG and pMTG separately. We hypothesized that if both IFG and pMTG are involved in gesture-speech integration, the increase in neural threshold of cathodal TMS stimulation would significantly decrease the amount of semantic gesture-speech integration. At the same time, decreasing the neural threshold via anodal stimulation would significantly increase the amount of semantic gesture-speech integration. As before, the gender congruency effect served as a control condition, in order to provide evidence that the increased and/or decreased neural threshold selectively influence only the semantic gesture-speech integration.

5.2. Experiment 5: tDCS over IFG

In Experiment 5, we intended to replicate the results from Experiment 1 using a different form of brain stimulation. We hypothesized that the increase of neural noise via cathodal tDCS would disrupt gesture-speech integration activity taking place in the IFG, resulting in a reduction of the semantic congruency effect. Meanwhile, we explored whether a decrease of neural noise via anodal tDCS would increase the gesture-speech integration activity taking place in IFG, resulting in an increased semantic congruency effect. In a within-subject design, participants underwent three sessions, where 20 minutes of tDCS stimulation was applied to the left IFG in either a cathodal condition or an anodal condition. To provide a baseline, we also applied 30s of tDCS stimulation as a sham condition. To blind participants from the experimental conditions, in the baseline, participants were told that they would undergo 20min of tDCS stimulation. The session order was counterbalanced across participants. After stimulation, which occurred at the beginning of each session, participants completed the reaction time task described above (section Pre-test 2). Thus, the full experimental design was a 3 (Stimulation Effect: Cathodal, Anodal, Sham) x 2 (Gender Congruency) x 2 (Semantic Congruency) factorial design. We predicted an interaction between Stimulation Effect and Semantic Congruency, indicated by a reduction of the semantic congruency effect between the Cathodal tDCS stimulation compared to the Sham condition, and an increase of the semantic congruency effect between the Anodal tDCS stimulation compared to the Sham condition. The factor of Gender Congruency was used as an additional control, to see whether brain stimulation specifically disrupts the processing of semantic (in)congruencies, or more generally interferes with task processing.

5.2.1. Method

5.2.1.1. Participants

Thirty English native speakers (8 male and 22 female, age 18-28, mean age 20.05, SE= .40) participated in Experiment 5. They all had normal or corrected-to-normal vision and were screened for tDCS suitability using a medical questionnaire. The Experiment was approved by the Ethics Committee of the Department of Psychology. Participants were remunerated at a rate of £8 or 1 hour of credit per session for taking part in the experiment.

5.2.1.2. Apparatus and stimuli

The same stimuli consisting of 256 videos as in Chapter 3 were used. tDCS stimulation at an intensity of 2000mA was delivered by a constant current stimulator (Version Edith DC-Stimulator-Plus, neuroConn). Two saline-soaked sponge electrodes were used. A stimulation electrode with the size of 2*2 cm was placed over IFG. To avoid simultaneously modulating cortical excitability in two brain areas at the same time, we intentionally placed the second reference electrode (size: 3x4 cm) on a non-cortical site (right shoulder), following the setup used by Accornero et al. (2007). Both electrodes were fixed by elastic bands.

The location of IFG with an MNI coordinates of (-62, 16, 22) was previously determined in the meta-analysis of FMRI studies. The best cortical projection in the international 10-10 system for this area is electrode F7 (Koessler et al., 2009). Accordingly, this electrode position was used as the centre point for placing the stimulation electrode. In Anodal condition and Cathodal condition, IFG was stimulated for 20 minutes with a 5s' fade in and 5s' fade out. In the Sham condition, a 5s' fade in was followed by only 30s of stimulation, followed by 20 minutes of no stimulation and 5s' fade out. Participants were blinded to stimulation conditions.

5.2.1.3. Procedure

Participants took part in three sessions, consisting of 20 minutes of tDCS stimulation over IFG in either Anodal condition, Cathodal condition, or Sham stimulation. The three sessions were counterbalanced among participants, with at least one week between each session. In each session, participants were guided to sit on a sofa to make themselves comfortable. The experimenter explained to the participant that they would first be given 20 minutes of tDCS stimulation. After the stimulation, they would be asked to perform an experimental task of simply responding to the gender of the speaker (see Exp.1). Their reaction time was recorded for final analysis.

5.2.2. Data analyses

All incorrect responses (702 out of the total number of 23040, 3.04% of trials) were excluded. To eliminate the influence of outliers, a 2.5% trimmed mean for every participant in each session was also calculated. We focused our analysis on the main effect of semantic congruency and its interactions with stimulation effects, as well as the gender congruency effect as a control effect.

We ran a 3 (Stimulation Effect: Sham condition, Anodal condition and Cathodal condition) * 2 (Semantic Congruency: semantic congruent vs semantic incongruent) *2 (Gender Congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs to see if tDCS stimulation caused any significantly reduced magnitude of gesture-speech integration.

5.2.3. Results

The 3*2*2 ANOVA on the RTs revealed a significant main effect of Semantic Congruency ($F(1, 29) = 22.24$, $p=.000$), reflecting longer RTs of semantically incongruent trials ($M = 554$, $SE = 14.4$) than congruent trials ($M = 544$, $SE = 13.6$). Furthermore, there was a significant main effect of Gender Congruency ($F(1, 29) = 226.15$, $p=.000$), with longer reaction times when speech and gesture were produced by different genders ($M = 562$, $SE = 13.9$) than the same gender ($M = 536$, $SE = 13.9$). The full pattern of results is shown in Table 5.1.

Besides, the ANOVA did not show a significant main effect of Stimulation Effect ($F(2, 28) = .06$, $p=.94$), nor was there a significant interaction of Stimulation Effect by Semantic Congruency ($F(2, 28) = .05$, $p=.95$), indicating that the magnitude of the semantic congruency effect was not significantly modulated by tDCS stimulation over IFG in either Anodal stimulation or Cathodal stimulation compared to the Sham condition (see Fig. 5.2). There was also no significant interaction of Stimulation Effect by Gender Congruency ($F(2, 28) = .95$, $p=.40$), indicating that no modulation was observed for the Gender Congruency effect either (see Fig. 5.3). The results of ANOVA failed to show any other significant interactions, either for the interaction of Gender Congruency by

Semantic Congruency (F (1, 29) = .10, p=.75), or the three-way interaction of Stimulation Effect by Semantic Congruency by Gender Congruency (F (2, 28) = 1.12, p=.34).

	Semantic congruent		Semantic incongruent	
	Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
Anodal IFG	528(14)	558 (14)	540 (15)	566 (15)
Cathodal IFG	532 (17)	560 (16)	545 (18)	567 (16)
Sham condition	532 (13)	554 (14)	540 (14)	568 (15)

Table 5-1 Mean reaction time in ms(SEM) for Experiment 5

Descriptive statistic showed that the semantic congruency effect was largest in the Sham condition (M=22.29, SE=5.44) compared with that in the Anodal condition (M=20.36, SE=5.85) and the Cathodal condition (M=20.61, SE=7.12) (see Fig. 5.2). The gender congruency effect was largest in the Anodal condition (M=56.08, SE=4.15) compared with that in the Sham condition (M=50.88, SE=5.94) and the Cathodal condition (M=49.47, SE=4.55) (see Fig. 5.3).

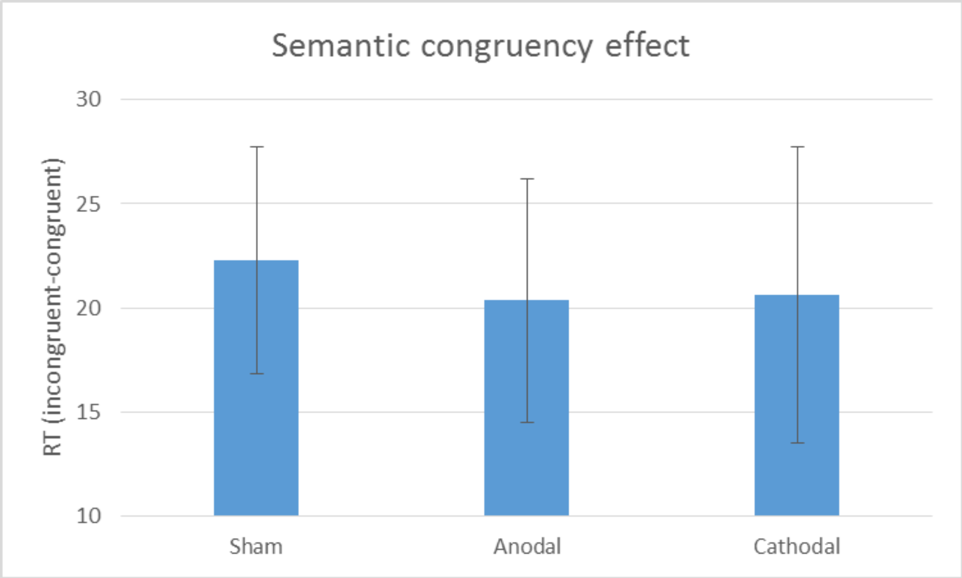


Figure 5.1 Magnitude of semantic congruency effects (ms) for Experiment 5. Errors show ± 1 standard error of the mean.

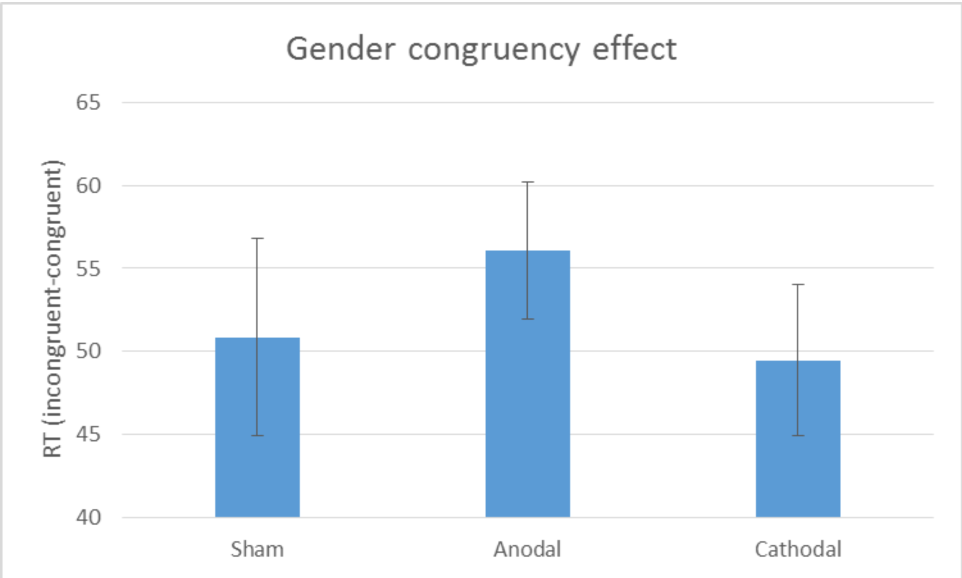


Figure 5.2 Magnitude of gender congruency effects (ms) for Experiment 5. Errors show ± 1 standard error of the mean.

5.2.4. Discussion

In Experiment 5, we applied Anodal tDCS stimulation and Cathodal tDCS stimulation over IFG with the assumption that anodal tDCS stimulation would significantly increase the semantic congruency effect in gesture-speech integration, and cathodal tDCS stimulation will significantly decrease the semantic gesture-speech integration. We also used the gender congruency effect as a control effect in order to test that tDCS stimulation selectively interferes with the semantic congruency effect in IFG as we have found in Chapter 3.

Results showed that the size of the gender congruency effect was not affected by tDCS when we applied either Anodal or Cathodal tDCS stimulation over IFG compared to the sham condition, as expected.

We failed to find a decreased semantic congruency effect when increasing the neural threshold by Cathodal TMS stimulation over IFG. We also did not show an increased semantic congruency effect when decreasing the neural threshold by Anodal TMS stimulation over IFG.

5.3. Experiment 6: tDCS over pMTG

In Experiment 6, we aimed to replicate the results of Experiment 2, by using a different form of brain stimulation—Transcranial Direct Current Stimulation (tDCS). By applying Cathodal tDCS stimulation over pMTG, we assumed that the increase in neural noise by Cathodal tDCS would result in a reduction of the semantic congruency effect. We also expected that a decrease of neural noise via Anodal tDCS would result in an increased semantic congruency effect. All the other study details were as described in Experiment 5.

5.3.1. Method

5.3.1.1. Participants

Thirty English native speakers (14 male and 16 female, age 19-54, mean age 25.19, SE= 1.63) participated in Experiment 6. They all had normal or corrected-to-normal vision and were screened for tDCS suitability using a medical questionnaire. The Experiment was approved by the Ethics Committee of the Department of Psychology. Participants were remunerated at a rate of £8 or 1 hour of credit per session for taking part in the experiment.

5.3.1.2. *Apparatus and procedure*

The same apparatus and stimuli as in Experiment 5 were used, except that in Experiment 6 the stimulation electrode was placed over pMTG (-50, -56, 10), as determined in the meta-analysis (see Chapter 2 for detail). The cortical project site of the pMTG position was determined as the midpoint between TP7, CP5, P7, and P5 (Koessler et al., 2009). All other experimental details were as described in Experiment 5.

5.3.2. Data analyses

All incorrect responses (702 out of the total number of 23040, 3.04% of trials) were excluded. To eliminate the influence of outliers, a 2.5% trimmed mean for every participant in each session was also calculated. We focused our analysis on the main effect of semantic congruency and its interactions with stimulation effects, as well as the gender congruency effect as a control effect.

We ran a 3 (Stimulate Effect: Sham condition, Anodal condition and Cathodal condition) * 2 (Semantic Congruency: semantic congruent vs semantic incongruent) *2 (Gender Congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs to see if tDCS stimulation caused any significantly reduced magnitude of gesture-speech integration.

5.3.3. Results

The 3*2*2 ANOVA on the RTs revealed a significant main effect of Semantic Congruency ($F(1, 29) = 26.51$, $p=.000$), reflecting longer RTs of semantically incongruent trials ($M = 573$, $SE = 15.9$) than congruent trials ($M = 561$, $SE = 14.8$). Furthermore, there was a significant main effect of Gender Congruency ($F(1, 29) = 79.00$, $p=.000$), with longer reaction times when speech and gesture were produced by different genders ($M = 578$, $SE = 15.6$) than the same gender ($M = 555$, $SE = 15.1$). The full pattern of results is shown in Table 5.2.

Besides, the ANOVA did not show a significant main effect of Stimulation Effect (F (2, 28) = 1.83, p=.18), nor was there a significant interaction of Stimulation Effect by Semantic Congruency (F (2, 28) = .10, p=.90), indicating that the magnitude of the semantic congruency effect was not significantly modulated by tDCS stimulation over pMTG in either Anodal condition or Cathodal condition compared to the Sham condition (see Fig. 5.5). There was also no significant interaction of Stimulation Effect by Gender Congruency (F (2, 28) = 2.56, p=.10), indicating that no modulation was observed for the gender congruency effect either (see Fig. 5.6). There was a significant interaction of Semantic Congruency by Gender Congruency (F (1, 29) = 6.10, p=.02), but no significant interaction of Stimulation Effect by Semantic Congruency by Gender Congruency (F (2, 28) = 1.12, p=.34).

