An Investigation on the Cognitive Effects of Emoji Usage in Text

Ousterhout, Thomas Kenneth

Publication date:
2017

Document version
Other version

An Investigation on the Cognitive Effects of Emoji Usage in Text

Thomas Ousterhout

Centre for Language Technology, Department of Nordic Research

University of Copenhagen

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

2017

ABSTRACT

Face-to-face communication is multimodal involving at least the auditory (speech) and the visual (gestures such as head movements, facial expressions and hand gestures) modalities. While multimodal signals are produced naturally in face-to-face communication, they are not so easily provided in written computer-mediated communication, and especially in instant messaging. The visual nonverbal cues are not available and there is a great potential for miscommunication. The growing use of emojis, pictures or short videos of facial expressions and symbols of various types, are a means to replace non-verbal cues.

Preceding studies have shown that emojis contribute to the semantics of the message, but their effect on reading and their potential uses as e.g. reading aids, are not thoroughly studied. The purpose of this dissertation is to investigate the cognitive and behavioral effects of emojis in text in order to determine how these stimuli complement the written text in a way that facilitates reading ability and/or comprehension. This is done through experiments aimed to measure cognitive and behavioral responses to visual presentations of semantically meaningful emojis in isolation and in conjunction with text. The emojis represent hand gestures and facial expressions. The method comprised electroencephalography, button press response times and accuracy. The main results of the experiments are the following. A commercial grade EEG equipment can be used to detect the N400 ERP in a natural environment. Simple words and emojis produce semantic priming despite being different channels within the same modality and the behavioral results corroborate with how we behave in face-to-face communication. Merging words with emojis also produced semantic congruity effects. When mixing emojis into sentences, the results of the experiments show that if the sentences are kept simple, they produce semantic integration. Integration of emojis and text worked best when hand gesture pictures were used instead of a picture of a man producing the same gesture. Furthermore, comparing emoji placement and emoji type in short sentences I found that that users had no problem integrating emojis in several positions of a sentence, but responded best to emojis of facial expressions rather than hand gestures, and to picture emojis rather than video emojis. The video emojis seemed to distract the readers from the main comprehension task, as the full man picture did in a preceding experiment. Therefore, salience and relevance of the emojis are also important elements when they are integrated into text.

RESUMÉ

Ansigt-til-ansigt kommunikation er multimodal, da mennesker både bruger tale (auditiv modalitet) og bevægelser (visuel modalitet), når de kommunikerer med hinanden. Disse multimodale signaler anvendes helt naturligt i ansigt-til-ansigt kommunikation, men de kan ikke umiddelbart udtrykkes i korte computer-medierede tekster. Dette kan medføre miskommunikation. Derfor, er brugere begyndt at anvende så-kaldte *emojis*, d.v.s. billeder og korte videoer af ansigtsudtryk og andre symboler, som supplement til kort tekst. Effekten af emojis på læseren er kun i mindre omfang blevet adresseret i forskning, selv om deres brug er stærkt voksende.

Formålet med denne afhandling er at undersøge, hvordan læsere reagerer påmeningsfulde emojis. Mere specifikt, indeholder afhandlingen adskillige eksperimenter, der har til formål at måle kognitive og adfærdsmæssige reaktioner på visuelle præsentationer af semantisk meningsfulde emojis tilføjet kort tekst. Disse emojis forestiller en gestus eller et ansigtsudtryk.

Læsernes reaktion til tekst som indeholder emojis er blevet målt via et kommercielt EEG-udstyr, og læsernes reaktionstid og præcision er også beregnet. Det første eksperiment demonstrerer at et kommercielt EEG-udstyr kan anvendes til at identificere N400 ERP i et simpelt kontor-miljø, og at semantisk priming er opnået med kongruente billeder og tekst. Et andet eksperiment indikerer, at emojis integreres bedst i teksten, hvis de repræsenterer et håndgestus end hvis billedet af en mand der udfører det samme håndgestus anvendes. Det menneskelige billede var dominerende, og virkede distraherende. Andre eksperimenter viser, at meningsfulde emojis bliver integreret i tekster uafhængigt af om de indeholder ny information i forhold til teksten eller om de forstærker dens indhold, og at responstid afhænger af forholdet mellem emoji og indhold og emojis placering i teksten. Endelig vises det at forskellige placeringer af emojis i teksten ikke påvirker reaktionstiden og korrekthed af resultater, og at responstiden var højere for ansigtsudtryks-emojis end for håndgestus-emojis. Disse resultater er et første skridt mod klargøring om hvordan forskellige typer emojis kan bruges i tekster fx for at hjælpe svage læsere, eller for at forbedre menneske-maskine kommunikation.

Acknowledgment

To my supervisor, Costanza Navarretta, for being the biggest supporter of my academic career and being the reason all of this worked out.

To my fellow PhD students Sigrid Klerke, Bjørn Nicola Wessel-Tolvig, and Maria Barrett for being there along the way.

To all of CST for having me as part of the team.

To everyone who volunteered their time to participate in my experiments.

To the Department of Psychology's Center for Visual Cognition at the University of Copenhagen for introducing me to EEG.

To Davis Morris for helping me build my first EEG lab.

To Daniel Spikol and his students at Malmö University and in particular Marcus Johansson for helping me with my survey.

Contents

List of Figures

List of Tables

1

Introduction

The purpose of this dissertation is to investigate the cognitive and behavioral effects of using pictures and moving pictures in text in order to determine how these stimuli complement the written text in a way that facilitates reading ability and/or comprehension. In this dissertation, the term "emoji" refers to all pictures and moving pictures which represent hand gestures, facial expressions and actions performed by them. Emojis can add information to the text or reinforce existing information. However, they can also be ambiguous or be interpreted in different ways given a certain context, and therefore might complicate or confuse comprehension. Furthermore variables such as culture, age, experience with using emojis as well as not optimal placement of emojis in the text can influence the way people process them.

Human-computer interaction and computer-mediated communication (which here comprises of any type of written communication through a digital device) are becoming increasingly prominent. Because of this, computer interfaces (e.g. button press, speech recognition and eye tracking), as well as communicative devices (e.g. speech recording and emojis) are implemented to facilitate the commutative ability of the human to interact with the computer efficiently or express him- or herself through the computer to the recipient human properly. It is equally important for the human-computer interface to provide messages by the computer to the human as naturally and intuitively as humans do. Thus, the goal of the present work is to determine how humans process text and meaningful emojis as a first step through understanding the effect on readers of the new writing styles which are developing as a consequence of the technological development. There is also a long term goal of being able to use emojis as didactic instruments for helping subjects with reading difficulties or as a means to be used in e.g. affective human-computer interaction.

Face-to-face communication is multimodal, meaning it involves many modalities, with the auditory and visual modalities being the most dominant ones. In the dissertation, when I speak about a modality with different channels within that modality I use the term *form*. For example, while vision is one modality, text and pictures are two forms of that one modality. Speech is typically the dominant information modality in face-to-face communication. However, visual information provided by body movements, facial expressions, and hand gestures, are quite important for fluent self-expression and comprehension of the interlocutors. Emojis might provide some of the same information to written text. The investigation of emojis is particularly relevant today where quickly typed messages via mobile phones or other devices in communities such as social media are replacing other forms of less impulsive communication such as readers' letters, opinion letters or auditory communication such as telephone calls. Short text messages lack social cues coming from body. The addition of pictorial information in written text can facilitate the interaction efficiency level by simplifying the reading requirement, making the reading speed faster, or by providing additional information to the content of the written text. Therefore,the use and perception of emojis needs further investigation.

The emojis could encompass any physical act of the human body that provides semantic information in face-to-face communication. An emoji that adds this type of semantic meaning instantaneously is ideal for this investigation because it will be what most closely resembles body behavior in face-to-face communication. Therefore, the use of common emojis such as smiley faces and sad faces, which can instantaneously provide semantic meaning of happiness or sadness, as well as culturally conventionalized hand postures such as the "thumbs up" or "thumbs down" gesture, which can instantaneously provide semantic meaning of positivity or negativity, will be investigated. Not only is the immediacy of the meaning of these emojis an advantage, but their natural usage in computer-mediated communication adds a level of intuitiveness. However, other emojis are also investigated to a lesser extent.

It's important that the emojis used in this investigation resemble signals used in face-to-face communication to the best of their ability. Thus far the research on emojis has primarily focused on usage frequency or on their contribution to the study of sentiment analysis, but not many investigations have been conducted on how to use them to improve comprehension. This dissertation addresses cognitively healthy readers, but future investigations should address whether emojis could improve text comprehension for people with reading difficulties.

I follow an evolutionary psychology perspective which argues that the origins of spoken language came from an earlier existing use of nonverbal communication (i.e. body language). This cognitive seed evolved into spoken language. Therefore, multimodal signals are a natural component in face-to-face communication and are missing in quick text messaging where there is no time to express nuances in meaning and affective states as in traditional written texts such as letters, essays, and novels.

The first hypothesis tested in this dissertation is that emojis which reflect the use of body in speech will be useful for readers' processing ability while emojis that are just added to text in an ambiguous or incongruent way will decrease comprehension and increase cognitive load. But not only the type of emoji is important, its placement in the message can also be relevant. Typically people place emojis at the end of sentences, but this thesis will also investigate different placements that might be useful for readers since people are starting to use emojis in all positions inside sentences. Therefore, the second hypothesis which we want to investigate is whether there will be differences in the benefits that emojis provide readers when they are placed in different locations.

Many people who converse digitally use emojis now, but are emojis just a stylistic supplement to text or are they adding useful information? Can they be used to help people in comprehending text better? If so, does the placement of emojis in sentences effect their perception? A number of experiments that address these questions are presented in the dissertation.

The method used in this work combines analyses of both behavioral and physiological responses to written and pictorial stimuli. Behavioral responses can be measured both by reaction time latency as well as accuracy of correct responses. Physiological responses are measured using electrophysiological (EEG) activity in the brain and event-relate potentials (ERPs). These ERPs are positive and negative going deflections in EEG activity in different regions of the brain at different time periods related to the onset of specific events.

The dissertation is organized as follows. First in Chapter 2, basic gesture theory, gesture types, and their functions are discussed. Chapter 3 covers modern trends of how emotions are expressed visually in text, which gives evidence through background literature for the feasibility of investigating the integration of emojis in sentences through the scientific method. Then, Chapter 4 discusses mirroring theories which provide biological evidence for the importance of gestures in communication supporting the need for the inclusion of emojis in text. After this, Chapter 5 provides a short introduction into EEG/ERPs, what they are, and how they work. Next, Chapter 6 provides an overview of how responses to using pictures of different gesture types can be measured neurophysiologically. Then, Chapter 7 introduces the experiments and the experimental setup. Successively my research articles, which have been published or submitted to international refereed conferences and journals follow.[1] They are presented in a form compatible to this dissertation layout and appendices are added to the articles with more information about each experiment.

[1] Information on this is provided for each article in the apposite chapters.

2

Communicative Gestures

The purpose of this chapter is to give a short overview of what communicative gestures are, how they are integrated in speech, how they are perceived, and why this is relevant to know when considering the placement of emojis in text. There are many theories of how gestures originated and their varying levels of involvement and relevance to speech. I primarily follow McNeil (1992)'s theory about the inherent integration of speech and gestures.

Co-speech gesture is found in people of all cultures, backgrounds, and ages. However, there is debate regarding the role of gestures in language comprehension. One side supports gesture as communicative, arguing that both speech and gesture are inherently integrated all the way down to the earliest stages of comprehension (Cassell et al., 1999; Kelly et al., 1999; McNeill, 1992). Conversely, Krauss (1998); Krauss et al. (1991) argue that gesture does not add significant communicative information to speech and that both modes are independent systems in which gesture potentially adds some information, but only after speech has been processed.

Gestures are such a part of our speech that most of us are often unaware of them as they take place. Speech and gesture occur universally and automatically together and while they occupy the same time, they share the same meanings and relationship to the context.

Gesture and speech are synchronously co-expressive and not redundant (McNeil, 1992). While they refer to the same concept at the same time, they contribute to this concept in their own unique way. One example is of a speaker describing a scenario where a man walks down a street. The speaker could say: “He went down that street.”

while the hand could indicate which street. These co-expressions happen simultaneously and contribute to each other while illustrating the same concept.

In the speech portion of the message, it is clear that there is a man moving down a street. Inversely, for the gestural portion, the hand shows along which street the man went. Here both spoken and gestural parts are simultaneous and co-expressive, without being redundant. Either type of information alone would not be enough to understand the intended message completely.

In face-to-face communication the visual and auditory modalities can overlap without conflict or confusion, but in the integration of emojis into text, all types of information are expressed visually and thus cannot occupy the same space in the written sentence. Gestures are actions and a picture is only an instant representation of a significant part of the gesture. Thus, there has to be some order in the placement of the pictures in the text. Since the temporal relation between speech and gestures differs depending on the gesture type and function, it is necessary to take into account these aspects when considering emojis representing different gesture types.

2.1 Types of gesture

While gesture in this chapter mainly refers to non-manipulative hand/arm movement that occurs during speech, it can also refer to any kind of body behavior which also includes facial expressions, head movements, gaze, and body posture. Kendon (1988a) classified gestures into different types. These kinds were then arranged into a continuum by McNeil (1992). There are four main groups in this continuum. The first is "Gesticulation" where the gesture means something which is nonredundant to its co-speech. It is the most common type of gesture and the one that is relevant for this thesis. It has multiple sub-categories being "deictic", "iconic", "metaphoric", "beats and "emblems".

Deictics are pointing movements indicating a specific person, object, place, or direction and can also be used to refer to unseen, abstract, or imaginary things. Deictics are conventionalized but their form can vary in different cultures (Kita, 2003).

Iconics are gestures used to help illustrate the referent by depicting a property of it. Forming the shape of a sphere with one or both hands can be used for referring to a ball for example. Iconics can include spatial relationships between objects, bodily actions, and pictures of objects.

Metaphorics, like iconics, illustrate content; however do so indirectly by a 3rd element acting as a metaphor. An example is speaking about a story, and using both hands to indicate a book, referring to its content.

Beats, also known as batons, are rhythmic movements that are produced simultaneously to speech. They emphasize speech and their rhythm is connected to prosody.

Emblems, also known as symbols, are gestures that can be used independent of speech since they have conventionalized meaning that can be directly translated into words. Emblems are culture specific. An example of an emblem which is common in many cultures is the "thumbs up" emblem meaning "yes", "good", or "approval". Emblems are often used when vocal communicative means are restricted due to noise or distance for example.

Adaptors are non-communicative movements involving self- and object touching, such as scratching or playing with an object. Although they are not related to the message content, and therefore some researchers don't include them in their gesture classes, they still can be informative e.g. showing nervousness or boredom and can be important when designing human-like agents (Kipp, 2005).

The hand gestures whose pictures are used in this dissertation are deictics, emblems and iconics.

2.2 Speech and Gestures: more functions and theories

Gestures also have other functions in communication and some of these are related to interaction management and comprise feedback and turn management functions (Allwood et al., 1992; de Kok and Heylen, 2011). For example, head nods and shakes as well as the thumbs up gesture are often used to give feedback. Emojis representing these gestures are also used in this thesis.

Sometimes gestures provide new information in respect to co-speech, other times they reinforce or emphasize the speech's meaning. Furthermore, gestures can give information about a speaker's emotions and attitudes. This is also true with some emojis.

In face-to-face communication, we perform intended actions many of which are not part of the verbal modality. These nonverbal actions can be used to express feelings, clarify disambiguations, meanings and contexts and therefore function as a resource the speaker can use simultaneously to the verbal modality in an effort to transmit the intended message as perfectly as possible (Esposito and Esposito, 2011; Kita and Özyürek, 2003).

In some cases, gestures are very effective in disambiguating speech and they are useful in noisy environments, for example, where speech cannot be heard clearly (Goldin-Meadow, 2005; Kendon, 1980, 1986, 1988b, 2004, 2010; McNeill, 1992, 2005; McNeill and Duncan, 1998; Thompson and Massaro, 1986).

Furthermore, gestures have been shown to support semantic cohesion of speech and can be found to be coordinated with prosody, pitch and tone (Esposito et al., 2007; Shattuck-Hufnagel et al., 2007; Yasinnik et al., 2004). The imagery that gestures add can also supplement phrasal content (Goldin-Meadow, 2005; Kähler et al., 2001; Kendon, 1980, 1988b; McNeill, 2005) and gestures have also been found to be produced in coordination with speech pauses (Butterworth and Beattie, 1978; Butterworth and Hadar, 1989; Esposito et al., 2003, 2001, 2002; Hadar and Butterworth, 1997). Because of these studies, it seems appropriate to regard gestures as having a close symbiotic relationship with speech as an expressive unit used to share thoughts (Kendon, 2010; McNeill, 2005; Ruiter, 2000).

McNeill (2005) discusses how the evolution of language was very much dependent on gestures and imagery. According to him, it was crucial that gestures came about because this type of communication behaved very much like a "seed" and activated a set of neural circuitry required for language to evolve and develop into its full complexity that it is today. This theory therefore states that gesture is not an "attachment" or "enhancement" to language, but it is a fundamental root that cannot be removed from the current language/speech system.

Neurological evidence supporting evolutionary co-development of gesture and language comes from anatomical areas know as Broca's area and Wernicke's area. Broca's area can be simplified as the center of the brain "where speech and gesture are orchestrated as motor actions" (McNeill, 2005). It is responsible for our ability to produce sequences of actions that share a goal, meaning, or agenda and Nishitani and Hari (2000) showed that its activity precedes motor activity by 250 ms. Binkofski et al. (2000) found that Broca's area is in control of the production of forelimb movements and shares similar neural mechanisms in assisting speech. While Bonda et al. (1994) investigated similar activity in Broca's area regarding arm and hand movements, Decety et al. (1994) discovered that this area activates simply with mental imagery of grasping. Further support was found by Horwitz et al. (2003) who claimed that Broca's area activates for not only limb movement, but also oral/laryngeal movement.

Wernicke's area is the area of the brain responsible for categorical content, which supports comprehension, verbal thought, and creating imagery for discourse content (McNeill and Pedelty, 1995). This means that Broca's and Wernicke's area work together to produce imagery-language dialectic constructions and generate semantic meaning.

Browman and Goldstein (1990) supplemented these studies by proposing a model on how gesture and speech work together. In their model, gestures are actions of the hands to produce something visual, while speech is action of the vocal tract articulators, which manipulates the flow of air producing acoustic fluctuations known as speech itself. The purpose of this model is to help explain how the brain is able to combine these two types of action for a shared purpose, and not just language-unrelated action in itself.

2.3 Gesture Perception

Along with the importance of knowing how and why gestures are produced, it is especially relevant for this thesis to address the perception of gesture through observation. Studies of people gesturing such as those made by Streeck (1993) give insight into the nature of gesture perception suggesting that it enhances communication. This enhancement can be considered supplementation as described by Melinger and Levelt (2004) who found that when listeners perceived part of the message through gesture which was not included in speech, they had a better understanding of the intended message. Hostetter (2011) claims through a meta-analysis that gesturing does produce a significant enhancement to listener's understanding of the communication, which depends on several factors including gesture topic, redundancy, and listener's proficiency level.

There are however contradictory perspectives (Kelly and Goldsmith, 2004; Krauss et al., 1995), which claim that a listener's comprehension level is no better when they can see a speaker's gesture in comparison to when they cannot. Krauss et al. (1996) suggest that the difference between these two conditions is insignificant because that additional semantic information provided in the gestures is rather small and thus not important enough to make a significant contribution to the speech. This contribution level can vary so much, that in the same study the benefits of gesture depends greatly on factors related to listeners (Sueyoshi and Hardison, 2005) and includes variables such as age, and cultural differences, or the complexity (McNeil et al., 2000) and content of the speech (Driskell and Radtke, 2003).

There are different opinions about the contribution of gesture to the perception of meaning. While Kendon (1994) concluded in favor of gestures being beneficial, he more importantly stated that it is most likely not a universal truth and therefore it would be pertinent to investigate when gestures communicate, rather then if they communicate.

This goes in favor of what Alibali (2005); Beattie and Shovelton (2002); Krauss (1998) found when showing that speakers are more likely to gesture when they are referring to spatial information in comparison to nonspatial information. Also, these representational gestures may be more likely to benefit the listener when the referent created by them is related to the content of the message. Therefore, it seems that spatial and motor related gestures are most communicative. In corroboration to this, Driskell and Radtke (2003) found that nonspatial and nonmotor related gestures provide very little benefit to the listener since they provide no additional information that the speech does not already give.

Gestures that do provide additional information which speech does not, are often called nonredundant or supplementary gestures (Church and Goldin-Meadow, 1986; Emmorey and Casey, 2002). According to many researchers (Broaders and Goldin-Meadow, 2010; Goldin-Meadow and Sandhofer, 1999; Goldin-Meadow et al., 1992; Kelly and Church, 1998; McNeill et al., 1994), listeners are very good at detecting this nonredundant information provided through gestures and supplementing the speech information with it.

Yet listeners may still pay attention to gestures even when they are redundant to the speech (Hostetter, 2011). This could be due to speech being difficult to understand and the gestures provide clarification. If listeners are not high in proficiency in the language being used, they might find that the gestures are far more useful than someone who has high proficiency. This is what Sueyoshi and Hardison (2005) found when their subjects with low proficiency demonstrated large improvements in speech comprehension with co-speech gesture in comparison to those who were fluent.

Another factor to consider is that gestures can improve comprehension simply by the fact that they are more likely to capture listener's attention and keep them engaged. Kelly and Goldsmith (2004) found that listeners liked the speakers who gestured in comparison to those who did not and Maricchiolo et al. (2009) found that listeners thought that speakers who gestured were more competent and composed in their speaking ability than speakers who did not gesture.

Not only do co-speech gestures provide attention benefits, but they seem to also provide memory benefits. Paivio (1991) investigated the long-term memory benefits that gestures provide and found that subjects were better at remembering messages with co-speech gestures in comparison to those who only heard the messages.

However, gestures and speech co-occur and involve two modalities, while text messaging and emojis involve two modes of the visual modality. In a study consisting of four experiments by Wu and Coulson (2014), video primes containing congruous and incongruous co-speech iconic gestures were paired with picture probes that were related or unrelated to the gesture portion of the video prime. In the first experiment, participants' responses to picture probes were more accurate and faster to congruent primes versus incongruent ones. Counter intuitively however, unrelated probes had a faster reaction time than the related ones. The authors explain this could be due to the fact that "judging related items required more fine grained analysis" than unrelated ones and therefore it took longer decision making time.

When adding visual-spatial working memory elements to the experiment (Wu and Coulson, 2014), participants were more accurate and faster with congruent primes versus incongruent ones. There was also better accuracy in the picture probe task in low memory tasks versus high memory tasks. However, there was faster response time with high load memory tasks showing that cognitive resources were prioritized to working memory, and not so much to congruency effects. Wu and Coulson (2014)'s third experiment included spoken numbers instead of a visual-spatial task and they found that the task of remembering digits engaged working memory and limited the cognitive capacity. Most importantly, unlike in the second experiment when visuo-spatial skills where under high cognitive load and no congruency effect was seen, in the third experiment when verbal working memory was under high load taxation, participants were still sensitive to the meaning of gesture and prime-probe congruency effects. The results of this experiment show that visuo-spatial working memory is important for mediating speech-gesture integration while auditory working memory is not as much. Also there were no significant error rate differences between the two additional tasks showing that the visual-spatial task didn't have a difference in difficulty level as compared to the verbal working memory task.

The results of the last experiment by Wu and Coulson (2014) show that participants classified picture probes that were related to congruent multimodal videos more accurately and faster than when the primes were incongruent suggesting that people integrate information presented in gestures with that conveyed in speech. Most importantly for this thesis however was that Wu and Coulson (2014)'s study shows increasing visuo-spatial working memory demands reduced the benefits of congruency effects. This supports the visuo-spatial resources hypothesis that states that visuo-spatial working memory serves as an important role in integrating speech and gesture. Finally, it corroborates the idea that iconic gestures function as image-based forms of the meaning of utterances. This can provide an explanation of why pictures of the same gestures can help with comprehension of texts even though the temporal relation between pictures and text is necessarily different from that of speech and gestures.

2.4 Discussion and Conclusion

It is easy to see that language is composed of two inseparable forms of communication involving gesture and speech. The ability to learn, contextualize, and produce these forms of communication comes from the evolutionary development of Broca's and Wernicke's area.

Conclusions such as those made by Melinger and Levelt (2004) support the use of emojis into text. Since in face-to-face communication speakers choose to have part of the message be visual and some be spoken, there might be similar needs in short messaging to communicate through more modalities and since this is not possible when writing, people are adding emojis to text. Since Alibali et al. (2001); Cohen and Harrison (1973) found that speakers are more likely to gesture when the listener can benefit from them, it seems likely that through digital communication, writers might be more inclined to use emojis to supplement their written text since they know that readers can benefit from their inclusion.

Yet similarly to how Kendon (1994) questioned that it is not so pertinent to investigate if gestures communicate, but when they communicate, it could be equally useful to study not if emojis supplement text, but when they do so. Therefore it is important to consider what Alibali (2005); Beattie and Shovelton (2002); Driskell and Radtke (2003) discussed regarding the importance of spatial and motor information provided in the gesture. If emojis become too abstract they might provide little information to the written text and thus not be very beneficial.

Therefore, as stated by some researchers (Broaders and Goldin-Meadow, 2010; Goldin-Meadow and Sandhofer, 1999; Goldin-Meadow et al., 1992; Kelly and Church, 1998; McNeill et al., 1994), it is important that the pictures of gestures, or emojis, provide additional information to text. The inclusion of these nonredundant emojis can be scripted such as Church et al. (2007) who presented scripted spoken sentences with nonredundant gestures and studied memory effects related to the information presented in the gestures. Other more natural, yet similar studies, such as Beattie and Shovelton (1999), presented naturally occurring co-speech gesture, and subjects were questioned again about information presented in gestures alone. Similar texts can be produced with the inclusion of nonredundant emojis, which provide completely new information not presented in speech, and participants can be questioned on that information alone.

Referring back to what Kendon (1994) stated about the importance of not if gestures communicate, but when they communicate, the results found by Sueyoshi and Hardison (2005) show that gestures are particularly useful when there is low proficiency. Therefore, the implementation of emojis in text could be far more useful for learners of a second language, than for native speakers. As discussed by Kelly and Goldsmith (2004); Maricchiolo et al. (2009), gesturing could help learners simply by the fact that they seem to engage listeners more. Thus, regardless of if the emojis included in the text are redundant or nonredundant, their inclusion alone could capture the interest level of the readers to a higher level that than of just text alone. Plus the additional long-term memory benefits of gesturing which Paivio (1991) discussed adds value to the idea that including emojis in text would have a similar effect on learning and information retrieval. However, as stated in Wu and Coulson (2014), visuo-spatial working memory demands cannot be increased too much since they decrease the communicative aid that congruent co-speech gestures produce, and therefore emojis should be as simple as possible and their placement should be as close to their related word in the text as possible.

However, Hostetter (2011), expresses caution that studies involving scripted gestures, such as those by Ping and Goldin-Meadow (2008); Valenzeno et al. (2003) which claim apparent benefits of gesture, lack external validity since they cannot be generalized to natural real-world communication. This investigation is the main endeavor of this thesis. Regardless of whether there is external validity and regardless of if a strategic implementation of scripted emojis in text reflects that natural usage of emojis in day-to-day communication, if this so called strategic implementation does produce benefits in communication, it would seem like a very valuable and useful asset to implement in a learning environment due to attention, comprehensive, and long term memory retrieval benefits. Furthermore, current uses of emojis in text messaging show that users do not restrict themselves to a certain tradition, but they show creativity in the way they mix text and emojis. Therefore, short text messaging is evolving.

3

Social Cues and Emotions in Communication

This chapter discusses how social cues and emotions are expressed in face-to-face communication and how emoticons and, more recently, emojis are a kind of substitute for them in computer-mediated written communication.

Body behavior does not only contribute to the semantics of a message, but, as also mentioned in Chapter 2.1, it also provides social cues that contribute to the development of the interaction (Allwood et al., 1992; de Kok and Heylen, 2011) and to the expression of emotions, affective states, and attitudes, henceforth emotions. During communication we express emotions through all the body, and especially facial expressions and tone of voice. These expressions happen during interaction as a function of an individual's underlying personal emotions as well as social display rules which are a standardized unspoken agreement controlling which expressions are socially appropriate in any given situation. Emotional expression is therefore thought of as a simultaneous combination of the component of internal emotions experienced and whatever social rules are defined (Fussell, 2002).

These rules can be adjusted depending on roles, gender, situation, and (sub)culture. For example, in many cultures it is common that women are more emotionally expressive than men which could be a direct result of them being more internally emotional than men as well as having more freedom defined by rules permitting emotional expression (Timmers et al., 1998). Another rule is that it is more acceptable to express emotion in a socio-emotional context than in a task-oriented one. Wagner and Lee (1999) find that emotion expressions are also more likely to occur when the interaction is between friends. Lee and Wagner (2002) show that people are more likely to express emotions in positive social contexts versus negative ones. This emotional expression can function in providing information, regulating interaction, and expressing intimacy (Ekman and Friesen, 1969; Harrison, 1989) so it is clearly an important factor in communication.

3.1 Text Messaging

Written computer-mediated communication does not provide multimodal social signals which are typical of face-to-face communication, and lacks social cues (Kiesler et al., 1984; Rutter, 1987) such as head movements expressing feedback and turn taking (Allwood et al., 1992, 2007). There are several theories that discuss the consequences of what happens in social environments where these social cues are not available. One is called the cuelessness model, which hypothesizes that the fewer social cues that are available to subjects engaged in communication, the larger their cuelessness is (Rutter and Stephenson, 1979). In this sense, written computer-mediated communication is very clueless which can lead to psychological distance between communicators (Rutter, 1987). This psychological distance is comparable to the saliency of social presence in physical interaction, or lack thereof (Short et al., 1976). Therefore, written computer-mediated communication can function as a filter removing nonverbal communicative devices rich with information and social presence causing communication to be based on task-oriented and superficial content (Walter, 1994).

Another theory is the cues–filtered out approach, which is similar to the previous theory except it focuses on lack of social context cues in written computer-mediated communication. In face-to-face communication the physical environment and speaker's nonverbal behavior all play an important role in controlling and dictating what rules and social norms should be followed. This theory discusses how without theses cues, people behave without considering the consequences which might be present in face-to-face communication. Such consequential examples include excited or uninhibited communication, as well as extreme and risky decision-making (Kiesler et al., 1984; Siegel et al., 1986; Sproull and Kiesler, 1986).

The Social Information Processing model proposed by Walther (1992) argues that since the limited expressive capacity of computer-mediated communication causes individuals to use only one modality, their ability to create an impression of the content is significantly inhibited in comparison to face-to-face communication. The consequences of this are impersonal communication and relatively negative evaluation of other people.

These theories show that with written computer-mediated communication, the nonverbal communicative tools normally used get reduced to a significant degree causing a lack of social presence and increased distance between speakers resulting in less friendly, less emotional, and less socio-emotional oriented interaction (Rice and Love, 1987). The popularity over the past decade of short text messaging via computer and other mobile devices has increased the need for adding social cues to this communication form.

3.2 Emoticons

As communication between people moves from the physical space to the digital, there is a need to compensate for the lack of natural nonverbal communicative devices. Emoticons are a relatively new usage of keyboard characters and punctuation marks combined to make up relatively obvious facial expressions which Walther and D'Addario (2001) showed to correlate with particular emotions, such as :-) for happiness, :-(for sadness, and ;-) for humor. There are many different styles and ways to make a number of different emoticons and it is believed that they function as nonverbal suggestive facial expression surrogates adding to the message emotions that would normally be present in face-to-face communication (Derks et al., 2007). This is why Thompsen and Foulger (1996) propose that these emoticons may enhance the information shared since they provide additional nonverbal social cues than what can be read in the text alone, thus improving the communication (Rezabek and Cochenour, 1998).

Derks et al. (2007) investigated the social interactions on the internet and how people use emoticons. They studied the use of emoticons by manipulating the context of both socio-emotional and task-oriented contexts as well as manipulating the valence to being both positive and negative in each. They found that people use more emoticons in socio-emotional contexts than in task-oriented ones. This supports social norms previously defined by Fussell (2002). According to these rules, it is more appropriate to express feelings to friends than to colleagues in face-to-face communication. What Derks et al. (2007) also discovered is that unlike in face-to-face communication where there is a display rule to only show positive emotions and not negative ones, online people showed just as much of both. This does not necessarily conflict with what Lee and Wagner (2002) found regarding rules for which emotions are appropriate to display, but rather that there is a difference in the rules when it comes to face-to-face communication and computer-mediated communication. One possible explanation for this is that there is a certain degree of anonymity behind internet communication, and this deindividualization can result in antinormative behavior as discovered by Postmes et al. (2001); Spears and Lea (1994).