	Semantic congruent		Semantic incongruent	
	Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
Anodal pMTG	549(16)	568 (15)	565 (16)	579 (16)
Cathodal pMTG	553 (17)	587 (19)	570 (19)	592 (18)
Sham condition	540 (16)	567 (17)	556 (18)	575 (17)

Table 5-2 Mean reaction time in ms(SEM) for Experiment 6.

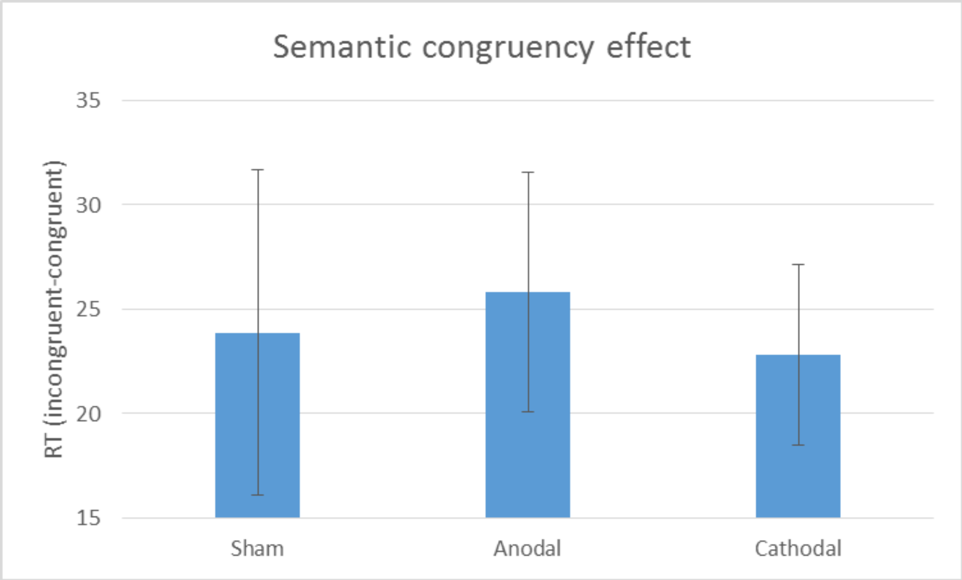


Figure 5.3 Magnitude of semantic congruency effects (ms) for Experiment 6. Error bars show M±SE.

Descriptive statistics showed that the semantic congruency effect was largest in the Anodal condition (M=25.82, SE=5.73) compared with the Sham condition (M=23.89, SE=7.78) and the Cathodal condition (M=22.80, SE=4.32) (see Fig. 5.5). The gender congruency effect was largest in the Cathodal condition (M=56.70, SE=8.38) compared with that in the Sham condition (M=47.04, SE=7.06) and that in the Anodal condition (M=32.89, SE=6.74) (see Fig. 5.6).

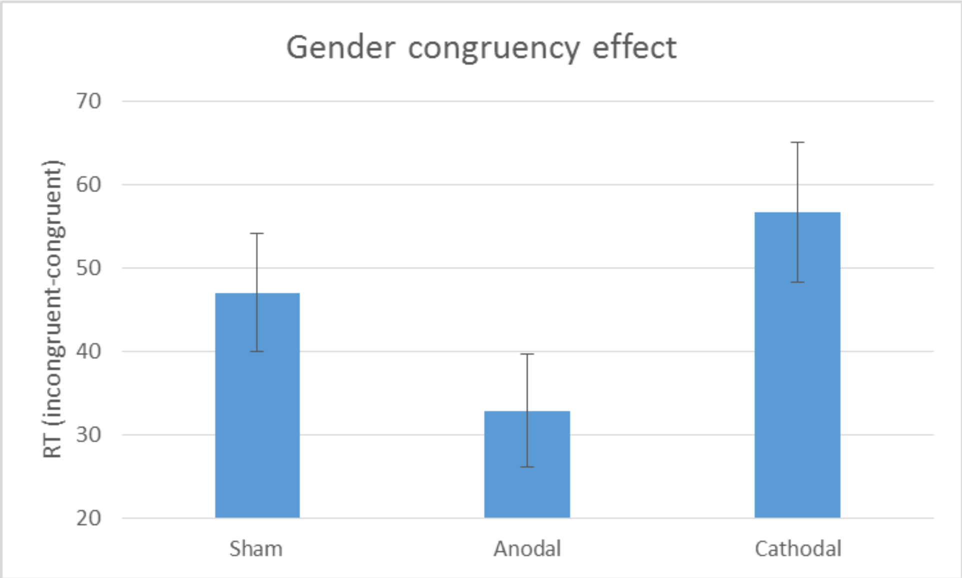


Figure 5.4 Magnitude of gender congruency effect (ms) for Experiment 6. Error bars show M±SE.

5.3.4. Discussion of Experiments 5 and 6

In Experiment 6, we applied Anodal tDCS stimulation and Cathodal tDCS stimulation over pMTG with the assumption that Anodal tDCS stimulation would significantly increase the semantic congruency effect and Cathodal tDCS stimulation would significantly decrease the semantic gesture-speech integration. We also examined the gender congruency effect as a control effect.

Results showed that the magnitude of semantic congruency was not modulated when either Anodal or Cathodal tDCS stimulation was applied over pMTG compared to the sham condition. As in Experiment 5, we failed to find a decreased semantic congruency effect when the neural threshold was increased by Cathodal TMS stimulation, nor did we find an increased semantic congruency effect when we decreased the neural threshold by Anodal TMS stimulation.

Taken together, when using tDCS as the method to modulate the cortical excitability of IFG and pMTG, we found no effect of brain stimulation on gesture-speech integration. This is in contrast to the findings of Chapter 3, where we observed clear evidence that disrupting activity in these areas impairs gesture-speech integration.

We will first consider why increasing cortical excitability in these areas via Anodal tDCS may not necessarily result in increased gesture-speech integration. Since the integration of gesture and speech is already automatic and obligatory, as has been shown in our Chapter 3, as well as by the results of Kelly et al. (2010a), the increase in neural threshold by Anodal stimulation may not further increase this integration effect. This may explain the failure of the significant effect in the Anodal condition of either IFG or pMTG.

As for the Cathodal condition, one possible explanation is that the decrease in excitability caused by tDCS Cathodal condition is not as strong as that caused by TMS stimulation (Miniussi et al., 2013). The weak decreased effect of Cathodal stimulation that is caused by an increased neural threshold may be not strong enough to 'block' the automatic integration of gesture and speech. In other words, even though the Cathodal stimulation has increased the neural threshold to 'stop' the neural membrane from being excited, the automatic and obligatory integration of gesture and speech may have a threshold beyond the one created by the stimulation. Finally, given that IFG and pMTG are anatomically connected (Friederici, 2009), the transcranial electrical stimulation over one area may be compensated by the activity of the other one (Whitney et al., 2011), resulting in a failure to find any significance in either increased or decreased brain excitability caused by tDCS.

5.4. Experiment 7: tRNS over both IFG and MTG

To test the hypothesis that IFG and pMTG may compensate for each other in the tDCS stimulation, in Experiment 7, we used transcranial random noise stimulation (tRNS) over both IFG and pMTG synchronously. tRNS is a relatively new technique of transcranial electrical stimulation (tES) which involves the application of a random electrical oscillation spectrum over the cortex, thus stimulating the two brain areas at the same time in a similar way. There are two types of stimulation in tRNS, with a high frequency of stimulation (101- 640 Hz) increasing the brain area(s) excitability and low frequency of stimulation (0.1- 100 Hz) decreasing the brain excitability (Terney et al., 2008). Thus, we hypothesized that high frequency tRNS over both IFG and pMTG would enhance the cortical excitability of these areas, leading to a significantly increased semantic congruency effect compared to the control condition. We also expected decreasing the excitability of LIFG and pMTG using low frequency tRNS would lead to a decreased semantic congruency effect, compared to the control condition. As before, we also used the gender congruency effect as an additional control, to see whether tRNS selectively interfered with semantic congruency information.

5.4.1. Method

5.4.1.1. Participants

Twenty-four English native speakers (12 male and 12 female, age 20-23, mean age 20.97, SE= .20) participated in Experiment 7. They all had normal or corrected-to-normal vision and were screened for tDCS suitability using a medical questionnaire. The Experiment was approved by the Ethics Committee of the Department of Psychology. Participants were remunerated at a rate of £8 or 1 hour of credit per session for taking part in the experiment.

5.4.1.2. Apparatus and stimuli

The same stimuli consisting of 256 videos as in Experiment 5 were used. tRNS stimulation was delivered by a battery-driven electrical stimulator (Version Edith DC-Stimulator-Plus, neuroConn). Two conductive saline-soaked sponge electrodes with the size of 2*2 cm were used, one placed over IFG with MNI coordinates of (-62, 16, 22) and the other placed on pMTG at MNI coordinates of (-50, -56, 10). Both electrodes were fixed by elastic bands.

A random level of current was generated from two types of frequency spectrum. One was a high-frequency spectrum with a frequency range between 101 and 640 Hz, the other was a low-frequency spectrum with a frequency range between 0.1 and 100 Hz. In each level, the current was randomly generated with a normal distribution. Previous research suggests that high-frequency current stimulation would cause an excitatory effect, while low-frequency current stimulation would cause an inhibitory effect (Terney et al., 2008). In both the high-frequency and low-frequency current stimulation conditions, there was 20 minutes' stimulation with a 5s' fade in and 5s' fade out at an intensity of 2mA. We also applied 30s of stimulation as the Sham condition. Participants were blinded to the stimulation conditions.

5.4.1.3. Procedure

There were three sessions: 20 minutes of tRNS stimulation over IFG and pMTG in a high-frequency stimulation condition, a low-frequency stimulation condition, and a Sham condition. The three sessions were counterbalanced among participants, with at least one week between each session. In each session, participants were told that they would undergo 20 minutes of tRNS stimulation. After the stimulation, they were asked to do the experimental task by simply responding to the gender of the speaker (see Exp.5).

5.4.2. Data analyses

All incorrect responses (702 out of the total number of 23040, 3.04% of trials) were excluded. To eliminate the influence of outliers, a 2.5% trimmed mean for every participant in each session was also calculated. We focused our analysis on the main effect of semantic congruency and its interactions with stimulation effects, as well as the gender congruency effect as a control effect.

We ran a 3 (Stimulation Effect: Sham, high-frequency tRNS, low-frequency tRNS) * 2 (Semantic Congruency: semantic congruent vs semantic incongruent) *2 (Gender Congruency: gender congruent vs gender incongruent) repeated measures analysis of variance (ANOVA) on the RTs to see if tRNS stimulation significantly affected gesture-speech integration.

5.4.3. Results

The 3*2*2ANOVA on the RTs revealed a significant main effect of Semantic Congruency ($F(1, 23) = 19.70$, $p=.000$), reflecting longer RTs of semantically incongruent trials ($M = 616$, $SE = 18.1$) than congruent trials ($M = 606$, $SE = 17.0$). Furthermore, there was a significant main effect of Gender Congruency ($F(1, 23) = 91.18$, $p=.000$), with longer reaction times when speech and gesture were produced by different genders (M = 622, SE = 17.3) than the same gender (M = 601, SE = 17.7). The full pattern of results is shown in Table 5.3.

Besides, the ANOVA did not show a significant main effect of Stimulation Effect (F (2, 22) = 3.21, p=.06), nor was there a significant interaction of Stimulation Effect by Semantic Congruency (F (2, 22) = .59, p=.56), indicating that when both IFG and pMTG were stimulated synchronously in a similar way, there was no significantly semantic congruency effect in any of the high frequency, low frequency, or sham conditions (see Fig. 5.8). There was also no significant interaction of Stimulation Effect by Gender Congruency (F (2, 22) = .69, p=.51), indicating that no modulation was observed for the gender congruency effect either (see Fig. 5.9). The results of ANOVA failed to show any other significant interactions, for the interaction of Gender Congruency by Semantic Congruency (F (1, 23) = .05, p=.82), the interaction of Stimulation Effect by Semantic Congruency by Gender Congruency (F (2, 22) = .82, p=.45).

	Semantic congruent		Semantic incongruent	
	Gender congruent	Gender incongruent	Gender congruent	Gender incongruent
High_frequency	613(22)	634 (21)	625 (22)	645 (23)
Low_frequency	595(21)	616 (19)	606 (22)	622 (19)
Sham condition	580(16)	601 (17)	588 (19)	613 (18)

Table 5-3 Mean reaction times in ms (SEM) for Experiment 7.

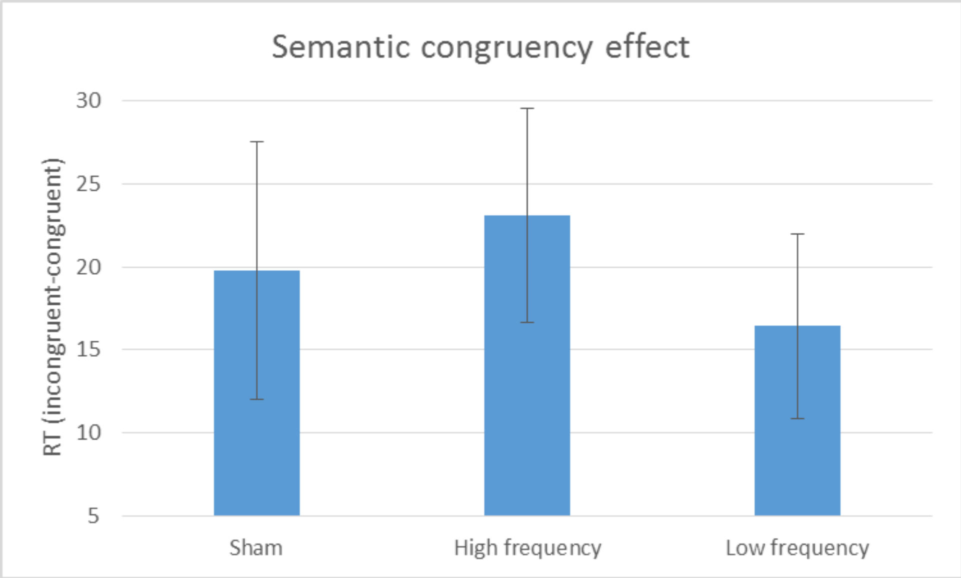


Figure 5.5 Magnitude of semantic congruency effect (ms) for Experiment 7. Errors show ± 1 standard error of the mean.

Descriptive statistics showed that the semantic congruency effect was largest in the High frequency condition (M=23.08, SE=6.41) compared with that in the Sham condition (M=19.78, SE=7.72) and the Low frequency condition (M=16.44, SE=5.55) (see Fig. 5.8). The gender congruency effect was largest in the Sham condition (M=45.68, SE=5.87) compared with that in the High frequency condition (M=39.86, SE=6.11) and the Low frequency condition (M=36.64, SE=8.95) (see Fig. 5.9).

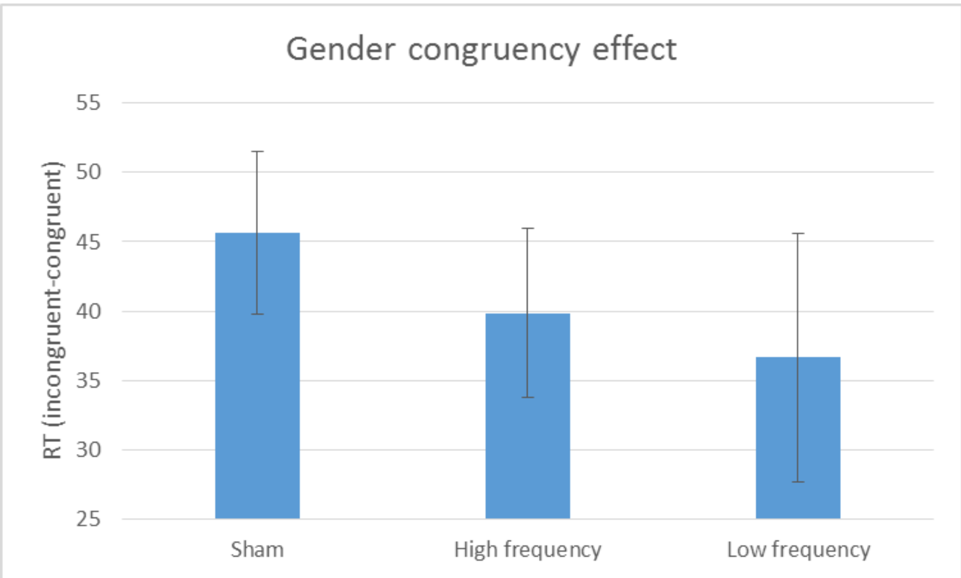


Figure 5.6 Magnitude of gender congruency effect (ms) for Experiment 7. Errors show ± 1 standard error of the mean.

5.4.4. Discussion of Experiment 7

One possible explanation for the null findings of Experiments 5 and 6 was that transcranial brain stimulation only results in subtle changes in cortical excitability, and subtle changes in one area (e.g., IFG) may be compensated by its well-connected counterpart in the anatomical network (e.g., pMTG). To test for this possibility, we simultaneously modulated cortical excitability in both areas at the same time in Experiment 7, by applying tRNS stimulation over both IFG and pMTG. We hypothesized that high frequency tRNS would increase the brain areas' excitability, leading to an increased semantic congruency effect. In contrast, low-frequency tRNS, which decreases the excitability of the two brain areas, was expected to result in a decreased semantic congruency effect.

Results showed no gender congruency effect when we applied either high frequency or low-frequency tRNS stimulation over both IFG and pMTG compared to the sham condition. Similar to the results of Experiment 5 and Experiment 6, we failed to find a decreased semantic congruency effect when we decreased the neural activity with high-frequency tRNS stimulation. We also did not find an increased semantic congruency effect when we increased the neural excitability with low-frequency tRNS stimulation.

5.5. General discussion

In the current chapter, we aimed to replicate the results from Chapter 3 with the use of Anodal and Cathodal tDCS stimulation over both IFG and pMTG separately. We hypothesized that increasing the neural threshold via Cathodal TMS stimulation would significantly decrease the amount semantic gesture-speech integration, and the decrease of the neural threshold of Anodal stimulation would significantly increase the amount of semantic gesture-speech integration. Results showed neither of these effects. Furthermore, to test the possibility that the anatomically well-connected IFG and pMTG could compensate for each other, we used tRNS stimulation, which can stimulate the two brain areas at the same time, in a similar way. We hypothesized that increased brain excitability with high-frequency tRNS stimulation would increase semantic congruency effect and decreased brain excitability with low-frequency stimulation would decrease semantic congruency effect. Again, the results failed to show any significance. Since we do found the involvement of both IFG and pMTG in the integration of gesture with speech in Chapter 3 as well as in Chapter 4, we attribute the null findings of this chapter to the techniques we used.

Even though tDCS has been proven to be an appropriate tool for monitoring cognitive functions (Kuo et al., 2012; Miniussi et al., 2013), the facilitatory effect of Anodal stimulation and the inhibitory effect of Cathodal stimulation seem to be valid only on the motor system (Nitsche et al., 2008). In cognitive tasks, the effect of tDCS stimulation seems to be quite complex and task-dependent. For example, in a study carried out by Schulke et al. (2017), they investigated the function of the left frontal lobe in the processing of metaphoric coverbal gesture and iconic coverbal gesture with the use of tDCS. The results only showed a significantly decreased reaction time for metaphoric gestures in the Anodal condition. There was no increase of reaction time in the Cathodal condition, nor was there any effect of brain stimulation for iconic gestures. In another study, Cohen-Maximov et al. (2015) presented participants with gestures as primes and verbal cues as targets, and asked participants to judge the semantic relatedness of the primes and targets after the stimulation of tDCS. They found that only when Anodal tDCS stimulation was applied over the right IFG and Cathodal tDCS stimulation over the left IFG, was there a faster response in terms of reaction time.

In our study, we investigated the automatic integration of gesture and speech by asking participants to pay attention only to the gender of the speaker, that paradigm may ‘hinder’ the semantic integration effect of gesture and speech, and caused no elicitation of signal when the neural activity was monitored with the weak non-invasive current. Future studies can investigate whether using a more explicit experimental task would lead to different results.

Chapter 6. General discussion

6.1. Overview

6.1.1. Summary of research aims

In summary, there are two aims in this thesis: one is to provide causal evidence for the involvement of IFG and pMTG in gesture-speech integration, the other is to investigate the time window for this integration. These questions were addressed in three chapters. Chapter 3 and Chapter 5 addressed at the first aim of locating the brain area(s) involved in gesture-speech integration. In Chapter 3, Transcranial Magnetic Stimulation was applied over IFG and pMTG to investigate whether the two brain areas are involved in the integration of gesture and speech. Semantic congruency effect (Rt (semantic incongruent) - Rt (semantic congruent)) was defined as the amount of semantic integration of gesture and speech. In Chapter 5, Transcranial Electrical Stimulation was used to provide further evidence for the involvement of IFG and pMTG in gesture-speech integration. Chapter 4 was concerned with investigating how gesture and speech integrate with each other, with an investigation of the time window for the priming effect of gesture on speech. In Chapter 4, by presenting gesture 200ms (mean identification point of all gestures 199.2ms, SE=9.96) ahead of speech, a prime paradigm of gesture on speech was created. The 400ms after the onset of speech was split into 10 time windows and double-pulse Transcranial Magnetic Stimulation applied over each time window to provide direct evidence for the time window where pMTG makes a critical combination to the integration process.

6.1.2. Summary of research findings

For all our experiments, the experimental paradigm of Kelly et al. (2010a) was used as the experimental task, in which participants were asked to pay attention only to the gender factor with manipulation of both gesture factor (gesture congruent vs gesture incongruent) and gender factor (gender congruent vs gender incongruent). All experiments were conducted with one type of non-invasive brain stimulation, either online (the rTMS used in Experiment 2 and the double-pulse TMS stimulation used in Chapter 4, where participants performed the experimental task under concurrent TMS stimulation) or offline (the theta-burst TMS used in Experiment 1 and the tES used in Chapter 5 where participants received a period of brain stimulation, followed by the experimental task).

6.1.2.1. Summary of Chapter 3

There were two experiments in Chapter 3. In Experiment 1, participants underwent a within-subject design with three sessions, where continuous theta burst stimulation (cTBS) was either applied to the left IFG, pMTG or a control site (vertex). Thus, the full experimental design was a 3 (Stimulation Site: IFG, pMTG, Vertex) x 2 (Gender Congruency) x 2 (Semantic Congruency) factorial design.

Results showed a significant Semantic Congruency by Stimulation Site interaction, indicating that the magnitude of the semantic congruency effect was modulated depending on the brain area stimulated. Simple effects analysis indicated that the size of the semantic congruency effect was significantly reduced when cTBS was applied to the left IFG relative to control site stimulation. A similar pattern was observed following stimulation of pMTG, although this comparison did not reach full significance. No such modulation was observed for the gender congruency effect, as indicated by a non-significant Stimulation Site by Gender Congruency interaction, indicating that cTBS selectively interrupted only the semantic congruency effect. From the above results, it was concluded that even if participants were not asked to pay attention to the gesture information, cTBS stimulation of the left IFG specifically disrupted the semantic congruency effect that was ongoing, which provides direct evidence for the involvement of left IFG in automatic gesture-speech integration.

In Experiment 2, a possible role for pMTG was further investigated with the use of online repetitive transcranial magnetic stimulation (rTMS), which perturbed the cortical activity that synchronized with the presentation of the experimental stimuli, thus enabling a more powerful statistical comparison of the effects of brain stimulation on gesture-speech integration. Experiment 2 employed a 2 (Stimulation Site: pMTG, vertex) by 2 (Gender Congruency) by 2 (Semantic Congruency) factorial design.

Results showed a significant interaction of Semantic Congruency and Stimulation Site, indicating that the magnitude of the semantic congruency effect was modulated by rTMS. No such effect of brain stimulation on the gender congruency effect was observed. Therefore, it was concluded that rTMS stimulation of the left pMTG specifically disrupts the semantic congruency effect, which illustrates that left pMTG is also involved in automatic gesture-speech integration.

It was concluded that the results from Chapter 3 provided evidence for a causal relationship of the involvement of both IFG and pMTG in the automatic integration of gesture-speech.

6.1.2.2. Summary of Chapter 4

There were two experiments in Chapter 4. In Experiment 3, the 400ms after the onset of speech was split into 10 time windows and a within-subject design was conducted with double-pulse TMS stimulation applied over the 10 time windows on either pMTG or the Vertex (control) site. The full experiment design was 2 (Stimulation Site: pMTG and Control) *10 (Time Window: 0/40ms, 40/80ms, 80/120ms, 120/160ms, 160/200ms, 200/240ms, 240/280ms, 280/320ms, 320/360ms, 360/400ms after the onset of the sound) *2 (Semantic Congruency: semantic congruent vs semantic incongruent) * 2 (Gender Congruency: gender congruent vs gender incongruent).

Results showed a significantly decreased RT in time window 3 (80-120ms after the onset of speech, and 280-320ms after the onset of gesture) when double pulse TMS stimulation was applied over pMTG compared to the same stimulation applied over the Vertex. However, there was no significant interaction of Stimulation Site by Time Window by Semantic Congruency, nor was there a significant interaction of Stimulation Site by Time Window by Gender Congruency. Therefore, it was not possible to tell whether the significantly decreased RT in the third time window when double pulse TMS was stimulated over pMTG compared to the Vertex was caused by the semantic factor or gender factor, or a combination of the two factors.

In Experiment 4, it was taken into account that some gestures may be processed faster than others and a gating study to find the identification point (IP) for each of the gesture (199.2±9.96ms) was conducted. Based on the IP, the time window was shifted into the IP_TW, in which all gestures have a clear understanding from the first IP_TW. Then the results of Experiment 3 were reanalysed by conducting a ANOVA of 2 (Stimulation Site: pMTG and Control) *10 (IP_TW: 0/40ms, 40/80ms, 80/120ms, 120/160ms, 160/200ms, 200/240ms, 240/280ms, 280/320ms, 320/360ms, 360/400ms after the IP of the gesture) *2 (Semantic Congruency: semantic congruent vs semantic incongruent) * 2 (Gender Congruency: gender congruent vs gender incongruent).

Results showed a significant interaction of Stimulation Site by Time Window by Semantic Congruency. Further analyses showed that this effect lay in the third IP_TW in the Vertex (control) condition, but the significant semantic congruency effect disappeared with double-pulse TMS applied over pMTG in the third IP_TW. It was concluded from the above results that the priming effect of gesture on speech takes place in the third IP_TW, which is 80-120ms after the IP of gesture and before the semantic analysis of speech (Friederici, 2002).

Taking the results from both Experiment 3 and Experiment 4 together, it was concluded that (1) experiment 4 showed a priming effect of gesture on speech, together with Experiment 2, it was hypothesized that this priming effect is how gesture and speech integrate with each other. (2) For gestures to have a 'priming' effect, they need to be clearly understood. (3) For speech to be influenced by the 'priming effect', it needs to be at a stage before the semantic analysis.

6.1.2.3. Summary of Chapter 5

There were three experiments in Chapter 5. In Experiment 5, the aim was to replicate the results from Experiment 1 with the use of transcranial direct current stimulation (tDCS). There were three sessions in this experiment: 20 minutes of Cathodal tDCS stimulation applied to the left IFG to decrease the excitability of the left IFG, 20 minutes of Anodal tDCS stimulation applied to the left IFG to increase the excitability of the left IFG, and 30s of tDCS stimulation as a sham condition. The full experimental design was a 3 (Stimulation Effect: Cathodal, Anodal, Sham) x 2 (Gender Congruency) x 2 (Semantic Congruency) factorial design.

Results showed no significant main effect of stimulation effect, nor was there a significant interaction of Stimulation Effect by Semantic Congruency, indicating that the magnitude of the semantic congruency effect was not significantly modulated by whether the IFG was stimulated in either Anodal condition or Cathodal condition compared to the Sham condition.

In Experiment 6, the aim was to replicate the results from Experiment 2 with tDCS. There were three sessions in this experiment: 20 minutes of Cathodal tDCS stimulation applied to the left pMTG to decrease the excitability of the left pMTG, 20 minutes of Anodal tDCS stimulation applied to the left pMTG to increase the excitability of the left pMTG, and 30s of tDCS stimulation as a sham condition. The full experimental design was a 3 (Stimulation Effect: Cathodal, Anodal, Sham) x 2 (Gender Congruency) x 2 (Semantic Congruency) factorial design.

Results showed no significant main effect of Stimulation Effect, nor was there a significant interaction of Stimulation Effect by Semantic Congruency, indicating that the magnitude of the semantic congruency effect was not significantly modulated by whether the pMTG was stimulated in either Anodal condition or Cathodal condition compared to the Sham condition.

In Experiment 7, transcranial random noise stimulation (tRNS) was applied over both IFG and pMTG synchronously to further test the role of IFG and pMTG play in gesture-speech integration. There were three sessions: 20 minutes of high-frequency tRNS stimulation over both IFG and pMTG to synchronously increase the excitability of the two brain areas, 20 minutes of low-frequency tRNS stimulation over IFG and pMTG to synchronously decrease the excitability of the two brain areas, and 30 seconds of tRNS stimulation as a Sham condition.

Results showed no significant main effect of Stimulation Effect, nor was there a significant interaction of Stimulation Effect by Semantic Congruency, indicating that when both IFG and pMTG were stimulated synchronously in a similar way, there was no significant semantic congruency effect in either the high frequency condition, low frequency condition or the sham condition.

6.1.2.4. Overall Summary

Taking the experiments together, Chapter 3 found a significantly reduced semantic congruency effect on application of either Theta-burst TMS stimulation over IFG or repetitive TMS stimulation over pMTG compared to the Vertex condition. With the definition that the semantic congruency effect (defined as the RT (semantic incongruent condition) - RT (semantic congruent condition)) represents the amount of gesture-speech integration, it was then concluded that both IFG and pMTG are involved in the integration of gesture with speech. Thus it can be concluded that Chapter 3 provide evidence as to ‘where’ the integration of gesture and speech took place. In Chapter 4, by splitting the 400ms after the onset of speech into 10 time windows and applied double-pulse TMS stimulation over each of the time window, as well as a further manipulation of the identification point of gesture, a significantly reduced semantic congruency effect was found 80-120ms after the IP of gesture, which is also during the phonological phase of speech processing. This led to the conclusion that gesture prime the lexical retrieval of speech during the phonological phase. Taking this together with the results of Chapter 3, it was concluded that Chapter 4 provided evidence of ‘how’ the two different modalities integrate with each other. The following section provides a detailed discussion of how the information from different modalities (gesture: visual, speech: audial), with different properties (gesture: global and synthetic, speech: linear-segmented), integrate with each other.