Similarly to what Fussell (2002) stated about people adjusting how they display their emotions based on the situation in face-to-face communication, Derks et al. (2007) found that online, people show more negative emoticons than positive ones in negative situations, and more positive emoticons than negative ones in positive situations. This shows that emoticon use depends very much on the context.

These studies demonstrate that communication and emotion expression are done very similarly over the internet and face-to-face communication. However, it should be noted that while emoticons are used as a means to provide nonverbal facial expressions as one would use in real life, they are not nonverbal behavior. This is because emoticons, in comparison to face-to-face communication, are much more voluntary, deliberate and actively produced. Since the ability to produce them over text is such a new technology, it is possible that they will become habitualized and less conscious in time. But it is unknown if they are interpreted as iconic and unconscious nonverbal facial expressions, or more like words which are then encoded as emotional supplements in communication (Marvin, 1995). This is similar to what Walther and D'Addario (2001) found in their research suggesting that the amount of contributions that emoticons provided were outweighed by text.

However, the fact that people do use emoticons shows a need or at least a certain desire to supplement text with signals which can replace nonverbal information. Similarly, when people use the phone, they still supplement speech with nonverbal gestures even though unable to see each other (Bavelas et al., 2008). However, as Walther and D'Addario (2001) pointed out, since there is a lack of nonverbal cues in written computer-mediated communication, more texts are required to be sent between two interlocutors resulting in more time spent on getting the correct message compared to what happens in face-to-face communication. This may also result in greater miscommunication (Erkens et al., 2002) and thus it would be ideal to have a greater diversity of available nonverbal communicative devices.

3.3 Emojis

Emojis are also a form of nonverbal communication, but with actual pictures. The use of mobile messaging applications is growing exponentially with companies such as Line, WhatsApp, WeChat, KakaoTalk and Facebook Messenger (Khalaf, 2014). Most of these applications provide specialized character emojis (which some call stickers). The use of these emojis is so popular that users generally use them instead of text when they need to express themselves and couples resolve arguments easier using these emojis to apologize (Olson, 2013). Tossell et al. (2012) write that emojis are generally used to display humor or sarcasm which corroborates with Dresner and Herring (2010)'s findings regarding emojis' primary usage for playfulness showing that text in the message shouldn't be taken too seriously.

While Tossell et al. (2012) showed that the usage rate of emojis in private mobile messaging is not very high, Line emojis are used much more often (Wang, 2015). This is because Line emojis depict body language and facial expression in the cartoon and character-driven emojis. The sophistication of the nonverbal expression in these images is more comprehensive towards what would be portrayed in face-to-face communication adding socio-emotional connectivity and saliency in the computer-mediated communication. Janssen et al. (2014) found that the better communication quality in emojis can lead to significant enhancements in the intimacy perceived by communicators in affective technology. Wang (2015) found that messages with Line emojis and text produce the highest level of intimacy between communicators versus only text and only emoji communication. These results show that these emojis function very similarly to standard emojis as a means to display nonverbal cues, which would normally be used to express warmth and affection in face-to-face communication. The findings suggest that there is a trend for people to have an affinity for ways to make computer-mediated communication more natural and provide nonverbal cues, as gestures do, in face-to-face communication, ultimately showing the importance of nonverbal communicative cues in any kind of human interaction.

3.4 Emoticons, Emojis, and Sentiment Analysis

Clearly, a very efficient way to provide sentiment in text is through the use of emoticons and emojis. Since the use of emojis is so common on the internet, sentiment classifiers used for linguistic analysis on social media are now including emojis in order to improve sentiment analysis.

Microblogging is a very popular communication tool on the internet with millions of users who share their opinions everyday. Since it is free with an easy accessible platform, there is a trend to use it as opposed to traditional communications tools such as blogs or mailing lists. Since people are posting their opinions about products, politicians, services and events, microblogging websites have become a valuable source of information regarding people's opinions and sentiments since they can be used for product development, marketing, political agendas and social studies.

Microblogs are ideal for sentiment analysis and opinion mining because microblogging is used by people independently of age, social class and nationality. The use of microblogging by big companies and politicians is also growing.

Pak and Paroubek (2010) collected a corpus of Twitter posts with three classes being positive sentiments, negative sentiments, and objective texts. They used Twitter posts using the following emoticons: happy emoticons: ":-)", ":)", "=)", ":D" etc. and sad emoticons: ":-(", ":(", "=(", ";(" etc. Since the rules of Twitter do not allow posts being longer than 140 characters, they are usually only one sentence long and the emoticon in it can be assumed to reflect the sentiment of all the words in the entire post. Using this system, Pak and Paroubek (2010) were able to create a sentiment classifier able to tag emotional text.

Similar work was done by Yang et al. (2007); Zhao et al. (2012) to build a corpus using emoticons in blogs as indicators of sentiment. Go et al. (2009); Read (2005) used emoticons in their classifiers so that they would be more reliable, and therefore less dependent, on the subject domain of the text.

One complication with these classifiers is that there is such a difference in socio-cultural settings for emoticon usage, which is a similar phenomenon to the production of emotion expression in face-to-face communication mentioned earlier. Janssen and Vogel (2008); Vogel and Janssen (2009) found one difference was that Swedes discussing politics were more likely to use positive emoticons than negative ones, while Italians were more likely to do the opposite. When discussing science, Germans, Italians, and English were more likely to use positive emoticons where Swedes used neutral ones.

3.5 Conclusion

People are trying to express themselves in written computer-mediated communication in the same way as they do in face-to-face communication. Emojis started as simple emoticons, and have evolved to more realistic looking emojis and even moving emojis. Not only are emojis growing with popularity and diversity, but they are intuitive and can easily be understood. Sentiment classifiers are greatly improved with the use of emoticons, which shows scientific evidence supporting the incorporation of emojis in text as well to improve the natural understanding of text more fluently. Emojis are often used to express sarcasm or jokes, but they also support memory and learning. Therefore it is important to scientifically investigate how emojis are processed in the mind and whether their use could improve reading. While in the beginning emojis were only used in the end of sentences, I have found examples showing that people use them differently such as in the middle of the sentence and as different parts of speech (see Figure 3.1). Further investigations could supplement this knowledge and in this thesis I will focus on theses issues such as the position of the emojis placement in the sentence and congruency/semantic effects.

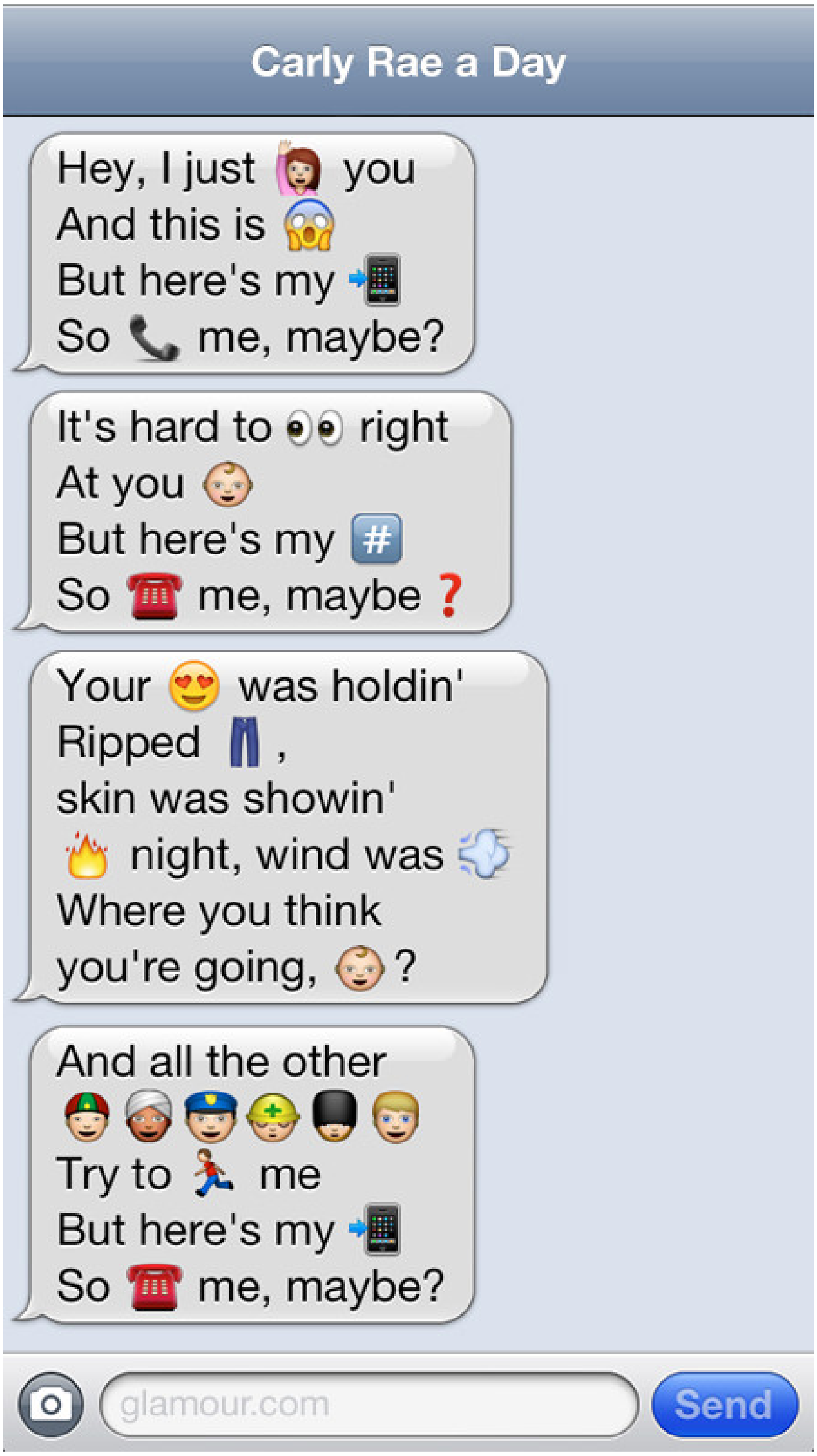


Figure 3.1: Emojis in different places and as different parts of speech. Taken from `http://media02.hongkiat.com/clever-funny-emoji/call-me-maybe.jpg`.

4

Mirroring

This chapter presents a short overview of the mirroring theory and other theories which relate mirroring to learning of inter alia gestures and social signals. It must be noted that the mirroring system is generally accepted, but its importance in the evolutionary development of humans is not shared by all researchers, see especially (Hickok, 2014). The critics of mirroring theories of evolutionary development of language and social skills focus especially on the fact that other brain areas than the mirroring system are involved in action learning. Furthermore, since some of the mirroring studies are based on the behavior of apes, Hickok (2014) doesn't believe that they can be generalized to humans. Hickok (2014) also argues that if it was true that we learn to do what we see, looking at people who commit criminal acts, would result in the viewers to commit the same criminal acts, e.g. killing. He also criticized the fact that all mirroring studies have focused on action learning since this might have prevented exploiting other functions of the mirroring system. However, since mirroring theories do not state that a) the mirroring system is the only mechanism behind action learning and the development of social behavior, and b) humans actually perform all the actions they see others perform, the above criticism does not affect the fact that mirroring has an effect on action learning and the mirroring studies on humans which I focus on in this chapter.

A necessary requirement for people interacting in any social environment is understanding what other individuals are doing, what they are intending, and how they are feeling. Not having the ability to do this would result in the loss of intended, and unintended messages between individuals, and any type of cooperative social system would fail. While there are several important parts of the brain that make this ability possible, one in particular is the mirror mechanism which has greatly been studied in monkeys and humans (Rizzolatti et al., 1996).

Mirror neurons have a rather crucial role in the evolutionary development of gesture and language as well. These neurons, are activated when the observation of others takes place. Interestingly, the mirror neurons are activated when action is observed as well as when the same action is performed (Goldenberg, 1999; Rizzolatti et al., 1996). Yet no mirror neurons are activated when observing passive hands or objects being used in actions (Nishitani and Hari, 2000). The mirror neuron system seems to be a network specifically designed for recognizing intentional goal-directed actions of one's own, or of others (McNeill, 2005).

McNeill (2005) argued that mirror neurons' ability to recognize actions of others was the precursor to the development of the human speech circuit involving Broca's and Wernicke's areas. This initial neurological phenomenon is the reason why language and gesture are centered in the same brain areas.

Mead (2009) discussed, more philosophically, how a gesture symbol is only a useful action if it implicitly produces understanding and meaning in the observer and that this meaning must have a mutual social acceptance. Therefore, it could be that the reason for the development of mirror neurons in human evolution was so that one's own gestural actions activated the same part of the brain as the observation of others' same intended actions. This could be one of the contributing reasons why gesture and speech are so inseparable.

4.1 Observation

A large variety of stimuli presented in the visual field showing the same action produce the same mirror effect. For example, the same cluster of mirror neurons that fire when watching a human hand grasping an object activate precisely the same cluster when watching a monkey performing the same action. Likewise, there is no difference in firing pattern if the action is observed at proximity or at a distance even though the size of the seen hand and object are quite different. Mirror neuron activation is also indifferent to whether or not there is a reward involved. (Rizzolatti and Craighero, 2004).

Pellegrino et al. (1992) observed that specific neurons will fire when observing someone perform meaningful goal-directed hand movements such as picking up or putting down objects on a table, taking food from others, and handling objects. This indicates that there is an observation/execution matching system. Whenever a motor action was observed, it was also represented in the mirror neurons even though the action was not executed. This mechanism is believed to play an important role in understanding the motor events performed by others. Most importantly, this matching system is found in humans, not only monkeys, and one of the regions is commonly referred to as Broca's area (Rizzolatti et al., 1996).

It was found that when healthy developing children observed an actor perform a goal, the mirror neurons responsible for firing for the last action, were already activated in the first part of the movement. This suggests that observers have a cognitive map of the entire action an actor intends to perform as soon as it starts, showing that they understand the entire movement and purpose (Cattaneo et al., 2007).

4.2 Understanding

Jeannerod (1994) proposes a theory that focuses more on learning and understanding through observation. According to this theory, a student can watch a teacher perform an action or task, while being still, and during the process, the brain forms a representation through the mirror neurons of how to do the action or task. Here the brain's representational motor image of the task activates the exact same neurons as when performing the task.

Jeannerod (1994); Rizzolatti et al. (1996) agree that the mirror neurons are vital in internal representation for learning, Rizzolatti et al. (1996) believe the focus is more on understanding motor actions. What they mean by understanding is the ability to recognize an action, differentiate observed actions, and to know how to respond accordingly. When an individual makes an action, it is usually with the intention and prediction of a specific outcome that is remembered through the senses. This action is stored neurologically along with its meaning.

Iacoboni et al. (2005) found that when participants were required to infer an actor's intention by the context, activation in the mirror-system was selectively increased. Brass et al. (2007); Kilner and Frith (2008) corroborate these findings by detecting mirror neuron activation when participants had to infer intentions in situations that were unusual, and when the task required mentalization.

Nelissen et al. (2005) showed video clips of actions done by humans performing actions as well as just their isolated hand, and also used robotic hands. Mirror neuron activation still occurred even when the subjects were fully aware that what they were looking at was not "real". More impressively, even with an "unreal" stimulus, and with an artificial device mimicking something that obviously is not a natural/biological limb, mirror neuron still fire. This supports very much the idea that mirror neurons could be activated from emoji presentations and thus help with viewers understanding intended actions.

4.3 Emotion Expression

Emotions are also crucial to express and detect in communication. Many researchers divide the spectrum of basic emotions into five categories of love, happiness, anger, sadness and fear; some also include disgust. This simplified classification system is used because it can be found in all humans of any race, gender, age and social class (Fabbri-Destro and Rizzolatti, 2008).

Disgust is an emotion that has often been studied neurophysiologically (Augustine, 1996; Royet et al., 2003; Schienle et al., 2002; Wicker et al., 2003). In particular, Wicker et al. (2003) conducted an fMRI study to investigate if the same areas in the insula that are activated when experiencing a disgusting sensation, would also be activated when observing faces of individuals experiencing/expressing disgust. Results show that the same area in the anterior insula was activated both by the experience of smelling disgusting odors as well as the observation of others experiencing it. Therefore it seems likely that the neurons in this part of the brain contain mirror mechanisms, however it is possible that a specific population of the neurons in the insula is responsible for sensing disgust while another is responsible for observing it.

Similar activations were noticed in the insula for a study involving pain exposure as well as observation. Singer et al. (2004) followed a similar paradigm as Wicker et al. (2003)'s experiments, but they tested pain. Two conditions were tested. In the first one, the participants were mildly shocked with painful electric stimulation from electrodes on their hand, and in the second one they watched videos of loved subjects experiencing the same shock on their hand. The results were that the same areas in the anterior insula are activated for both the sensation and observation to pain exposure suggesting a mirror mechanism precisely similar to that involved in disgust.

Even artistic pictures have the ability to produce emotional responses which can be detected through the mirror mechanism. This concept was discussed in Freedberg and Gallese (2007) who studied the emphatic engagement of emotions in others, produced by pictures and sculptures. The authors found that when participants looked at particular pieces of art, some experienced similar agony (physical empathy) in the same body parts as those represented in the art form and that there was a sense of bodily resonance. The same occurred in studies with emotional pieces of art, where participants experienced empathy for the subjects in the art and thus shared their emotions. The authors suggest that this response is due to mirror neurons and their activation involved when observing art which is possible since mirror neurons can be activated from static images. In fact, the authors account for three phenomena which can occur when looking at pictures "(i) the feeling of bodily engagement with the gestures, movements and intentions of others; (ii) the identification of the emotions of observed others; and (iii) a feeling of empathy for bodily sensations."

The notion that the observation/recognition of emotions in others activates the same neurological structure as when the same emotion is experienced by themselves is a hypothesis tested and developed by several studies (Calder et al., 2000; Carr et al., 2003; Damasio, 2003; Freedberg and Gallese, 2007; Gallese et al., 2004; Goldman and Sripada, 2005). The main points of these studies comment on the fact that emotional understanding can also function similarly as in somatosensation (touch) in that emotion recognition activates cortical regions, which are represented in the body as well. The fact that there is activation in the anterior insula, yet no activation in somatosensory regions for emotion processing strongly supports the idea of a non-sensorial structure for emotion recognition. Since emotions play a role in the mirror mechanism, and it has been found that pictures and sculptures activate the mirroring system(Freedberg and Gallese, 2007), it is likely that emojis can activate emotional mirror neurons with equal effectiveness.

4.4 Discussion and concluding remarks regarding the need for Emoji implementation

As Rizzolatti and Craighero (2004) showed, a wide variety of visually presented stimuli depicting the same action will result in the same mirror neurons firing. Since there is such a variety of emojis, which depict a relatively small range of emotions or messages, their potentially large visual differences will result in relatively no difference in the reactive firings of mirror neurons. For example, there are many emojis showing happiness, but that does not mean that there are also equally countless firing patterns for each individually different happy emoji.

In short written messages it is often difficult to express the mood of the message. Adding a facially expressive emoji can dramatically change the message e.g. indicating that the sender is joking. Emotion understanding is obviously a very important part in individual or group communication. If while communicating, people were not able to detect the specific emotion a speaker was feeling, and instead only interpreted the literal words he/she were expressing, mixed messages would certainly occur. The studies done by Augustine (1996); Royet et al. (2003); Schienle et al. (2002); Singer et al. (2004); Wicker et al. (2003) help shed insight into the importance of emotion observation and understanding and how mirror neurons play a significant role in this function. Since the observation of facial expressions showing an emotion activate the same mirror neurons used for sensing it, it seems likely that the emojis portraying the same emotions will also activate these mirror neurons in the observer helping them understand exactly what the messenger is feeling.

Meister et al. (2003); Rizzolatti and Craighero (2004); Seyal et al. (1999); Tokimura et al. (1996) demonstrated that the hand motor cortex neuronal activity increased during reading and speech, suggesting that there is a neurophysiological link between hand gestures and reading aloud. Therefore, when people are producing a short text message, and thus are not communicating naturally since they are not using their hands to produce gestures, there could be an inhibition in their ability to produce the written form of what they are wanting to communicate verbally and visually. Therefore, including emojis into short text messages is potentially not only beneficial for the receiver of the message, but also facilitates the message sender's ability to produce the intended message accurately.

As discussed by Brass et al. (2007); Cattaneo et al. (2007); Iacoboni et al. (2005); Kilner and Frith (2008), the ability of the mirror neuron system to function not only as a system that understands what individuals are doing, and the goals they are trying to accomplish, but more importantly why they are performing them. If this visual information could be harnessed in emojis then it could add to a much deeper and intuitive understanding of why people were writing their short texts rather than just what the message was in itself. The prediction ability could help with sentence understanding and in determining congruency and incongruency based off of cloze probability.

Most importantly for the support of this dissertation is the work by Freedberg and Gallese (2007); Nelissen et al. (2005) who demonstrated that mirror neurons respond to a scenario that is known to be artistic or unreal, such as in a video, and when actions are produced by actors who are not real looking, such as a robotic limb. This means that even if the stimulus is in a medium that is not real life, and looks far from the intentioned object, but functions as the object should, it can still be processed like the natural stimulus would. Therefore, emojis, despite their obvious artificial nature, still can contribute to a neurological comprehension of what is going on in an intended message and what the message sender is trying to communicate.

Mirror neurons in summary function as a system that takes sensory information from others' actions, into a motor format through the mirror mechanism to produce a large variety of cognitive functions and responses including learning, competition, understanding intentions, prediction, imitation, goal related actions, emotion recognition and understanding, speech production and communicative cooperation. All of these cognitive functions come from observing others perform bodily or vocal gestures.

In this chapter there has been a discussion on how mirror neurons play a fundamental role in effective communication showing why gesturing is important. Gesturing plays a necessary role in communication to produce cooperation in any social system. The method in which we communicate is evolving faster through technology than we are biologically. This evolution includes the production of digital messages being shorter and sent more casually than previously done. Therefore we have to come up with a system of technological communication that adapts to this shift.

Including emojis into written messages seems to function as a simple and popular surrogate to bodily gestures to supplement the missing information which would otherwise be there on a subconscious, natural, and intuitive level. The purpose of this dissertation is to not only investigate if this digital artifact does produce the same neurological response as other similar bodily communicative devices in face-to-face communication do, but more specifically, if there is a strategic way to implement them so that they supplement the short text most effectively.

5

Short Introduction to Electroencephalography

This chapter presents an introduction into electroencephalography (EEG) which I use as my main method to investigate the hypotheses of this dissertation. This entire chapter summarizes the work of Luck (2014) who shares extensive knowledge useful for beginners and experts in the subject.

There are different kinds of neuro-physiological recording techniques. One type is hemodynamic measurement, which includes positron emission tomography (PET) and function magnetic resonance imaging (fMRI). A second type is electromagnetic measurement, which includes EEG (Luck, 2014).

PET and fMRI provide non-invasive means of localizing spatial changes in blood flow that are triggered by overall changes in neural activity, but blood flow changes too slowly to permit the measurement of most cognitive processes in real time (Luck, 2014). Inversely, EEG provides non-invasive means of localizing temporal voltage fluctuations along the scalp that are triggered by changes in neural activity, but the scalp distributes the internal electrical activity too much to accurately detect them spatially. These technologies share each others' contrasting strengths and weaknesses. PET and fMRI are best for spatial measurements, and EEG is best for temporal measurements.

The first report of EEG was in 1929 when Hans Beger demonstrated that he could measure electrical activity of the human brain by placing an electrode on the scalp. Within the EEG recordings are a conglomeration of hundreds to thousands of neurons communicating and responding to specific sensory, cognitive and motor events, which when averaged together are called event-related potentials (ERPs). These waveforms consist of a sequence of positive and negative voltage deflections, which are called components. Components are labeled with the prefix "P" or "N" to indicate positive going and negative going peaks (Luck, 2014). Components are also labeled with a number indicating the peak's temporal position in the waveform. P300 and N400, for example, indicate positive-going and negative-going components that can be found at 300 and 400 ms after the onset of their designated stimuli. Often times, including in this dissertation, ERP names are abbreviated from P200 and P300, to simply P2 and P3.

Almost all ERPs are grand average waveforms, which are created by averaging the average of an individual subjects' waveforms. This has advantages because the lack of variability between subjects makes it easier to see similarities, and disadvantages because grand averages may not accurately reflect the results.

Due to individual differences in the cortical folding and functional relationships of the brain in specific locations in regard to gyri and sulci, huge differences can be expected in EEG activity. It is quite normal for voltage peaks to be positive for some people and negative for others and therefore the grand average can appear to be smaller than in individual subject performances.

However there are of course advantages to using ERP measurements. For example, if one wants to study the reaction time to a particular stimulus, the magnitude of variances involved in the cognitive processes can be difficult to accurately measure simply by looking at the behavioral response. For example, does the Stroop Test cause a slowing in the perceptual process or in the response process? With ERP measurements, such as the P300, it can quite conclusively be seen that the delay in response time is due to perceptual processes and these tools can be used to see in which stage specifically cognitive processes are being manipulated in experiments. A second major advantage to this technique is that it can measure processing of stimuli covertly, which means even when there is no behavioral response. An example of this is measuring the response to attended versus ignored stimuli, and is something often done in language studies.

Disadvantages include the necessity of inference of ERP latencies in contrast to understanding the significance of behavioral responses. For example, a timed response to a stimulus through a keyboard input means it took X amount of time longer to push the keyboard with one stimuli over another. But with ERP latency in condition A that differs than with condition B, no specific conclusions can necessarily be made and a significant amount of assumptions and inferences from previous research are required.

Another huge disadvantage is the number of trials required to accurately measure ERPs. Since ERPs are so small, it requires fifty, hundreds or even thousands of trials per subject for accurate analyses. In contrast, behavioral response differences such as reaction time can be accurately observed with twenty to thirty trials per subject.

5.1 Types of ERP

When looking at ERP components, it's important to know that although components such as P1 and N1 give information regarding their polarity and position in a waveform, components from different modalities are not necessarily related meaning that an auditory N1 might not have anything to do with a visual N1.

Interestingly enough, the first major visual ERP component is called the C1. This component is not identified with a P or N like most ERPs because its polarity can vary. It is found in the primary visual cortex and since the part of this area that codes the lower visual field is on the upper bank of the cerebral fissure (called the calcarine fissure) and inversely the part that codes the upper visual field is on the lower bank, voltage recorded from the scalp is positive for stimuli in the lower field and negative for stimuli in the upper field.

The next visual component is the P1 and is also largest at the occipital region. Due to temporal overlapping with the C1, it can be quite difficult to measure accurately especially since there are at least thirty isolatable visual areas that are activated within the first 100 ms, many of which contribute to the P1 and C1. Variables that affect the P1 include stimulus contrast, direction of spatial attention, and state of arousal.

The visual N1 is quite complicated since there are several components within the component itself, one being anterior and two posterior (one parietal and one lateral occipital) arising later. Factors that affect these components are spatial attention and discrimination/detection tasks, which give insight to discriminative processing.

The visual P2 is located at anterior and central scalp regions and is larger when stimuli contain target features, especially when infrequent. Although functionally similar to the P3, it seems to occur only when the target has relatively simple features, where the P3 works for complex target categories as well. Due to overlappings with the N1, N2 and P3 in the posterior sites, the P2 can be quite difficult to isolate and therefore not much is known about it.

The N170 and the Vertex Positive Potential seem to be opposite sides to a spectrum involving face and non-face stimuli. While the VPP can be found between 150 and 200 ms at the central midline, the N170 can be found in a response to facial stimuli at lateral occipital sites especially on the right side. The N170 is also larger and/or later for inverted faces than upright ones.

For auditory sensory responses, there is a set of very early components. It is possible to detect ERPs within the first 10 ms which originate from the brainstem auditory pathways called brainstem evoked responses (BER), and can be used for assessing auditory pathology. The midlatency component occurs after 10-50 ms and arises in the medial geniculate nucleus and primary auditory cortex. This can be used as the first component to accurately measure auditory attention. This is followed by the auditory P1, located in the frontocentral region.

The auditory N1 has three subcomponents including a frontocentral component, which arises from the superior temporal lobes, a vertex-maximum component with unknown origin, and a laterally distributed component originating from the superior temporal gyrus. All of these are affected by attention.

The mismatch negativity (MMN) functions similarly to the P3 and can be detected if there is a repetitive sequence of identical stimuli and then an occasional mismatching stimulus. This negativity peaks between 160 and 220 ms and is largest at central midline sites. The MMN can be observed even if subjects are not paying attention to the auditory stimuli, such as if they are reading a book for example with noises in the background. This component is indicative of automatic processing that compares incoming stimuli with sensory memory of preceding stimuli.

There are somatosensory ERP components however they are not studied nearly as much as visual or auditory ERP. The N10 reflects action potentials rather than postsynaptic potentials, which are typical for ERP, and they arise from the peripheral nerves. There are subcortical components from 10-20 ms and short/medium latency cortical components from 20-100 ms. There is also an N1 and P2 which combined are often called the vertex potential. Olfactory and gustatory ERP are rare and difficult to read because precisely timed stimuli are difficult with these modalities. But there are some more recent studies that use them.

The N2 has been so thoroughly studied that there is an entire N2 family. The basic N2 is when in a series of repetitive stimuli, a nontarget "deviant" stimulus will create an N2 deflection. If the deviant is a task-irrelevant tone, it will create a mismatch negativity response, sometimes called the N2a. However if the deviant is task-relevant, a slightly more latent N2 can be observed called the N2b, which is larger for infrequent targets and can be indicative of categorization process. While the N2b can be created for both auditory and visual task-relevant deviants, it is largest over central sites for audition and posterior sites for vision. The N2pc is an abbreviation for posterior contralateral meaning that the stimulus creates a component on the opposite side of the target location and reflects spatial attention.

There is also a P3 family. Almost always, when researchers use the term P3/P300, it is in reference to the P3b component that is parietally maximal. This component can be observed when a particular target or target type is infrequent, but still expected. P3a occurs when a task-irrelevant stimuli is unexpected and even surprising and is frontally located. The P3 component is sensitive to a target's probability meaning that as the probability goes down, the response amplitude goes up. The amplitude also increases when preceded by more and more nontargets. Also, if subjects devote more effort to the task, the amplitude of the P3 will go up indicating resource allocation, yet inversely if a target is difficult to determine or is confusing, the amplitude will go down.

There is also a group of language related ERP components. The most researched one is the N400, which is located over the central parietal region and is a response to semantic expectancy violations. This component can also be found in nonlinguistic stimuli such as pictures like line drawings if there are inconsistencies within the semantic context such as a story. Since this is the ERP used for investigating the hypothesis in this dissertation, a waveform and scalp form example of it can be seen in Figure 5.1 and 5.2.The P600 is a response to syntactic violations and the N280 can be observed for function words such as "to" and "with" but not for content words such as nouns and verbs.

The error-related negativity (ERN) is the brain's response following the detection of an error. It is a negative-going deflection located at frontal and central regions and rises right after the response. It can also be called the Ne and the component is often times followed by an inversely positive going deflection called the Pe. The ERN can be elicited when receiving negative feedback due to an incorrect response as well as by

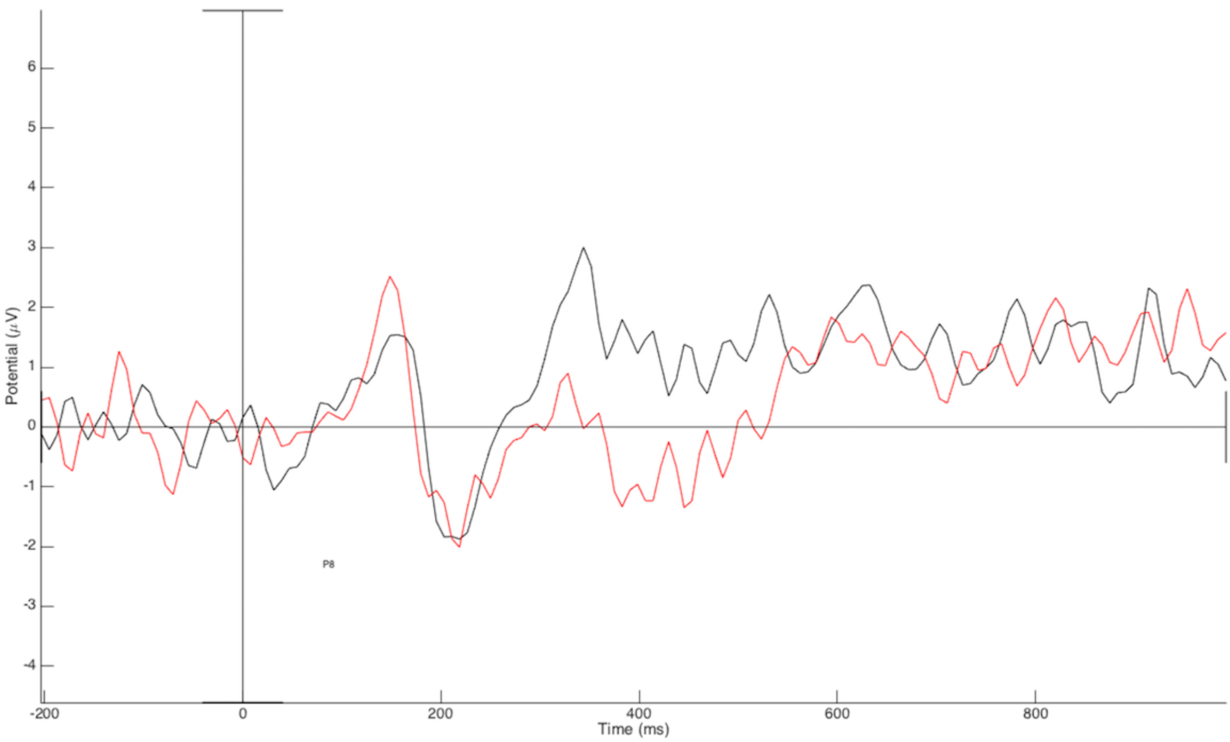


Figure 5.1: A comparison of the ERP waveform responses at electrode position P8 to a congruous (black) and incongruous (red) probe stimuli. Notice the negative deflection between 300 and 500 ms which peaks around 400 ms. Waveforms are displayed in microvolts.