6.2. An integrated system

6.2.1. A common area for both the processing of speech and gesture

As two traditional language areas, IFG (Broca’s area) and pMTG (Wernicke’s area) have been shown to be involved in both language production and comprehension. Moreover, studies have shown that IFG and pMTG are also involved in both action perception and production.

6.2.1.1. IFG (Broca’s area) in speech and gesture processing

It has long been suggested by previous researchers that Broca’s area is associated with both language comprehension and production (Burnstine et al., 1990; Ojemann, 1991). In a study carried out by Fridriksson et al. (2009), they investigated the neural recruitment of the cortical area during either imitating or observing speech movement using fMRI. They found that the greatest frontal lobe activity in Broca’s area was triggered not only during production of speech but also during perception of the speech movement. Based on the results, they concluded that Broca's area, which involved in the execution of speech movements is also recruited in the perception of the same movements by other speakers.

The involvement of Broca's area in speech production has been shown in both lesion studies and functional imaging studies, at various levels of processing: phonological (Bohland et al., 2006; Fiez et al., 1996; Fridriksson et al., 2009), semantic (Amunts et al., 2004; Friederici et al., 2003b; Thompson-Schill et al., 1997) and syntactic (Ben-Shachar et al., 2004; Friederici et al., 2003a; Heim et al., 2003). Additionally, speech does not need to be produced aloud to involve Broca's area. Bookheimer et al. (1995) found activation of Broca's area during both silent and oral object naming. In language perception, several studies have indicated the involvement of Broca's area in phonological encoding during speech processing (Gold et al., 2005; Gough et al., 2005; Moineau et al., 2005). There are also studies that found the activation of Broca's area during lexical speech comprehension (Frost et al., 1999; Moineau et al., 2005; Zhuang et al., 2011).

In action perception, Rizzolatti et al. (1996) found the activation of the posterior part of the inferior frontal gyrus (Broca's area) during grasping observation. Grafton et al. (1996) found the activation of the left inferior frontal cortex during grasp observation. They also found activation of left inferior frontal and middle frontal cortex during imagined grasping. Furthermore, the activation of Broca's area has also been found during the execution of self-ordered movement (Bonda et al., 1995), during the preparation to make a copied movement (Krams et al., 1998), and even at the sight of a picture of the hands (Parsons et al., 1995).

In action comprehension, a study carried out by Decety et al. (1997) found strongly activated left frontal regions when participants were presented with meaningful actions (e.g. opening a bottle, hammering a nail) compared to meaningless ones (derived from the American Sign Language, which participants were unacquainted with). In another study, Husain et al. (2009) manipulated the level of processing of gestures by asking participants to do either a discrimination task or a category task. While the first one only engaged in a phonological level of processing, the latter one engaged in the semantic level of processing. They found that in contrast to the discrimination task, the category task activated the left middle and inferior frontal gyrus. They concluded that these results illustrate the semantic-processing of gestures in the brain in inferior frontal gyrus.

6.2.1.2. pMTG (Wernicke's area) in speech and gesture processing

Speech processing begins in the superior temporal gyrus. There are plenty of studies demonstrating the activation of the superior temporal gyrus in word perception (Fiez et al., 1996; Mazoyer et al., 1993; Petersen et al., 1988), in the translation of orthographic symbols to phonemic representations (Xu et al., 2001), as well as the activation of semantics from speech sounds (Hickok et al., 2004; Roder et al., 2002).

Wernicke's aphasia, which involves an impairment of language comprehension, also involves a prominent speech production impairment. An example is the phonological encoding impairment found by Wilshire et al. (1996), in which they provided evidence for the involvement of Wernicke in phonological processing using functional neuroimaging methods. Wise et al. (1991) found that during the verb generation task, Wernicke's area was engaged in the processing of both vocalization and silent word generation, as has been found by some other studies with the use of positron emission tomography (PET) (Fiez et al., 1996; Small et al., 1996). Activations of the Wernicke's area has also been reported during semantically mediated word retrieval tasks by Warburton et al. (1996).

During action processing, Rizzolatti et al. (1996) found the activation of the superior temporal sulcus during grasping observation with the use of positron emission tomography. Grafton et al. (1996) found the activation of the left rostral superior temporal sulcus during grasp observation. In another study, Husain et al. (2009) manipulated the content of gestures-either meaningful or meaningless. The results showed that unlike meaningless gestures, meaningful gestures activated the middle and superior temporal gyrus. The activation of the temporal regions in meaningful actions rather than meaningless ones has also been found by Decety et al. (1997).

6.2.1.3. Summary

From above, we can conclude that both IFG and pMTG are involved in the production of language and action, as well as in the comprehension of the information from both sources. This makes it possible for the integration of information from gesture and speech to take place in both areas, as has been shown in Chapter 3.

6.2.2. A unified origin of gesture and speech

Previously in this thesis, evidence was summarized for the bidirectional influence of gesture and speech (Bernardis&Gentilucci, 2006), as well as that gesture always co-occur with speech (Beattie et al., 1979; Habets et al., 2011; Krauss et al., 1991; Morrelsamuels&Krauss, 1991; Obermeier&Gunter, 2015; Obermeier et al., 2011). In Chapter 3, evidence was also provided that the integration of gesture and speech takes place in both IFG and pMTG, both of which are involved in both the processing of gesture as well as the processing of speech. Based on this, it is suggested that gesture and speech are not separate communication systems, as has been assumed by some researchers (Hadar&Krauss, 1999; Hadar et al., 1998; Levelt et al., 1985). Instead, there is a common origin in which there is a synthesis of opposite modes of thought – global-synthetic and instantaneous imagery with linear-segmented temporally extended verbalization, as has been suggested by both McNeill (1992) and Morrel-Samuels and Krauss (1991).

6.2.2.1. *McNeill's growth point theory*

According to McNeill (1992, 2005), thinking is dialectic of both global imagery and linguistic category that are realized in both speech and gesture. A growth point is the smallest unit of the codified meaning independent of modality. During the utterance realization, thoughts undergo continuous changes, thus, thinking is shaped during speaking. For example, in the realization of the visuospatial context, the underlying visuospatial thinking in gesture will be brought into the system of categories of the language. Originating from the same growth point, and realized in different modalities in synchrony, there is a mutual influence between the imagery and linguistic category. Context is the background from which the growth point is differentiated. The speaker shapes the context background in a certain way to realize the intended growth point. The context and the growth point are mentally constructed together. According to McNeill, such a differentiated growth point is validated by the temporal connection of gesture stroke with the peak of acoustic output in speech. As has been stated by McNeill (1992): the growth point is seen in the gesture stroke, together with the linguistic segment with which it co-occurs, plus a word that follows this segment if this word preserves semantic and pragmatic synchrony.

6.2.2.2. *Goldin-Meadow's framework*

Goldin-Meadow (2003) states that gesture and speech draw upon a single set of representations. Some of the representations can be articulated through both gesture and speech, while some of them can only be expressed through gesture. A gesture-speech match will occur for those representations that can be accessed by both gesture and speech, but for those that can only be expressed through gesture, a gesture-speech mismatch (i.e., gesture and speech contain additional information) will occur when trying to articulate these representations. However, this single set of representations of gesture and speech does not exist from birth. A study carried out by Butcher et al. (2000) investigated the relationship between gesture and speech in a longitudinal study of six children aged between 12 and 27 months, the period in which the children make the transition from one-word speech to two-word utterance. They examined both the relation of the production of gesture to meaningful words, as well as that to uninterpretable sounds. They found that for the most part, children produced gesture on its own without any speech. When gestures are combined with sounds, they are either uninterpretable or asynchronous. Goldin-Meadow concluded that when the gesture is first produced, it does not form a single integrated representation system with speech. The two forms become integrated later in development, probably earlier than children begin to pronounce words in combination with other words. From that point forward, information from gesture and speech integrate with each other and form a single representation system.

6.2.2.3. Summary

In the above section, two theories of a unified system for gesture and speech have been presented. According to McNeill (1992, 2005), the origin of gesture and speech is thinking. Thinking is assumed to be dialectic of both global imagery and linguistic category, with growth point as the smallest unit of the codified meaning that can be realized in both speech and gesture. Goldin-Meadow (2003) states that gesture and speech draw upon a single set of representations, some of which can be articulated through both gesture and speech, and some of them can only be expressed through gesture. Gesture provides a ‘window on the mind’ to describe some representations that cannot be described through speech (Goldin-Meadow, 2003).

6.2.3. A connected production system

The hand and mouth are closely coordinated with each other and are controlled by the same motor system, as has been suggested by Hostetter et al. (2008, 2010). According to this viewpoint, speech is articulatory mouth movement that generates successive speech sounds (Kotz et al., 2010; Liberman et al., 2000). The motor system that controls the movement of the mouth also controls the movements of the hands, so activation of the mouth for speaking automatically activates the hand to gesture (Kelly et al., 2002).

In a study carried out by Floel et al. (2003), they presented participants with five different tasks: ‘listening to well-known fairy-tales’, ‘listening to short simple sentences’, ‘reading aloud of the fairy-tales’, ‘reading silently of the fairy-tales’ and a task without linguistic property, ‘listen to white noise’. Participants completed all the tasks with transcranial magnetic stimulation (TMS) over the primary motor cortex. MEPs of the first dorsal interosseous muscle were recorded. Results showed significantly greater MEPs in the ‘listening to well-known fairy-tales’, ‘listening to short simple sentences’, ‘reading aloud of the fairy-tales’, and ‘reading silently of the fairy-tales’ conditions compared to the baseline. Floel et al. argued that the results demonstrated that both productive and perceptive linguistic tasks would facilitate the hand motor cortices. In another study, Fadiga et al. (2002) presented participants with action sounds of three types: hand action sounds (e.g. typing or tearing paper), leg action sounds (e.g. walking), and control sounds (e.g. thunder). By applying transcranial magnetic stimulation (TMS) over the hand muscles they found a greater motor corticospinal excitability in the hand action sound condition compared to the other two conditions. Fadiga et al. argued that the results showed that action coding may be a precursor of language.

Gentilucci et al. (2001) designed a set of experiments to test the relationship between hand movement and mouth movement. They found that in cases when participants were asked to reach and grasp objects different sizes with either the hand or the mouth, the size of the object had an influence on the initial kinematics of both the mouth and finger opening, and this effect was not due to the synchrony between the opening of the mouth and the finger. When participants were asked to grasp the object with the hand while pronouncing a syllable at the same time, the opening of the mouth and sound production was affected by the size of the object. Gentilucci et al. hypothesised that there exist double motor commands simultaneously sent to both arm and mouth, and Broca's area is the neural substrate that underlies this unique action planning.

The motor theory of speech perception assumes that speech is a system of a series of articulatory mouth movements that generate successive speech sounds. According to this theory, a listener understands the speaker because presentation of the speaker's articulatory gesture activates that of the listener (Kotz&Schwartze, 2010; Liberman&Whalen, 2000). In other words, the perception of the speech is accompanied by exploiting the motor commands that produce the speech--the basic logic of the motor theory of speech perception.

The motor theory of speech perception has been supported by the results showing the activation of the same motor (Fadiga et al., 2002) and premotor (Wilson et al., 2004) areas during listening to phonemes and syllables, during the production of these phonemes in overt production (Pulvermuller et al., 2006), as well as the findings of the activation of tongue primary motor cortex during both visual and auditory speech perception (Sato et al., 2009). Using transcranial magnetic stimulation (TMS), there is a dissociation in speech sound discrimination when TMS is applied over motor cortex controlling lips and tongue (D'Ausillo et al., 2009), and a disruption of the ability to perform a phonetic discrimination task with rTMS stimulation over the premotor cortex (Meister et al., 2007).

6.2.4. An integrated system of information

6.2.4.1. A common ground of semantic information

Both speech and gesture are arbitrary signs, the comprehension of which needs a mapping from the form onto the meaning.

Language comprehension is a process of matching words onto meaning. According to the two-system architecture sketched by Levelt (2001), conceptual-semantic meaning and the corresponding phonological/ orthographic form are two steps that are processed sequentially. In word comprehension, the meaning is accessed only when the phonological/ orthographic form encoding process is completed. In word production, the phonological/ orthographic form encoding process begins only after the meaning based form is completed. There is a bi-directional influence between the two steps, phonological form-based processing can modulate conceptual meaning-based processing, and meaning processing can also have an influence on phonological processing. Spivey et al. (2005) presented participants with pictures of two objects that were either phonologically related (e.g., /candle/ and /candy/) or phonologically unrelated (e.g., /picture/ and /candy/). Participants were instructed to click the mouse to select the item according to a speech file. The mouse movement trajectories were linked to the lexical representation by a computational TRACE model. Results showed that mouse trajectories exhibited attraction towards the alternative incorrect response, with respect to control trials. Spivey et al. (2005) explained that the results showed an "intrusion" and an influence of the phonological information on the competition between lexical stimuli. Furthermore, they reported evidence for automatic processing of phonological information during lexical tasks. In another study carried out by Peleg et al. (2016), participants were presented with either unambiguous words or ambiguous words in types of homonyms, homophones, and homographs. They found that the semantic-conceptual process was influenced by the activated lexical form.

Rizzolatti et al. (2004) hypothesized that the activation of brain areas during action observation reflects the mapping between observed actions and motor representations stored in the brain. An observed action can only be understood or imitated if it is part of representation that the observer has with regard to the nervous system (Decety et al., 1997). Actions that are not in the observer's motor repertoire, for example, a dog's barking, will not lead to the activation of the mirror neurons (Buccino et al., 2004). By mapping the observed action onto their own motor representation, which is stored in the brain in terms of memories or experiences, the observer can understand the action, as well as the goal of that action. The perception and execution of action are not independent of each other, there is an interactive relationship between them. On one hand, the process of action perception is constrained, or guided, by implicit knowledge of the movement it is capable of producing (Viviani et al., 1992). For instance, Viviani and Stucchi (1992) found that the perceptual judgement of the velocity of a moving dot as uniform or not depends on the radius of the trajectory's curvature, participants tend to judge the velocity of the moving dot as uniform if the trajectory follows the motor rule of movement production, otherwise the velocity of the moving dot will be misjudged. On the other hand, action production can shape the information being perceived, eye, hand, and body movement together shapes visual perception (Gibson et al., 1979).

From the above, we can see that language is a process of mapping words onto meaning, and action is a process of mapping kinetic movement onto representation. The linkage of production and perception in language processing suggests common coded words and meaning in language processing, the linkage of production and perception in action processing suggests a common coded kinetic movement and representation in action processing. With the unified system for gesture and speech illustrated by McNeill (1992, 2005) and Goldin-Meadow (2003), it can be concluded that the common coded meaning in language is of the same representation/thinking system as the common coded representation in action processing. Moreover, this unified representation/thinking system of gesture and speech is located in IFG and pMTG, as has been illustrated in part 6.2.1, as well as proved by Chapter 3.

As has been reported in the introduction (part 1.2.3, page 14-19), there is a bidirectional influence of gesture and speech in terms of external form. Multimodal voice spectra (especially formant 2, F2) are enhanced by gestures, whereas multimodal gesture parameters (like gesture time, maximal height, hand oscillation amplitude, maximal peak velocity, the number of velocity peaks) are reduced by words (Bernardis&Gentilucci, 2006). There is also a bidirectional influence of speech and gesture at the semantical level, in that gestures used to express motion events are influenced simultaneously by how the features of motion events are expressed in language (Kita&Ozyurek, 2003). Using a double-prime paradigm, Kelly et al. (2010b) showed that gestures influenced speech comprehension, and speech influenced gesture comprehension in a comparable way. The remaining question is how this unify-originated, bi-directionally influenced information system articulates effectively into two modalities (analytic speech or spatial-motoric gesture).