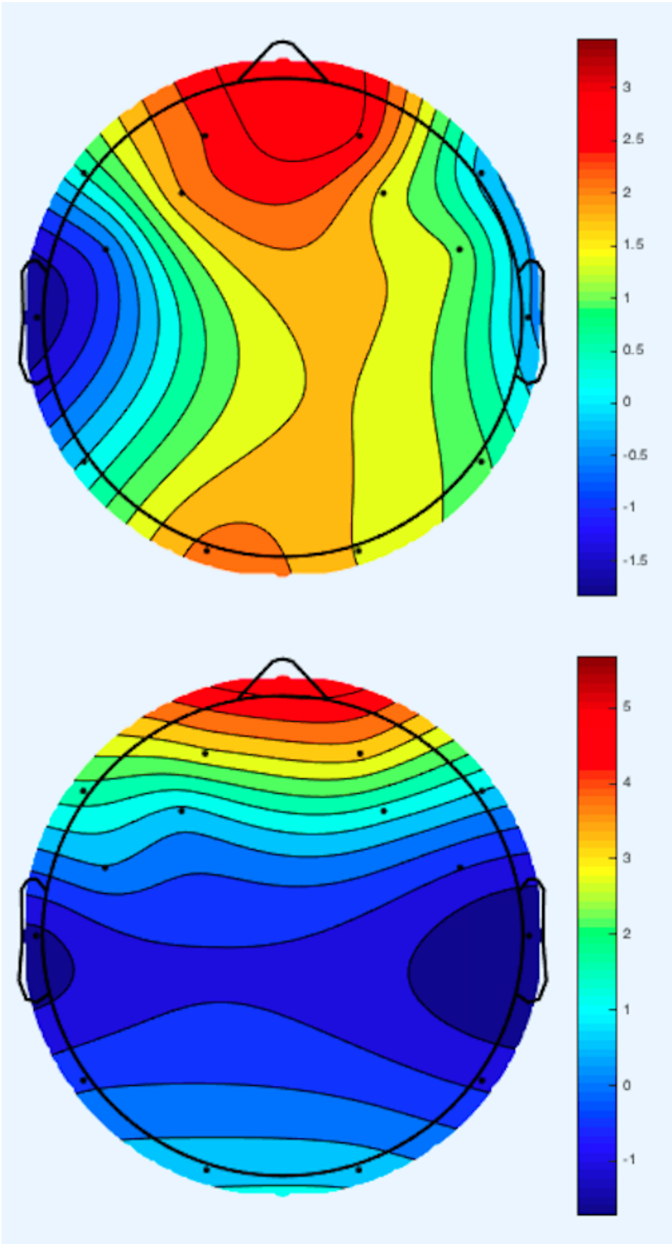


Figure 5.2: A comparison of the ERP scalp map responses to a congruous (top) and incongruous (bottom) probe stimuli at 400 ms. Notice the large and broad negative deflection distributed at the central parietal region. Scalp maps are displayed in microvolts.

observing someone else make an incorrect response. It is believed that this component is sensitive to conflict between intended and actual response and is indicative towards a system that monitors responses.

When participants are instructed to make manual responses without any influence from any kind of stimulus, there is a slow negative shift in frontal and central electrode sites up to a second before the physical response, which can be called the readiness potential (RP). The RP is lateralized depending on which side of the body will be used and this component is called the lateralized readiness potential (LRP).

5.2 Limitations

Although there is much information to be gathered in ERPs, there are a number of important limitations with interpreting ERP results. A waveform consists of a number of peaks and valleys, which are the sum of multiple independent latent components. These latent components are difficult to isolate and measure independently. It is important to distinguish between the observable peaks and valleys and the unobservable latent components (Luck, 2014). There are three fundamental problems with EEG interpretation and distinguishing waveform peaks with latent ERP components.

1. The first problem is that observable voltage peaks are not special. Two different combinations of unobservable components can result in the same observable waveform (Luck, 2014). This happens because multiple voltages are being presented simultaneously and summed together. Since peaks and components are not the same thing, the ERP waveforms are not necessarily good reflections of the latent components and therefore there is nothing special about the point at which the voltage reaches a local maximum.

2. A second problem is that peak shapes are not the same as component shapes. Since components come in all shapes and sizes, the measurable peak can be completely distorted by the other components (Luck, 2014). Due to the little resemblance the breadth a component has with its size in two waveforms, it is impossible to estimate the time course or peak latency of a latent ERP component by looking at a single ERP waveform. There may be no obvious relationship between the shape of a local part of the waveform and the underlying component. Since change in one independent peak can affect the latency and amplitude of the others in the same waveform, one cannot

interpret an effect that occurs for several peaks in a waveform as indication that there are changes in each component, even though it may very well appear to be that way.

This is why it is also problematic to compare a difference between two waveforms during an experiment with either of the original waveforms. Since each of the two original waveforms consist of several overlapping components, the difference between those in comparison to either of the originals only shows that the experimental manipulation did not affect the amplitude of all components proportionally. This is also why a difference in a peak's amplitude does not necessarily have anything to do with the component's size and why differences in peak latency does not necessarily reflect a difference in a component's timing.

3. Distortions caused by averaging are very common in a great majority of ERP experiments (Luck, 2014), including the ones presented in this dissertation. The purpose of signal-averaging procedures is to isolate ERP waveforms. However due to the trial by trial differences the averaged form may be very distorted. This means that the averaged ERP waveform doesn't necessarily accurately represent the individual waveforms that were averaged together. This can be seen most clearly with the onset and offset in the averaged waveform since it will represent the earliest and latest offsets from the individual trials that were averaged together also resulting in a lower amplitude.

5.2.1 Avoiding Limitations

Despite the previously mentioned limitations, there are some helpful strategies to get around them (Luck, 2014, pages 96-98).

1. "It is important to create an experimental design that focuses on just one or possibly two components otherwise things can get too complicated and confusing.

2. Using well-studied experimental manipulations is helpful to examine ERPs because when the conditions are as similar as possible to another experiment, it is easier to characterize the ERPs.

3. It can be helpful to focus on large components such as the P3 or N4 since their significant amplitude in comparison to the others will not be sensitive to distortions of other components.

4. It is possible to isolate components with different waves by varying the way an experiment is set up to elicit an ERP. With more variety in stimuli the more types of a single component will be created.

5. If possible, it can be advantageous to focus on components that are easily isolatable. However this is usually easiest for components that are lateralized where one hemisphere responds very different then the other.

6. Component-independent experimental designs involve those that are not necessarily looking for a specific ERP but rather comparing differences between two different types of stimuli. If the waveforms are identical until a certain point, or return to being identical after a certain point, it could be possible to infer that during that non-identical part is where the variances of the stimuli were being processed.

7. It's a good idea to compare ERPs that are a response to the same stimuli while varying only the psychological conditions. If doing a word study on word class such as verbs and nouns, it is important to assume that ERPs are not a response to the length of the words instead of the class.

8. Avoid physical stimulus confounds by using the same stimuli across different conditions, such as different order.

9. Don't assume a small physical stimulus difference cannot account for an ERP effect.

10. Use the same number of trials when comparing ERPs.

11. Be cautious when the presence or timing of motor responses differ between conditions.

12. Experimental conditions should be varied within trial blocks and not between them."

5.3 Basic Principles

Electrode positioning is clarified by the standard 10/20 system which has its name due to the electrode placement at 10 and 20 percent points along the latitude and longitude lines. The electrode names use letters to identify their position such as Fp = frontal pole, F = frontal, C = Central, P = Parietal, O = Occipital, and T = Temporal. The number indicates the distance from the midline with odd in the left and even in the right hemisphere. "z" is used to indicate zero, since the number 0 looks too much like the letter O. A few exceptions to this include "Nz" which is the depression between the eyes at the top of the nose, "Lz" which is the bump at the back of the head, and the left and right pre-auricular points, which are behind the middle of the pinnae labeled A1 and A2.

5.4 Averaging

Averaging is a necessity for analyzing ERPs, which is typically done by taking a number of waveforms time-locked to a stimulus onset. This is done because the waveform is assumed to have a lot of random noise along with the ERP under investigation, and averaging out all the waveforms isolates the ERP while averaging out the noise. Although this only works in perfect 100% repeatable ERP waveforms where the EEG is the same on every trial in relation to the time-locking event, which none are, this system is still acceptable.

Although this trial-by-trial averaging method is not too problematic, it has the possibility to significantly ruin data. When latencies are very different, averaging can significantly change the amplitude of the ERP. The latency differences can make an enormous impact on the peak amplitude of the average ERP. This means that when comparing two ERP averages, a significant difference in peak amplitude can look like participants had a different responses, but latency could be the only variable. If the latencies are different enough, the average could look like absolutely nothing.

Fortunately there are ways to minimize the latency variable effects. The only one that will be mentioned will be the method used in all of the experiments used in this dissertation which involves area measurements. The area under the curve in an average of several trials and is always equal to the average of the areas under the curves in each of the individual trials, which means that there will be absolutely no effect by any latency variability. However this only works when the latency range used to measure the area spans the entire latency of the component, meaning that if there are multiple ERP components in the waveform, and the measuring window needs to be minimized, variability will have an affect. This method can also be problematic if an ERP has both negative and positive components, because their areas can be canceled out.

Another thing to look out for is when overlap from preceding and subsequent stimuli occurs. This happens when the response to the previous stimulus has not reached a baseline before the current stimulus or when a subsequent stimulus is displayed before the response to the current stimulus has ended. Because waveforms can last for several seconds, overlap distortion can be subtle but sometimes quite obvious. The best way to reduce this effect is to ensure that there is enough delay between stimuli to minimize the overlap as much as possible. There are other methods to eliminate the overlap, which include high-pass filter, subtracting waveforms that have been induced from an absence of stimuli, and subtracting estimated overlap.

Data can be contaminated with worse things than overlap however. Artifacts include blinks, eye movements, muscle activity, and skin potentials and these are usually very large in comparison to ERP signals meaning that they can significantly ruin the signal to noise ratio. Also, some can be systemic artifacts meaning they are related to the time-locked stimulus in some way as opposed to being random which means that averaging signals does not eliminate them. There are two ways to eliminate artifacts the first being artifact rejection, which is excluding contaminated trials from waveforms that have been averaged together, and the second is subtracting the estimated contribution of the artifact which is called artifact correction. The experiments in this dissertation used a type of artifact rejection called moving peak-to-peak rejection, which is where the waveforms are analyzed in a particular window (in this case a width of 200 ms) and examines steps along the waveform (in this case steps of 50 ms) and used a voltage threshold of 100 microvolts, meaning that anytime the waveform had an amplitude deviation of over 100 microvolts in the 200 ms window, it would be considered an artifact, and thus rejected.

5.5 Filters

Filters are necessary during data acquisition and can technically refer to wide variety of data manipulations, but this section will focus on what ERP researchers typically use to attenuate specific ranges of frequencies. One of the main goals to filtering is to reduce the amount of noise that every EEG signal has. Some of this noise has a significantly higher or lower frequency than the ERP under investigation and therefore easy to eliminate. For example, most ERPs have a frequency between 0.01 Hz and 30 Hz, which means that anything above or below this is most likely noise. Muscle movement leads to artifacts which for the most part have frequencies of above 100 Hz, which means that they can easily be eliminated. Likewise noise below .01 Hz, comes from events such as sweating which causes skin potentials and drifts in electrode impedance can also be removed quite easily. Unfortunately neurological alpha waves produce noise artifacts that are around 10 Hz, which means they are very difficult to filter out without distorting the ERP waveform.

There are four main types of filters: (1) low-pass filters, which diminish high frequencies and pass low frequencies, (2) high-pass filters, which diminish low frequencies and pass high frequencies, (3) bandpass filters, which diminish both low and high frequencies while passing only a middle range, and (4) notch filters, which diminish a specified narrow band of frequencies while passing everything else. While these filters are important, it is also important to remember that they can significantly distort waveforms by changing the amplitude and timing of the components as well as adding artificial peaks. Not only can artificial peaks be created, but shifts in waveform latency become more pronounced.

5.6 Plotting, Measurement, and Analysis

According to the Society for Psychophysiological Research (SPR), it is mandatory that researchers present the averaged ERP waveform in their research, which illustrates the principal phenomena under investigation. The SPR has several recommendations that all should follow. The first is the value of showing data from multiple electrode sites because it allows experts to determine the underlying component and waveform. The second is the importance of showing the voltage and time scales so that readers can evaluate amplitudes and latencies from a single frame of reference, as well as know in which direction the scale is polarized since both ways are acceptable. The third is that it is important to show a sufficient amount of pre-stimulus amplitude, such as -200 ms of the onset for example. This is important so readers can evaluate the level of noise in the data as well as the presence of overlap from previous components. It is important that there is a relatively steady baseline before the onset of the stimulus.

Since most ERP studies focus on the amplitudes and latencies of one of more ERP components, accurate measurements of these are critical. One of the most common ways to do this is by defining a time window and for each waveform under investigation, find the maximum amplitude in that window which is called the peak amplitude. The second most common is by using a time window by calculating the mean voltage for each waveform in that window, which is called the mean amplitude.

In these time windows, it is important to not necessarily look for the maximum voltage, since that can be part of a preceding or subsequent component. Instead it can be more accurate to look for maximum point surrounded by smaller points, which is called the local peak amplitude and should be defined as having a greater voltage than the average of the three to five points of either side of it. It is never valid to compare peak amplitudes from averages of different numbers of trials or from time windows with different lengths. There is also a disadvantage to averaging many waveforms with varying component latencies because it will be smaller than the single amplitude. It might seem advantageous to use a smaller time window, but due to general and expected latency variations, the component under investigation might be excluded with this method.

However when using the mean amplitude method, a narrower time window can be used because it doesn't matter if the maximum is outside. The narrower the window, the less distortion from overlapping components will be found on the data measurements. Also with this method, there is less sensitivity to high frequency noise because a range of time points is used rather than just one. Another advantage is that the measures don't become biased when noise level increases or windows broaden. Therefore, unlike with peak amplitude measures, it is acceptable to compare mean amplitude measures from waveforms with different numbers of trials. Also, since mean amplitude is a linear measure, it is possible to measure the mean of a component of each subject, and use the mean of all of these to be equivalent to the mean amplitude of the component from the grand-average. However there are still problems with this approach including artifacts from overlapping components and choosing a time window can cause a researchers to "fish" for significant results.

When measuring the mean amplitude or peak amplitude, it will always be in relation to the baseline, which should be 100-200 ms before stimulus onset. This baseline should be as close to zero as possible but it is important to consider that there can be overlap from previous stimuli especially if the baseline is longer than 200 ms, and there can be influence from a preparatory process.

When measuring peak latencies, the latency of the maximum amplitude within a time window is a measurement of the latency of the component. A local peak latency measure can be used where it's not considered a peak unless three to five points on each side have smaller values. The problem with this however is that high frequency noise can cause the maximum voltage to be far from the middle of the peak. This can cause the noise-related peak to be far in time from the true peak, but close to the true peak's amplitude. Also, as noise increases, the peak latency will shift closer to the center of the time window due to all the averaging. This method, unlike mean amplitude, is nonlinear, which mans that a grand average will typically not be the same as the average of the peak latencies of single-subject waveforms; and the peak latency from the average of a single-subject will typically not be the same as the single-trial peak latencies.

Once all the data is collected and amplitude and latency measures are finalized, statistical analysis is required to find significance and possibly a main effect. Therefore to assess the interaction through a cross factorial design, ANOVA is the typical approach done in ERP experiments. Things to include in the factors include stimulus types (target vs. non target), experimental conditions such as brightness (bright vs. dark), electrode position (frontal, central, parietal, left, midline, right), and so forth.

Further in this dissertation, when the experiments are presented, this information will be useful in understanding the method, data acquisition, and data analysis sections.

6

Gestures, Emojis and Neuroscience

The studies related to gesture and speech so far mentioned in this dissertation have mostly relied on behavioral measurements. This chapter reviews further the investigation on how various gestures and emojis are perceived and understood through the use of EEG. So far, there is evidence from a behavioral perspective that nonverbal cues such as gesturing and facial expressions can support a more natural and effective communicative paradigm. However, there is also neurological evidence that supports the importance of these cues in communication and more importantly, why they should be incorporated via emojis into our text messages. As said in the introduction, the word *emojis* refers to both emoticons produced with punctuation marks, and emoji pictures. When studies refer to emoticons specifically, I will use this term.

6.1 Neuroscientific Studies of Iconic Gestures

To investigate if iconic gesture facilitates communication during language comprehension, Kelly et al. (2004) studied if gestures influence ERPs to spoken probes, and if this is the case, when do they do so. In this study, participants watched videos of actors sitting behind a tall, thin glass, and a short, wide dish. The actor spoke one of these four salient words (tall, thin, short or wide) describing a glass or dish while performing iconic gestures, and the audio and video were arranged in four conditions. 1) The matching condition consisted of the actor gesturing to one of the objects and speaking with the same description dimension, e.g. said tall and gestured tall to the tall, thin glass. 2) The complementary condition had the actor gesture to an object combined with a complementary dimension of the same object, e.g. said tall but gestured thin to the tall, thin glass. 3) The mismatching combination had gestures to the object and contained a different dimension respect to speech, e.g. the actor said tall but gestured short to the short, wide dish. 4) The fourth relationship was just speech with no gesture. Participants were instructed to push one button when the speech referred to the glass and another for the dish.

The results showed that there was a main effect with a large negative deflection in late semantic processing (324-648 ms) across all electrodes for the no gesture condition compared to the other three. There was also a difference in the bilateral frontal regions of the brain versus the other sites for this condition as well with larger negativity. This suggests that the brain processed speech differently when it was accompanied by gesture versus when it was not. The results are consistent with typical N400 ERPs that are lateralized to the right hemisphere.

There were also main effects during early pre-semantic processing stages. In the time window of 148-352 ms for frontal sites, the mismatch condition produced a larger positivity than the matching and no gesture conditions, but not the complementary condition. It was concluded that this effect was most likely a P2 ERP reflecting phonological processing of speech information. It was discussed that incongruent non-verbal information such as faces, produce larger P2 ERPs to linguistic information than congruent information that is non-verbal.

Also, for the 72-168 ms time window at frontal sites, the complementary condition resulted in a much larger positive deflection than match and no gesture conditions, but not the mismatch condition. This suggests that the P1 and N1 ERPs were activated reflecting auditory processing of the speech. There was also an effect in the time window of 0-92 ms for occipital sites where matching resulted in a smaller negativity than to the complementary and mismatch condition, but not to the no gesture condition. This is strong evidence against the theories claiming that gesture does not play an important role in the processing of speech, since the difference occurred hundreds of milliseconds before semantic processing occurred.

These results show that gesture does affect the brain's processing of speech. The most fundamental finding was that there were significantly different ERPs for speech that was combined with gesture versus that which was not. More importantly, the type of gesture played a role. There were different responses when the gesture and speech conveyed the same meaning, as in the matching condition, compared to when they were different, as in the complementary and mismatching conditions. This showed that it was not just the presence of a gesture itself that produced the differences in cognitive processing but that the type of gesture is also important. Furthermore, the complementary effect supports the idea that communication relies on gesture since the only difference between complementary and matching gesture conditions was the representational content. This is a large amount of evidence in supporting the importance of gesture in communication and its inherent connection with speech.

The early effects involving the P1, N1, and P2 ERPs showed that mismatching and complementary conditions were processed differently from matching conditions. This suggests that there may be a cross-modal effect regarding speech processing involving high-level visuospatial cues produced by hand gestures. This further supports the idea that gesture is closely integrated with speech.

The late effects involving the N400 provide more insight into how the different conditions were processed. The complementary condition did not produce a semantic processing effect like the mismatching condition did. Both conditions used different representational cues than speech, but only the mismatching condition had incongruent indexical cues. This shows that for semantic processing, the complementary information was processed partially in congruence with the speech since there was still reference to the same object. But when the cues were referencing a different object, they were processed as the most semantically incongruent and therefore the largest N400 was produced.

Therefore, it can be seen that complementary gestures, which are slightly incongruent, and mismatching gestures which are very incongruent, both affect early cognitive processing stages of speech. However, only those gestures that are very incongruent have an effect on the late semantic processing. This gives insight into the cognitive processes of the brain highlighting the importance of the integration of both gesture and speech in functional communication.

Moreover, this cognitive processing has been shown to happen incrementally by continuously adding to the global semantic context by building a representation of the discourse. Several studies have investigated the on-line representation of language comprehension using ERPs to show that words are integrated semantically into the context word by word to create a global overall message representation with each upcoming word (Berkum et al., 1999, 2003; Hagoort, 2003a,b; Kutas and Hillyard, 1980; Osterhout et al., 1997).

Özyürek et al. (2007) took this idea of on-line processing of speech and incorporated the natural element of simultaneously occurring gestures into the sentence context. Their aims were to study the global integration of information from gestures into previously spoken sentences and how that compared to word integration. Namely, are both speech and gesture integrated simultaneously, or is one before the other, and is there a difference in gesture integration to previous context relating to the global message, compared to gesture integration to the simultaneous context relating to the local message?

Özyürek et al. (2007) manipulated the semantic congruency of the critical verb in speech and/or gesture relating to the preceding part of the sentence stimuli affecting global integration, and the congruency of the simultaneous gesture/speech in the subsequent portion of the sentence stimuli affecting local integration. This was done by aligning iconic gesture video segments to the speech to produce the following four conditions: a) a correct condition where both gesture and language matched the preceding sentence, b) a language mismatch condition where only the spoken part was incongruous to the preceding sentence, c) a gesture mismatch condition where only the gesture was incongruous to the preceding sentence, and d) a double mismatch condition where both gesture and language were incongruous to the preceding sentence, yet congruous with each other locally.

The results of the experiments showed a statistically significant difference in all 3 mismatching conditions in comparison to the correct condition with an N400 effect peaking around 480 ms with an anterior distribution. Thus, co-occurring speech and gesture were integrated simultaneously (350-550 ms after gesture and word onset) in relation to preceding sentence context.

Since in most cases the N400 effect occurred before the end of the spoken or gestural information, it was an indication of the immediacy of the integration of both. And since the distribution effects were identical for both gesture and word effects, the authors suggest that the cognitive processing used to integrate the semantics of both are very similar.

Also, it is important to note that the double mismatch (local match) condition did not affect the N400's latency in comparison to the language and gesture mismatches (local mismatch). Nor did the local mismatch conditions have any effect on the global mismatching condition. The results suggested that verbs and gestures, even though occurring simultaneously, are not integrated as one unit before being integrated with the preceding content. Instead, it seems integration occurs in parallel immediately, rather than in stages or steps of semantic organization.

Due to the fact that the effect was frontally distributed and also occurred close to the 300 ms time range, and the fact that visual stimuli involving images were used, it is not unlikely that there was also an N300 effect. Since all three mismatching conditions produced a similar N300 effect with similar topographical distributions for both modalities, it could be suggested that semantic integration for both visual and verbal semantics uses overlapping neuronal sources. This conclusion supports the idea that language comprehension uses semantic information from several modalities simultaneously as well as using the same neurological sources which very much highlights the issue of a tightly interconnectedness of speech and gesture.

To test if gestures alone influence comprehension ability when in the absence of speech, Wu and Coulson (2005) used iconic gestures to study their semantic processing. By using iconic gestures that were congruent/incongruent with the context of previously presented cartoons, they expected to detect an N400 effect. They also expected to find an N300 effect due to the fact that the experiment involved visual semantic information. The test involved showing a short cartoon, then a congruous or incongruous gesture video clip, and then a related or unrelated probe word. Participants were simply asked to respond to if the gesture was describing the cartoon or not.

The participants responded to congruous gestures statistically significantly faster than to incongruous ones. As expected, there were ERP congruency effects at 300 ms (N300) post onset and peaks around 458 ms which Wu and Coulson (2005) called the N450. The peaks were larger for incongruous gestures. The authors also found a positive going deflection around 740 ms called the LPC (late positive component), which was elicited by congruous gestures. The probe words produced N1 and P2 ERPs followed by N400s, which were more negative for unrelated words.

The N450 effect was found since it is produced similarly to the N400, but since it was in response to videos of actions, which Wu and Coulson (2005) claim take longer to process, the peak occurred later. Nevertheless, the effect suggests semantic processing of gestures similarly to standard N400 effects for verbal stimuli. In a follow up experiment participants were only asked to indicate if the probe word was related or unrelated. In this case there was still an N450 effect for the gestures, but no LPC deflection. This indicated that while there was no overt attention to the gestures, they created a semantic response anyway. Since there was no LPC, Wu and Coulson (2005) suggested that there was a dissociation between the N450 and LPC and that the dissociation was task driven in the first experiment.

A second finding was that there was a modulation of word comprehension when preceded by congruous gestures. Even though related words required the same button press, when preceded by incongruous gestures (which was not task related) there was a larger N400. This did not happen for unrelated probe words. This ordered congruous gesture-word pairing advantage is typical for lexical priming studies where primes are used to activate perceptual features shared by the probe. Examples include names of concrete objects preceded by pictures of those objects which produced attenuated N400s in comparison to cases with unrelated pictures. In conclusion, the presentation of gestures before words affected the processing of the related words because they activated stored knowledge and created expectations about the words that were going to follow them creating a facilitation effect.

To investigate how speech and gesture affect the interpretation of cross-modal communication in real time, Wu and Coulson (2007) studied the contribution of these two modalities on discourse comprehension. They showed videos of discourse involving descriptive speech and iconic gestures and then presented a probe picture that was congruous or incongruous to the spatial information provided from the video. There were four types of probe pictures: cross-modal congruency with both speech and gesture, speech-only congruency (not related to gesture), and two unrelated probes pairing each probe with an incongruous discourse prime video. In each of these pairings, the speech and gesture were complementary and nonidentical or redundant. For example, when the speaker said "Two throw pillows", the gesture was about their location on the opposite ends of a couch. The results showed a much larger N300 and N400 for unrelated items in comparison to related ones, and much larger ERPs for the cross modal stimuli in comparison to the speech-only condition.

It is also important to note that the N300 effect was larger on the anterior right hemisphere than the left one. There was no N300 for speech-only probes suggesting that the N300 was a visual response to pictures and that co-speech gesture affected image-specific semantic processing. This shows that the N300 can be used to measure how iconic gestures facilitate the identification of multimodal communication.

The N400 was much larger for the cross-modal condition than the speech-only condition. Inversely, the cross-modal related probes reduced the N400 amplitude the most. This suggests that cross-modal probes fit the semantic context of the dialog more accurately versus speech-only probes. It also shows that cross-modal communication is much easier to understand than speech-only. The smaller amplitude of the N400 for cross-modal related stimuli suggests that gestures provide semantic information related to the content of the speech enforcing the robustness of the intended message. Therefore it can be concluded that semantic activity in the brain engendered by speech and iconic gestures cooperatively contributes to conceptual understandings of the discourse.

There is also an interaction, or at least an overlap, between object recognition and comprehension of depictive gestures, which is what Wu and Coulson (2011) wanted to look at directly. When observing pictures of objects, stored knowledge is required to process them, but with depictive gestures, such as iconic gestures, a higher cognitive process requiring supportive context, such as concurrent speech is needed for the listener to form a meaning. The hypothesis of this study was that the listener, in order to understand the depictive gesture's semiotic elements, used some of the same cognitive features used to integrate photographs of objects. Since the N300, like the N400, is affected by context congruity, yet is not affected by degrees of relatedness (Hamm et al., 2002; McPherson and Holcomb, 1999), while reflecting image based processing in long-term memory (Schendan and Kutas, 2002, 2007; Schendan and Maher, 2009; West and Holcomb, 2002), it seemed ideal to use it in comparison to the N400.

This experiment used the same stimuli as Wu and Coulson (2005) but the authors also used static gesture freeze frames from the dynamic gesture video clips. Therefore the cartoon clips were paired with congruent and incongruent gesture videos and pictures. Yet to compare neurological activity from static gestures as well as visual representations such as objects, a second experiment was done using related and unrelated photographs of objects. ERPs showed an N300 effect only for static gestures, and an N400 effect in both dynamic and static gesture conditions. Both pictures of objects as well as depictive gestures produced an N300 and N400 effect when unrelated which were anteriorly located and lateralized to the right. This finding of the N300 effect, in response to static gestures, showed that the object recognition process in the brain is part of gesture recognition; otherwise the N300 would not be modulated by relatedness. Therefore, this study suggested that the cognitive process involved in object recognition is also affected by the semantic properties of iconics. More specifically, recognizing and categorizing percepts of a static gesture picture is modulated by the properties of the gesture. The configuration of hands when performing iconic gestures are more or less similar to the contours and shapes of objects in a picture, and therefore both are processed using overlapping cognitive areas.

Iconic gestures do not only contribute to the conceptual meaning, but can also help with disambiguation as shown by Holle and Gunter (2007). Since iconic gestures are somewhat dependent on (speech) content, they cannot be used alone. Therefore, in situations where the speech has the potential for ambiguity, iconic gestures can help solve the ambiguity. In Holle and Gunter (2007)'s experiment, a speaker uttered the first half of a sentence, which included an ambiguous homonym and one of two co-speech gestures (one referring to the dominant meaning of the homonym and one referring to the subordinate meaning), followed by a disambiguating sentence, which was congruent with either the dominant or subordinate meaning of the homonym. Therefore there were four possible combinations of preliminary sentence and co-speech gesture pairs with each concluding sentence with a target word identifying exactly what the homonym was referring to (dominant (D) or subordinate (S)).

The reaction time was longer for incompatible gesture-target word pairs and had more errors (DS and SD compared to DD and SS). The N400 was largest when a dominant target word followed a subordinate gesture and visa-versa. This shows that the meaning of the homonym changed considerably in accordance with what the gesture was. Since the reaction time was longer, the accuracy was lower, and the N400 amplitude was larger for incompatible pairs, the experiment results showed that much more cognitive load was required to process it. This also showed that subjects used the gesture to create meaning out of something otherwise unknown and that iconic gestures created strong contextual cues which modulated processing of subsequent words. Furthermore, this study showed that gestures that support the subordinate homonym, can actually facilitate the processing of the related target word during compression. Therefore, in a situation with ambiguity, a listener can save energy by using the gestural information to facilitate his/her ability to understand it. Also, this potentially saved neural resources which can be used for something else such as attention to body posture or prosody for example, which would have otherwise been missed due to the cognitive load of trying to disambiguate the speech.

6.2 Neuroscientific Studies of Emblematic Gestures

As mentioned in Chapter 2, emblems have a specific meaning for a given culture. Theories regarding the ontogenesis of emblems suggest that they could have begun as iconic gestures, which were ritualized, and overtime became a code for a specific meaning (Kendon, 1981).

While some fMRI experiments studied emblems by modifying their affective content (Gallagher and Frith, 2004; Knutson et al., 2008) or by comparing them to other gestures/conditions (Montgomery and Haxby, 2008; Villarreal et al., 2008), Xu et al. (2009) wanted to investigate if symbolic gestures and spoken language were processed by the same system in the brain. They compared the neurological response of emblematic gesture videos to their direct equivalent in speech. They wanted to determine if there was much overlapping brain activity or if emblems and speech had their own networks. The results showed that there were many overlapping patterns of connection between the frontal and temporal regions of interest during the process of spoken language and symbolic gesture production.