6.2.4.2. Information packaged with a common coded representation

Hommel et al. (2001) developed the theory of event coding (TEC) which claimed that there exists an internal representation of external events. That internal representation is a common coded representational domain of perceived events and action generated events, both of which are stored in the brain in terms of feature codes (f1, f2 etc). The feature codes can be as simple as colour or shape, or as complex as ‘sit-on-ableness’ that represent time and changes. Feature codes are not restricted to one modality, they can form a temporal composition from multiple resources to make an ideal response. As summarized in figure 6.1, as long as there are the same feature codes in the common coding system (f1, f2), there is an interaction between the two systems (the Sensory system and the Motor system). It is the feature codes that link the sensory system and motor system together.

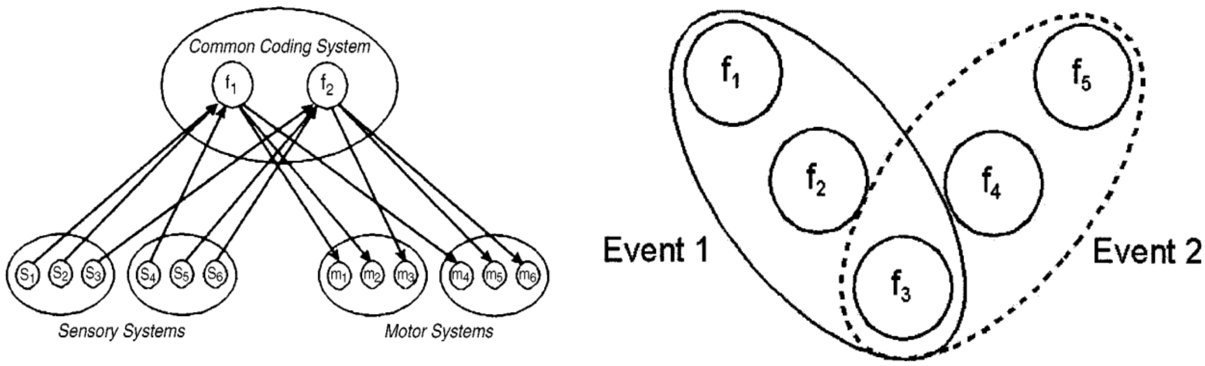


Figure 6.1 An illustration of feature codes and common coding system (cited from Hommel et al., 2001)

An event code consists of several feature codes. For example, event 1 consists of three feature codes (f1, f2, f3), and event 2 also consists of three feature codes (f3, f4, f5) with overlapping feature codes of f3. The activation of one feature code (f3) will prime all the events of which it is part of (Event 1 and Event 2), leading to the synchronous activation of feature codes banded together by each event. The activated feature codes may come from different systems, as depicted in Figure 6.1, where f1 is from the sensory system, while f2 is from the motor system. This synchronous activation of feature codes from different systems is what Hommel et al. (2001) called integration of the two systems. Hommel et al. (2001) also assumed that there are two phases for perceptual coding. In the first phase, there is a parallel activation of all feature codes that relate to the stimulus processing. In the second phase, the activated feature codes will be integrated, to prevent any of them being involved in other concurrent coding processes.

Based on the common coded representation hypothesis that put forward in sub-session (part 6.2.3.1), I would like to promote the theory of event coding proposed by Hommel et al. (2001) from an explanation of a linkage of perception and production of action to a theory of an explanation of a linkage of speech with gesture. It is hypothesized that a common coded representation of both gesture and speech is where the information from the two different modalities 'integrates' with each other, given that gesture conveys information relevant to speech (Goldin-Meadow&Sandhofer, 1999; Kelly&Church, 1998). The common coded representation is equivalent to the 'growth point' proposed by McNeill (1992) and the packaged information proposed by Kita (2000). In the present study, the common coded representation is defined as the smallest unit for the integration of gesture and speech, after the common coding system proposed by Hommel et al. (2001).

Under a common coded representation, features of both speech and gesture gather together—that is, the feature codes of gesture and the feature codes of speech. The feature codes of gesture contain spatial-motoric gestural information that is global and synthetic, while the feature codes of speech contain analytic lexical information that is linear-segmented. The two different kinds of feature codes are related to each other, and contain either supplementary or redundant information. In the following two sections it will be discussed in detail how the feature codes of gesture and speech relate to each other.

6.2.4.3. Two aspects of a single thinking system with certain independence

Gesture and speech are two modalities with different properties (gesture: global and synthetic, speech: linear-segmented). From above, we learned that gesture and speech are driven by a unified original system, a common thinking, as has been proposed by McNeill (1992, 2008), or a single set of representations, as has been proposed by Goldin-Meadow (2003). Therefore, in the common coded representation, there are actually two kinds of thinking—spatial-motoric thinking of gesture and analytic thinking of speech, as has been illustrated by Kita (2000). The difference between this thesis and Kita's (2000) account is that, whereas Kita refers to two modes of thinking of gesture and speech, in this study, these are expressed as two aspects of thinking of a single representation/thinking.

In Experiment 3, when double-pulse TMS stimulation was applied over the 10 time windows that spread 400ms after the onset of speech, a significantly reduced reaction time was found when double pulse TMS was applied over pMTG compared to that applied to the Vertex in the third time window (80-120ms after the onset of speech), but there was no interaction between the stimulation site and the semantic congruency effect in the third time window. In Experiment 4, after considering the identification point of each of the gesture into consideration, the time window was shifted into the identification point time window (IP_TW) to make sure all gestures were at the same stage of comprehension from the first IP_TW. Then a significant interaction was found between the stimulation site and the semantic congruency effect in the third IP_TW.

From these findings, it can be concluded that gesture needs to be comprehended to integrate with speech, in which the information conveyed in gesture is encoded as a whole (McNeill, 1992). The phenomenon that action is encoded based on its goal, rather than the individual movement of the action form, has been suggested by several researchers (Rizzolatti et al., 2009; Umilta et al., 2008). For example, in the study carried out by Umilta et al. (2008), they observed the activation of the cortical motor neurons, in cases when a monkey was asked to grasp with the hands, to grasp with normal pliers (requiring a close hand), and also when the monkey was asked to grasp with reversed pliers (requiring an open hand). It was argued that since gestures are encoded as a whole and based on their goal, there would be no significant interaction between stimulation site with the semantic congruency effect until the identification point of the gesture been taken into consideration, as has been shown in Experiment 3 and Experiment 4.

It is further concluded that there is a certain independence in the information conveyed in gesture, as it was shown in Experiment 4 that once the identification point has been identified, the comprehension of gesture will no longer be influenced by further occurrence of speech.

To explain in terms of the common coded hypothesis, it can be hypothesised that feature codes f1, f2, f3 comprised gesture event, while feature codes f3, f4, f5 comprised speech event, with an overlapping feature code f3. We hypothesised that the overlapping feature code f3 is the semantic meaning, which represents in gesture in a visuospatial way, while represents in speech in a linguistic way. The presents of the visuospatial feature code f3 activates the gesture event, which lead to the activation of the equivalent representation of speech (speech event) that shares a common coding system with the gesture event, thus lead to the activation of the feature code f3 in the linguistic way. However, the further presents of feature code f1 and/or f2 would not have an influence on activation of f3 in the visuospatial way, neither can the presents of feature code f4 and/or f5. In other words, the temporary composition of feature code of gesture (f1, f2, f3) and feature code of speech (f3, f4, f5) are independent of each other. However, they are under a single representation and share a common coding system, so the feature of one code will have an influence on the form of the feature of the other. In the following sub-section, the operative of the cross-model priming effect will be discussed.

6.2.4.4. Two aspects of single thinking that primed interactively

In Experiment 4, when we applied double-pulse TMS over pMTG, we found a significantly reduced semantic congruency effect in the third IP_TW (80-120ms after the IP of gesture) compared to the Vertex condition. These results demonstrate a ‘priming’ effect of gesture on speech. According to growth points in thinking-for-speaking, the processing of lexical semantic information takes place 300-500ms after the onset of the speech. Therefore, the ‘priming’ effect of gesture on speech happens before speech’s semantic analysis phase. In fact, gesture will have an influence on the semantic retrieval of speech.

Krauss et al. (2000) proposed a lexical gesture process model to explain the facilitation effect of gesture on speech. According to this model, information stored in memory is encoded in multiple ways, including both the visuospatial format of information arising from gesture and the conceptual format of information of speech. The access of one format of information (gesture) will activate its representation that is stored in the working memory, the activated representation (of gesture) will result in the spreading of activation of the related representations (speech), thus leading to a cross-modal priming. In gesture production, the motor planner organizes the spatial/dynamic format of information that stored in memory into a motor programme, which provides the motor system introductions for execution of a gesture. The motoric feature of the gesture will be picked up by the kinesthetic monitor, which will monitor the features of the source information represented in gesture, and then transfer the picked up information into the phonological encoder of speech in where it facilitates the lexical retrieval of the word. Additionally, the fact that speech is impeded after gesture has been restricted also shows that gesture helps lexical retrieval of speech (Rauscher et al., 1996; Rime et al., 1984).

The results in Chapter 4 provided direct evidence for the ‘priming’ effect of gesture during the phonological phase of speech. More importantly, our results showed that instead of monitoring the kinesthetic information of gesture, gesture is monitored as a whole and based on its goal. In other words, gesture first has to be ‘fully understood’, before it can have a ‘priming’ effect on speech.

It was further hypothesized that this priming effect is how gesture and speech 'integrate' with each other. There is a cross-modal priming effect between gesture and speech, in the same way of a bi-directional influence. To be more specific, on one hand, the representation of gesture acts as an equivalent source of analytic information for lexical selection of speech. On the other hand, the representation of speech will also act as an equivalent source of spatial-motoric information for motoric retrieval of gesture. In terms of the common coded hypothesis, the explanation would be that under a common coding system, there is a cross-modal priming effect between feature codes of gesture and feature codes of speech.

6.2.4.5. Summary

In summary, in this part, this section has provided a description of how the results of the present thesis speak to the concept of an integrated system of gesture and speech. Based on a linkage of production and perception in both language processing and gesture processing, as well as the bi-directional influence between gesture and speech in both the outer form and the semantic level, it is assumed that a common coded meaning/representation is stored in the brain regardless of its modality. It is argued that this common coded representation is where the information from the two different modalities of gesture and speech is 'integrated' with each other. Thus, the 'integration' of gesture and speech is defined in terms of two aspects of information which prime each other interactively under a single representation.

6.2.5. Summary

Taken together, the results from Chapter 3 provided direct evidence for the involvement of both IFG and pMTG in the integration of gesture and speech. Together with evidence that both IFG and pMTG are involved in the production of language and action, as well as in the comprehension of the information from both sources, this suggests that there is a unified system of gesture and speech. According to McNeill (1992, 2008), the origin of gesture and speech is thinking. Goldin-Meadow (2003) states that gesture and speech draw upon a single set of representation. The results from Chapter 4 provided evidence that gestures prime the lexical retrieval of speech after the gesture has been clearly understood, before the semantic analysis of speech. Together with the evidence that there is a linkage of production and perception in language processing, a linkage of production and perception in gesture processing, a linkage of speech and gesture in terms of outer form, as well as a linkage of speech and gesture in terms of the semantic meaning. It is also suggested that there is a common coded representation of gesture and speech that integrate the feature codes of gesture with the feature codes of speech for a proper response. Furthermore, it is suggested that there exists a cross-modal priming between the two kinds of feature codes with a certain degree of independence.

6.3. Limitations and Suggestions

6.3.1. Different roles of IFG and pMTG in gesture-speech integration?

Even though both IFG and pMTG are involved in gesture speech integration, the role they play during the integration may be different. pMTG is known for its involvement in the storage and retrieval of semantic information (Hagoort, 2005; Hickok et al., 2007), while IFG is associated with the unification that can unify semantic information retrieved from memory (Hagoort, 2005). Tesink et al. (2009) made a functional distinction between the ‘integration process’ and the ‘unification process’. They define ‘integration’ as the convergence of different sources of information on a common memory representation, for example, the sight of a cat and the sound of meow. The ‘unification’ is a constructive process where a semantic representation is build up out of components stored in memory.

The work in Chapter 3 only showed an involvement of IFG and pMTG in gesture-speech integration, without distinguishing the role these two areas play. Future work could investigate the activation of brain areas in the different semantic relationship between gesture and speech, for example, the effect of the semantic difference between different gesture types and speech (i.e. the integration of iconic gestures with speech, and the integration of pantomime gestures with speech). Pantomimes are re-enactions or demonstrations of an action without using an object, therefore pantomimes can ‘stand on their own’ in conveying information. Iconic gestures refer to the concrete content of sentences (for example, the drop of the right hand to illustrate the concrete feature of the sentence “the man goes down the hill”). Such gestures do not have conventional or unambiguous meanings in the absence of speech (Feyereisen et al., 1988; Krauss et al., 1991). The meaning of iconic gesture has to be generated online based on gesture form and the co-speech context in which the gesture is observed. It has been assumed by Willems et al. (2009) that this different semantic relationship may have an influence on the neural underpinnings of gesture-speech integration.

In pantomime-speech integration, a joint activation of IFG and pMTG has been shown in the study of Willems et al. (2009). Nevertheless, when the gesture is ambiguous, and does not have a clear meaning unless it is accompanied by speech, as in the case of iconic gestures, there will be no direct retrieval of information from memory, but a unification process to build up a ‘new representation’. For example, a pushing forward movement has been stored in memory with various meanings (pushing someone/something, stretching one’s arms, punching etc.), the word ‘mow’ has also been stored in the memory with its meaning. When the pushing forward movement is accompanied with the word ‘mow’, a representation of mowing has been established. By this logic, only the IFG will be activated in the iconic condition, as has been suggested in the study of Willems et al. (2009).

Future research can make use of Transcranial Magnetic Stimulation to provide evidence of a direct, causal role of IFG and pMTG in the integration of iconic gesture-speech and/or pantomime gesture-speech.

6.3.2. Implicit paradigm vs explicit paradigm?

In all our studies, we used an implicit paradigm that asked the participants to pay attention only to the speech information. While speech and gesture integrate automatically, the interaction can also be modulated by some controlled processes. For example, Holle, &Gunter (2007) showed that the presence of non-meaningful gestures (e.g., grooming behaviours) modulates the extent to which meaningful gestures set up semantic expectations for a speech later in a sentence. Obermeier et al. (2011) found that when gestures did not temporally overlap with the homonyms and when participants were not explicitly asked to pay attention to gestures, speech- gesture integration did not occur. Kelly et al. (2007) demonstrated that if the two modalities are perceived as not intentionally coupled (i.e. gesture and speech being produced by two different persons) the integration is less strong (difference between incongruent speech-gesture pairs in terms of N400 effect) compared with when they are perceived as being intentionally coupled (i.e. gesture and speech being produced by the same person).

Further studies can investigate whether the implicit or explicit paradigm has an effect on the involvement of the brain area in gesture speech integration.