However, there were also some differences so Xu et al. (2009) concluded that the anterior and inferior temporal regions may be part of a high level extracting system dependent on the type of modality in which the information is encoded from, i.e. whether speech or symbolic gesture. Therefore, these modality specific regions may extract salient information from the communicative devices from where they are then sent to a modality-independent communicative system where they are converted into conceptual representations. So while the posterior temporal regions connect the communicative modalities with their semantic features, the inferior frontal regions use higher cognitive processing to guide the selection and integration process with world knowledge that can be understood in the brain. In short, the meaning, or that which is signified, is paired with symbols, which are the signs from words, gestures, images, sounds, objects, and so forth.

This investigation was corroborated by the study done by Andric et al. (2013) who studied the different and overlapping brain regions involved with processing emblematic gestures and goal-oriented hand action. They did this by showing videos of actors producing emblematic gestures, uttering synonymous phrases to emblematic gestures, and producing goal-oriented hand actions such as reaching to grab an object. They found that processing emblematic gestures used similar brain responses involving processing the meaning of language, as well as processing hand actions. The lateral temporal and inferior frontal areas were activated when meaning was conveyed in speech and hand gestures. The parietal and premotor regions responded to hand actions for both emblematic gestures and action-oriented ones. Therefore, there are overlapping yet distinct brain responses that organize the recognition of perceived actions as well as the interpretation of symbolic meaning and emblematic gestures.

The source of this higher level processing produces evoked responses which Lau et al. (2008) argue are very similar to the N400. In order to study if emblematic gestures are processed in the brain similarly as words, but using EEG to focus on temporal resolution, Gunter and Bach (2004) conducted a study using both meaningful and meaningless emblem hand gestures. It had been previously shown that pseudowords produce an N400 effect in comparison to words (Bentin, 1986; Bentin et al., 1985) and therefore it was assumed that the same effect would occur with emblems and pseudo-emblems. Also, since concrete words, or words referring to picturable objects produce an anterior N400 effect rather than the central-parietally located N400 which abstract words produce (Holcomb et al., 1999), emblems were thought to produce the same ERP output as concrete words. Finally, anomalous pictures processed semantically in sentences produced an anteriorly distributed N300 and N400 (Federmeier and Kutas, 2001), and therefore it was assumed that pictures of emblematic gestures processed semantically would do the same.

In the experiment by Gunter and Bach (2004), semantic processing was required to differentiate the meaningful and meaningless hand postures in a semantic categorization task. Since the experiment used pictures, an N300 effect was expected for the meaningless hand postures and since the emblems were abstract, they also expected a classical N400 effect in response to the meaningless ones.

The results showed an effect for meaning in the 300-400 ms time window lateralized to the right anterior, which was identified as an N300. There was also an effect for meaning in the 450-550 ms time window in the posterior region without any lateralization, which was identified as a classical N400. This experiment showed that meaningless hand postures produce an N300 and N400 in comparison to meaningful emblem gestures, indicating that their processing is very similar to that of abstract words. Therefore, it seems reasonable to assume that emblems and words share a common semantic representation in the brain.

With the exact same hand gesture stimuli as Gunter and Bach (2004) and using the same meaningful versus meaningless paradigm, yet with MEG (magnetoencephalography), Nakamura et al. (2004) investigated which brain regions process emblematic gestures. The results showed synchronized neuronal activity in the inferior parietal, superior temporal sulcus, and inferior occipito-temporal regions when participants engaged in emblem recognition. Gunter and Bach (2004) concluded that the primary visual, mirror neuron, social recognition and object recognition systems were all activated and involved in this recognition.

Husain et al. (2012) further investigated the neurological effects of emblematic gestures in the brain but using fMRI instead and compared regular healthy hearing people with deaf sign language users. This study used “thumbs up” and “thumbs down” for meaningful gestures and the participants had to perform a category discrimination task. While the deaf participants used more bilateral auditory processing regions in the temporal cortex, more importantly, the hearing participants used mirror neuron system in the premotor cortex as well as the inferior parietal lobule. The fMRI highlighted regions for emblematic responses are not dissimilar to the regions where the N300 can be found in the frontal cortex lateralized to the right side, and the N400 found in central-parietal region.

As a means to investigate the interaction between words and symbolic gestures, Fabbri-Destro et al. (2014) primed verbs with congruent or incongruent emblems. This was to verify if symbolic gestures primes behave equally as when they are probes. 20 videos lasting 2 seconds were used each paired with a congruent and incongruent verb. Fabbri-Destro et al. (2014) found that there was a longer response time for the incongruent gesture/verb pairs since there was no semantic priming, and there was also an N400 effect for incongruent pairs.

To address the issue of how spontaneous gesture comprehension is processed in the brain, Proverbio et al. (2015) primed gestures such as emblems, deictics, and iconic gestures with congruent and incongruent short sentences. Congruent picture probes produced a posterior P300 and incongruent ones produced an anterior N400 effect.

Willems and Hagoort (2007) addressed the question of how the presence or absence of co-speech gesture affects the neural processing of speech using an fMRI. In this study subjects watched an actor tell a story with natural gestures, with 1) 'self-adaptor' movements (e.g. scratching, touching face, moving glasses), 2) arms at rest, or 3) no visual information. The activated regions involved Broca's area and the mirror neuron system. Most interestingly, when speech was accompanied by natural gestures, Broca's area used the least amount of activity since the amount of semantic information provided by the co-speech gestures made comprehension for the listeners easy so that less semantic control was required. The increased sources of information from the natural gesture condition, in comparison to the no gesture and self-adaptor conditions, decreased the need for semantic control making the work load easier for proper comprehension.

It is therefore clear that the visual aspect of gesture is so crucial to spoken language. Likewise it is easy to see why including these gestures in written text could significantly help our reading ability since they add the physical and visual dimension missing from physical communication. It is also why there is a high demand for emojis in text and this can explain the continuous expansion of new emojis.

6.3 Neuroscientific Studies of Emojis

Evidence towards a specific mechanism in the brain that is used for facial recognition goes back almost half a century ago. Yin (1969) tested subjects' ability to remember faces and other objects which were "mono-oriented", meaning they were typically found in one type of orientation. He found that remembering faces presented upside down was significantly more difficult than remembering other inverted objects showing that not only does the brain have a familiarity with mono-oriented objects, but there is a special mechanism related specifically to faces. Collishaw and Hole (2000); Farah et al. (1998) contributed to this understanding showing that this effect occurs with both familiar and unfamiliar faces.

Rossion and Gauthier (2002) hypothesized that this face inversion effect had to do with our brain's perceptual encoding of images. Since inverting an image changes the configuration of the features that make up a stimulus, humans' configural processing ability is disrupted even if the features that make up the configuration are easily identifiable. Therefore, since this inversion effect causes more of a disruption with faces than objects, it seems that faces are perceived through configural processes while other objects are perceived through featural processes. Yet when both are inverted, they are processed featurally (Maurer et al., 2002).

This effect can be observed through ERPs. The N170 ERP is a negative going EEG deflection occurring 170 ms after the onset of the stimulus in the occipito-temporal region (Bentin et al., 1996). This ERP shows a reliable response to the inversion effect in that its latency and amplitude are increased for inverted faces (Eimer, 2000; Rossion et al., 2000).

Allison et al. (1999); McCarthy et al. (1999) identified more ERP negative going activity in the occipito-temporal regions around 200 ms which are larger in amplitude in response to whole faces than parts of faces or inverted faces. These neurological regions could therefore be primarily involved in the configural processing of faces. However laterally to these areas, McCarthy et al. (1999) found cortical regions which produce large amplitudes in response to parts of faces instead of whole faces, meaning they could be part of the feature processing system.

It is therefore likely that the N170 is due to both configural and featural processing but the featural information produces more of an affect to the component (Bentin et al., 1996). Sagiv and Bentin (2001) added to this proposing that the increased latency and amplitude of the N170 is due to the fact that upright faces are processed configurally, while inverted faces activate the feature processing system contributing to the amplitude and the time taken to process this.

Because of the reliability of the N170 and its usefulness as a metric in knowing if something is processed configurally or featurally, Churches et al. (2014) investigated the processing of emoticons. Traditional emoticons are made up of typographic symbols representing eyes, a nose and mouth, and when they are together they look like a sideways face. To study if emoticons are processed configurally and not featurally, the inversion of the emoticon should revert the meaning of the punctuation marks to be their default typographic meaning of colon, hyphen and closed parenthesis, and thus produce a reduced N170 since no face is recognized. Therefore Churches et al. (2014) investigated the neurological response to inverted emoticons, natural faces and strings of typographic characters.

The results of their experiments showed that inverted emoticons produce reduced N170s indicating that neither the configural face processing nor featural face processing regions were activated. This meant that the inverted emoticons were less actively perceived as faces. These results were contrary to natural face responses, where inverted faces produced standard increased amplitude and latency effects to the N170. These results further supported the fact that faces activate the configural processing regions of the brain which produces a smaller and earlier N170 than the lateral feature specific regions, which respond to inverted faces since configural processing can no longer function with the face image. There was no inversion effect to the strings of typographic characters supporting the notion that non face-like stimuli do no contribute to the N170 since neither the configural nor featural face processing regions were activated. This study showed that emoticons are perceived as faces only through configural processes, since once they are inverted and the configuration is disrupted, the parts of the emoticon are not perceived as facial features and therefore not processed featurally.

Neural activity regarding emoticons has also been investigated using fMRI as in the study of Japanese emoticons by Yuasa et al. (2006). Since Japanese emoticons are oriented upright instead of to the side like western emoticons are, they are more representational of real faces and thus closer in appearance to real faces and useful for neurological analysis. In this study three experiments were conducted: 1.) comparing face images and non-face images 2.) comparing emoticons and non-emoticons and 3.) comparing sentences with and without emoticons, which were either congruent or incongruent to the preceding sentence. Under the three different experiments, the activated areas of interest were those that have been shown to be related to the emotional valence decision tasks on faces in previous studies. The right fusiform gyrus was activated in response to photographs of faces in the first experiment but not when emoticons were presented in the second experiment. Since the right fusiform gyrus has been known to be activated during the perception of faces, it follows that it should be activated when photos of faces are presented, but the facial representation of emoticons was not significant enough to activate this area. Interestingly, this area is not far from the N170 facial recognition ERP located at occipital-temporal sites. However, the right inferior frontal gyrus, which is associated with the emotional valence decision task, was activated for both experiments. Therefore, Yuasa et al. (2006) inferred that seeing emoticons activated emotional valence detection without being perceived as faces. The right middle frontal gyrus and right inferior parietal lobule were both activated during the face discrimination tasks in the two experiments. The right inferior frontal gyrus and the right middle frontal gyrus were activated in all experiments but slightly different in the third experiment, which involved congruent and incongruent emoticons. Yuasa et al. (2006) believed that this area is involved with working memory. This is because the subjects read a sentence, memorized the content, and then saw if the following emoticon was congruent or not. Interestingly, these emoticons also activated seemingly similar areas in the brain such as the N400 semantic congruency ERP and N300 object identification ERP did regarding semantic processing and picture processing.

Yuasa et al. (2011) obtained similar results that suggested that emoticons stimulate the brain very similarly as faces do and are perceived as nonverbal communicative devices. They therefore can be assumed to enrich digital communication. Yuasa et al. (2011) investigated how we perceived emoticon enriched sentences using highly abstract emoticons. Subjects had to interpret semantic meaning of text along with the facial expressions of emoticons composing a basic verbal and nonverbal message similar to real life. The study used a number of sentences without emoticons and some with emoticons at the conclusion of the sentences where the emoticons were either congruent or incongruent to the sentence.

Much like the results from Yuasa et al. (2006), activation was detected near the right inferior frontal gyrus which is responsible for emotion discrimination, the left inferior frontal gyrus where Broca's area is and is responsible for text comprehension and syntactic judgment, and the anterior cingulate gyrus which is responsible for facial expressions. These results show that while emoticons activated areas responsible for facial recognition, the regions required for emotional indication similar to other nonverbal communication were also activated. Also, the fact that both left and right inferior frontal gyri were activated suggests that verbal and nonverbal communication were divided so adding emoticons to text activated these regions more significantly with verbal processing on the left and nonverbal on the right.

Emoticons have recently evolved in nature to be emojis, which are now more realistic looking with shapes, colors, vertical in orientation, and sometimes moving. Because of the previous literature, one can see the potential for neurological enhancement with using emojis in short messaging in that it could possibly produce faster and more affective processing.

Comesaña et al. (2013) used emojis in a masked priming experiment to investigate if there would be a masked affective priming effect. What this effect means is that presenting a prime stimulus with a mask (or a distraction presented before the stimuli of interest that disrupts the visibility of it), with an affective valence (positive or negative), will affect the behavioral and physiological responses to a related or unrelated probe presented thereafter. The study's primes varied between type (emoji or word) and valence (positive or negative) with many probes thereafter. Affective priming occurs when the response to the probe stimulus is different when it is congruent or incongruent to the prime.

This study found a masked affective priming effect with emoji primes but not with words, which was analyzed with two ERPs being the N2 and LPC. The N2 effect was observed with emoji primes congruous to negative probes where the amplitude was larger than with incongruent pairs. This N2 priming effect occurring only to negative emojis and not words is consistent to research suggesting an enhanced processing of negative stimuli since there is greater saliency to threatening stimuli.

The LPC effect was observed at 450-600 ms after prime onset (not probe onset as is done in most studies). Just like with the N2, the effect only occurred for emojis but inversely, the effect occurred with incongruous pairs instead of congruous ones. The big difference here is that the LPC was greatest with positive stimuli and much less for negative ones.

The dissociation between priming negative words for early components such as the N2 and positive words for late components such as the LPC implies different brain regions being activated for saliency purposes. This is important when considering the saliency and threatening-ness of stimuli. It is very important to respond quickly (automatically) to threatening stimuli, which is why negative stimuli created the effect at the early stages of processing, and once the stimuli were evaluated as not threatening, positive incongruous stimuli produced the most neural activity. This study showed the automatic processing of emojis and their saliency over words. This effect supports the idea that emoji processing functions faster than word processing in some circumstances.

Jolij and Lamme (2005) performed a study investigating the subconscious processing of emojis and their affective valence evaluation. They did this by using transcranial magnetic stimulation to create virtual lesions in the brain which temporarily deactivated the visual cortex. When people were presented emojis with a positive or negative valance (happy or sad face), while their visual cortex was deactivated, even though they couldn't see the face, they were still able to guess the correct affect. This remarkable ability is called "blindsight" and while normally witnessed with patients with real lesions in their primary visual cortex, this study used healthy people. This study showed that people were able to detect emotions from emojis, without the conscious processing of facial features. This further supports the notion that emojis stimulate the same automatic processes that faces do.

6.4 Conclusion

Emojis (and emoticons) produce emotional responses just like their physical representations do in face-to-face communication. These emotional responses produce a supplemental response to that of processed written language alone. As Kendon (2004) discusses, gestures contribute to the experience that a listener has with the utterance, allowing it to be more informative, produce better imagery, or evoke a more emotional response than it would have been without the gesture. While gestures have shown to be used to disambiguate speech and facilitate the processing of target words during sentence comprehension, there is ample evidence suggesting that pictures of gestures similarly supplement reading comprehension and in turn, emojis can do the same. This supplementation could not only produce more accurate and more comprehensible text, but could potentially also increase speed of reading, as well as saving cognitive resources for other tasks. Therefore the implementation of visual information in text such as pictures of gestures and facial features could augment readers' ability in regard to speed, accuracy, or cognitive load if done strategically and properly.

7

Introduction to Experimental Set Up and Follow Through

This chapter describes the experimental set up of the experiments which are presented in the following chapters. Each of these chapters describes a single experiment presented in article form. The set up for each experiment involved a data acquisition computer (See Figure 7.1), which recorded the EEG from the participants and the stimulus presentation computer (See Figure 7.2) which showed the participants the stimulus material. The stimulus material for each experiment is explained and shown in its chapter (in the Method and Appendix sections) as are the corresponding EEG waveforms (in the Results section). There are statistical evaluations of these waveforms which are explained and presented in each experiment, as well as behavioral results describing participant accuracy and reaction time when performing the experiment. Even though EEG is being recorded throughout the duration of the experiment, a behavioral task such as pushing a correct/incorrect button is essential to measure behavioral phenomena as well as ensure that the participants are engaged in the tasks.

At the start of my PhD project, I wanted to work with a system that could be used at a commercial/user level meaning it had to be simple, cheap, and wireless so I chose the Emotiv 16-electrode headset (Emotiv) which is shown in Figure 7.3. The Emotiv headset has 16 electrodes, but only 14 are used as channels for waveform analyses, while the remaining two are reference electrodes. Experiments were performed in my office because I wanted to use non-medical grade EEG equipment that was user friendly. I also wanted that my experimental environment was natural and not medical. Partic-

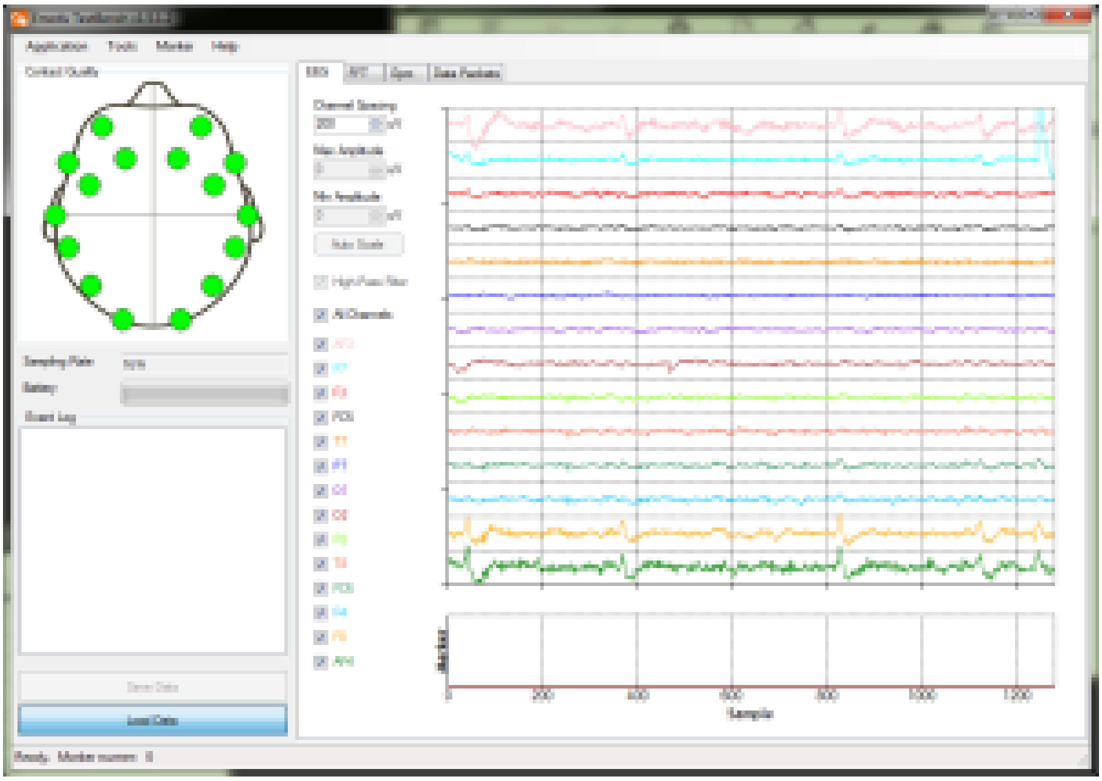


Figure 7.1: A screen shot from the data acquisition computer of the Emotiv software which shows the position of the electrodes on the left, and the raw EEG signal for each electrode on the right.

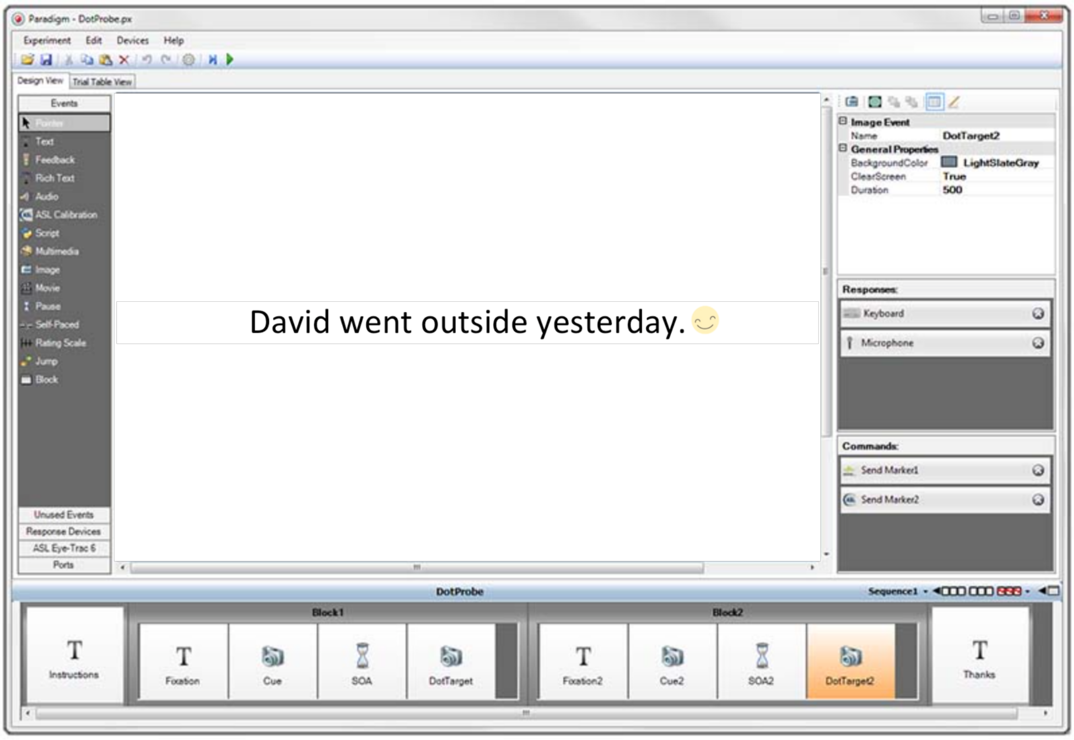


Figure 7.2: A screen shot from the stimulus presentation computer.

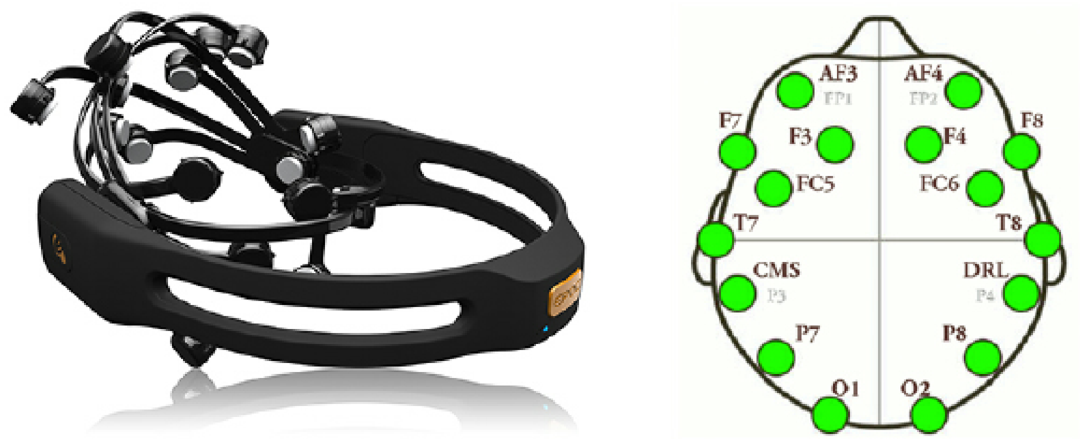


Figure 7.3: A picture of the Emotiv headset as well as the positions of its 16 electrodes. Taken from `https://www.researchgate.net/profile/Nicola_Catenacci_Volpi/publication/284031383/figure/fig1/AS:296815892156418@1447777831163/Fig-1-The-Emotiv-EPOC-and-the-electrodes-location.png`.

ipants were recruited through university classes, networking, and friendships and all signed informed consent contracts indicating their awareness of their tasks as an unpaid volunteer.

All EEG waveforms are presented in the unit measurement of micro-vaults which is not shown in the article pictures. An example is in Figure 7.4.

Also it must be noted that the electrodes which were used in measuring changed through the entire experimentation process. At first I used all electrodes to get a grand average, then I selected T7, FC5, F3, F4, FC6, and T8 because these 6 seemed more centrally located, but then upon further research experience of looking at scalp maps, I realized that P7, O1, O2, and P8 were in fact closer to the N400 origin at the Pz position. To test if this was true, I re-analyzed my results from Chapter 9. In the article I used all 14 electrodes, while in the test I only used the back 4 electrodes (P7, O1, O2, and P8) and my results of statistical significance improved from ($F(1,15) = 5.36$, $p = 0.035$) to ($F(1,15) = 5.405$, $p = 0.026$) confirming my assumption that the back electrodes are closer to the N400 origin at the Pz position.

It must also be noted that the order in which the articles are presented in this thesis is not in accordance to how they were conducted chronologically. Chapters 9 and 11, were finished before Chapter 8 and then the final adjustment to use the back 4 electrodes was implemented. This is why Chapter 8 uses 6 electrodes, Chapters 9 and 11 use all 14

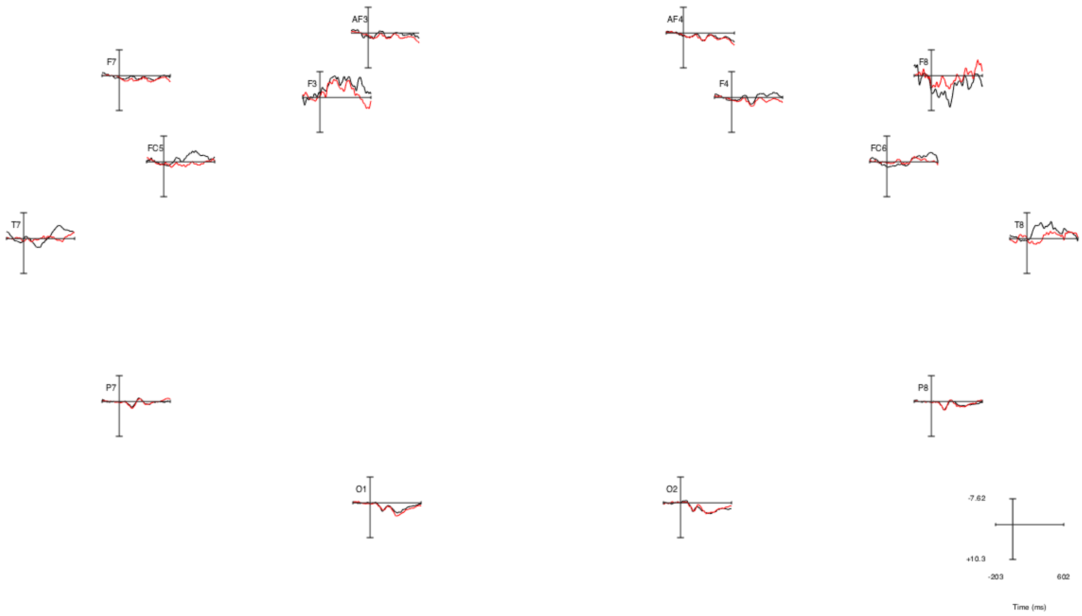


Figure 7.4: Wavelengths from the time window of -203 to 602 ms, where negativity is plotted upwards, and with congruous (red) and incongruous (black) conditions. Waveforms are displayed in microvolts.

electrodes, and only the last two EEG experimental Chapters 12 and 13 use the more accurate back 4 electrodes.

Reaction time was provided by giving both stimuli and button press onsets specific time stamps and this deference was automatically recorded through MatLab's ERPLab. Physiological data was also measured in ERPLab using the mean under the curve of the wavelength in a specific time window and computing a grand average. The output from the Emotiv are so-called edf files which I imported in ERPLab via a MatLab script. All statistical significance testing was done with a p level being 0.05. Student's T-Test and repeated measure ANOVAs were done on the behavioral data for button press reaction times as well as the physiological EEG data in response to experimental probes. The ANOVAs were performed on SPSS and you can see an example of the results from Chapter 11 in Figure 7.5 where a 2x2x3 ANOVA was done. The reaction time statistics for the congruency test are "Congruency (F(1,237) = 289.67, p = 5.5865E-43)". The degrees of freedom for Congruency are the number of types for congruency (congruity and incongruity) minus 1 and the total number of participant responses (238) minus 1.

In the conducted EEG experiments I test different aspects of the relationship be-

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
con	Sphericity Assumed	57602457.3	1	57602457.3	289.677	.000
	Greenhouse-Geisser	57602457.3	1.000	57602457.3	289.677	.000
	Huynh-Feldt	57602457.3	1.000	57602457.3	289.677	.000
	Lower-bound	57602457.3	1.000	57602457.3	289.677	.000
Error(con)	Sphericity Assumed	47127632.3	237	198850.769		
	Greenhouse-Geisser	47127632.3	237.000	198850.769		
	Huynh-Feldt	47127632.3	237.000	198850.769		
	Lower-bound	47127632.3	237.000	198850.769		
mode	Sphericity Assumed	693608.714	1	693608.714	5.977	.015
	Greenhouse-Geisser	693608.714	1.000	693608.714	5.977	.015
	Huynh-Feldt	693608.714	1.000	693608.714	5.977	.015
	Lower-bound	693608.714	1.000	693608.714	5.977	.015
Error(mode)	Sphericity Assumed	27501107.3	237	116038.427		
	Greenhouse-Geisser	27501107.3	237.000	116038.427		
	Huynh-Feldt	27501107.3	237.000	116038.427		
	Lower-bound	27501107.3	237.000	116038.427		
position	Sphericity Assumed	1579846.06	2	789923.030	5.196	.006
	Greenhouse-Geisser	1579846.06	1.977	799067.503	5.196	.006
	Huynh-Feldt	1579846.06	1.994	792440.193	5.196	.006
	Lower-bound	1579846.06	1.000	1579846.06	5.196	.024
Error(position)	Sphericity Assumed	72060993.9	474	152027.413		
	Greenhouse-Geisser	72060993.9	468.576	153787.345		
	Huynh-Feldt	72060993.9	472.494	152511.863		
	Lower-bound	72060993.9	237.000	304054.826		
con * mode	Sphericity Assumed	51277.284	1	51277.284	.380	.538
	Greenhouse-Geisser	51277.284	1.000	51277.284	.380	.538
	Huynh-Feldt	51277.284	1.000	51277.284	.380	.538
	Lower-bound	51277.284	1.000	51277.284	.380	.538
Error(con*mode)	Sphericity Assumed	31945478.4	237	134791.048		
	Greenhouse-Geisser	31945478.4	237.000	134791.048		
	Huynh-Feldt	31945478.4	237.000	134791.048		
	Lower-bound	31945478.4	237.000	134791.048		
con * position	Sphericity Assumed	826088.490	2	413044.245	2.820	.061
	Greenhouse-Geisser	826088.490	1.997	413627.520	2.820	.061
	Huynh-Feldt	826088.490	2.000	413044.245	2.820	.061
	Lower-bound	826088.490	1.000	826088.490	2.820	.094
Error(con*position)	Sphericity Assumed	69422247.5	474	146460.438		

Figure 7.5: 2x2x3 ANOVA for Congruity, Mode and Position. Using the Sphericity Assumed test there was an effect for Congruency ($F(1,237) = 289.67$, $p = 5.5865E\text{-}43$), Mode ($F(1,237) = 5.97$, $p = 0.015$), and Position ($F(2,474) = 5.196$, $p = 0.006$), Congruency x Mode ($F(1,237) = .38$, $p = 0.538$), Congruency x Position ($F(2,474) = 2.82$, $p = 0.061$). No other interactions were significant.

tween emojis and text. First, I conduct a replication experiment so that I can ensure that the technology I use, which is far more simplistic than traditional medical grade EEG systems, can measure the same neurological responses from the replicated experiment. Then, I investigate semantic priming with the stimuli from the first experiment, to test if mixing text with emojis can produce semantic congruity. Thirdly, I use the same data from the preceding experiment to comment on how behavioral responses to this system of using text and emojis, functions comparably to how gesture stimuli in face-to-face communication do. In my fourth experiment, I merge the forms of text and emoji, from prime and probe stimuli, into a single probe stimuli, in order to determine in which formation/order the union works best. The fifth experiment is an elaboration on this so that the insights from the fourth experiment are used to place emojis into a complete sentence to determine whether emojis provide supplemental meaning to the text. After that, an online survey is presented investigating where in a sentence emojis are most commonly understood and what they refer to in their different placements, which is then used to investigate semantic congruity of sentences with moving emojis. Finally, I test the response to sentences with three types of emojis (hand gesture, pictures of faces, and moving faces) along with three different placements (before the sentence, after the object the emojis modify, and at the end of the sentence). All of this is in a effort to establish some kind of pattern or system that can be used to scientifically supplement written text with visual stimuli that produces the best response.

8

Replication of meaningful gesture study for N400 detection using a commercial BCI

Thomas Ousterhout. Abstract of article accepted for publication in 4th European and 7th Nordic Symposium on Multimodal Communication.

8.1 Abstract

In an effort to test the ability of a commercial grade EEG headset to effectively measure the N400 ERP, a replication study was conducted to see if similar results could be produced as that which used a medical grade EEG. Pictures of meaningful and meaningless hand postures were borrowed from the original author and subjects were required to perform a semantic discrimination task. The N400 was detected indicating semantic processing of the meaningfulness of the hand postures. The results corroborate those of the original author and support the use of some commercial grade EEG headsets for non-critical research applications.

KEYWORDS: EEG, ERP, N400, Semantics, Congruency, Gestures, Emotiv

8.2 Introduction

This study was designed to promote and validate the functionality of commercially available and user friendly neuroimaging technology as a brain-computer interface (BCI) and

Electroencephalography (EEG) research tool. Developments in cognitive technologies are allowing researchers and users to access cognitive information in a cost effective manner. EEG is a measurement tool used to detect and measure the electrical signals in the brain when neurons communicate with each other. While invasive, cortically-implanted electrodes, allow for a more precise method of measuring brain activity, non-invasive scalp electrodes allow for a much more appropriate scientific method for the average researcher and user (Lin et al., 2008).

BCIs have shown an incredible ability to allow those with mobility disabilities to control medical devices such as prosthetic limbs (Farwell and Donchin, 1988; Guger et al., 1999; Müller-Putz and Pfurtscheller, 2008; Nunez and Srinivasan, 2006), wheelchairs (Barea et al., 2002a,b, 2003; Chowdhury and Shakim, 2014; Rebsamen et al., 2006, 2007) and robots (Chowdhury et al., 2014; Neto et al., 2006; Tripathy and Raheja, 2015). One reason this is possible is due to the ability to predict voluntary human movement more than a second before it occurs (Bai et al., 2011; Funase et al., 1999; Morash et al., 2008).

However, BCI's are not just used for mind controlled vehicles or devices using cognitive thought, they can also perform as diagnostic tools to detect driver fatigue (Jap et al., 2009; Lin et al., 2008; Zhao et al., 2011) and drowsiness (Eoh et al., 2005; Khushaba et al., 2011; Lin et al., 2010; Rosario et al., 2010). This shows that applications using BCIs range from medical purposes for people who are locked in a vegetative state and helping them communicate with the world, to gaming/recreational purposes for healthy users who want to enhance their lives with smart technology.

While many EEG and BCI systems use medical grade technology as a data acquisition tool, the relatively cheap and wireless BCI system called the Emotiv EPOC is a cost effective consumer grade EEG unit with only 14 channels and this system has proven effective in several studies (Campbell et al., 2010; Debener et al., 2012; Ousterhout and Dyrholm, 2013; Vos et al., 2014a,b). However the technology is still controversial as there are some studies that do not support its use fully (Duvinage et al., 2012, 2013; Liu et al., 2012; Stytsenko et al., 2011), stating that the system, being significantly worse than standard medical grade EEG, should only be used in noncritical applications. One noncritical application could certainly be communication.

When people communicate face-to-face, they typically engage in multimodal communication which simultaneously uses both modes of auditory and visual information.

Auditory information normally only consists of speech, and visual information can include things like body behavior such as facial expressions, hand and arm gestures, and body posture. Visual information is also used inter alia to disambiguate context by providing supplemental information to the dominantly used vocal information, can be used instead of speech, and can change the meaning of speech (Goldin-Meadow, 1999; Kelly et al., 1999; Kendon, 2004; McNeill, 2005). While the auditory modality provides the most information content in face-to-face communication, thus typically being the dominant modality of communication, the visual cues are very important and sometimes necessary to understand fully what the intended message is (Clark, 1996).

Hand gesturing, for example, is an integral part of our daily communication paradigm. Hand gesture types can be categorized into several groups, while simultaneously being part of a larger continuum. For example, one type is called an emblem, which is a hand gesture requiring no verbal supplement, and has a conventionalized meaning in a particular culture, such as the "thumbs up" gesture in western cultures. These gestures can be useful in face-to-face communication because one gesture alone can give a complicated message to the recipient instantaneously (McNeill, 1992). Therefore emblems can be considered unspoken words or phrases, since there is a strong relationship between the gesture and its meaning. Another type are iconic gestures, which are used to symbolize something, such as putting one's hands in the shape of a ball when talking about a ball. There are also deictic gestures, which comprise pointing hand postures.

Gestures thus contribute to the semantics of the dialogue in face-to-face communication. Semantics can be measured with EEG by looking at event-related potentials (ERPs), which are amplitude deflections in the brain produced in response to certain events or stimuli. One ERP has been studied for over thirty years to measure semantic processing (Duncan et al., 2009; Gunter and Bach, 2004; Kutas and Federmeier, 2000, 2011) it the N400 ERP which is a negative deflecting component occurring 400 ms after the onset of a auditory or visual stimulus.

This ERP is used to measure semantics because when there is an incongruous stimulus, in relation to a congruous one, or one that is expected, there is a much larger negative deflection. Thus, the semantics of sentences, videos, or any other stimulus type can be measured to see if they are congruous or incongruous with the preceding context by looking at this ERP amplitude deflection. But not all electrode positions can measure this ERP since the responses to abstract words in semantic processing are typically found in centro-parietal sites, while concrete words, such as ones referring to picturable objects, have a frontal distribution (Holcomb et al., 1999).

The N400 has also shown utility in its ability to measure the amount of cognitive load required for an individual in semantic memory retrieval. This is because the ability to process the information from probe stimuli is highly dependent on one's ability to recall previous relevant stimuli from any of the multimodal channels such as images or sounds. This difficulty, or cognitive load, is associated with memory representations and cues from previous content priming the meaningful probe stimulus Federmeier and Kutas (2001); Lau et al. (2008); Petten and Luka (2006). Therefore, when a difficult stimulus requires more effort to process, thus having more cognitive load, the N400's amplitude deflection is larger than when it is easy. It is therefore that the N400 is larger for rarely used words and when semantically incongruent or unrelated to previously acquired content (Laszlo and Federmeier, 2011; PETTEN, 1995).

A study done by Gunter and Bach (2004) investigated the N400 effect using pictures of semantically meaningful and meaningless hand postures. Pictures of 11 common and well-known emblematic, iconic, and deictic gestures were used as the meaningful stimuli along with 11 similarly positioned yet meaningless hand positions. During the pictorial semantic categorization task, subjects were required to identify, through a button press response, if each randomly displayed picture was meaningful or meaningless. They found that in comparison to meaningful hand positions, the meaningless ones produced a larger negative going amplitude deflection in the centro-parietal region, which they classified as the N400.

This current study presumes to replicate precisely the study done by Gunter and Bach (2004) in an attempt to also find the N400, despite the fact that the Emotiv has no electrodes positioned that can measure the centro-parietal region. The hypothesis is that since the N400 is such a large ERP, even with the poor resolution of the 14-channel Emotiv, in comparison to 59-channel medical grade EEG scalp cap, the Emotiv will still be able to detect the N400 from the meaningless hand postures.

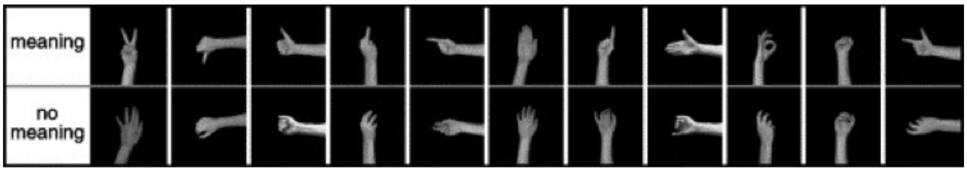


Figure 8.1: The 11 meaningful and similarly positioned yet meaningless hand postures provided by Gunter and Bach (2004).

8.3 Method

8.3.1 Participants

This study used 16 participants who were native English or fluent English speaking adults at the University of Copenhagen. Their ages ranged from 20-37 years (mean = 26.9), 9 were males and all were right handed. All participants signed an informed consent form ensuring their understanding of the experiment to be conducted. All participants had normal or correct-to-normal vision with no reported psychiatric, neurological, or reading disorders that could disrupts this study's efficacy.

8.3.2 Stimuli

Participants were presented with stimuli courtesy of Gunter and Bach (2004) which consisted of 66 meaningful and 66 meaningless grey-scale hand posture photos. Each of the 11 meaningful and meaningless hand postures seen in Figure 8.1 were photographed by six different people and all 132 pictures were shown in 3 cycles.

8.3.3 Procedure

Using Paradigm stimulation software, a trial of the discrimination task progressed first with a random hand posture for 700 ms, then a blank screen for 500 ms, and finally a "GO" signal was presented indicating that the participant had to input with a button press if they judged the hand posture as meaningful or meaningless. This lasted approximately 20-25 minutes for each participant. Since the Emotiv is designed for real world applications, the study was done in a closed university office with normal lighting conditions and the possibility for auditory noise outside.

8.3.4 Electrophysiological Acquisition

For EEG acquisition, the 14-channel Emotiv was used which has electrodes at the International 10/20 system at AF3, F7, F3, FC5, T7, P7, O1,O2, P8, T8, FC6, F4, F8, AF4 with two left and right mastoid references at P3 and P4. The data was filtered offline from 0.1 to 30 HZ and sampled continuously at 128 HZ. To support the use of the Emotiv in the real world involving noisy environments in real time, no artifact rejection or correction was applied, however only correct responses were used. ERPs were identified and measured off-line using Matlab's ERPLab with a baseline averaged from the -200 to stimulus onset interval window and average ERPs lasted 1000 ms after the onset of the probe.

8.4 Results

8.4.1 Electrophysiological Results

The ERPs were measured using a repeated measures ANOVA with a 2 x 6 design (Meaningfulness x Electrode) using only the 6 electrodes F3, FC5, T7, T8, FC6 and F4 since they were closest to the PZ electrode position which is typically used to measure the N400. The mean amplitude for these 6 electrodes was calculated within the time window of 300-500 ms after stimulus onset. Figure 8.2 shows the grand average wavelengths of meaningful and meaningless stimuli and Figure 8.3 shows the scalp map distribution in the measurement time window in 50 ms intervals. The ANOVA Sphericity Assumed test showed an effect for meaningfulness ($F(1,15) = 5.36$, $p = 0.035$) and thus was identified as an N400. No other significant effects were found.

8.5 Discussion

In summary, this study investigated the N400 effect regarding meaningless hand gestures compared to meaningful hand gestures made up of emblem, iconic, and deictic gestures. This study also replicated the paradigm and reproduced the results of Gunter and Bach (2004) regarding N400 detection. Most importantly, this study gives further evidence to support the use of a simple and affordable BCI as a research and user tool for noncritical EEG/ERP/BCI applications.

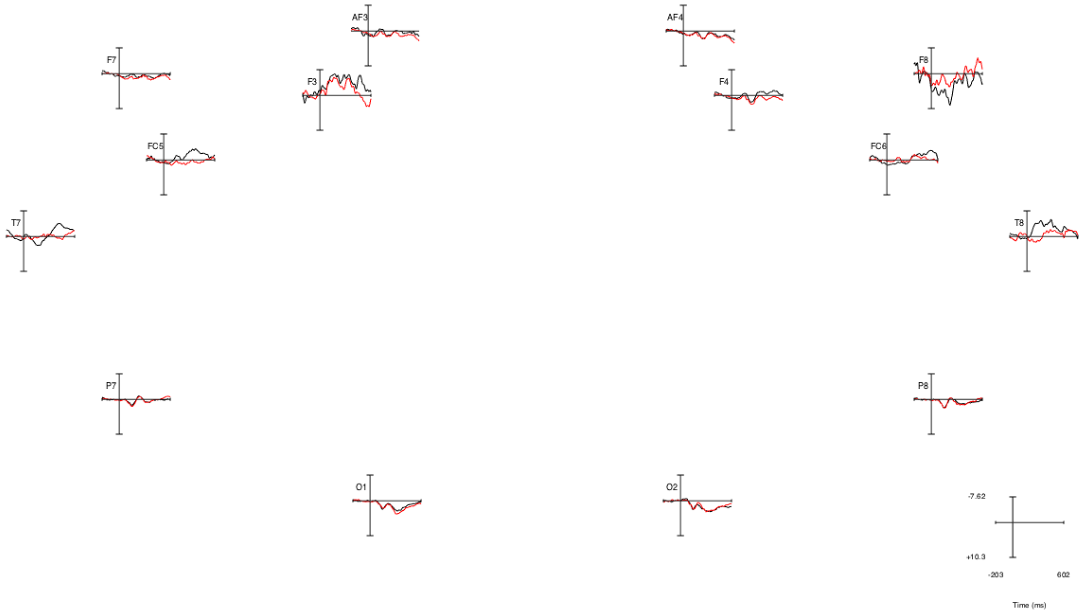


Figure 8.2: Wavelengths from the time window of -203 to 602 ms, where negativity is plotted upwards, and with congruous (red) and incongruous (black) conditions. Waveforms are displayed in microvolts.

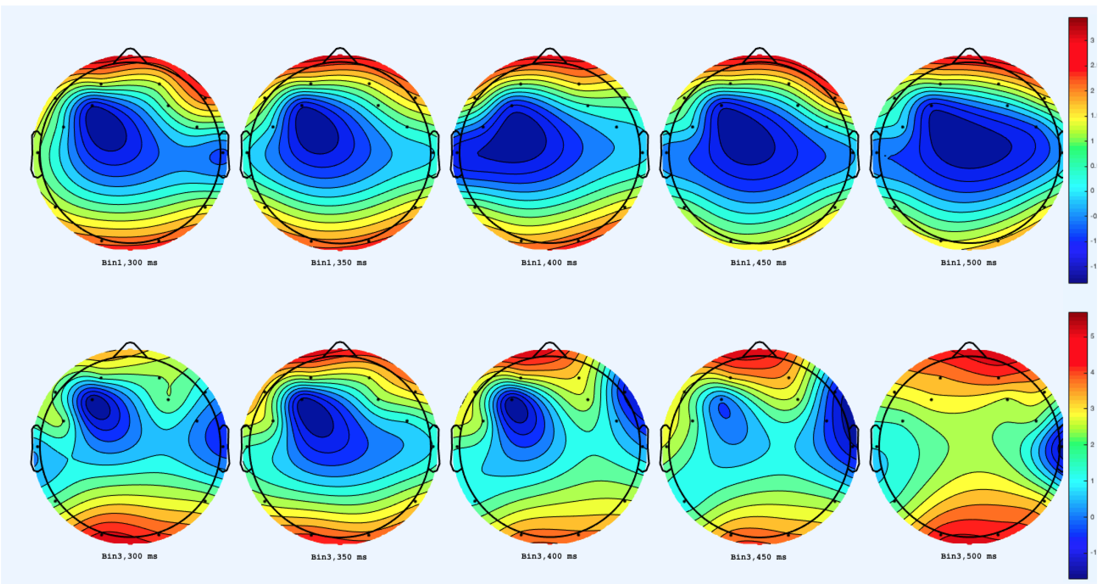


Figure 8.3: Scalp maps of incongruous (top) and congruous (bottom) conditions from 300 to 500 ms in 50 ms intervals displayed in microvolts.

This study further corroborates with previous research regarding the issue of if some meaningful hand postures, such as emblems, are lexicalized and thus processed in the brain like words are. The theory was that since the comparison between meaningful words and similar yet false pseudo words produces an N400 effect, the same would go for meaningful hand gestures and similar yet false meaningless hand gestures. The increased N400 of the meaningless hand postures in comparison to the meaningful ones is similar to results shown by Bentin (1986); Bentin et al. (1985).

The only difference between the result of this study and those done by Gunter and Bach (2004) are that there was an N300 effect with right-frontal distribution in that study which is indicative of picture processing and thus should have also been seen in this study (Barrett and Rugg, 1990; Federmeier and Kutas, 2001; McPherson and Holcomb, 1999; West and Holcomb, 2002). This current study found a greater negativity lateralized towards the left. The cause of this difference is unknown but will be investigated. However, Gunter and Bach (2004) did mention that the N300 and N400 effects were relatively small and could have been due to the large repetition of stimuli in the experiment and also could potentially have been facilitated with priming. Another potential explanation for the difference could be that Gunter and Bach (2004) used 22 native-German speaking students, where this study used 16 students from countries all over the world. This cultural difference could have had a dramatic effect on the semantic processing of the meaningful hand postures.

However, the most important part of this study is the demonstration that a simple, cost-effective, 14-channel EEG headset can detect the N400 in a similar manner that medical grade 59-channel EEG systems can. Even more impressive is that the system worked without any electrodes covering the source of the N400, and the data acquisition was done in a regular room with real world auditory and visual distractions, and no type of artifact rejection or correction was done. This further supports the usefulness of the commercially available EEG equipment, such as the Emotiv, as a research tool for ERP detection and user interface for BCIs.

8.6 Acknowledgements

I would like to thank Thomas Gunther for the permission to use his stimuli in my study.

9

N400 congruency effects from emblematic gesture probes following sentence primes

Thomas Ousterhout. N400 congruency effects from emblematic gesture probes following sentence primes. In Intelligent Engineering Systems (INES), 2015 IEEE 19th International Conference on, pages 411–415. IEEE, 2015b.

9.1 Abstract

Emblematic gesture pictures were presented to subjects as probes in relation to semantically congruent and incongruent sentences to investigate if there is a similar cognitive processing network for congruity as there is with words. Subjects had to perform a simple discrimination task while undergoing EEG recordings. The ERPs elicited by semantically incongruent gestures produced larger N400 and possibly N300 components. While the N400 is indicative of semantic processing showing that emblematic gestures are lexicalized, the N300 likely reflects the cognitive processing of picture stimuli.

KEYWORDS: Emotiv, ERP, N400, N300, Semantics, Congruency

9.2 Introduction

In multimodal communication, gesture is a visual modality that can help support context by assisting with disambiguation or potentially providing supplemental information

(Goldin-Meadow, 1999; Kelly et al., 1999; Kendon, 2004; McNeill, 2005). Although the auditory modality has the highest information content in face-to-face communication, the visual modality cues such as gestures and lip movements are also important (Clark, 1996). One type of gesture, called an emblem, has a very clear meaning such as the "thumbs up" gesture. These gestures are very useful in communication because one simple hand position can provide a very complex message instantaneously (McNeill, 1992). Since emblems can be considered unspoken words or phrases, they are lexicalized. Note that there is a linguistic difference between emblematic gestures, and sign language, which use syntactic and morphological information (Supalla, 1986).

The N400 event-related potential (ERP) is an electrophysiological component that has been thoroughly studied for over thirty years as being indicative of semantic processing (Duncan et al., 2009; Kutas and Federmeier, 2011). N400s have a negative amplitude deflection around 400 ms after stimulus onset for both auditory and visual modalities (Kutas and Federmeier, 2000) and the more incongruous the stimulus is with preceding events, the larger the deflection. Electrophysiological responses to abstract words in semantic processing can be found in centro-parietal electrode cites, where concrete words, or those referring to picturable objects, have a more anterior distribution (Holcomb et al., 1999).

The N400 has been used as a measurement device to index how much cognitive load is required in memory retrieval to process the knowledge previously acquired that is associated with the relevant multimodal stimuli (i.e. images, sounds) and this task's difficulty is entirely dependent on memory representation as well as cues from previous content leading up to the meaningful probe stimulus (Kutas and Federmeier, 2000; Lau et al., 2008; Petten and Luka, 2006). Because of the cognitive load required to process more difficult stimuli, N400s are larger for rarely used words, and when semantically incongruent or unrelated to the previous content (Laszlo and Federmeier, 2011; PETTEN, 1995).

However the N400 isn't the only component involved with semantic processing. Anomalous pictures at the conclusion of sentences create a frontal N300 effect as well as a standard anteriorly distributed N400, as seen in the studies by Federmeier and Kutas (2001); Gunter and Bach (2004). The N300 is a component which appears to be isolated to picture processing (Barrett and Rugg, 1990; Federmeier and Kutas, 2001; McPherson and Holcomb, 1999), indicating that picture specific semantic processing, and not the processing of physical characteristics of the subject or object in the picture, activates a neurological system different from what is used for verbal semantic processing. The N300 has been suggested to indicate nonverbal semantic processing (West and Holcomb, 2002) and semantic categorization of visual objects into semantic groups (Hamm et al., 2002).

Gunter and Bach (2004) used pictures of meaningful emblematic hand gestures as well as very similarly positioned, yet completely meaningless, hand gestures to investigate N300s and N400s during a pictorial semantic categorization task. They found in comparison to meaningful emblems, meaningless hand postures elicited both frontal N300s and centro-parietal N400s. It was noted that the effects were relatively small, and therefore they suggested that a priming paradigm most likely would have facilitated the effect.

Word and picture semantic facilitation effects were found by Bajo (1988) for cross-form (word-picture and picture-word) and within-form (word-word and picture-picture) stimuli. In their study, semantic facilitation occurred during category verification for both picture and word probes for all form combinations.

This study used a similar cross-form paradigm where sentences were presented as primes to facilitate the semantic congruency/incongruency categorization task for the emblem picture probes. The hypothesis is that the incongruent probes of emblem pictures will be semantically processed as any other lexical modality and elicit an anterior N400 effect similar to concrete and picturable words as well as N300 effects similar to picture probes. Also I plan to see faster reaction time to semantically related versus unrelated probes.

While many EEG studies use medical-grade data acquisition systems, the 14-channel Emotiv has been successful in several ERP studies (Campbell et al., 2010; Debener et al., 2012; Ousterhout and Dyrholm, 2013; Vos et al., 2014a,b), as well as a modified Emotiv system (Zich et al., 2014). Other studies have not been so supportive of the Emotiv (Duvinage et al., 2012, 2013; Liu et al., 2012; Stytsenko et al., 2011), indicating that it is significantly worse than medical grade EEG systems in regard to the accuracy and speed of the data acquisition, and technical issues as well, and therefore should only be used for noncritical applications.

9.3 Method

9.3.1 Participants

36 native English or fluent English speaking adults volunteered for their participation, all of who had completed collegiate level education. Their ages ranged from 19-57 years (mean = 31.2), 19 were female and 33 were right handed. All participants had normal or corrected-to-normal vision and reported no history of psychiatric or neurological disorders or current use of medications thought to affect the central nervous system. The experiments were conducted with the understanding and written consent of each participant.

9.3.2 Stimuli

Stimuli consisted of 31 short and simple sentences or meaningful sentence fragments, ranging from 1-8 words long. Each sentence was paired with two probes being a congruous and incongruous emblematic hand gesture image. See Figure 9.1. The 11 different emblematic images were provided courtesy of Gunter and Bach (2004). A preliminary test was done by five master students at the University of Copenhagen who read each sentence and paired it with the emblem they though was most congruous and incongruous. These five did not participate in the EEG experiment. Only sentences that were paired with at least an 80% agreement for congruency with a particular emblem were used from the original 43 sentences (18 had 100% and 13 had 80% agreement). There was little-to-no consistency with which emblem was selected as most incongruous to sentences (only one had 80% agreement and 11 had 60% agreement, the remaining 19 were selected from one of the 5 possible incongruous emblem matches provided).

9.3.3 Procedure

The 31 sentences were presented in full for as long as the participant wanted. This was to ensure comprehension and as a means for the subjects to take breaks as needed. Once the sentence was read and fully understood, subjects were instructed to push any key after which a fixation cross would appear for a random duration between 750 and 1250 ms. Then randomly, either of the two selected emblem probes would appear, where the participants were instructed to push a key accordingly for congruent or incongruent as

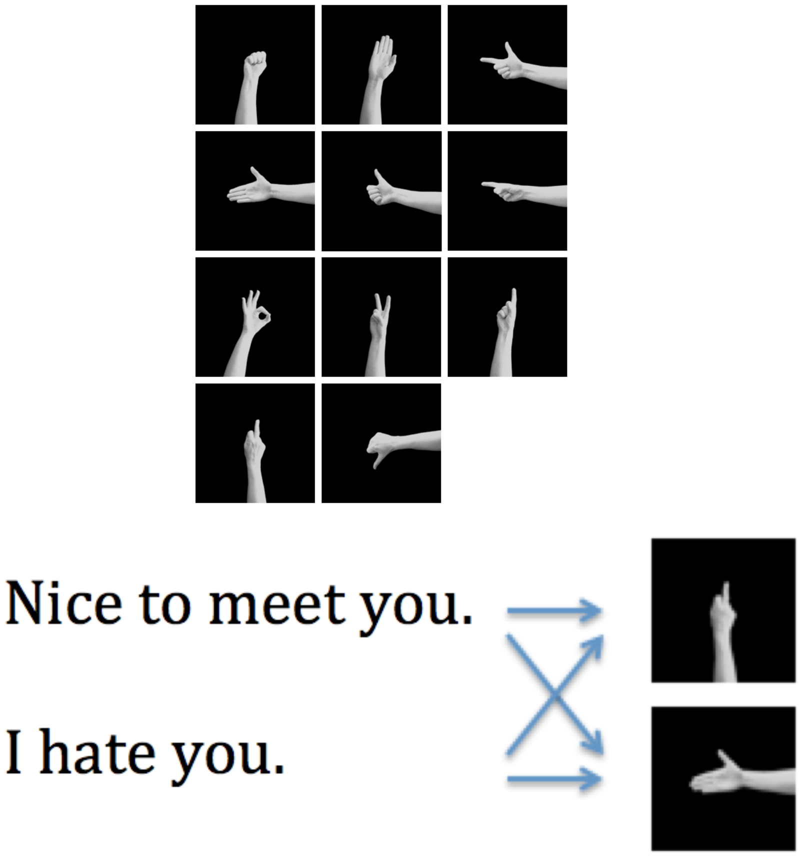


Figure 9.1: All 11 emblematic gestures and an example of 2 sentences each paired with a congruous and incongruous gesture.

fast and accurately as possible. Each cycle consisted of all 62 trials for each sentence-congruous/incongruous pair and each cycle was completed 3 times taking roughly half an hour for each participant. Participants were randomly instructed to use either their left or right hand throughout the experiment's duration.

9.3.4 Electrophysiological Acquisition

For EEG acquisition the Emotiv was used, which has 14 EEG channel locations based off of the International 10/20 system at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 with left and right mastoid references (P3, P4). The data was filtered at 0.01 to 30 Hz and sampled continuously at 128 Hz. Since the Emotiv was used and in an effort to support EEG technology in noisy environments in real time, no artifact rejection or correction was applied. Incorrect responses were not used in ERP analysis. ERP components were identified and measured off-line with reference to the average baseline voltage from the interval window of -200 ms to stimulus onset. Average ERPs starting from -200 ms and lasting 1000 ms after probe onset were computed for each electrode for each position. The data were processed and analyzed using Matlab, Matlab's EEGLab and ERPLab, and SPSS.

9.4 Results

9.4.1 Behavioral Results

The subjects' mean reaction time to congruent data was 909.4 ms with a standard deviation of 539.96 ms and an accuracy of 95.4%. Their reaction time to incongruent data was 985.67 ms with a standard deviation of 544.06 ms with an accuracy of 96.8%. A paired two-tailed t-test resulted in a difference that was significant ($p = 2.91E\text{-}12$).

9.4.2 Electrophysiological Results

The ERPs were analyzed using a repeated measures ANOVA with a 2 x 14 design (Congruency x Electrode) and using the mean amplitude under the curve for each electrode within the time window of 300-500 ms after probe stimulus onset. Figure 9.3 shows the grand average for each electrode as a response to congruent and incongruent hand gestures. The ANOVA Sphericity Assumed test shows an statistically significant effect for semantic congruency ($F(1,35) = 4.90$, $p = 0.03$) with larger negative deflecting

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Congruency	Sphericity Assumed	580.442	1	580.442	4.902	.033
	Greenhouse-Geisser	580.442	1.000	580.442	4.902	.033
	Huynh-Feldt	580.442	1.000	580.442	4.902	.033
	Lower-bound	580.442	1.000	580.442	4.902	.033
Error(Congruency)	Sphericity Assumed	4143.907	35	118.397		
	Greenhouse-Geisser	4143.907	35.000	118.397		
	Huynh-Feldt	4143.907	35.000	118.397		
	Lower-bound	4143.907	35.000	118.397		
Electrode	Sphericity Assumed	3882.123	13	298.625	6.770	.000
	Greenhouse-Geisser	3882.123	3.470	1118.843	6.770	.000
	Huynh-Feldt	3882.123	3.898	995.873	6.770	.000
	Lower-bound	3882.123	1.000	3882.123	6.770	.013
Error(Electrode)	Sphericity Assumed	20070.145	455	44.110		
	Greenhouse-Geisser	20070.145	121.442	165.266		
	Huynh-Feldt	20070.145	136.437	147.101		
	Lower-bound	20070.145	35.000	573.433		
Congruency * Electrode	Sphericity Assumed	78.394	13	6.030	.619	.838
	Greenhouse-Geisser	78.394	4.627	16.944	.619	.673
	Huynh-Feldt	78.394	5.418	14.469	.619	.698
	Lower-bound	78.394	1.000	78.394	.619	.437
Error (Congruency*Electrode)	Sphericity Assumed	4429.555	455	9.735		
	Greenhouse-Geisser	4429.555	161.936	27.354		
	Huynh-Feldt	4429.555	189.630	23.359		
	Lower-bound	4429.555	35.000	126.559		

Figure 9.2: Raw data of the ANOVA.

amplitudes for incongruent probes versus congruent ones. The ANOVA Sphericity Assumed test shows an effect for electrode position (F(13,455) = 6.77, p = 5.9778E-12) with the greatest difference between conditions in electrode F8. The explanation for how these statistics were produced was given in Figure 7.5 and Figure 9.2 shows the raw data.[1] Table 9.1 shows the mean amplitude within the time window of 300-500 ms for each electrode position under congruent and incongruent conditions with their respective difference.

[1]The degrees of freedom for Congruency are the number of types for congruency (congruity and incongruity) minus 1 and the total number of participants (36) minus 1. The degrees of freedom for Electrode effect are the number of electrodes (14) minus 1 and the product of the first DF number and the number of participants (36) minus 1 (13 x 35 = 455).

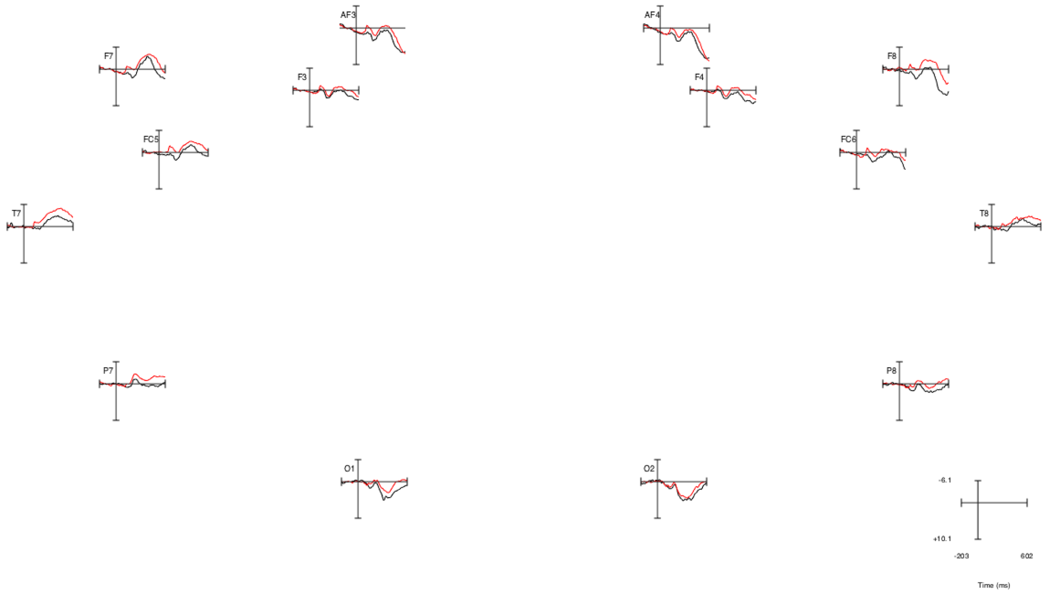


Figure 9.3: Emotiv electrode positioning from time window -203 to 602 ms where negativity is plotted upwards and with congruous (black) and incongruous (red) waveforms. Waveforms are displayed in microvolts.

	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4
Congruent	2.14	-1.803	0.61	-1.145	-2.451	0.451	3.692	4.321	1.947	-1.4	0.614	1.548	1.52	2.959
Incongruent	0.412	-3.504	-0.485	-2.6	-4.438	-1.464	1.64	3.256	0.437	-2.455	-0.536	0.055	-1.789	1.587
Difference	1.728	1.701	1.095	1.455	1.987	1.915	2.052	1.065	1.51	1.055	1.15	1.493	3.309	1.372

Table 9.1: Mean amplitude in microvolts within the time window of 300-500 ms for each electrode position under congruent and incongruent conditions with their respective difference.