6.3.3. Prime effect of speech on gesture?

In chapter 4, there was evidence of a significant priming effect of gesture on lexical retrieval of speech during the phonological phase. Together with the bidirectional influence between gesture and speech, it was concluded that this cross-modal priming effect is how gesture and speech integrate with each other. However, to fully support this hypothesis, there should also be evidence for a priming effect of speech on the motoric retrieval of gesture.

Furthermore, in chapter 4, the evidence indicated that the priming effect of gesture on the lexical retrieval of speech takes place in pMTG. Given that both pMTG and IFG are involved in the interaction of gesture with speech, as has been shown in Chapter 3, it is concluded that both of the two areas involved in the cross-modal priming effect between gesture and speech. Future studies can investigate the cross-modal priming effect between gesture and speech in IFG.

Lastly, given that there may be a division of labour between the IFG and pMTG in the integration of gesture and speech, future studies can also examine the division of labour of pMTG and IFG in the cross-modal priming effect.

Specifically, it would be interesting to test three effects: first of all, the priming effect of speech on gesture in pMTG, as well as the gestural phase (preparation phase, stroke phase, or retraction phase) this priming effect took place. Since gesture has a priming effect on the lexical retrieval of speech, it is hypothesized that speech would also have a priming effect on the semantic retrieval of gesture. Hence, the priming effect of speech should take place before a clear comprehension of gesture, in other words, this priming effect of speech on gesture should be in the preparation phase of gesture. Secondly, the priming effect of gesture on speech in IFG. Given a possible division of labour of IFG in the integration of gesture and speech, it would be interesting to test whether this priming effect of gesture on speech also takes place in the phonological phase, as has been found in pMTG. Thirdly, the priming effect of speech on gesture in IFG, and the comparison of the gestural phase of this priming effect of speech on gesture in IFG with that occurring in pMTG could be investigated.

6.4. Conclusion

In summary, it is concluded from three chapters that, first of all, a direct evidence for the involvement of left IFG and pMTG in automatic gesture-speech integration was provided. Secondly, the automatic gesture-speech integration in pMTG takes place 80-120ms after the identification of gesture, and before the semantic analysis of speech.

Reference

Accornero, P., Martignani, E., Macchi, E., & Baratta, M. (2007). Hepatocyte growth factor exerts multiple biological functions on bovine mammary epithelial cells. *Journal of Dairy Science, 90*(9), 4289-4296. doi:10.3168/jds.2007-0179

Alibali, M. W., Flevares, L. M., & GoldinMeadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology, 89*(1), 183-193. doi:10.1037/0022-0663.89.1.183

Amunts, K., Weiss, P. H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J. M., . . . Zilles, K. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space - The roles of Brodmann areas 44 and 45. *NeuroImage, 22*(1), 42-56. doi:10.1016/j.neuroimage.2003.12.031

Anand, S., & Hotson, J. (2002). Transcranial magnetic stimulation: Neurophysiological applications and safety. *Brain and Cognition, 50*(3), 366-386. doi:10.1016/s0278-2626(02)00512-2

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28*(2), 105-+. doi:Doi 10.1017/S0140525x05000038

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). INTERACTIVE GESTURES. *Discourse Processes, 15*(4), 469-489.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica, 123*(1-2), 1-30. doi:DOI 10.1515/semi.1999.123.1-2.1

Beattie, G. W., & Butterworth, B. L. (1979). CONTEXTUAL PROBABILITY AND WORD-FREQUENCY AS DETERMINANTS OF PAUSES AND ERRORS IN SPONTANEOUS SPEECH. *Language and Speech, 22*(JUL-), 201-211.

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *NeuroImage, 21*(4), 1320-1336. doi:10.1016/j.neuroimage.2003.11.027

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*(2), 178-190. doi:10.1016/j.neuropsychologia.2005.05.007

Bikson, M., Inoue, M., Akiyama, H., Deans, J. K., Fox, J. E., Miyakawa, H., & Jefferys, J. G. R. (2004). Effects of uniform extracellular DC electric fields on excitability in rat hippocampal slices in vitro. *Journal of Physiology-London, 557*(1), 175-190. doi:10.1113/jphysiol.2003.055772

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage, 32*(2), 821-841. doi:10.1016/j.neuroimage.2006.04.173

Bonda, E., Petrides, M., Frey, S., & Evans, A. (1995). NEURAL CORRELATES OF MENTAL TRANSFORMATIONS OF THE BODY-IN-SPACE. *Proceedings of the National Academy of Sciences of the United States of America, 92*(24), 11180-11184. doi:10.1073/pnas.92.24.11180

Bookheimer, S. Y., Zeffiro, I. A., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping, 3*(2), 93-106. doi:10.1002/hbm.460030206

Broaders, S. C., & Goldin-Meadow, S. (2010). Truth Is at Hand: How Gesture Adds Information During Investigative Interviews. *Psychological Science, 21*(5), 623-628. doi:10.1177/0956797610366082

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language, 89*(2), 370-376. doi:10.1016/s0093-934x(03)00356-0

Burnstine, T. H., Lesser, R. P., Hart, J., Uematsu, S., Zinreich, S. J., Krauss, G. L., . . . Gordon, B. (1990). CHARACTERIZATION OF THE BASAL TEMPORAL LANGUAGE AREA IN PATIENTS WITH LEFT TEMPORAL-LOBE EPILEPSY. *Neurology, 40*(6), 966-970.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In D. McNeill (Ed.), *Language and Gesture* (pp. 235-258). Cambridge: Cambridge University Press.
Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport, 14*(17), 2213-2218. doi:10.1097/00001756-200312020-00016
Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(11), 649-657. doi:Doi 10.1016/S0960-9822(00)00513-3
Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris, 98*(1-3), 191-205. doi:10.1016/j.jphysparis.2004.03.018
Church, R. B., & Goldinmeadow, S. (1986). THE MISMATCH BETWEEN GESTURE AND SPEECH AS AN INDEX OF TRANSITIONAL KNOWLEDGE. *Cognition, 23*(1), 43-71.
Cohen-Maximov, T., Avirame, K., Floel, A., & Lavidor, M. (2015). Modulation of Gestural-verbal Semantic Integration by tDCS. *Brain Stimulation, 8*(3), 493-498. doi:10.1016/j.brs.2014.12.001
Crowder, E. M. (1996). Gestures at work in sense-making science talk. *Journal of the Learning Sciences, 5*(3), 173-208. doi:10.1207/s15327809jls0503_2
D'Ausillo, A., Pulvermueller, F., Saimas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology, 19*(5), 381-385. doi:10.1016/j.cub.2009.01.017
Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., . . . Fazio, F. (1997). Brain activity during observation of actions - Influence of action content and subject's strategy. *Brain, 120*, 1763-1777. doi:10.1093/brain/120.10.1763
Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science, 15*(2), 165-180. doi:10.1111/j.1467-7687.2011.01100.x
Dick, A. S., Mok, E. H., Beharelle, A. R., 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. doi:10.1002/hbm.22222
Dockery, C. A., Hueckel-Weng, R., Birbaumer, N., & Plewnia, C. (2009). Enhancement of Planning Ability by Transcranial Direct Current Stimulation. *Journal of Neuroscience, 29*(22), 7271-7277. doi:10.1523/jneurosci.0065-09.2009
Drijvers, L., & Ozyurek, A. (2017). Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension. *Journal of Speech Language and Hearing Research, 60*(1), 212-222. doi:10.1044/2016_jslhr-h-16-0101
Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-Based Activation Likelihood Estimation Meta-Analysis of Neuroimaging Data: A Random-Effects Approach Based on Empirical Estimates of Spatial Uncertainty. *Human Brain Mapping, 30*(9), 2907-2926. doi:10.1002/hbm.20718
Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15*(2), 399-402. doi:10.1046/j.0953-816x.2001.01874.x
Fertonani, A., Pirulli, C., & Miniussi, C. (2011). Random Noise Stimulation Improves Neuroplasticity in Perceptual Learning. *Journal of Neuroscience, 31*(43), 15416-15423. doi:10.1523/jneurosci.2002-11.2011
Feyereisen, P., Vandewiele, M., & Dubois, F. (1988). The Meaning of Gestures - What Can Be Understood without Speech. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition, 8*(1), 3-25.
Fiez, J. A., Raichle, M. E., Balota, D. A., Tallal, P., & Petersen, S. E. (1996). PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cerebral Cortex, 6*(1), 1-10. doi:10.1093/cercor/6.1.1

Floel, A., Ellger, T., Breitenstein, C., & Knecht, S. (2003). Language perception activates the hand motor cortex: implications for motor theories of speech perception. *European Journal of Neuroscience, 18*(3), 704-708. doi:10.1046/j.1460-9568.2003.02774.x

Frens, M. A., Vanopstal, A. J., & Vanderwilligen, R. F. (1995). SPATIAL AND TEMPORAL FACTORS DETERMINE AUDITORY-VISUAL INTERACTIONS IN HUMAN SACCADIC EYE-MOVEMENTS. *Perception & Psychophysics, 57*(6), 802-816. doi:10.3758/bf03206796

Fridriksson, J., Moser, D., Ryalls, J., Bonilha, L., Rorden, C., & Baylis, G. (2009). Modulation of Frontal Lobe Speech Areas Associated With the Production and Perception of Speech Movements. *Journal of Speech Language and Hearing Research, 52*(3), 812-819. doi:10.1044/1092-4388(2008/06-0197)

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84. doi:10.1016/s1364-6613(00)01839-8

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences, 13*(4), 175-181. doi:10.1016/j.tics.2009.01.001

Friederici, A. D., & Kotz, S. A. (2003a). The brain basis of syntactic processes: functional imaging and lesion studies. *NeuroImage, 20*, S8-S17. doi:10.1016/j.neuroimage.2003.09.003

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003b). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170-177. doi:10.1093/cercor/13.2.170

Frost, J. A., Binder, J. R., Springer, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Language processing is strongly left lateralized in both sexes - Evidence from functional MRI. *Brain, 122*, 199-208. doi:10.1093/brain/122.2.199

Fuhrmann Alpert, G., Hein, G., Tsai, N., Naumer, M. J., & Knight, R. T. (2008). Temporal characteristics of audiovisual information processing. *J Neurosci, 28*(20), 5344-5349. doi:10.1523/JNEUROSCI.5039-07.2008

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: A kinematic study on healthy subjects. *Journal of Neurophysiology, 86*(4), 1685-1699.

Gibson, J. K., Korn, N. J., Counsell, R. E., & Lucchesi, B. R. (1979). TISSUE DISTRIBUTION AND ANTI-ARRHYTHMIC ACTION OF N-DIMETHYLPROPRANOLOL (UM-272) IN THE DOG. *Federation Proceedings, 38*(3), 697-697.

Gold, B. T., Balota, D. A., Kirchhoff, B. A., & Buckner, R. L. (2005). Common and dissociable activation patterns associated with controlled semantic and phonological processing: Evidence from fMRI adaptation. *Cerebral Cortex, 15*(9), 1438-1450. doi:10.1093/cercor/bhi024

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*(11), 419-429. doi:10.1016/s1364-6613(99)01397-2

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science, 2*(1), 67-74. doi:10.1111/1467-7687.00056

Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience, 25*(35), 8010-8016. doi:10.1523/jneurosci.2307-05.2005

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography .2. Observation compared with imagination. *Experimental Brain Research, 112*(1), 103-111.

Grant, H. M., Bredahl, L. C., Clay, J., Ferrie, J., Groves, J. E., McDorman, T. A., & Dark, V. J. (1998). Context-dependent memory for meaningful material: Information for students. *Applied Cognitive Psychology, 12*(6), 617-623. doi:10.1002/(sici)1099-0720(1998120)12:6<617::aid-acp542>3.0.co;2-5

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108*(3), 1197-1208. doi:10.1121/1.1288668

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. doi:10.1002/hbm.20753

Gunther, F., Dudschig, C., & Kaup, B. (2015). LSAfun--An R package for computations based on Latent Semantic Analysis. *Behav Res Methods, 47*(4), 930-944. doi:10.3758/s13428-014-0529-0

Habets, B., Kita, S., Shao, Z. S., Ozyurek, A., & Hagoort, P. (2011). The Role of Synchrony and Ambiguity in Speech-Gesture Integration during Comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845-1854. doi:10.1162/jocn.2010.21462

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica, 115*(1-2), 147-172. doi:10.1515/semi.1997.115.1-2.147

Hadar, U., & Krauss, R. K. (1999). Iconic gestures: the grammatical categories of lexical affiliates. *Journal of Neurolinguistics, 12*(1), 1-12. doi:10.1016/s0911-6044(99)00001-9

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture - Implications for gesture classification and function. *Journal of Language and Social Psychology, 23*(2), 204-214. doi:10.1177/0261927x04263825

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. doi:10.1006/brln.1997.1890

Hagoort, P. (2004). How the brain solves the binding problem for language. *International Journal of Psychology, 39*(5-6), 349-349.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9*(9), 416-423. doi:10.1016/j.tics.2006.07.004

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1481), 801-811. doi:10.1098/rstb.2007.2089

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature, 406*(6792), 147-150. doi:10.1038/35018000

Heim, S., Opitz, B., & Friederici, A. D. (2003). Distributed cortical networks for syntax processing: Broca's area as the common denominator. *Brain and Language, 85*(3), 402-408. doi:10.1016/s0093-934x(03)00068-3

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci, 27*(30), 7881-7887. doi:10.1523/JNEUROSCI.1740-07.2007

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99. doi:10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). Opinion - The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402. doi:10.1038/nrn2113

Hogenraad, R., & Orianne, E. (1981). IMAGERY VALUES FOR 1,130 NOUNS FROM SPOKEN FRENCH. *Psychologica Belgica, 21*(1), 21-30.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*(7), 1175-1192. doi:10.1162/jocn.2007.19.7.1175

Holle, H., Gunter, T. C., Rueschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage, 39*(4), 2010-2024. doi:10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage, 49*(1), 875-884. doi:10.1016/j.neuroimage.2009.08.058

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849-+. doi:10.1017/s0140525x01000103
Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495-514. doi:10.3758/pbr.15.3.495
Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63*(2), 245-257. doi:10.1016/j.jml.2010.04.003
Hostetter, A. B., & Alibali, M. W. (2011). Cognitive skills and gesture-speech redundancy Formulation difficulty or communicative strategy? *Gesture, 11*(1), 40-60. doi:10.1075/gest.11.1.03hos
Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron, 45*(2), 201-206. doi:10.1016/j.neuron.2004.12.033
Husain, F. T., Patkin, D. J., Thai-Van, H., Braun, A. R., & Horwitz, B. (2009). Distinguishing the processing of gestures from signs in deaf individuals: An fMRI study. *Brain Research, 1276*, 140-150. doi:10.1016/j.brainres.2009.04.034
Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*(6708), 228-228. doi:10.1038/24300
Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language, 28*(2), 325-349.
Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40*(4), 577-592. doi:10.1006/jmla.1999.2634
Kelly, S. D., & Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*(1), 85-93. doi:10.2307/1132072
Kelly, S. D., Creigh, P., & Bartolotti, J. (2010a). Integrating Speech and Iconic Gestures in a Stroop-like Task: Evidence for Automatic Processing. *Journal of Cognitive Neuroscience, 22*(4), 683-694. doi:DOI 10.1162/jocn.2009.21254
Kelly, S. D., Iverson, J. M., Terranova, J., Niego, J., Hopkins, M., & Goldsmith, L. (2002). Putting language back in the body: Speech and gesture on three time frames. *Developmental Neuropsychology, 22*(1), 323-349. doi:10.1207/S15326942dn2201_1
Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253-260. doi:10.1016/s0093-934x(03)00335-3
Kelly, S. D., Ozyurek, A., & Maris, E. (2010b). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260-267. doi:10.1177/0956797609357327
Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language, 101*(3), 222-233. doi:10.1016/j.bandl.2006.07.008
Kendon, A. (1997). Gesture. *Annual Review of Anthropology, 26*, 109-128. doi:10.1146/annurev.anthro.26.1.109
Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162-185). Cambridge: Cambridge University Press.
Kita, S., & Ozyurek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*(1), 16-32. doi:10.1016/s0749-596x(02)00505-3
Kobayashi, M., & Pascual-Leone, A. (2003). Transcranial magnetic stimulation in neurology. *Lancet Neurology, 2*(3), 145-156. doi:10.1016/s1474-4422(03)00321-1

Koessler, L., Maillard, L., Benhadid, A., Vignal, J. P., Felblinger, J., Vespignani, H., & Braun, M. (2009). Automated cortical projection of EEG sensors: Anatomical correlation via the international 10-10 system. *NeuroImage, 46*(1), 64-72. doi:10.1016/j.neuroimage.2009.02.006

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences, 14*(9), 392-399. doi:10.1016/j.tics.2010.06.005

Krams, M., Rushworth, M. F. S., Deiber, M. P., Frackowiak, R. S. J., & Passingham, R. E. (1998). The preparation, execution and suppression of copied movements in the human brain. *Experimental Brain Research, 120*(3), 386-398. doi:10.1007/s002210050412

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*(2), 54-60. doi:10.1111/1467-8721.ep13175642

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), *Language and Gesture* (pp. 261-283). Cambridge: Cambridge University Press.