9.5 Discussion

This experiment compared the processing of congruent and incongruent emblem hand gesture probes to sentence primes. The purpose of this investigation was to see if they were processed in the same manner in which semantically congruent and incongruent words are processed. To do this, an N400 effect where incongruent stimuli produced a large negative going deflection would have to be found. This seemed likely since emblematic gestures are theorized to be lexicalized, meaning that they are essentially words and thus processed similarly as words.

Although N400s are largest at centro-parietal sites (Kutas and Federmeier, 2011), and the Emotiv not only does not have central electrodes, but also has a deficiency of parietal electrodes, large negative going amplitude deflections were found at almost all 14 electrode cites. This statistically significant difference in amplitude deflection within the 300-500 ms time window is considered to be a standard N400 effect. N400 effect detection was manageable because pictures, instead of words, produce more of an anteriorly distributed N400 effect (Ganis et al., 1996; Gunter and Bach, 2004; McPherson and Holcomb, 1999). Not only was there an effect with congruency, but also electrodes sites, where the greatest deflection was seen for electrode F8 which is located at the anterior right side of the scalp and could be related to the N300. Because of these two effects, there is neurological corroboration that emblem gestures have symbolic meaning as words do and create a similar semantic response.

The reaction time for the emblems was statistically significantly shorter for congruency versus incongruency. This supports the semantic priming effect and shows that the response time was facilitated by semantically related meaning. Interestingly enough, the accuracy was higher for incongruent probes versus congruent ones meaning it was easier to detect incongruities than congruities.

It is important to note that in this study, there were different levels of semantic congruity and not all the emblems categorized as semantically congruent to the priming sentence had the same relationship. While some were supplemental, such as "Nice to meet you." followed by a "handshake" probe which could be expected to come after the sentence was spoken in dialog, others were more reinforcing, such as "Go to the left." followed by a "pointing left" probe, and some could be considered judgment such as "The concert was too long." followed by a "thumbs down" emblem. While all of these were more or less congruent, especially in relation to the incongruent probes, the semantic processing of each type of congruency could be different. An example of a sentence gesture pair that could provide additional meaning could be "Look at that man." followed by the "middle finger" gesture adding a sense of dislike.

These semantic differences could be a contributing factor to why the standard deviation in reaction time was so high. While some types of semantically congruent/incongruent pairs were easy and therefore faster to differentiate, some pairs could have taken much longer to process due to higher or more abstract levels of processing. Another factor that could have contributed to the large deviation in response time was cultural background. Participants were from many countries and while they were instructed to interpret the emblematic gestures from an "American" perspective, this controlled attention could have resulted in longer processing time and possibly confusion.

Since the time window used in averaging the N400 amplitudes was from 300-500 ms, the results could very well include N300 effects which would corroborate what is already known about semantic processing of pictures (Barrett and Rugg, 1990; Federmeier and Kutas, 2001; Gunter and Bach, 2004; McPherson and Holcomb, 1999; West and Holcomb, 2002). Detection of the N300 is further supported by the fact that the largest difference between congruent and incongruent conditions was at the F8 electrode which is located anteriorly and lateralized to the right. The potential N300 effect further confirms that semantic processing of pictorial stimuli involves a different neurological process than those with semantically processing words. Since N400 and possibly N300 effects were found using mostly frontally located electrodes, these findings support the notion that semantic representation is modulated and processed in the human brain with separate and distinguishable components.

Although there is a large discrepancy of opinion regarding the Emotiv headset, in this particular study it proved itself competent in its performance. Although having a significant disadvantage in lack of electrodes, the study was conducted intentionally in a more real-world environment with no particular outside noise elimination procedures and with no artifact rejection. This was to test and show that this technology could in fact be used at home or school for reading tests for individuals with cognitive reading abnormalities or deficiencies.

9.6 Conclusion

This study investigated if neurological evidence for the cognitive processing of emblematic gestures produced the same effect to that of words in regard to semantic congruity. The results indicated that there is a similar processing network between the forms of gestures and words. While semantic processing for various types of stimuli induce an N400 effect indicating a shared network of amodal semantic representations (Nigam et al., 1992), there are also independent modal semantic networks for pictures and words (Hamm et al., 2002). This study corroborates the findings of Gunter and Bach (2004) showing neurological evidence that emblematic gestures are lexicalized and processed similarly as words are.

Future work could involve creating groups of semantic congruity and incongruity to differentiate emblems that added supplemental information, those that created reinforcement, and those that involved judgment. Also, it could be interesting to compare the same kind of sentence-gesture pairs with sentence-word pairs with directly translated gesture to word combinations. If this was done, it might show a different frontal N400 due to the lack of the N300 effect for word probes.

10

Reaction time for two types of semantically related gestures and words

Thomas Ousterhout and Costanza Navarretta. Reaction time for two types of semantically related gesture and sentence pairs. In Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on, pages 499–503. IEEE, 2015.

10.1 Abstract

The aim of this study was to investigate in an EEG experiment the reaction time of the participants deciding on the semantic congruency of sentences with subsequent pictures of symbolic and iconic hand gestures. In this study, we investigate the reaction time employed by the participants to determine the semantic congruence of sentences and gesture pictures which were semantically related in different ways. In one case the gesture conveyed the same content as the sentence, in the second case it provided additional content to the sentence.

The hypothesis that we wanted to test is that the reaction time would be shorter in the latter case than in the former since preceding corpus-based studies have indicated that speech and gestures conveying the same information are more temporally synchronous than speech and gestures which convey complementary information. Therefore, we expect that it will be harder to process semantic congruency for gestures which

usually are synchronous to speech if they follow the sentences they are related to than in the case where gestures provide additional information. Our study confirms the hypothesis showing that not only the participants reaction time was shorter in the case of gestures which provided additional information to sentences, but also that the accuracy in the semantic congruency task was higher.

10.2 Introduction

Human face-to-face communication is multimodal since people communicate via multiple modalities especially the auditory (speech) and visual (gestures) channels. The relation between speech and gestures is complex but their synchrony and semantic parallelism suggest a common cognitive base (McNeill, 1992, 2005).

The semantic relations which connect gestures and co-speech are many and they partly depend on the semiotic type of the gesture (Allwood et al., 2007; Kendon, 2004; McNeill, 1992). In this study, we focus on symbolic and iconic hand gestures. These gestures convey meaning. Symbolic gestures, also called emblems, have their own meaning and can be used as words, an example being the "thumbs up" gesture (Gunter and Bach, 2004). Iconic gestures are strongly related to the semantics of the entities they are related to (Beattie and Shovelton, 1999; Kendon, 2004; Lis, 2014; McNeill, 1992), an example being moving the index and medium fingers illustrating the act of walking.

Also the temporal relation between co-speech gestures and speech partly depends on the type and function of the gestures. However, previous research has shown that hand gestures often precede the speech with which they co-occur, inter alia Kendon (2004); Loehr (2007, 2012). Bergmann et al. (2011) analyzed the temporal relation of iconic gestures having the same content of speech or having other content, and they found that the asynchrony in onset is larger in the latter case.

Determining how humans process meaning conveyed by various modalities or by the same modality via different modi, is important for both understanding how humans communicate and then integrating this knowledge for presenting and exchanging information through different channels or modalities in human-computer systems. In this paper, we focus on the reaction time used by participants to process written sentences and pictures of hand gestures which are semantically congruent, but in different ways. The reaction time is extrapolated from the results of an EEG experiment which was set up to study semantic congruency between sentences and pictures of symbolic and iconic gestures. In the EEG study, sentences were primes to semantically congruent and incongruent emblematic and iconic gestures. The task was for the participants to engage in this semantic congruency categorization task by simply reading the prime at their leisure, and once they indicated comprehension with a button press, a fixation cross would appear. Then, either a semantically congruent gesture, or a very obvious incongruent one would appear and participants had to push different buttons accordingly as quickly and accurately as possible. The EEG study's hypothesis was that semantically congruent gesture probes would produce a neurological N400 ERP effect, as well as produce a faster reaction time in relation to incongruent probes, both of which occurred.

In the present work, we want to test the hypothesis that the reaction time for semantically congruent gestures following sentence primes will be shorter if the gestural information provided new information to the previous sentence, which we call *information addition*, while it will be longer if the gestural information conveys synonymous meaning as that conveyed by the sentence prime, which we call *information reinforcement*. This is because humans are used to receive *reinforcement* information in synchrony with speech (Bergmann et al., 2011) and thus have to backtrack when this information follows the sentence meaning, while *addition* information can more naturally follow the textual meaning.

The paper is organized as follows. In section 10.3 we discuss related literature and in section 10.4 we describe the experiment and present the results of the study in section 10.5. In section 10.6 we discuss the results, and finally we conclude and present future work in section 10.7.

10.3 Related Literature

In the literature, different but related classifications of the semiotic types of gesture have been proposed inter alia Allwood et al. (2007); McNeill (1992), all inspired by Peirce's work on semiotic signs (Peirce, 1974). In this paper, we focus on pictures of symbolic gestures, or emblems, and iconic gestures. Symbolic gestures are established by means of an arbitrary conventional relation and their function is similar to that of words, while iconic gestures denote their objects by similarity. Iconic gestures include in some classifications metaphoric gestures. The distinction between symbolic and iconic gestures is not clear-cut, see the Kendon's continuum (Kendon, 2004), and in specific contexts, iconic gestures can be used alone to represent meaning.

Gestures can be related to the content of speech in many ways. For example, they can a) convey the same meaning as their co-speech, e.g. if the speaker says "Hi" and waves with the hand to greet the interlocutor, b) provide extra information about what is said by speech, e.g. if the speaker says "take the box" and then shows with her hand that the box is very small, or points to a specific box, c) give information about the speaker's attitude towards speech e.g. showing that she is ironic and that the content uttered by speech should not be understood literally.

Poggi and Caldognetto (1996) distinguish between the following five categories of gesture-speech semantic relation:

1. *repetition*, if gesture and speech bear the same meaning,
2. *addition* if the gesture adds information to word meaning,
3. *substitution* if the gesture has a meaning that is not uttered at all,
4. *contradiction* if the gesture communicates something opposite to what said by words,
5. *no relationship*, if the gesture makes part of a different communicative plan.

Bergmann et al. (2011) call gestures that have the same meaning of speech, *redundant* and gestures that have different content of speech *complementary*. In this paper, we investigate gestures and sentences related by the relations of *repetition* and *addition* according to the classification by Poggi and Caldognetto (1996) or *redundancy* and *complementarity* in the more grain-corned classification by Bergmann et al. (2011).

In the following, we call the two relations for *reinforcement* and *addition*, respectively, to stress that gestures always provide an important contribution to communication. The fact that gestures introduce new information to communication and that the process is similar to that of adding new information through speech is suggested by Esposito et al. (2001, 2002) who analyzed silent pauses and gesture holds in audio- and video-recorded conversations, concluded that the two phenomena have parallel functions. Silent pauses indicate mental activation processes aimed at replacing old information with new information while gesture holds are markers of mental activation processes aimed at replacing old gestural information with new.

Researchers who have investigated the temporal relation between speech and co-speech hand gestures have found that usually hand gestures slightly precede the speech with which they co-occur. These studies have investigated both batonic, deictic and iconic gestures, inter alia Bergmann et al. (2011); Kendon (2004); Loehr (2007, 2012). In particular, Bergmann et al. (2011) find that iconic hand gestures that convey the same meaning of speech are more synchronous to it than gestures that convey another content. They propose two possible explanations to this phenomenon. Either the realization of what they call "redundant" gestures is faster because the redundant aspects of meaning are already activated for the purpose of gesture generation or "redundant" gestures adapt more strongly to the flow of speech.

The perception studies by Leonard and Cummins (2011) have similar results. In these studies, the temporal relation between hand-gestures and speech was artificially modified and participants were asked to rate whether they found the gesture natural or not. The results of the experiments indicate that the participants accepted larger temporal asynchrony if co-speech batonic hand gestures preceded speech than if they followed it.

In EEG studies, the N400 is an event-related potential (ERP) that produces a negative going deflection peaking around 400 milliseconds after the onset of a stimulus (Kutas and Federmeier, 2000). This N400 effect is a neurological response that has been thoroughly studied for over 30 years and can be used to identify semantic congruity since there is a larger amplitude deflection for incongruous stimuli over congruous ones (see Duncan et al. (2009); Kutas and Federmeier (2011) for review). Although this ERP has predominantly been used in studying words, it can be found in many different modes such as actions in videos (reviewed in Sitnikova et al. (2008)), mathematics (Galfano et al., 2004, 2009; Niedeggen et al., 1999), hearing (Mäkelä et al., 2001; Praamstra and Stegeman, 1993) and semantic memory (Kutas and Federmeier, 2000; Salisbury, 2004).

EEG studies are becoming a good means at confirming gesture studies by providing neurological evidence to support experiments' conclusions. Özyürek et al. (2007) investigate semantic congruence between iconic gestures and speech in ERP and find that incongruence between gestures and words are processed similarly in the brain as incongruence between words. In an other ERP study, Habets et al. (2011) find that speech and iconic gestures are integrated most efficiently when the differences in onsets do not exceed a certain time span. Gunter and Bach (2004) found that meaningful and meaningless hand postures produced similar properties to those of abstract words, in regard to neurological comprehension and responses involving semantic representation. This means that studies can involve both words and pictures of gestures to study semantic congruity and use the same neurological response to make the experimental conclusions.

In our work, we investigate the reaction time used to process semantic congruence between sentences and subsequent pictures of iconic or symbolic hand gestures. We compare two types of semantic congruence: congruence between sentences and pictures of gestures conveying the same information (reinforcement) and pictures of gestures conveying new information (addition).

10.4 Method

In the following, we shortly describe the EEG experiment whose reaction time is used in the present study and the results of the study. The experiment was run using Paradigm and Emotiv which is a 14-channel low cost EEG system.

10.4.1 Participants

36 native English or fluent speaking volunteers who had completed collegiate level education participated in the EEG experiment. The age range for the participants was 19-57 years (mean = 31.2), 19 were female and 33 were right handed. Everyone had normal or correct-to-normal vision and no one had any known neurological or psychological disorder that could interfere with the results of the experiment. All participants signed a written consent form ensuring their understanding of the experiment.

10.4.2 Stimuli

31 short and simple sentences or meaningful sentence fragments were used ranging from 1-8 words long. Each sentence prime was paired with both an obvious congruent and incongruent hand gesture image probe. The 11 different emblematic/iconic images were provided from a previous study by Gunter and Bach (2004). Figure 10.1 shows the given gestural pictures and two combinations of sentences with congruent or incongruent

gestures. The first sentence *Nice to meet you.* is congruent with the iconic gestures of wanting to shake hands, while it is incongruent with the middle finger gesture, while the second sentence *I hate you.* is congruent with the latter picture and incongruent with the former one.

44 sentences were originally created by the authors with the intention that four would be most congruous for each of the 11 different gesture types. To test this, a pilot test was run and involved five university master students at the University of Copenhagen who had to pair each sentence with the picture of the gesture they thought most congruous and most incongruous out of the 11 options. They did not participate in the actual experiment later and only sentences that were paired with at least 80% agreement between the five of them in congruency were used from the original 44 sentences. In total, 18 sentence-picture pairs received 100% agreement and 13 pairs received 80% agreement between participants meaning only 31 of all the sentence-picture pairs were used. [1]

10.4.3 Procedure

The 31 sentences were randomly presented in full for as long as the participant needed for comprehension. Once ready, a button press would initiate a fixation cross for a random duration between 750 and 1250 ms. Then either the congruent or incongruent picture probe would appear and the participants had to push a button as accurately and fast as possible indicating whether they found the picture congruent or incongruent with the sentence. The test was composed of three cycles of 62 congruous and incongruous combinations of sentences and pictures. To complete, it took around half an hour. Participants were selected randomly to use their right or left hand throughout the entire experiment.

10.5 Results

In the present study, we used 20 pairs of sentences and pictures of gestures connected by a *reinforcement* or an *addition* relation, and we have only used probes with a correct response since we are comparing different types of semantic congruence. However, we also looked at the accuracy achieved by the participants in the congruence task and

[1] In the appendix, a subset of the examples is given.

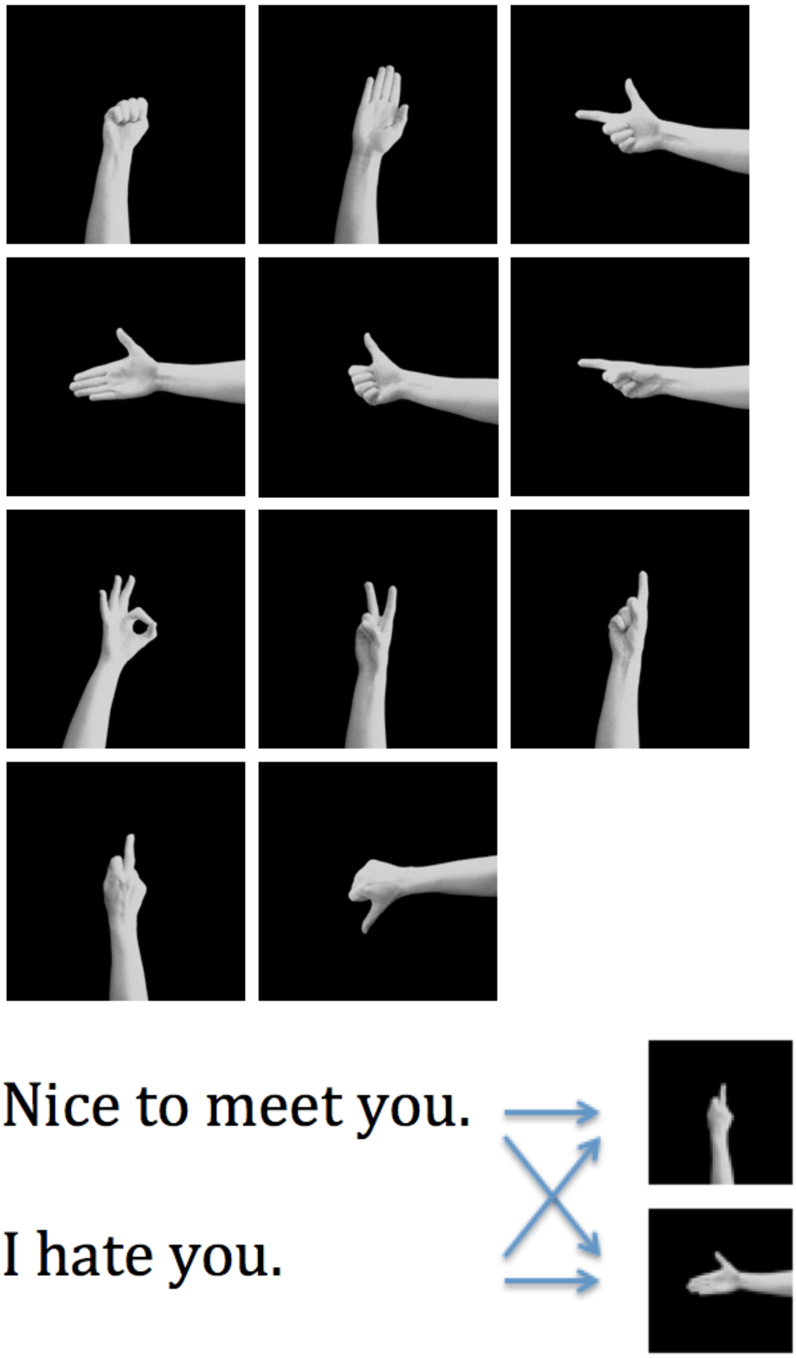


Figure 10.1: All 11 emblematic gestures and an example of two sentences each paired with a congruous and incongruous gesture.

for the *reinforcement* related gestures it was 95.3%, while for the *addition* relation was 96.9%.

A grand mean reaction time for the *reinforcement* type congruity was 934.5 ms with a standard deviation of 559.1 ms. For the *addition* type congruity the grand mean reaction time was 884.9 ms with a standard deviation of 453 ms. Using a two tailed T-test the difference was marginally significant ($p = .055$). [1]

10.6 Discussion

The results of this study indicate that the accuracy of the semantic congruency task was lower if the gestures conveyed the same meaning as the sentence than if they conveyed additional content. Furthermore, the reaction time for deciding whether gestures and sentences were congruent or not was lower in the case of *addition* than in the case of *reinforcement* relation with a statistical significance of 0.055. Thus, our initial hypothesis that the reaction time in identifying semantically congruent pictures of symbolic and iconic gestures following sentences would be shorter in the case of gestures conveying additional information to the text than in the case of gestures conveying the same meaning as it, is confirmed by this study. These perception study results are parallel to those obtained by Bergmann et al. (2011) in a study of the temporal relation between speech and hand gestures conveying meaning in a multimodal corpus of German route-description dialogs. In fact, Bergman et al. find that the gestural onset of co-speech gestures conveying the same meaning as speech is nearer to the onset of the related speech in that corpus than it is the case when the gestures provide information complementary to speech, that is additional but congruent information.

It is natural to assume that humans have more difficulties in processing asynchronous multimodal information that usually is presented to them synchronously. It seems that this is the case independently from the language and from the medium. In fact, Bergman et al. analyzed German face-to-face communication in which two modalities (auditory and visual) were involved while in our EEG study the language was English and the information presented to the participants was unimodal but involved two channels texts

[1] Not all the picture probes are the same in the pairs of *addition*/*reinforcement* sentences primes, but a large subset of them is. A post test of the difference in reaction times of both the *reinforcement* type and the *addition* type in relation to the same congruous picture probe were tested after the article publication and the magnitude of the significance between the two was confirmed.

and pictures of iconic and symbolic hand gestures and text. Thus, our study seems to indicate that the way we present information in face-to-face communication involving auditory and visual modalities influences the way we process the same type of information when it is computerized even if it is presented via other channels. However, our results should be tested on more data and more languages.

10.7 Conclusion

In the article, we tested the hypothesis that the reaction time in a congruency task between sentences and pictures of symbolic and iconic gestures following the sentences would be higher when the pictures show gestures which convey additional semantically congruent information to the sentences than when the gestures convey the same information as the related sentences. The results were extracted from an EEG experiments in which the participants had to decide whether the picture of an hand gesture following a sentence was semantically congruent or not.

Only semantically congruent symbolic and iconic gestures which were correctly identified by the participants were included in the study. The results confirm our hypothesis and are parallel to the results in Bergmann et al. (2011) which show that gestures conveying the same content of speech are more synchronous with the related speech than gestures conveying additional information. Since we work on information presented via two channels in the same modality (visual), our results indicate that face-to-face communication is the natural communication form and also influences the way we process information presented in computerized form. Since the study in Bergmann et al. (2011) is done on an other language and our study concerns restricted types of gesture and semantic relation, the results must be tested on more data. In the future, we will also extend the EEG experiments to other modalities such as various combinations of speech, videos, tests and pictures.

10.8 Appendix

Reinforcement

1. Left.

2. The movie was good.

3. Look to the left.

4. Shake hands.

5. The boy won the game.

6. The show was nice.


Addition

1. There.

2. Sally won the lottery.

3. Look over there.

4. Nice to meet you.

5. The boy ordered drinks.

6. She saw the show.


11

Cross-form facilitation effects from simultaneous gesture/word combinations with ERP analysis

Thomas Ousterhout. Cross-form facilitation effects from simultaneous gesture/word combinations with ERP analysis. In Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on, pages 493–497. IEEE, 2015a.

11.1 Abstract

The aim of this study was to investigate if simultaneous redundant forms of stimuli involving gesture and word combinations would produce a facilitation effect in reaction time as well as produce a neurological semantic incongruity effect called an N400. Event-related potentials (ERPs) were recorded from 14 scalp electrode sites while subject performed reaction-time decision tasks. For each trial, subjects had to indicate if the single form or cross-form stimuli were oriented upwards or downwards, or if they contradicted each other (only applicable in cross-form stimuli were there were two stimuli). The uni-form stimuli were defined by a picture of a man pointing up or down, a picture of just a hand pointing up or down, or various word combinations such as up/down, high/low, above/below, and top/bottom. The cross-form stimuli were combinations of the words and pictures. The results show that while there was no N400 effect, there was however a facilitation effect regarding the positioning of the cross-form

stimuli indicating that people have a tendency to pay more attention to gestures than words for information.

KEYWORDS: Emotiv, ERP, N400, Semantics, Congruency, Facilitation

11.2 Introduction

Unimodality studies (e.g. two visual or two auditory stimuli) have been used in many semantic facilitation studies investigating if several types of stimuli can facilitate identical/synonymous subsequent stimuli of the same modality. These can be either within-form (e.g. word-word or picture-picture) or cross-form (e.g. picture-word or word-picture) and some examples of them include Bajo (1988); Durso and Johnson (1979); Ousterhout and Navarretta (2015) who found that the results were highly dependent on the task requirement. Name verification, where subjects had to determine if a subsequent stimulus had the same name as the previous one, facilitated picture targets. Category verification, where subjects had to determine if a subsequent stimulus was in the same category as the previous one, facilitated both pictures and words (Bajo, 1988).

Redundancy gain, also referred to as the redundant-signal effect, is a common phenomenon found in behavioral studies where performance is enhanced with the presentation of simultaneous stimuli requiring the same response compared to either stimulus alone. The redundancy gain can enhance reaction time (Grice et al., 1984; Hershenson, 1962; Miller, 1982, 1986), increase accuracy (Baird and Burton, 2008; Mohr et al., 1994a, 1996, 2002), as well as increase response forcefulness (Giray and Ulrich, 1993; Mordkoff et al., 1996).

Redundancy gain can reflect probability summation (also known as the race model or separate-activations model) (Guzman et al., 2008; Miller, 1982), where the redundant stimuli (e.g. the response to same target on left, right, or simultaneous visual fields) are presented and the one detected first produces the fastest response. Since the channel processed first produces the fastest response, statistical facilitation has been achieved. The second way to explain redundancy gain is by coactivation (Miller, 2007; Mordkoff and Yantis, 1993) where reaction to stimuli defined by different features (e.g. shape/color) or modalities (e.g. audio/visual) occurs and the coactivation of both is combined to produce a naturally faster response than either alone (Guzman et al., 2008).

The majority of redundancy gain experiments have focused on and demonstrated the effect in cognitively low level tasks such as target detection (Savazzi and Marzi, 2002; Schwarz and Ischebeck, 1994; Veldhuizen et al., 2010). There have however been some redundancy gain experiments that utilized cognitively high level processing which include fame decision judgment for faces (Baird and Burton, 2008; Mohr et al., 2002; Schweinberger et al., 2003) emotion recognition tasks (Collignon et al., 2008, 2010; Tamietto et al., 2006, 2007) and lexical decision tasks (Mohr et al., 1994a,b, 1996; Mullin and Egeth, 1989).

In lexical decision tasks, participants were required to make word/non-word distinction responses. The results showed that they responded faster and more accurately to redundant stimuli versus single-stimulus trials. These results were explained with Hebbian cell assemblies (Hebb, 1949), which is a neurobiological model of language where words are represented neurologically as an assembly of cells in the cortex. Presentation of a specific word activates its designated assembly of cells allowing for a response and when two redundant copies of the word are presented, there is an enhancement in the assembly's activation creating a faster response time.

Words are not the only objects that have been used in lexical decision tasks. Gunter and Bach (2004) did a variation of the word/nonword discrimination task and used meaningful and meaningless hand posture pictures. There is a specific type of hand gesture called emblems (McNeill, 1992), that have a very specific meaning such as the "thumbs up" gesture. This can be understood with no supplementary speech and thus can also be produced at a distance or in a noisy environment. Since these gestures can provide a very complicated meaning in isolation, or along with auditory information in the form of disambiguation (Goldin-Meadow, 1999; Kelly et al., 1999) they are considered as unspoken words, and theoretically lexicalized (McNeill, 1992).

Since emblems are lexicalized, they can be used in semantic processing to produce an N400 effect in EEG studies. The N400 is an event-related potential (ERP) with a negative deflection that peaks around 400 milliseconds after the onset of a stimulus (Kutas and Federmeier, 2000). It is a thoroughly researched ERP that indicates semantic processing (cf. Duncan et al. (2009); Kutas and Federmeier (2011) for review) which can be seen by larger amplitude deflections from incongruous stimuli, versus congruous ones. This effect works with concluding words in sentences, and also works with pictures of hand gestures as demonstrated by Gunter and Bach (2004), who used meaningful and meaningless hand gestures in a categorization task to investigate the N400.

This study further investigated unimodal processing by using simultaneous redundant cross-form stimuli in a category verification task. In this study the cross-form stimuli consisted of synonymous words and emblems, thus being redundant, to see if a facilitation effect would occur with redundant stimuli, and also if an N400 effect would be produced, when cross-form stimuli were incongruous.

The hypothesis was that redundant gesture/word stimuli would produce a facilitation effect and therefore a faster reaction time to their categorization, as well as produce an N400 effect when the stimuli are incongruent.

This study used the Emotiv headset for acquisition instead of medical-grade EEG systems. The Emotiv is a 14-channel low cost system that has proven effective in ERP studies (Campbell et al., 2010; Debener et al., 2012; Ousterhout and Dyrholm, 2013; Vos et al., 2014a,b), including a modified Emotiv setup (Zich et al., 2014). However, it is important to make note of some studies which have not been so supportive of the system (Duvinage et al., 2012, 2013; Liu et al., 2012; Stytsenko et al., 2011), stating that there are significant problems in comparison to medical-grade EEG systems regarding poor accuracy and speed of the data recording, and technical issues, concluding that the system should only be used for applications that are noncritical.

11.3 Method

11.3.1 Participants

36 participants volunteered who had an English proficiency of at least college degree. Their age ranged from 19-57 years old (mean = 31.2) with 19 being female and 22 being right handed. All participants had normal or corrected-to-normal vision with no reported history of cognitive disorders or medication that could affect heir neurological behavior or reading ability. Each participant signed a written form of consent and had full understanding of what the experiment required of them.

11.3.2 Stimuli

Stimuli consisted of a combination of ten words, four pictures of emblem gestures and males gesturing, or both, and all were directionalized upwards or downwards. The

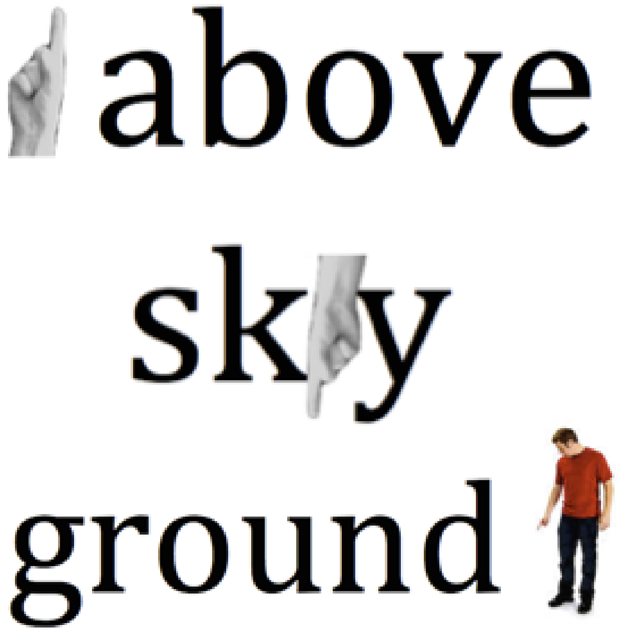


Figure 11.1: Here are three examples of the stimuli, the first of a congruous hand gesture on the left, the second of an incongruous hand gesture in the middle, and the third of a congruous man on the right.

upward words were "up", "top", "high", "above" and "sky" and the downward words were "down", "bottom", "low", "below" and "ground". The picture modes consisted of emblems with an index finger pointing, or a man pointing up or down. The words and pictures were combined so that not only was every word paired with every picture mode in a congruous and incongruous fashion, but also the picture's position was presented to the left of the word, in the middle of the word, and to the right of the word resulting in 120 combinations. See Figure 11.1 for more details.

11.3.3 Procedure

Before the experiment, participants were shown random examples of the stimuli to familiarize themselves with the test. The experiment consisted of an instruction screen and as soon as the participant was ready, instructed to push any button to continue. A fixation cross would appear for a random interval between 750 and 1,250 ms and then a stimulus would be presented where participants needed to respond as quickly and accurately as possible. Participants were instructed to push the Up arrow key if a single or both stimuli were upwards directional, the Down key if a single or both stimuli were downwards directional and the 0 number key if they were incongruous. Participants were randomly selected to use their right hand for pressing 0 on the right part of the keyboard where the numbers are in a 3x3 square, and using their left for the 0 along the horizontal row of numbers at the top of the keyboard, then they would use their other hand for the arrow buttons. They seemed to have no difficulty with this regardless of if they were left or right handed. They completed three cycles of all uni-form and cross form combinations.