Krauss, R. M., Dushay, R. A., Chen, Y. S., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology, 31*(6), 533-552. doi:10.1006/jesp.1995.1024

Krauss, R. M., Morrelsamuels, P., & Colasante, C. (1991). Do Conversational Hand Gestures Communicate. *Journal of Personality and Social Psychology, 61*(5), 743-754. doi:Doi 10.1037/0022-3514.61.5.743

Krishnan, C., Santos, L., Peterson, M. D., & Ehinger, M. (2015). Safety of Noninvasive Brain Stimulation in Children and Adolescents. *Brain Stimulation, 8*(1), 76-87. doi:10.1016/j.brs.2014.10.012

Kuo, M. F., & Nitsche, M. A. (2012). Effects of Transcranial Electrical Stimulation on Cognition. *Clinical Eeg and Neuroscience, 43*(3), 192-199. doi:10.1177/1550059412444975

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470. doi:10.1016/s1364-6613(00)01560-6

Kutas, M., & Hillyard, S. A. (1980). READING SENSELESS SENTENCES - BRAIN POTENTIALS REFLECT SEMANTIC INCONGRUITY. *Science, 207*(4427), 203-205. doi:10.1126/science.7350657

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920-933. doi:10.1038/nrn2532

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human Brain Mapping, 19*(4), 213-223. doi:10.1002/hbm.10112

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23), 13464-13471. doi:10.1073/pnas.231459498

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). POINTING AND VOICING IN DEICTIC EXPRESSIONS. *Journal of Memory and Language, 24*(2), 133-164. doi:10.1016/0749-596x(85)90021-x

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187-196. doi:10.1016/s1364-6613(00)01471-6

Luber, B., & Lisanby, S. H. (2014). Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). *NeuroImage, 85*, 961-970. doi:10.1016/j.neuroimage.2013.06.007

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., . . . Mehler, J. (1993). THE CORTICAL REPRESENTATION OF SPEECH. *Journal of Cognitive Neuroscience, 5*(4), 467-479. doi:10.1162/jocn.1993.5.4.467

McNeill, D. (2008). *Gesture and Thought*: University of Chicago Press.

McNeill, D., Cassell, J., & McCullough, K.-E. (1994). *Communicative Effects of Speech-Mismatched Gestures* (Vol. 27).

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Lacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17*(19), 1692-1696. doi:10.1016/j.cub.2007.08.064

Miniussi, C., Harris, J. A., & Ruzzoli, M. (2013). Modelling non-invasive brain stimulation in cognitive neuroscience. *Neuroscience and Biobehavioral Reviews, 37*(8), 1702-1712. doi:10.1016/j.neubiorev.2013.06.014

Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. *Journal of Speech Language and Hearing Research, 48*(4), 884-896. doi:10.1044/1092-4388(2005/061)

Moliadze, V., Atalay, D., Antal, A., & Paulus, W. (2012). Close to threshold transcranial electrical stimulation preferentially activates inhibitory networks before switching to excitation with higher intensities. *Brain Stimulation, 5*(4), 505-511. doi:10.1016/j.brs.2011.11.004

Morrelsamuels, P., & Krauss, R. (1991). GESTURES PRECEDE SPEECH - A TEMPORAL INTERVAL PREDICTED BY WORD FAMILIARITY. *Bulletin of the Psychonomic Society, 29*(6), 506-506.

Morrelsamuels, P., & Krauss, R. M. (1992). WORD FAMILIARITY PREDICTS TEMPORAL ASYNCHRONY OF HAND GESTURES AND SPEECH. *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(3), 615-622. doi:10.1037/0278-7393.18.3.615

Naumer, M. J., Doehrmann, O., Mueller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical Plasticity of Audio-Visual Object Representations. *Cerebral Cortex, 19*(7), 1641-1653. doi:10.1093/cercor/bhn200

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., . . . Pascual-Leone, A. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation, 1*(3), 206-223. doi:10.1016/j.brs.2008.06.004

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *Journal of Physiology-London, 527*(3), 633-639. doi:10.1111/j.1469-7793.2000.t01-1-00633.x

Nitsche, M. A., Seeber, A., Frommann, K., Mein, C. C., Rochford, C., Nitsche, M. S., . . . Tergau, F. (2005). Modulating parameters of excitability during and after transcranial direct current stimulation of the human motor cortex. *Journal of Physiology-London, 568*(1), 291-303. doi:10.1113/jphysiol.2005.092429

Nyffeler, T., Cazzoli, D., Hess, C. W., & Muri, R. M. (2009). One session of repeated parietal theta burst stimulation trains induces long-lasting improvement of visual neglect. *Stroke, 40*(8), 2791-2796. doi:10.1161/STROKEAHA.109.552323

Nyffeler, T., Cazzoli, D., Wurtz, P., Luthi, M., von Wartburg, R., Chaves, S., . . . Muri, R. M. (2008). Neglect-like visual exploration behaviour after theta burst transcranial magnetic stimulation of the right posterior parietal cortex. *Eur J Neurosci, 27*(7), 1809-1813. doi:10.1111/j.1460-9568.2008.06154.x

Nyffeler, T., Wurtz, P., Luscher, H. R., Hess, C. W., Senn, W., Pflugshaupt, T., . . . Muri, R. M. (2006). Repetitive TMS over the human oculomotor cortex: Comparison of 1-Hz and theta burst stimulation. *Neuroscience Letters, 409*(1), 57-60. doi:10.1016/j.neulet.2006.09.011

Obermeier, C., & Gunter, T. C. (2015). Multisensory Integration: The Case of a Time Window of Gesture-Speech Integration. *Journal of Cognitive Neuroscience, 27*(2), 292-307. doi:10.1162/jocn_a_00688

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What Iconic Gesture Fragments Reveal about Gesture-Speech Integration: When Synchrony Is Lost, Memory Can Help. *Journal of Cognitive Neuroscience, 23*(7), 1648-1663. doi:10.1162/jocn.2010.21498

Ojemann, G. A. (1991). CORTICAL ORGANIZATION OF LANGUAGE. *Journal of Neuroscience, 11*(8), 2281-2287.

Oldfield, R. C. (1971). THE ASSESSMENT AND ANALYSIS OF HANDEDNESS: THE EDINBURGH INVENTORY. *Neuropsychologia, 9*(1), 97-113. doi:10.1016/0028-3932(71)90067-4

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology, 112*(4), 713-719. doi:10.1016/s1388-2457(00)00527-7

Ozyurek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369*(1651), 20130296. doi:10.1098/rstb.2013.0296

Ozyurek, 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. *J Cogn Neurosci, 19*(4), 605-616. doi:10.1162/jocn.2007.19.4.605

Parkin, B. L., Ekhtiari, H., & Walsh, V. F. (2015). Non-invasive Human Brain Stimulation in Cognitive Neuroscience: A Primer. *Neuron, 87*(5), 932-945. doi:10.1016/j.neuron.2015.07.032

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., . . . Lancaster, J. L. (1995). USE OF IMPLICIT MOTOR IMAGERY FOR VISUAL SHAPE-DISCRIMINATION AS REVEALED BY PET. *Nature, 375*(6526), 54-58. doi:10.1038/375054a0

Pascual-Leone, A., Walsh, V., & Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience - virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology, 10*(2), 232-237. doi:10.1016/s0959-4388(00)00081-7

Pascualleone, A., Vallssole, J., Wassermann, E. M., & Hallett, M. (1994). RESPONSES TO RAPID-RATE TRANSCRANIAL MAGNETIC STIMULATION OF THE HUMAN MOTOR CORTEX. *Brain, 117*, 847-858. doi:10.1093/brain/117.4.847

Peleg, O., Edelist, L., Eviatar, Z., & Bergerbest, D. (2016). Lexical factors in conceptual processes: The relationship between semantic representations and their corresponding phonological and orthographic lexical forms. *Memory & Cognition, 44*(4), 519-537. doi:10.3758/s13421-015-0576-5

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). AUDITORY PSYCHOMOTOR COORDINATION AND VISUAL-SEARCH PERFORMANCE. *Perception & Psychophysics, 48*(3), 214-226. doi:10.3758/bf03211521

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). POSITRON EMISSION TOMOGRAPHIC STUDIES OF THE CORTICAL ANATOMY OF SINGLE-WORD PROCESSING. *Nature, 331*(6157), 585-589. doi:10.1038/331585a0

Pine, K. J., Reeves, L., Howlett, N., & Fletcher, B. (2013). Giving cognition a helping hand: The effect of congruent gestures on object name retrieval. *British Journal of Psychology, 104*(1), 57-68. doi:10.1111/j.2044-8295.2011.02098.x

Pulvermuller, F., Huss, M., Kherif, F., Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(20), 7865-7870. doi:10.1073/pnas.0509989103

Rafal, R. (2001). Virtual neurology. *Nature Neuroscience, 4*(9), 862-864. doi:10.1038/nn0901-862

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107*(6), 1556-1563. doi:10.1152/jn.00985.2011

Rauscher, F. H., Krauss, R. M., & Chen, Y. S. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*(4), 226-231. doi:10.1111/j.1467-9280.1996.tb00364.x

Rime, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). EFFECTS OF RELATIVE IMMOBILIZATION ON THE SPEAKERS NONVERBAL BEHAVIOR AND ON THE DIALOG IMAGERY LEVEL. *Motivation and Emotion, 8*(4), 311-325. doi:10.1007/bf00991870

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fabbri-Destro, M., & Cattaneo, L. (2009). Mirror neurons and their clinical relevance. *Nature Clinical Practice Neurology, 5*(1), 24-34. doi:10.1038/ncpneuro0990
Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in humans by PET .1. Observation versus execution. *Experimental Brain Research, 111*(2), 246-252.
Robertson, E. M., Theoret, H., & Pascual-Leone, A. (2003). Studies in cognition: The problems solved and created by transcranial magnetic stimulation. *Journal of Cognitive Neuroscience, 15*(7), 948-960. doi:10.1162/089892903770007344
Roder, B., Stock, O., Neville, H., Bien, S., & Rosler, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance Imaging study. *NeuroImage, 15*(4), 1003-1014. doi:10.1006/nimg.2001.1026
Ruzzoli, M., Marzi, C. A., & Miniussi, C. (2010). The Neural Mechanisms of the Effects of Transcranial Magnetic Stimulation on Perception. *Journal of Neurophysiology, 103*(6), 2982-2989. doi:10.1152/jn.01096.2009
Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language, 111*(1), 1-7. doi:10.1016/j.bandl.2009.03.002
Schulke, R., & Straube, B. (2017). Modulating the assessment of semantic speech-gesture relatedness via transcranial direct current stimulation of the left frontal cortex. *Brain Stimulation, 10*(2), 223-230. doi:10.1016/j.brs.2016.10.012
Schwarzkopf, D. S., Silvanto, J., & Rees, G. (2011). Stochastic Resonance Effects Reveal the Neural Mechanisms of Transcranial Magnetic Stimulation. *Journal of Neuroscience, 31*(9), 3143-3147. doi:10.1523/jneurosci.4863-10.2011
Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRl and. PET. *Neuroscience Research, 47*(3), 277-287. doi:10.1016/S0168-0102(03)00214-1
Slamacazacu, T. (1993). NONVERBAL COMPONENTS OF ORAL COMMUNICATION AND THE CONCEPT OF SO-CALLED MIXED-SYNTAX - A SYNTHESIS. *Degres-Revue De Synthese a Orientation Semiologique*(74), E1-E24.
Small, S. L., Noll, D. C., Perfetti, C. A., Hlustik, P., Wellington, R., & Schneider, W. (1996). Localizing the lexicon for reading aloud: Replication of a PET study using fMRI. *Neuroreport, 7*(4), 961-965. doi:10.1097/00001756-199603220-00027
Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America, 102*(29), 10393-10398. doi:10.1073/pnas.0503903102
Straube, B., Green, A., Bromberger, B., & Kircher, T. (2011). The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum Brain Mapp, 32*(4), 520-533. doi:10.1002/hbm.21041
Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A Supramodal Neural Network for Speech and Gesture Semantics: An fMRI Study. *Plos One, 7*(11). doi:10.1371/journal.pone.0051207
Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661-699. doi:10.1111/j.0023-8333.2005.00320.x
Summerfield, Q. (1992). LIPREADING AND AUDIOVISUAL SPEECH-PERCEPTION. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 335*(1273), 71-78. doi:10.1098/rstb.1992.0009
T. Dittmann, A., & G. Llewellyn, L. (1969). *Body Movement and Speech Rhythm in Social Conversation* (Vol. 11).
Terney, D., Chaieb, L., Moliadze, V., Antal, A., & Paulus, W. (2008). Increasing Human Brain Excitability by Transcranial High-Frequency Random Noise Stimulation. *Journal of Neuroscience, 28*(52), 14147-14155. doi:10.1523/jneurosci.4248-08.2008

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of Speaker and Meaning in Language Comprehension: An fMRI Study. *Journal of Cognitive Neuroscience, 21*(11), 2085-2099. doi:DOI 10.1162/jocn.2008.21161

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America, 94*(26), 14792-14797. doi:10.1073/pnas.94.26.14792

Umilta, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., . . . Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences of the United States of America, 105*(6), 2209-2213. doi:10.1073/pnas.0705985105

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28*(2), 187-204. doi:10.1016/s0361-476x(02)00007-3

Vallar, G., & Bolognini, N. (2011). Behavioural facilitation following brain stimulation: Implications for neurorehabilitation. *Neuropsychological Rehabilitation, 21*(5), 618-649. doi:10.1080/09602011.2011.574050

Viviani, P., & Stucchi, N. (1992). BIOLOGICAL MOVEMENTS LOOK UNIFORM - EVIDENCE OF MOTOR-PERCEPTUAL INTERACTIONS. *Journal of Experimental Psychology-Human Perception and Performance, 18*(3), 603-623. doi:10.1037//0096-1523.18.3.603

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*(1), 73-79. doi:10.1038/35036239

Walsh, V., & Rushworth, M. (1999). A primer of magnetic stimulation as a tool for neuropsychology. *Neuropsychologia, 37*(2), 125-135.