11.3.4 Electrophysiological Acquisition

The Emotiv headset was used for EEG acquisition which has 14 channels at the International 10/20 system at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 locations and two left and right mastoid reference electrodes at P3 and P4. The data was sampled continuously at 128 Hz and filtered between 0.01 and 30 Hz. The system was used in a standard university office with potential noise disturbances to support research with this device in more realistic and natural environments. Therefore, no artifact rejection was conducted and only incorrect responses were eliminated from analysis. Component measurement was done offline with averaged baselines from the time window of -200 ms to the onset of the stimulus. Waveforms from -200 to 1000 ms in reference to stimulus onset were computed for each electrode. The data were processed using Matlab, Matlab's EEGLab and ERPlab, and then SPSS for statistical analysis.

11.4 Results

11.4.1 Behavioral Results

See Table 11.1 for behavioral statistics comparing response times for unimodal with congruous and incongruous stimuli. Using a repeated measures ANOVA with a 2x2x3 design (Congruency x Mode x Position) on the cross-form data, with a Sphericity Assumed test there was an effect for Congruency ($F(1,237) = 289.67$, $p = 5.5865E-43$), Mode ($F(1,237) = 5.97$, $p = 0.015$), and Position ($F(2,474) = 5.196$, $p = 0.006$), but no effect for Congruency x Mode ($F(1,237) = .38$, $p = 0.538$), or Congruency x Position

Table 11.1: Behavioral Responses

Stimulus Type	Reaction Time (ms)	Standard Deviation (ms)	Accuracy
Uni-form	708.2	267.7	98.6%
Congruous	892.5	302.6	97.4%
Incongruous	1,160.4	440.4	81.1%

Table 11.2: Congruency Effect

Stimulus Type	Reaction Time (ms)	Standard Deviation (ms)	Accuracy
Congruous	892.5	302.6	97.4%
Incongruous	1,160.4	440.4	81.1%

($F(2,474) = 2.82$, $p = 0.061$) nor were there any other significant interactions.[1] See Tables 11.2, 11.3 and 11.4.

11.4.2 Electrophysiological Results

The EEG data were analyzed using a repeated measures ANOVA with a 2x2x3x14 design (Congruency x Mode x Position x Electrode) and using a mean amplitude under the curve within the time window of 300-500 ms after stimulus onset. There was no effect for Congruity, so all other effects were ignored since N400 ERP semantic effects only measure differences in congruity/incongruity. However, the waveforms and scalp maps can be seen at Figure 11.2 and Figure 11.3.

[1]The degrees of freedom for Congruency are the number of types for congruency (congruity and incongruity) minus 1 and the total number of participant responses (238) minus 1. The same was done for Mode. The degrees of freedom for Position are the total number of positions (3) minus 1, and the number of positions minus 1 multiplied by the number of participant responses minus 1 (237 x 2 = 474). The same was done for the Congruency x Position interaction.

Table 11.3: Mode Effect

Stimulus Type	Reaction Time (ms)	Standard Deviation (ms)	Accuracy
Hand Gestures	1,014.1	349.8	89.0%
Man Gesturing	1,038.7	393.2	89.4%

Table 11.4: Position Effect

Position	Reaction Time (ms)	Standard Deviation (ms)	Accuracy
Left	1,021.0	179.4	81.5%
Middle	1,029.6	120.3	92.7%
Right	1,028.6	167.7	93.5%

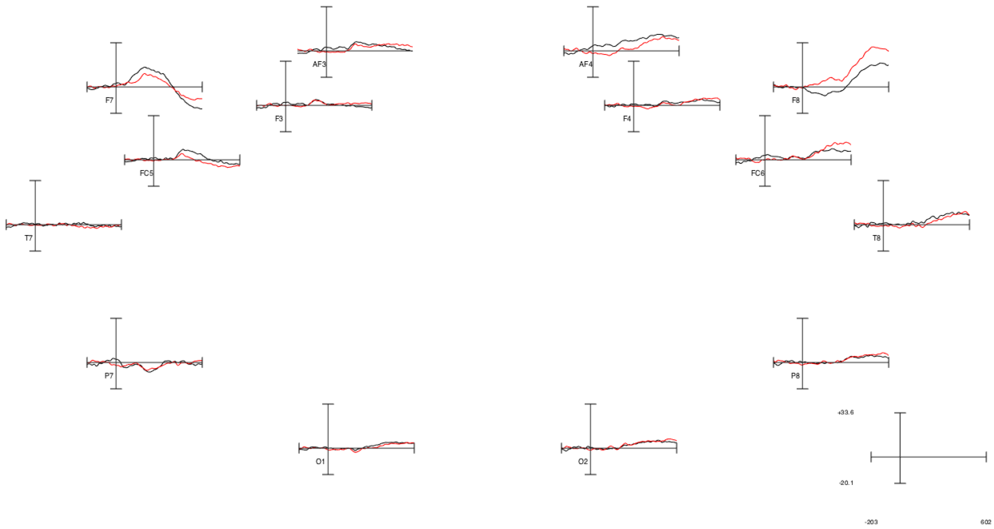


Figure 11.2: The electrode positioning from the time window of -203 to 602 ms, where negativity is plotted downwards, with congruous (black) and incongruous (red) waveforms. Waveforms are displayed in microvolts.

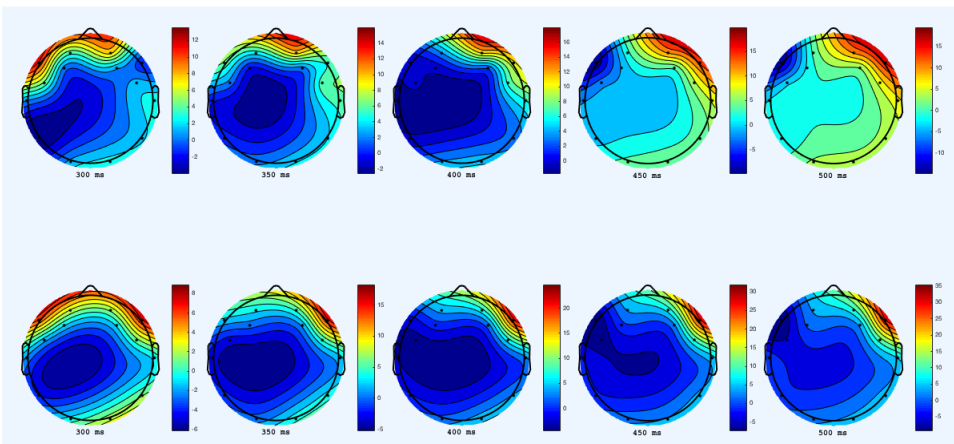


Figure 11.3: Scalp maps of grand average responses to congruous (top) and incongruous (bottom) probes from 300 to 500 ms after onset and displayed in microvolts.

11.5 Discussion

This experiment was designed to test if redundancy gain involving semantic categorization at a high level of processing for cross-form stimuli would produce faster and more accurate behavioral responses, as well as an N400 ERP response.

The behavioral responses to the data produced mixed results. While gesture position did have a statistically significant effect on reaction time, the difference in reaction time for the three positions was not much. Having the gesture on the left gave slightly faster response time than in the middle or the right. Also, the accuracy was far worse for left positioning than for middle and right. This could mean that having gesture on the left was processed first and thus more salient evoking a faster response, but this faster response was wrong more often since it overrode the participants' ability to process the second stimuli causing them to make a guess. Also, having the emblem in the middle of the word created the longest reaction time, as well as the worst accuracy, most likely due to how unnatural it was.

These results for faster reaction time but less accuracy for gesture on the left of the word seem to corroborate with previously established literature. Research has investigated temporal relationships with speech and co-speech hand gestures and concluded that usually the co-speech gestures are produced slightly before the speech that they are congruous with. Some of these studies include Bergmann et al. (2011); Kendon (2004); Loehr (2007, 2012) which investigated batonic, deictic, and iconic gestures. Leonard and Cummins (2011) found further supporting evidence regarding participants perceiving and being more accepting of temporally asynchronous co-speech gesture when it preceded speech as opposed to following it. These studies could give a possible explanation for the results of the present experiment where gestures which are read first, on the left, seemed more natural and salient than the other two positions which are processed later, but the parallel between pictures of gestures and real gestures which is proposed here needs further investigation.

There was also an effect for mode with pictures of just the hand gestures being significantly faster than those including the entire man. This could be due to the fact that the hand pictures in isolation are more salient than that of the entire man pictures and thus require less processing time. This information is useful for considering incorporating semantically relevant pictures of gesture with text, and isolating the visual information that is useful in understanding what the gesture is expressing.

Finally, while there was a main effect for congruency where congruent stimuli were both faster and more accurate then incongruent ones, uni-form stimuli were the simplest and thus quickest to process. The reaction time to uni-form stimuli was over 300 ms faster than that of the cross-form stimuli. This could have been due to the fact that with traditional redundant stimuli effects, most stimuli refer to the same concept, while one of them is sometimes meaningless. This means that the participant responds to the first stimuli that is conceptually processed without needing to analyze the second. While in this study, both were always meaningful but sometimes they were congruent, and sometimes they were incongruent. This meant that no matter what, the participant always had to take the time to process both stimuli. This extra processing task seemed to have taken about 300 ms extra time in comparison to the uni-form stimuli which required no second image processing.

The fact that there was no N400 effect could be because there was no priming effect or probe to induce the neurological response. Since subjects had to look at one of the two simultaneously presented stimuli, and then the other to see if they were the same, there could potentially have been an N400 effect after the second stimulus was processed meaning that it may have happened at 600 ms post stimulus onset for example. However, it would be impossible to isolate an N400 effect to such a stimulus without eye tracking because then it would be impossible to know when to time lock the cognitive processing onto the eye fixation of the second stimulus.

11.6 Conclusion

This study investigated if redundant congruent and incongruent cross-form stimuli would produce a facilitation effect when congruent, and an N400 effect when incongruent, neither of which occurred. There was no facilitation effect because participants had to process both stimuli every time, and there was no N400 effect since there was no priming. However, results of the combination types of the cross-form stimuli seem to support literature regarding the importance of order with pairing the two together and the saliency of having gesture proceed text, which most likely comes from our use of co-speech gesture beginning before speech. However this hypothesized relation between pictures of gestures and text should be investigated further. Future studies could also involve using cross-form stimuli but having a prime before the cross-form probe, and having the cross-form probe use meaningful and meaningless emblems and words so that participants could process just one stimulus to respond correctly. This paradigm could produce an N400 effect since a prime would be congruent or incongruent to the probe.

12

N400 investigation of semantic priming effect to symbolic pictures in text

Thomas Ousterhout. The abstract for this article was accepted for publication in 4th European and 7th Nordic Symposium on Multimodal Communication and the article is under assessment to be published in the post-proceedings.

12.1 Abstract

The purpose of this study was to investigate if incorporating meaningful pictures of gestures and facial expressions in short sentences of text could supplement the text with enough semantic information to produce an N400 effect when probe words incongruent to the picture were subsequently presented. Given that there is a new trend for users to use emojis in different parts of the sentences, sometimes as replacement for words, we have tested whether an N400 effect could be produced when the emojis are placed inside the sentence. Event-related potentials (ERPs) were recorded from a 14-channel commercial grade EEG headset while subjects performed congruent/incongruent discrimination tasks. Since pictures of meaningful gestures have been shown to be semantically processed in the brain in a similar manner as words are, it is believed that pictures will add supplementary information to text just as the inclusion of their equivalent synonymous word would. The hypothesis is that when subjects read the text/picture mixed sen-

tences, they will process the images and text in a similar way as they process gestures in speech and therefore probe words incongruent to the image will produce an N400. The behavioral results showed that facilitation did take effect in both the accuracy and response time with congruent responses in comparison to incongruent ones. Likewise, a scalp map showed a negatively going broad distribution in the posterior regions of the brain which was interpreted as the N400 deflection being more negative for the incongruent probes versus the congruent ones.

KEYWORDS: EEG, ERP, N400, Semantics, Congruency, Facilitation, Emotiv

12.2 Introduction

In face-to-face communication, when people speak, they often also use their hands to communicate. These hand movements can add supplemental information to the speech, such as iconic gestures where the fingers of the hand represent that a person is walking, or can be socially conventionalized and understood on their own, such as emblems where someone might give the "the peace/victory" sign (Burling, 1999; Gunter and Bach, 2004).

It is becoming more and more commonplace for people to use semantically meaningful pictures (commonly called emojis) in their text messages to supplement the text with pictures of communicative body behavior. From an evolutionary psychology perspective, body behaviors such as hand gestures, body language, and facial expressions, have been around long before spoken language came about (Corballis, 2003; Hewes et al., 1973; Rizzolatti and Arbib, 1998; Tomasello, 2000). While semantically meaningful pictures attempt to replicate this body behavior in text messages, it is unclear if they are a sufficient substitute.

The semantic priming effect can be used to study semantic processing because the response to a target word is facilitated, and thus the reaction time is decreased, if a semantically related prime word is previously presented (e.g. "cat-tiger" instead of "napkin-lion") (Meyer and Schvaneveldt, 1971; Neely, 1977). This behavioral response latency is inhibited when words are unrelated in comparison to when they are related. A corresponding neurophysiological phenomenon occurs in relation to semantic priming, which involves event-related potentials (ERPs). The N400 is a negative going EEG deflection with a posterior yet broad scalp distribution that usually occurs between 200 and 600 ms post stimulus onset of a critical/target word and is affected by semantic relations (Bentin et al., 1985; Holcomb, 1988; Kutas and Hillyard, 1984; Rugg, 1985). Just like facilitated reaction time with related pairs, there is a smaller N400 effect when related versus unrelated pairs are presented as well, however the negative deflection is not large (Holcomb and Anderson, 1993; Kounios and Holcomb, 1992). The more unrelated the probe word is, the larger the N400 deflection will be.

While the majority of N400 priming effect studies have been conducted to observe linguistic phenomena (for a review see Kutas and Federmeier (2011)), this effect still occurs for any kind of meaningful stimuli where congruency/relatedness is established such as with pseudowords (Borovsky et al., 2012, 2013), visual information such as pictures of semantically related objects (Barrett and Rugg, 1990; Ganis et al., 1996; Holcomb and McPherso, 1994), faces (Barrett et al., 1988; Barrett and Rugg, 1989; Olivares et al., 1999), gestures (Gunter and Bach, 2004; Ousterhout, 2015b; Proverbio et al., 2015), and even other sensory modalities such as environmental sounds (Petten and Rheinfelder, 1995).

There have been a tremendous amount of N400 component studies investigating the cognitive functions occurring when informational context progresses from perceptual processing to semantic processing. When the component was first discovered (Kutas and Hillyard, 1980) during an investigation of the predictability of sentence-final words, where semantically incongruent conclusions produced the largest negative wave deflection, the interpretation was that the N400 reflected semantic processing of sentence reading. Brown and Hagoort (1993) suggested that the component's amplitude deflection is indicative of the amount of working effort required during semantic integration of a word with its preceding context. The N400 has also been interpreted as evidence towards long-term semantic memory facilitation in that it is easier to predict and access words that fit with the content stored in memory (Federmeier et al., 2007; Kutas and Federmeier, 2000; Lau et al., 2008).

Several semantic priming N400 studies investigated how gestures are processed in relation to verbal material (Holle and Gunter, 2007; Kelly et al., 2004; Ousterhout, 2015b; Özyürek et al., 2007; Wu and Coulson, 2005, 2007). Due to the fact that symbolic gestures and words are reciprocal when they are semantically congruent (Barbieri et al., 2009; Bernardis and Gentilucci, 2006), Fabbri-Destro et al. (2014) used symbolic gestures as primes and found in their investigation that they perform just as well as verbal material in producing the N400 modulation.

Since it has been shown that N400 effects can be created with the probe words in relation to the semantic content to previous sentences, and it has also been demonstrated that primes in this previously presented semantic content can be a symbolic gesture, we expand on these ideas. More specifically the hypothesis of this experiment is that a picture, similar to an emojis, depicting a symbolic gesture or emotion, presented in the middle of a sentence, will function as the salient prime in the sentence and produce an N400 effect to subsequently presented semantically incongruent probes.

12.3 Method

12.3.1 Participants

This study used 11 participants (6 females), 10 were right handed, and they were aged between 26 and 57 years old (mean = 36 years). They all had an English proficiency at university level and some of them were native speakers. All participants had normal or corrected-to-normal vision and none had any reported history of cognitive disorders or medication that could impact their reading ability. All participants filled out and signed a written form of consent stating they had full understanding of the experiment, what was required of them, and what the data would be used for.

12.3.2 Stimuli

The study used 22 sentences primes, each with a congruous and incongruous subsequent probe sentence resulting in a total of 44 total combinations presented in a random order. Half (11) of the prime sentences used emojis made up of gestures which were borrowed from Gunter and Bach (2004), and the other half were simple smiley/sad emojis taken from the internet. Each prime sentence was short with only three or four words and each had a picture of either an emblematic gesture, or a happy/sad face emoji between the concluding salient word and the word prior. It was decided to put the emoji on the left of the word the emoji referred to as if the emoji replaced an adjective. See Figure 12.1 for examples. The probe consisted of a three word sentence, where one word was displayed at a time, and the final word would be directly congruous or incongruent with the picture. The words used were mostly only one syllable long and very common so that they could be read at equal speed regardless of English proficiency. Moreover, the longer and slightly less frequent probes were matched equally for congruent and incongruent conditions in order to balance out.

12.3.3 Procedure

Participants were allowed to read the prime sentence for as long as was required to fully process it and would continue with a button press on a keyboard. Then there would be a fixation cross for a random duration between 750 and 1,250 ms. Afterwards the three words of the probe sentence would appear in the center of the screen, each for 500 ms, and on the final probe word, which was always the 3rd word, the participant was required to make a button press response for whether the word was congruent or incongruent with the picture/emoji in the prime sentence. This experiment consisted of 2 cycles of all 22-sentence pairs summing up to 88 stimuli sentences and taking about 20 minutes to complete. The experiment was conducted in a standard university office with potential for noise disturbances outside to support continued research with this device in more natural environments.

12.3.4 Electrophysiological Acquisition

This experiment used the Emotiv headset for EEG acquisition which has 14 channels at electrode positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 according to the International 10-20 locations (plus CMS/DRL references at P3/P4 positions). These data were sampled continuously at 128 Hz and was low-pass filtered between 0.01-30 Hz. For artifact rejection a moving window peak-to-peak system was used with a full width of 200 ms, a window step of 50 ms, and a voltage threshold of 100 microvolts. Component measurement was conducted offline where waveforms were baselined from -200 ms to probe onset time. Waveforms from -200 to 1000 ms were computed for analysis from the electrodes. These data were processed with Matlab, Matlab's EEGLab and ERPLab, and SPSS.

12.4 Results

12.4.1 Behavioral Results

Responses to congruent probes had an accuracy of 88.5% and incongruent ones had 86.7% accuracy. Congruent probes had a mean reaction time of 1,028.6 ms with a SD

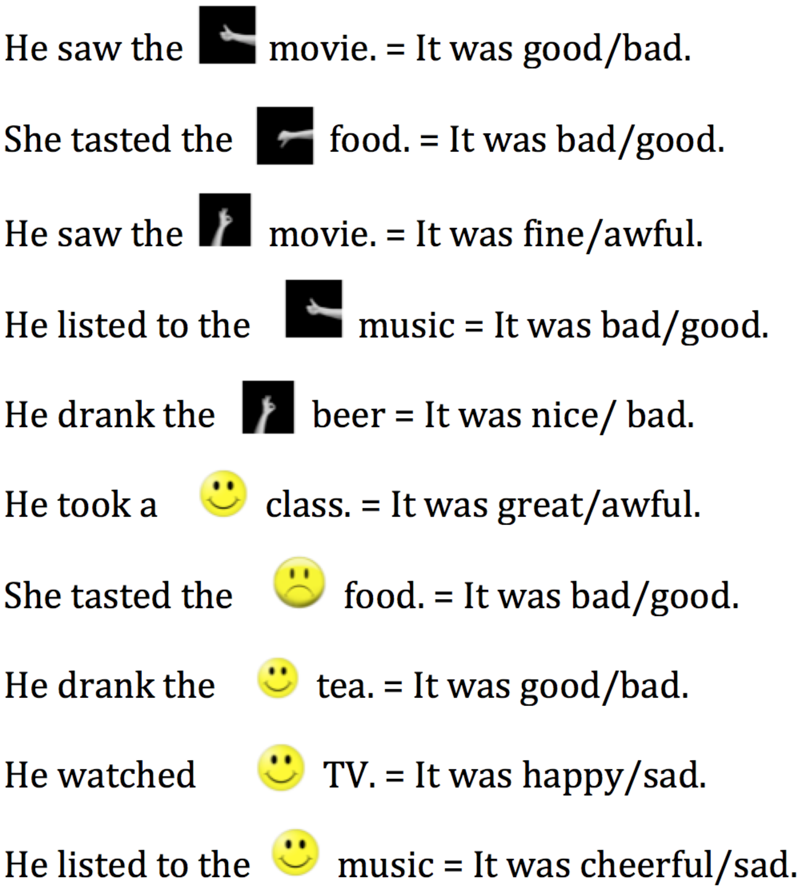


Figure 12.1: Here are some examples of the stimuli. The prime sentences on the left have emojis such as the "thumbs up" gesture and a smiley face. The probes on the right are short sentences where one word was presented at a time so that they could be time-locked. It was always the final 3rd word which was congruent or incongruent with the prime.

of 529.8 ms, and incongruent probes had a mean reaction time of 1,128.6 ms with a SD of 1,265.7 ms. Using a paired t-test, these reaction time means were statistically significantly different with $p = 0.0006$. A supplementary analysis was done measuring the congruent sentences with a gesture emoji where the accuracy was 80.4% and the reaction time was 1138.31 ms, to those with face emojis where the accuracy was 90.1% and the reaction time was 1130.53 ms. A paired t-test with these two sub-groups was conducted and showed that there was not a statistically significant difference in reaction time for congruency between emoji types since $p = 0.35$.

12.4.2 Electrophysiological Results

Visual inspection of the grand average scalp maps and waveforms for both congruous and incongruous tasks showed a posteriorly distributed negativity that was larger for incongruous than congruous conditions in the time window of roughly 350-450 ms. See Figure 12.2. The topography of this negative latency seems analogous with literature describing the N400 in semantic priming studies ((Bentin et al., 1985; Holcomb, 1988; Kutas and Hillyard, 1984; Rugg, 1985), and for a review see Kutas and Federmeier (2011)). Using only the back four electrodes being the two parietal (P7, P8) and two occipital electrodes (O1, O2), a 2x4 repeated measures ANOVA investigating Congruency x Electrode position, there was a main effect for Congruency ($F(1,10) = 5.262$, $p = 0.045$) showing that there was a statistically significantly different mean amplitude in the N400 region where incongruent probes were more negative than congruent ones. See Figure 12.3. No other significant effects were found.

12.5 Discussion

In this study a semantic priming paradigm was implemented to investigate if semantically meaningful images of gestures or facial expressions placed in a sentence would facilitate the processing of subsequently presented probe words, which were congruent or incongruent to the image. The behavioral results showed that facilitation did take effect in both the accuracy of the decision-making where congruent probes had a higher accuracy than the incongruent ones and as well with response time, where congruent responses were faster than incongruent ones. Upon further analysis, although the congruent sentences that had an emoji of a face had a higher accuracy than those with a

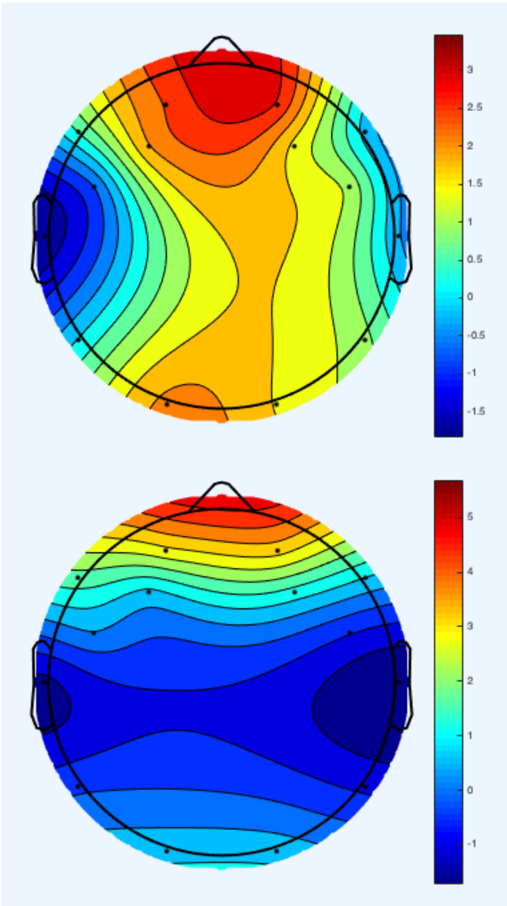


Figure 12.2: Scalp maps of congruous (top) and incongruous (bottom) conditions at 420 ms and displayed in microvolts.

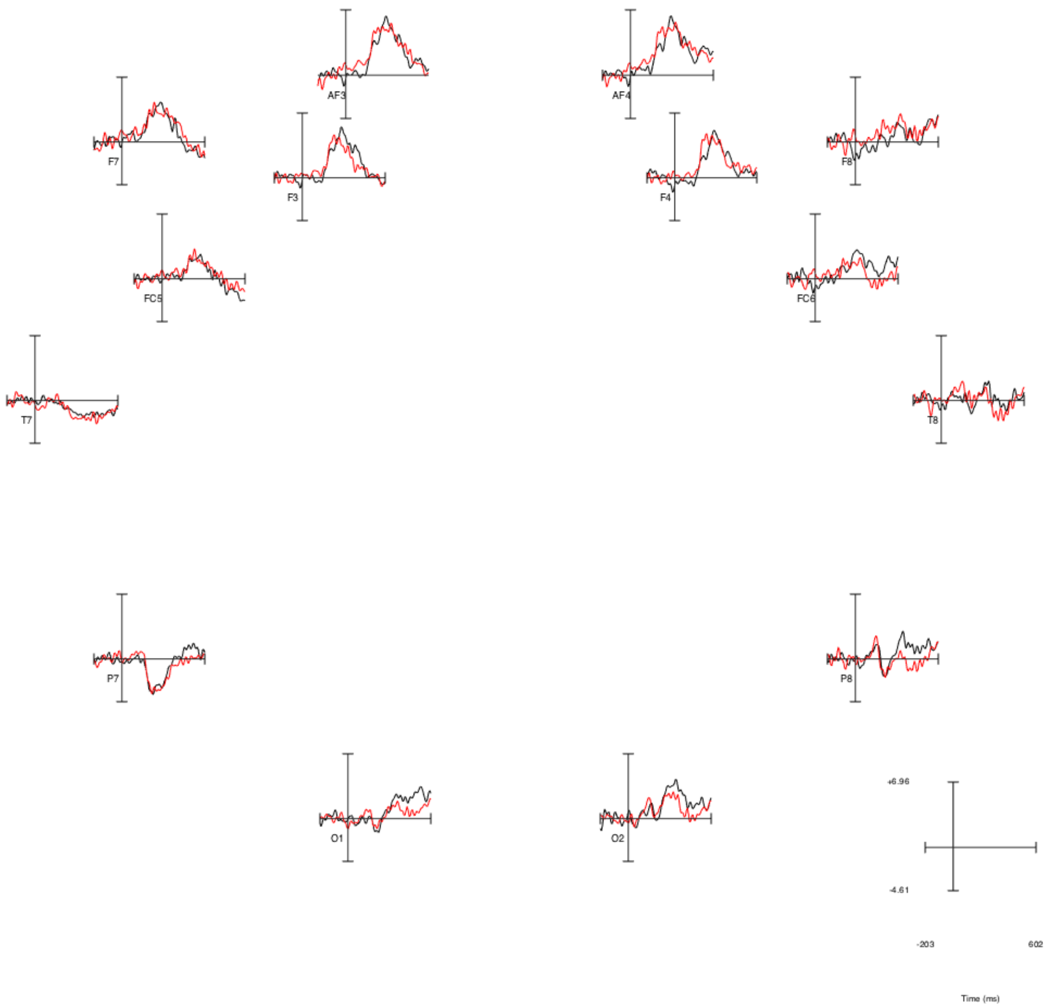


Figure 12.3: The electrode positioning from the time window of -203 to 602 ms, where negativity is plotted downwards, with congruous (black) and incongruous (red) waveforms. Waveforms are displayed in microvolts.

gesture, there was no significant reaction time improvement. These results show that using emojis in sentences does facilitate accuracy and response time, but as long as the emojis used are simple enough to understand, it does not matter which type of emoji they are.

Likewise, a scalp map showed a negatively going broad distribution in the posterior regions of the brain and analyses of the 4 posterior Emotiv electrodes showed a significant N400 deflection which was more negative for the incongruent probes versus the congruent ones. Traditionally N400 deflections have been shown to be elicited with words in sentences, but it seems that they are produced with any meaningful stimuli that can be considered congruent or incongruent to a prime (see Lau et al. (2008) for a review), which includes words in isolation (Bentin et al., 1985), as well as psuedowords (Rugg and Nagy, 1987), but more importantly pictures (Barrett and Rugg, 1990), pictures of faces (Barrett and Rugg, 1989), and pictures of semantically meaningful gestures (Fabbri-Destro et al., 2015; Gunter and Bach, 2004; Ousterhout, 2015b).

These semantically meaningful gestures can produce N400 effects even when they are not directly task relevant. In the study by Wu and Coulson (2005) consisting of two experiments, one where gesture comprehension was overt and the second where it was covert, both produced an N400 effect when gestures were presented in relation to an incongruous context. The second covert task had an additional subsequent stimulus which was a probe word related or unrelated to the previous context. This study, as well as many others have observed processing advantages (either by observing physiological N400 effects (Pratarelli, 1994), or decreased reaction times (Carr et al., 1982; Coney and Abernethy, 1994; Hines, 1993; Vanderwart, 1984)) in lexical priming cross-modal studies using pictures and words where perceptual features were shared by both the prime and probe.

It seems to be that gestures can activate features which are stored in working memory thus providing facilitated processing in following words related to objects with those features. These results are consistent with experiments that investigate high cloze sentences, which are those where there is a preferred ending which attenuates the N400 amplitude, versus low cloze sentences which have unlikely endings and thus produce large N400 amplitude deflections (Kutas and Hillyard, 1984). What happens in these situations is that the stimuli activate long term memory which produces expectations about what should follow in upcoming stimuli presentations. When what is expected is presented to the viewer, semantic processing is already primed and therefore the high cloze ending is easily processed which produces a shallower N400 deflection (Wu and Coulson, 2005). Since pictures of meaningful gestures have been shown to be processed semantically similarly to words (Gunter and Bach, 2004; Ousterhout, 2015b), it follows that their implementation in sentences creates additional semantic meaning to the sentence which thus also produces high cloze expectations with what should conclude as the ending.

This experiment's results do support the view that incorporating semantically meaningful pictures of relevant body behavior in text can supplement the written text in a similar way as physical gesturing does in face-to-face communication. Emojis give communicators a user-friendly way of adding another form of communication to supplement written text. Most importantly, the use of these emojis requires no formal learning since they have been created to be as literal, diverse, and intuitively representative of the missing physical body language as possible.

12.6 Conclusion

In this article, we showed that emojis of symbolic hand gestures and facial expressions are processed in sentences as words confirming previous research that emblems have the same functions as words and can substitute them. This article also supports the findings by Wu and Coulson (2005) who demonstrated that visual stimuli of gestures can covertly produce N400 effects. There was a difference in reaction time between congruency tests and there was no difference when processing hand gesture emojis and facial emojis. This indicates that when the emojis replace words in sentences there is no difference whether they are facial expressions or hand gestures. Since there is also a trend of using moving emojis in text, this should be tested in the future.

13

Investigation of the semantic priming effect with the N400 using symbolic pictures in text

Thomas Ousterhout. Submitted to Australian Journal of Intelligent Information Processing Systems.

13.1 Abstract

In face-to-face communication, a large portion of communicative devices rely on the visual modality of bodily behaviors which include facial expression and hand gestures. However through the use of digitally mediated communication which is becoming increasingly prevalent with advances in technology, people are evolving their way to communicate. Texts become shorter and the use of emojis are changing. Facial emojis are symbols for human faces that have become increasingly popular with communicative devices. The original and still most frequent use of emojis is to provide a comment to the text which they follow. However, the latest trend is also to use emojis in the middle of sentences replacing words or adding information to the text. Through the use of EEG and the N400 ERP component, this study investigates which objects emojis refer to via an internet survey and a EEG semantic priming test in which moving emojis in sentences are paired with congruous and incongruous probes. The results of both the survey and the EEG test indicate that there is no preference for particular positions of

the emojis and that some of the unusual emojis were ambiguous and did not add to comprehension.