Warburton, E., Wise, R. J. S., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., & Frackowiak, R. S. J. (1996). Noun and verb retrieval by normal subjects studies with PET. *Brain, 119*, 159-179. doi:10.1093/brain/119.1.159

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology, 114*(4), 591-600. doi:10.2307/1423612

Whitney, C., Kirk, M., O'Sullivan, J., Ralph, M. A. L., & Jefferies, E. (2011). The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex, 21*(5), 1066-1075. doi:10.1093/cercor/bhq180

Willems, R. M., Ozyurek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex, 17*(10), 2322-2333. doi:10.1093/cercor/bhl141

Willems, R. M., Ozyurek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage, 47*(4), 1992-2004. doi:DOI 10.1016/j.neuroimage.2009.05.066

Wilshire, C. E., & McCarthy, R. A. (1996). Experimental investigations of an impairment in phonological encoding. *Cognitive Neuropsychology, 13*(7), 1059-1098.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), 701-702. doi:10.1038/nn1263

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). DISTRIBUTION OF CORTICAL NEURAL NETWORKS INVOLVED IN WORD COMPREHENSION AND WORD RETRIEVAL. *Brain, 114*, 1803-1817. doi:10.1093/brain/114.4.1803

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., . . . Nitsche, M. A. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. *Clinical Neurophysiology, 127*(2), 1031-1048. doi:10.1016/j.clinph.2015.11.012

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654-667. doi:10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*(3), 234-245. doi:10.1016/j.bandl.2006.12.003

Wu, Y. C., & Coulson, S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain and Language, 119*(3), 184-195. doi:10.1016/j.bandl.2011.07.002

Xu, B., Grafman, J., Gaillard, W. D., Ishii, K., Vega-Bermudez, F., Pietrini, P., . . . Theodore, W. (2001). Conjoint and extended neural networks for the computation of speech codes: The neural basis of selective impairment in reading words and pseudowords. *Cerebral Cortex, 11*(3), 267-277. doi:10.1093/cercor/11.3.267

Xua, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences of the United States of America, 106*(49), 20664-20669. doi:10.1073/pnas.0909197106

Yap, D. F., So, W. C., Yap, J. M. M., Tan, Y. Q., & Teoh, R. L. S. (2011). Iconic Gestures Prime Words. *Cognitive Science, 35*(1), 171-183. doi:10.1111/j.1551-6709.2010.01141.x

Zaehle, T., Rach, S., & Herrmann, C. S. (2010). Transcranial Alternating Current Stimulation Enhances Individual Alpha Activity in Human EEG. *Plos One, 5*(11). doi:10.1371/journal.pone.0013766

Zhuang, J., Randall, B., Stamatakis, E. A., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The Interaction of Lexical Semantics and Cohort Competition in Spoken Word Recognition: An fMRI Study. *Journal of Cognitive Neuroscience, 23*(12), 3778-3790.

Appendixs

Appendix A Experimental stimulus

pair	video	sound-c	sound-inc	sound-c-f	sound-c-m	sound-i-f	sound-i-m
1	break.avi	break.wav	dial.wav	f-break.wav	m-break.wav	f-dial.wav	m-dial.wav
1	dial.avi	dial.wav	break.wav	f-dial.wav	m-dial.wav	f-break.wav	m-break.wav
3	close.avi	close.wav	unscrew.wav	f-close.wav	m-close.wav	f-unscrew.wav	m-unscrew.wav
3	spread.avi	spread.wav	punch.wav	f-spread.wav	m-spread.wav	f-punch.wav	m-punch.wav
4	cut.avi	cut.wav	stir.wav	f-cut.wav	m-cut.wav	f-stir.wav	m-stir.wav
4	stir.avi	stir.wav	cut.wav	f-stir.wav	m-stir.wav	f-cut.wav	m-cut.wav
5	flip.avi	flip.wav	sew.wav	f-flip.wav	m-flip.wav	f-sew.wav	m-sew.wav
5	sew.avi	sew.wav	flip.wav	f-sew.wav	m-sew.wav	f-flip.wav	m-flip.wav
6	hammer.avi	hammer.wav	spray.wav	f-hammer.wav	m-hammer.wav	f-spray.wav	m-spray.wav
6	spray.avi	spray.wav	hammer.wav	f-spray.wav	m-spray.wav	f-hammer.wav	m-hammer.wav
7	iron.avi	iron.wav	whisk.wav	f-iron.wav	m-iron.wav	f-whisk.wav	m-whisk.wav
7	whisk.avi	whisk.wav	iron.wav	f-whisk.wav	m-whisk.wav	f-iron.wav	m-iron.wav
8	knock.avi	knock.wav	write.wav	f-knock.wav	m-knock.wav	f-write.wav	m-write.wav
8	write.avi	write.wav	knock.wav	f-write.wav	m-write.wav	f-knock.wav	m-knock.wav
10	light.avi	light.wav	steer.wav	f-light.wav	m-light.wav	f-steer.wav	m-steer.wav
10	steer.avi	steer.wav	light.wav	f-steer.wav	m-steer.wav	f-light.wav	m-light.wav
12	open.avi	open.wav	wash.wav	f-open.wav	m-open.wav	f-wash.wav	m-wash.wav
12	wash.avi	wash.wav	open.wav	f-wash.wav	m-wash.wav	f-open.wav	m-open.wav
13	peel.avi	peel.wav	throw.wav	f-peel.wav	m-peel.wav	f-throw.wav	m-throw.wav
13	throw.avi	throw.wav	peel.wav	f-throw.wav	m-throw.wav	f-peel.wav	m-peel.wav
14	saw.avi	saw.wav	type.wav	f-saw.wav	m-saw.wav	f-type.wav	m-type.wav
14	type.avi	type.wav	saw.wav	f-type.wav	m-type.wav	f-saw.wav	m-saw.wav
15	pull.avi	pull.wav	scrub.wav	f-pull.wav	m-pull.wav	f-scrub.wav	m-scrub.wav
15	scrub.avi	scrub.wav	pull.wav	f-scrub.wav	m-scrub.wav	f-pull.wav	m-pull.wav
17	vacuum.avi	vacuum.wav	weigh.wav	f-vacuum.wav	m-vacuum.wav	f-weigh.wav	m-weigh.wav
17	weigh.avi	weigh.wav	vacuum.wav	f-weigh.wav	m-weigh.wav	f-vacuum.wav	m-vacuum.wav
18	swipe.avi	swipe.wav	zip.wav	f-swipe.wav	m-swipe.wav	f-zip.wav	m-zip.wav
18	zip.avi	zip.wav	swipe.wav	f-zip.wav	m-zip.wav	f-swipe.wav	m-swipe.wav
19	shake.avi	shake.wav	turn.wav	f-shake.wav	m-shake.wav	f-turn.wav	m-turn.wav
19	turn.avi	turn.wav	shake.wav	f-turn.wav	m-turn.wav	f-shake.wav	m-shake.wav
20	sharpen.avi	sharpen.wav	wipe.wav	f-sharpen.wav	m-sharpen.wav	f-wipe.wav	m-wipe.wav
20	wipe.avi	wipe.wav	sharpen.wav	f-wipe.wav	m-wipe.wav	f-sharpen.wav	m-sharpen.wav

Appendix B Duration of the videos and audios used as experimental stimulus

Audio_F	ms	Video_F	ms	Audio_M	ms	Video_M	ms	Audio_Mean	Video_Mean
f-break	370	f-break	1520	m-break	410	m-break	1680	390	1600
f-close	604	f-close	1840	m-close	577	m-close	1960	591	1900
f-cut	472	f-cut	1320	m-cut	439	m-cut	2000	456	1660
f-dial	478	f-dial	1920	m-dial	501	m-dial	2120	490	2020
f-flip	452	f-flip	1400	f-flip	510	f-flip	2040	481	1720
f-hammer	393	f-hammer	1520	m-hammar	461	m-hammar	1720	427	1620
f-iron	530	f-iron	2040	m-iron	491	m-iron	2040	511	2040
f-knock	510	f-knock	1480	m-knock	475	m-knock	1200	493	1340
f-light	603	f-light	2120	m-light	555	m-light	1720	579	1920
f-open	563	f-open	1520	m-open	529	m-open	2240	546	1880
f-peel	488	f-peel	2040	m-peel	539	m-peel	3120	514	2580
f-pull	375	f-pull	1160	m-pull	464	m-pull	1560	420	1360
f-saw	570	f-saw	1600	m-saw	623	m-saw	1880	597	1740
f-scrub	557	f-scrub	1800	m-scrub	624	m-scrub	1680	591	1740
f-sew	535	f-sew	1880	m-sew	633	m-sew	2280	584	2080
f-shake	702	f-shake	2120	m-shake	629	m-shake	1640	666	1880
f-sharpen	687	f-sharpen	1320	m-sharpen	754	m-sharpen	2160	721	1740
f-spary	628	f-spary	1480	m-spary	799	m-spary	2000	714	1740
f-spread	652	f-spread	1600	m-spread	795	m-spread	2240	724	1920
f-steer	587	f-steer	1440	m-steer	693	m-steer	2200	640	1820
f-stir	626	f-stir	1600	m-stir	779	m-stir	2160	703	1880
f-swipe	607	f-swipe	1680	m-swipe	757	m-swipe	2320	682	2000
f-throw	653	f-throw	1120	m-throw	595	m-throw	1680	624	1400
f-turn	610	f-turn	1360	m-turn	608	m-turn	1920	609	1640
f-type	824	f-type	1600	m-type	464	m-type	1840	644	1720
f-vacuume	867	f-vacuume	1640	m-vacuume	768	m-vacuume	2720	818	2180
f-wash	617	f-wash	1680	m-wash	634	m-wash	2720	626	2200
f-weight	485	f-weight	1680	m-weigh	581	m-weigh	2560	533	2120
f-whisk	618	f-whisk	1800	m-whisk	497	m-whisk	2160	558	1980
f-wipe	574	f-wipe	1560	m-wipe	632	m-wipe	2320	603	1940
f-write	681	f-write	1560	m-write	633	m-write	1640	657	1600
f-zip	681	f-zip	1480	m-zip	545	m-zip	1920	613	1700
Mean±SD(ms)	581±113		1621±255		594±112		2045±389	588±99	1833±259

Appendix C Practicing experimental stimulus

pair	video	sound-c	sound-inc	sound-c-f	sound-c-m	sound-i-f	sound-i-m
22	squeeze.avi	squeeze.wav	wring.wav	f-squeeze.wav	m-squeeze.wav	f-wring.wav	m-wring.wav
22	wring.avi	wring.wav	squeeze.wav	f-wring.wav	m-wring.wav	f-squeeze.wav	m-squeeze.wav

Appendix D Deleted experimental stimulus

pair	video	sound-c	sound-inc	sound-c-f	sound-c-m	sound-i-f	sound-i-m
2	chop.avi	chop.wav	mow.wav	f-chop.wav	m-chop.wav	f-mow.wav	m-mow.wav
2	mow.avi	mow.wav	chop.wav	f-mow.wav	m-mow.wav	f-chop.wav	m-chop.wav
9	lift.avi	lift.wav	stab.wav	f-lift.wav	m-lift.wav	f-stab.wav	m-stab.wav
9	stab.avi	stab.wav	lift.wav	f-stab.wav	m-stab.wav	f-lift.wav	m-lift.wav
11	slap.avi	slap.wav	lock.wav	f-slap.wav	m-slap.wav	f-lock.wav	m-lock.wav
11	unscrew.avi	unscrew.wav	close.wav	f-unscrew.wav	m-unscrew.wav	f-close.wav	m-close.wav
17	sweep.avi	sweep.wav	tear.wav	f-sweep.wav	m-sweep.wav	f-tear.wav	m-tear.wav
17	tear.avi	tear.wav	sweep.wav	f-tear.wav	m-tear.wav	f-sweep.wav	m-sweep.wav

Appendix E Brain coordinates

Holle et al. 2010, table1, GSgood∩Gsmod, n=16,MNI

51	-33	3	rMTG/STG
-48	-39	3	LMTG

Green et al. 2009,table1, German unrelated>German iconic,n=16,MNI

-60	16	24	LIFG
-60	-60	-4	LMTG
-44	-36	40	l inferior parietal lobule
48	32	12	r inferior frontal gyrus
52	-36	36	r supramarginal gyrus
8	20	56	r/l vsupplementary motor area
44	12	48	r middle frontal gyrus
8	-36	56	r paracental lobule
-20	-12	-16	I hippocampus

Straube et al. 2010, table3,(IC>S∩IC>GI)∩G∩S,n=17.MNI

-56	-52	12	LMTG

56	-36	12	RSTG

Willems et al. 2006,table3 Gmismatch>Gcorrect(global), n=16, Talairach

-46.46	28.67	26.54	LIFG
-32.32	-48.96	31.15	left intraparietal sulcus(anterior)
-19.2	-66.52	31.32	left intraparietal sulcus(posterior)

Willems et al. 2006,table3 Lmismatch>Lcorrect(global), n=16, Talairach

-43	12	24	l inferior frontal sulcus
-33	-65	35	l intraparietal sulcus(posterior)
-52	-50	4	l superior temporal sulcus

Willems et al. 2009,table4, 0<G-video<G-bimodal>G-audio>0, n=20, MNI

-52	-56	6	LpSTS
58	-36	18	RpSTS
-46	10	24	LIFG
-46	22	24	LIFG
14	-92	-2	r inferior occipital sulcus
4	-92	-2	r inferior occipital sulcus
-24	-76	-10	l inferior occipital sulcus

Willems et al. 2009, table4, 0<pant-video<pant-bimodal>pant-audio>0, n=20, MNI

-48	-54	8	l pSTS
62	-36	18	r pSTS
58	-52	10	r pSTS
-40	10	26	l IFG

Holle et al. 2008, table2, Dominant>Grooming, n=17, MNI

-8	45	36	left medial middle frontal gyrus
49	3	36	right precentral sulcus
-47	3	33	left precentral sulcus

			right inferior parietal
58	-36	30	lobule
-59	-36	33	left inferior parietal lobule
34	-39	42	right fusiform gyrus
-50	-54	15	left posterior STS
			lateral middle occipital
-38	-75	24	gyrus
37	-48	-6	right fusiform gyrus
-8	-96	9	left medial middle occipital gyrus

Holle et al. 2008, table2, Subordinate>Grooming, n=17,MNI

-47	6	27	left precentral sulcus
55	-24	39	r inferior parietal lobule
-56	-36	33	left inferior parietal lobule
43	0	27	r precentral sulcus
-41	-51	-9	l fusiform gyrus
-53	-72	12	occipito-temporal junction
40	-48	-9	right fusiform gyrus
-47	-57	12	posterior STS

Dick et al. 2012,table1,(G+nonspecific L(supplemental)>G+specific L(redundant))>(nons L>s L), n=17,Talairach

-54	-59	1	left MTG
-14	-95	-12	l Lingual gyrus
-53	20	18	LIFG