KEYWORDS: EEG, ERP, N400, Semantic Priming, Congruency, Emojis

13.2 Introduction

In order to interact with people, no matter what the social circumstance or environment is, it is absolutely necessary to understand what others are doing, intending and feeling. Without the ability to understand others, intended and unintended messages would be lost in the process of communicating and cooperation would be unproductive. There are certainly many parts of the brain that are responsible for effective communication, one part in particular is called the mirror mechanism (Rizzolatti and Fabbri-Destro, 2008).

Mirror neurons, which were first discovered in monkeys, have also been found in humans. Whenever an individual sees an individual performing a motor action, mirror neurons activate a part of the brain that also fire when the observer executes the exact same action themselves. Jeannerod (1994) believes this is for learning purposes. As many are familiar with, students watching a teacher, rather than just listening, help with the learning process of performing the action or task. This is because while watching the agent perform the task, the mirror neurons encode a representation of the action itself, which it just has to repeat when executing that action.

Rizzolatti et al. (1996) theorize that the mirroring system contributes to understanding motor actions through recognition, differentiation, and knowing how to respond appropriately. Simply put, when an agent performs an action, it can be assumed that there is an intention with a prediction of a specific outcome. An observer can learn quickly how to produce specific outcomes from observation alone rather than practice through this mirror mechanism and more importantly what the meaning of those actions represent.

Interestingly, there is a large amount of generalization when it comes to the type of stimuli in which mirror neurons respond to. Rizzolatti and Craighero (2004) found that the same mirror neuron cluster fires when a human grasps an object as well as when a monkey does the exact same performance. This same firing pattern also happens when watching the action from a distance or at proximity.

While mirror neurons are very important in action learning and understanding, another theory by MacNeilage (1998) proposes that human speech evolved from monkey open-close jaw movements such as when they perform "lipsmacks". These simple facial manipulations created a type of faciovisual communication, also known currently as facial expressions or gestures. This suggests that communication began as a visual modality which later was supplemented with sounds. These are all proposals of why mirror neurons play a role in matching observed and executed actions and why they are important in understanding each other's behavior. Also mirror neurons are found in Broca's area, which is significantly responsible for understanding speech, and furthermore several theories address how speech evolved from visual/gestural communication (Armstrong et al., 1995; Corballis, 2002; Rizzolatti and Arbib, 1998).

Emotion expression and comprehension are also crucial communicative devices needed for effective message transmission. Singer et al. (2004); Wicker et al. (2003) both performed experiments in which participants experienced an emotion such as disgust or pain, and then watched another go through the same experience. Both studies showed that similar neurological activity was produced in experience and observation conditions suggesting a mirror mechanism involved.

It seems however that observing an action performed by an agent is not the only factor activating the mirror neurons, but rather perceiving the action performed by the agent. Observation is typically required in order to produce some kind of understanding, however Kohler et al. (2002); Umilta et al. (2001) removed the ability of the participants to produce visual observation and yet still were able to measure the variable of understanding in isolation. They set up an experiment where visual observation was not possible due to a blocking screen, and found that when participants could hear distinct action sounds such as paper ripping, the mirror neurons still fired. This showed that mirror neurons do not necessarily respond to visual stimuli specifically, but rather the understanding that usually comes with visual stimuli.

Finally, Nelissen et al. (2005) performed a study showing that mirror neurons do not fire to biological agents only. They set up a study where stimuli consisted of video clips of several objects performing the actions 1. humans 2. just human hands, and 3. robotic hands. The results showed that the mirror neurons fired in all conditions demonstrating that an artificial stimulus such as a video is sufficient for mirror neuron activation, the entire actor's body is not required for mirror neuron activation since only seeing a hand was enough, and most importantly, non-biological agents, such as robotic limbs, and potentially also emojis, activate mirror neurons.

Face-to face-communication is multimodal since it consists of at least the auditory and the visual modalities. While auditory information comprises speech and is considered most dominant in message transmission, the visual modality which includes inter alia head movements, facial expressions and hand gestures, can disambiguate the speech content, emphasize it, change its meaning or substitute for it (Goldin-Meadow, 1999; Kelly et al., 1999; Kendon, 2004; McNeill, 2005). Due to how the mirror neurons function and the type of stimuli that activates them, seeing people communicate is an inherent part of our face-to-face communication system and the visual modality is extremely important in comprehension.

There is no doubt, however, that speech or text are very powerful in transmitting messages, which is why they have been used in numerous semantic priming studies. Semantic priming studies are useful for studying semantic processing because a priming word or sentence can activate the brain in a way where the response to a probe, which is subsequently presented, will be facilitated with a faster reaction time, when the probe is related instead of unrelated. An example of this would be *cat -tiger* being processed faster than *napkin-lion* (Meyer and Schvaneveldt, 1971; Neely, 1977).

Another type of priming is affective priming. This works by presenting a priming stimulus with either a positive or negative affective valence and then measuring the behavioral and psychological response to a related or unrelated probe thereafter. The way the affective priming works is if the probe is affectively congruent to the prime instead of incongruent. This was demonstrated using emojis and just words as primes(Comesaña et al., 2013). More importantly, this task used masked primes, meaning that a distraction was presented right before the prime so that the participants were unaware of the emoji prime. Not only did the priming effect occur even though the stimuli were covertly processed, the results show that the priming effect occurred more significantly for the emoji than it did for the words. Comesaña et al. (2013) conclude that the results occurred due to the automatic processing of the saliency of the facial expressions being more significant that the words.

Both of these types of priming can be measured physiologically through EEG. This is done through event-related potentials, which are amplitude deflections in the EEG waveform that are related to a specific event. One ERP in particular is called the

N400, which is a negative going deflection approximately 400 ms after the onset of the stimulus (Holle and Gunter, 2007; Kelly et al., 2004; Kutas and Federmeier, 2011; Kutas and Hillyard, 1980; Ousterhout, 2015b; Özyürek et al., 2007; Wu and Coulson, 2005, 2007). The way it is measured and useful, is that when a probe is incongruous to the prime, instead of congruous, the amplitude is much more negative.

There are several types of EEG recording devises ranging in a large number of electrodes used for medical and consumer purposes. One consumer grade EEG system in particular is called the Emotiv headset, which utilizes 14 EEG channels. Although this system is much simpler than standard medical grade EEG devices, there have been a number of studies that support its efficacy in ERP research (Badcock et al., 2013, 2015; Boutani and Ohsuga, 2013; Ekanayake, 2010; Kawala-Janik et al., 2015; Mayaud et al., 2013; Ousterhout, 2015a,b; Ousterhout and Dyrholm, 2013).

According to the mirroring theories and other theories which address the importance of gestures in face-to-face communication (Kendon, 2004; McNeill, 1992), multimodality is an essential aspect of the way in which humans communicate. Communicating by written texts involves the visual modality only and all the discourse content is expressed by words. Short messaging is often a replacement for oral communication, it is quick and it can therefore be difficult to express one's personality, affective state, irony, or emphasis in it. This is why emojis are becoming so popular.

This study aims to investigate the use of moving emojis in different positions of short sentences to see whether they aid or supplement text adding elements usually expressed by body behavior. A preliminary survey was conducted online in which short sentences had emojis placed in different locations and participants had to respond to which subject and/or object they thought the emojis were referring to. Successively, this information was used to place emojis in short sentences and test whether they produce enough semantic priming (N400 effect), when an incongruous probe stimuli is presented.

In the next section, an explanation of how the study was conducted is provided. This includes the pre-test survey, the description of the participants and stimuli, the procedure of the entire follow up experiment, and the method used for analyzing data. Then a summary of the results is given, explaining how the participants performed behaviorally as well as physiologically. Following this is a discussion of the results commenting on why the participants performed the way they did and what this means.

Then there is a short conclusion discussing future work. Finally there is an appendix providing the pre-test survey and EEG test sentences.

13.3 Method

13.3.1 Pre-test: Survey

To figure out which position in a sentence an emoji would be best used to refer to an element in that sentence, a survey was created with sentences where emojis in different locations had the potential to refer to multiple items in that sentence (see Appendix). Each of the sentences had 2 or 3 questions asking to which element the emoji was mostly related.

The survey was answered by 72 participants and show mixed results. Since there seems to be little pattern, and some examples directly contradict each other, it was decided to use the sentence examples where there was the most unanimity among participants. Therefore only examples with answers above 70% agreement were looked at. In most of these examples, the participants thought that the emojis referred mostly to the element (person, object or animal) which the emoji followed. In a minority of examples, in which the emojis preceded the subject of the sentence, the emoji was found to refer to the subject.

13.3.2 Participants

For this experiment 19 participants were used that had an English University speaking level. The mean age was 31.6 years of age with a standard deviation of 8.9 years. 10 were males, and 16 were right handed. None reported any cognitive or reading problems and everyone had good or correct-to-good vision. They all signed an informed consent document explaining that they knew what the experiment was about and that their data would be published yet individually they would remain anonymous. Due to artifacts in the EEG which involved too much noise in the signal quality, 2 participants' data were eliminated.

13.3.3 Stimuli

The study consisted of 45 prime sentences each with a congruent and incongruent probe stimulus resulting in 90 stimuli examples in total. When the emojis function as verbs,

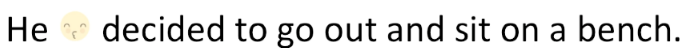


Figure 13.1: Sentence example with [He][was][whistling.] as congruous probe and [He][was][sad.] as incongruous.

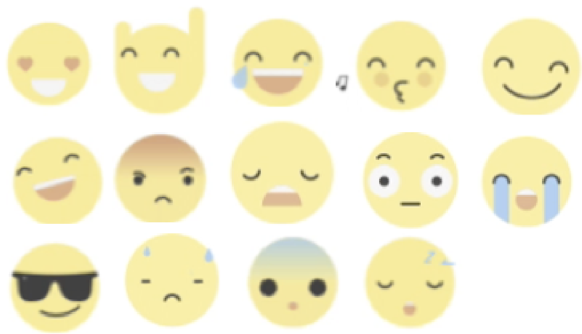


Figure 13.2: Pictures of all 14 moving emojis which were taken from `https://cdn.dribbble.com/users/43762/screenshots/1925708/emojis.gif`.

the incongruent probe word is inconsistent (e.g. whistling to yelling, and laughing to crying), when the emojis are adjectives the probe words are antonyms (happy to sad, and excited to bored). See Figure 13.1 for a sentence example and more are in the Appendix. Each prime sentence had one of 14 different moving emojis, see Figure 13.2 for emoji examples. These emojis would play on a replay-loop since they each only lasted 1-3 seconds. When each prime sentence was displayed, subjects had as much time as they needed to read the sentence and the experiment would continue with a button press. The sentences could be complex. Then a sequence of three slides with a single word would appear at a time, which in culmination made a phrase where the third word was always directly congruent or incongruent to the moving emoji. The third word, which was the probe, was typically only one syllable long and very common so that it could be read at equal speed regardless of English proficiency. Moreover, the longer and slightly less frequent probes were matched equally for congruent and incongruent conditions in order to balance out.

13.3.4 Procedure

Participants were allowed to read the prime sentence for as long as was required to fully process it and would continue with a button press on a keyboard. Then there would be a fixation cross for a random duration between 750 and 1250 ms. Afterwards the three words of the probe sentence would appear in the center of the screen, each for 500 ms, and on the final probe word, which was always the 3rd word, the participant was required to make a button press response for whether the word was congruent or incongruent with the emoji in the prime sentence. This experiment consisted of 1 cycle of all 45-sentence pairs summing up to 90 stimuli sentences and taking about 20 minutes to complete. The experiment was conducted in a standard university office with potential for noise disturbances outside to support continued research with this device in more natural environments. After the experiment was concluded, participants were asked about their opinion of the experiment as a whole, if the sentences were coherent, and their thoughts about the emojis themselves.

13.3.5 Electrophysiological Acquisition

This investigation was conducted using the Emotiv headset for EEG acquisition which is a commercial grade system using 14 electrode channels at positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 and CMS/DRL references at P3 and P4 according to the International 10-20 locations. The sampling rate for the acquisition was done at 128 Hz and was filtered between 0.01-30 Hz. The artifact rejection method was done using a moving peak-to-peak system where the full width of the window was 200 ms, the window step was 50 ms, and the voltage threshold was 100 microvolts. Components were measured offline where waveforms were baselined from -200 ms to probe onset time. Waveform total duration were set from -200 to 1000 ms and were computed for analyses with Matlab, Matlab's EEGlab and ERPLab, and SPSS.

13.4 Results

13.4.1 Behavioral Results

The behavioral results show an average accuracy of 87% for all congruent and 89% for incongruent responses. Also, there was an average reaction time of 1,116.4 ms for all congruent stimuli with a 1,046.8 ms response time for incongruent stimuli. A paired two-tailed t-test of these reactions times resulted in a statistically significant difference of 0.034.

Further analyses was conducted dividing the primes into two groups where the emoji referred to an object or subject before its placement in the sentence, and after its placement. The accuracy for before placement was 86.1% and for after its placement it was 88.8%. The mean reaction time for before placement was 1222.38 ms where for after its placement it was 1114.3 ms. A paired two-tailed t-test of these reactions times resulted in no statistically significance difference of 0.13.

13.4.2 Electrophysiological Results

Using a repeated measures ANOVA with 2 x 4 for congruency and electrode position (P7, O1, O2, P8) using a culmination of all congruent and incongruent data, there was no effect $F(1,16) = 16.0$, $P = 0.781$. See Figure 13.3 for the waveform and Figure 13.4 for the scalp map. However, performing the same repeated measures ANOVA upon analyzing the individual emojis, there was a significance for the "scared" emoji $F(1,16) = 5.0$, $p = 0.034$ where the incongruent responses were more negative going than congruent ones. See Figure 13.5 for the waveforms and Figure 13.6 for the scalp map. No other significant effects were found.

Further analyses involving the two subgroups where the emoji referred to an object or subject before its placement or after was also conducted. A 2 (congruency) x 2 (referent direction of before/after) repeated measures ANOVA was conducted with no effect for congruency $F(1,16) = 0.276$, $p = 0.607$, and no effect for direction $F(1,16) = 0.003$, $p = 0.954$.

13.5 Discussion

The discovery of mirror neurons has demonstrated the importance of observation when trying to understand what others are doing (Jeannerod, 1994; Rizzolatti and Fabbri-Destro, 2008; Rizzolatti et al., 1996).This observation allows viewers and learners to understand why agents perform a certain action and thus allow them to predict a certain outcome (Rizzolatti et al., 1996). This mechanism is very robust and the same neural area will be activated despite if a human or monkey performs an action and also regardless of distance or proximity (Rizzolatti and Craighero, 2004). This means the activation will occur to stimuli that do not look exactly human, and when they are very far away, and thus very small. Furthermore, Freedberg and Gallese (2007) found that

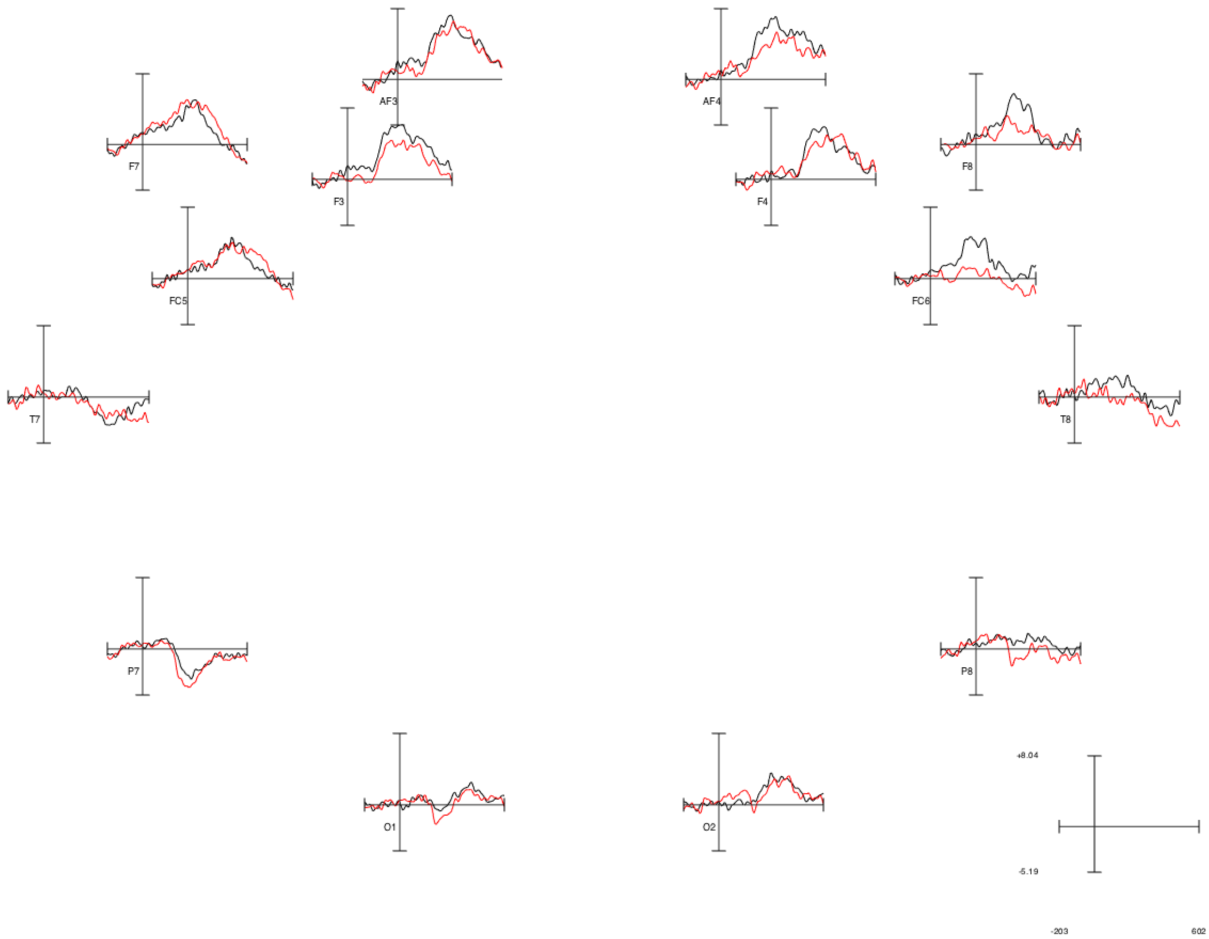


Figure 13.3: Waveforms of grand average responses to congruous (black) and incongruous (red) probes to emoji prime sentences. Waveforms are displayed in microvolts.

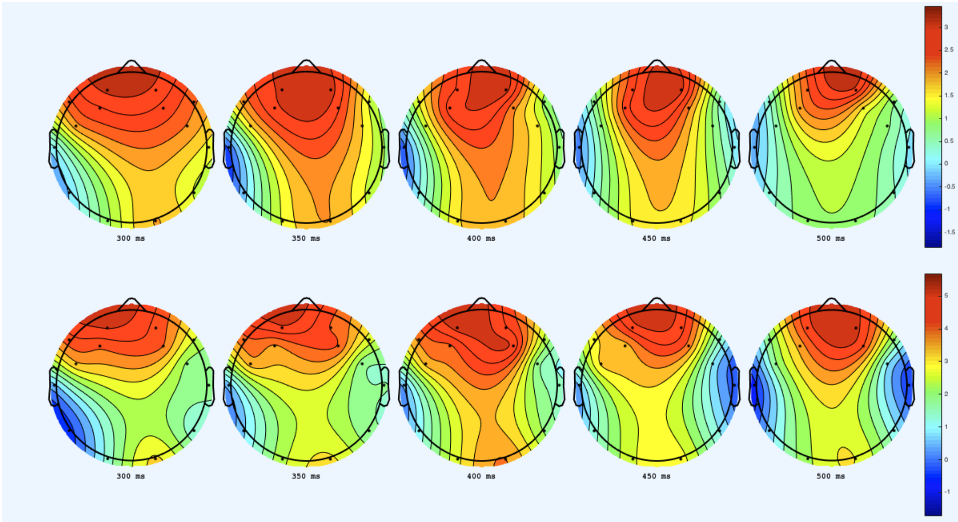


Figure 13.4: Scalp maps of grand average responses to congruous (top) and incongruous (bottom) probes to emoji prime sentences displayed in microvolts.

the mirror neurons are also fired when humans look at pictures of humans. Emojis, although not identical to human faces, are similar to them and therefore their presence can be assumed to activate mirror neurons as well and thus facilitate communication, understanding and learning. Not only that, but since mirror neurons have been shown to be activated during emotion expression(Singer et al., 2004; Wicker et al., 2003), emojis seem like an ideal method to express emotions digitally. Lastly, since Nelissen et al. (2005) showed that mirror neurons fired even when a robot hand was showed on a video screen, there seems to be no limit to how artificial or abstract an emoji can be where it would be unable to transmit an intended message that could activate the mirror neuron system in the observer, which is the source for the generation of this study.

Studies such as Chu and Kita (2011) have shown how the incorporation of gesture can help with tasks such as spatial problem solving, and Broaders et al. (2007); Goldin-Meadow et al. (2009) have demonstrated how gesturing during problems solving helps with learning through visualization techniques. Therefore, this paper attempted to show the need for co-speech gestures in text reading and how the implementation of emojis could do so. The first step was to find out where in a sentence people typically found emojis to be most descriptive when describing a specific subject. With this information, the study investigated if the implementation of various moving emojis placed strategically in sentences, would supplement reader's understanding of prime

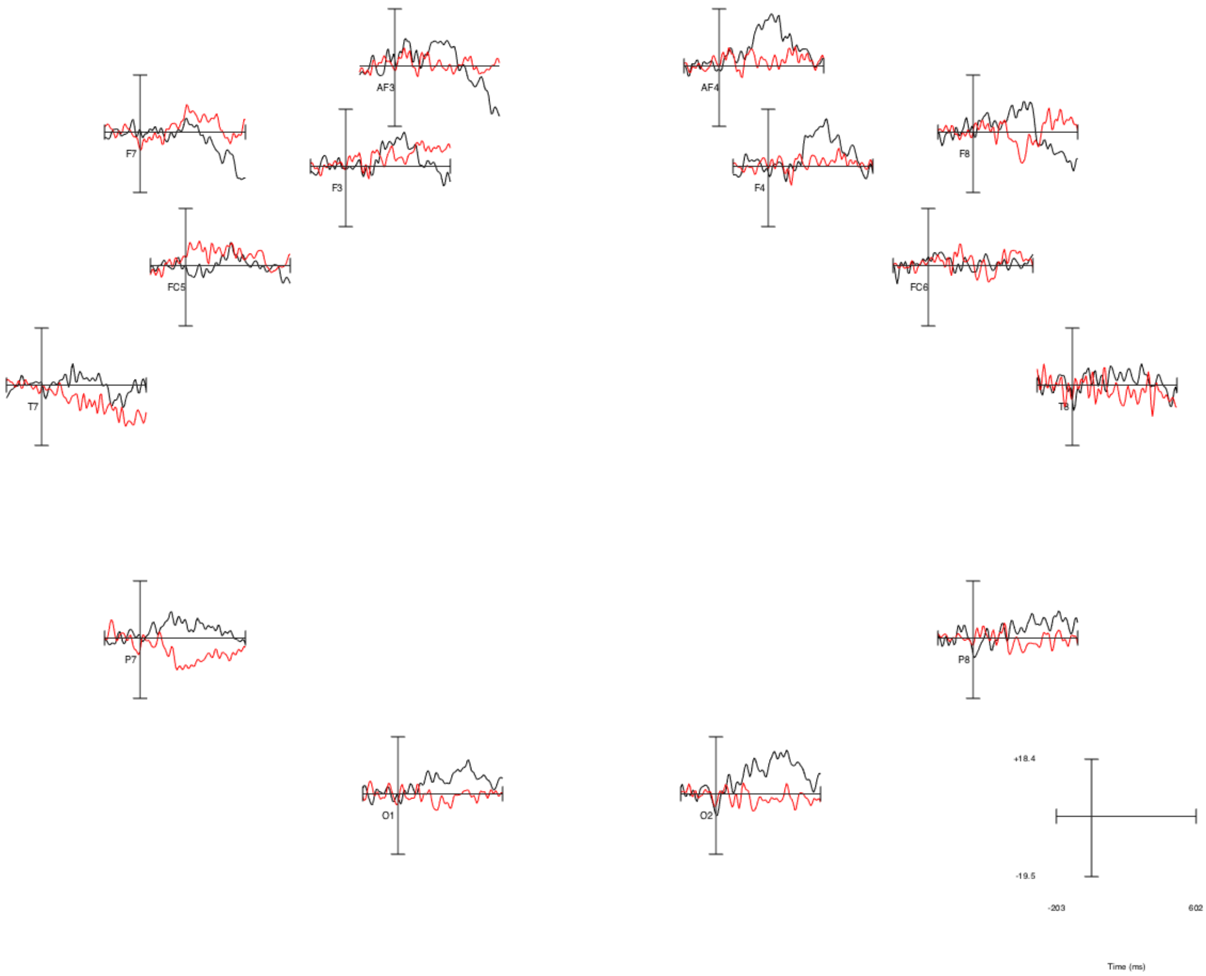


Figure 13.5: Waveforms of responses to congruous (black) and incongruous (red) probes to "scared" emoji prime sentences. Waveforms are displayed in microvolts.

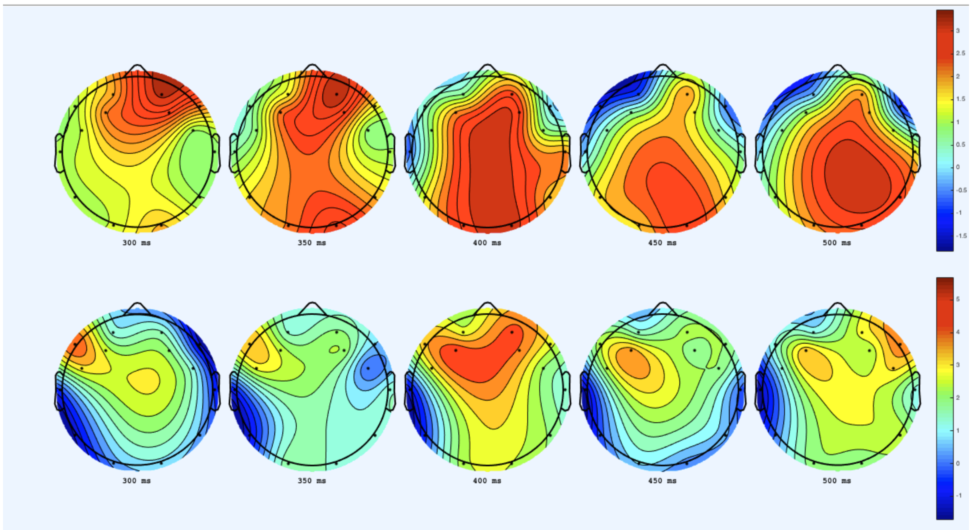


Figure 13.6: Scalp maps of responses to congruous (top) and incongruous (bottom) probes to "scared" emoji prime sentences displayed in microvolts.

sentences by adding emotional information to the context of the sentence. To test this, semantic priming was studied through N400 production. The reason for this entire investigation was to make the first steps towards creating a universal multimodal reading system that would supplement text with images of body behavior since that is natural and crucial in our face-to-face communication.

The results of the pre-study showed that people had a very broad and varying opinion of what an emoji referred to when placed in different sentence positions. However, it seemed that most participants thought that an emoji following an element would refer to this preceding element. Therefore, using only the sentences with the most agreement among participants, which meant over 70% agreement, it was decided to place emojis directly after what they were referring to. Despite this, reaction times were slower for congruous responses than incongruous ones, and more inaccurate. Also, the grand average of all the responses to the congruous versus incongruous probes were unable to produce an N400 effect that was statistically significant. In fact, both grand averages for congruous and incongruous probes look almost identical on the scalp map and extremely similar on the waveforms. This means that the semantic supplementation of emojis was confusing on a grand average scale resulting in the same general cognitive response to both congruous and incongruous probes to the same sentences. To investigate further if any emojis in isolation produced an N400 effect, all 14 different emoji types and the sentences they are in were averaged in groups and resulted in only one of them producing a statistically significant N400 effect, which was the emoji depicting fear. While there were no predictions for individual emoji responses, the results for the scared one follow with the general hypothesis that there would be a larger negative N400 effect for incongruous emojis versus congruous ones.

The additional analyses of the behavioral and physiological reactions to the two types of sentence probes where the emoji referred to an object or subject before its placement in the sentence or after it resulted in no significant effects. The accuracy was lower and insignificantly slower when the emoji referred to something before it in the sentence than after meaning that participants had no real response difference regarding where the emoji was referring to in the sentence. This could mean that subjects are equally comfortable retaining the content of the sentence with the context of the emoji in the sentence at the same time on a more global level where the semantic processing effect of the subsequent probe results in no real difference. This potentially allows for greater freedom in the placement of the emoji in a sentence, as long as the reader can make sense of it.

There can be several reasons why there was no success in a kind of universal emoji complementation in sentences. The first can be seen from the pre-study, where only few sentence examples had over 70% participant agreement on the same configuration of text and emoji. This shows that many people had extremely different opinions about where emojis should be located in sentences which could be a result of their country of origin, country they currently live in, background, age, education level, and familiarity and experience with using emojis. So even with the placement that was most agreed upon, many participants certainly still found their location to not be ideally placed. This placement problem could have caused participants to think the emojis did not refer to the subjects or objects in the sentence, but to the whole sentence holistically. Another reason for the lack of results may be due to the fact that in the internet survey the stimuli used static emojis while in the EEG experiment moving emojis were used. In the future, differences in processing the two types of emojis should be investigated. A third reason could be that the sentence primes were quite complicated and longer than in previous EEG experiments.

The lack of definitive results can be explained by many reasons, such as cultural differences of emoji experiential usage in everyday life, which is not isolated to variables of age, gender or nationality. Another explanation for the results, which was discovered during the post experiment interview process was that some subjects reported intentionally ignoring the emojis in sentences since they were under the assumption that they were there as some kind of "trick" that they did not want to be susceptible to. Without the semantic priming of the emoji in the sentence, there could neither be a congruous or incongruous probe response. This issue could not be avoided because during the participant instruction phase of the experiment, they were simply told to read the sentence as they normally would, and there was no emphasis on paying particular attention to the emoji.

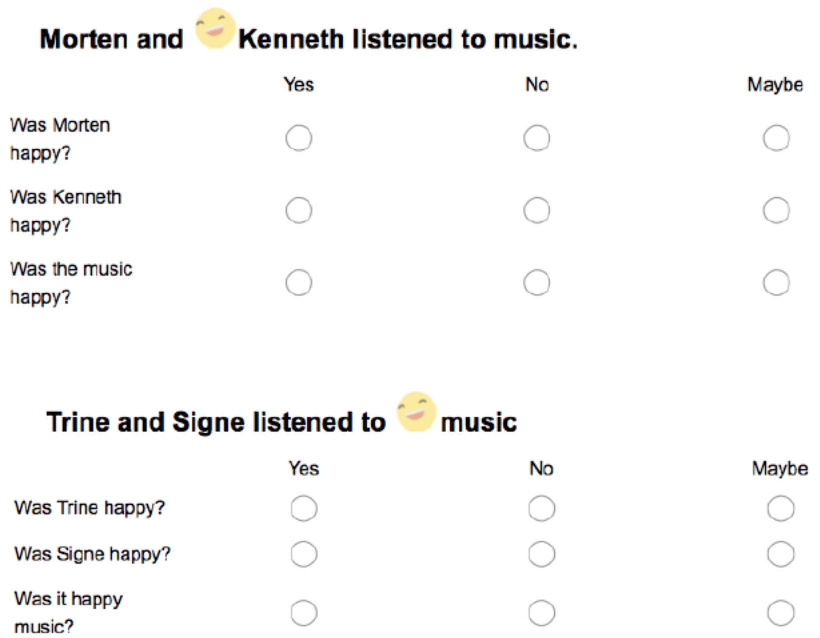
Finally, there was a large discrepancy between participants regarding which emotion the emoji was supposed to portray. One notable example was that two people reported one emoji to be "flirtacious" while the other claimed it was "angry". With such a large discrepancy on the semantic and emotional meaning and content of the emojis, there would also be a likewise disagreement in which probes were congruent and incongruent. These, along with potentially other variables account for why the grand average congruous and incongruous responses looked identical. These results are in alignment with Miller et al. (2016) who found that emojis are very open to interpretation with large variability in opinion regarding both sentiment and semantics and thus may lead to communication error. One solution to this would be some kind of standardization of emojis for particular emotions and expressions, however as Miller et al. (2016) explains, the same type of emoji is displayed differently on different company devices and platforms.

While the results of this study are inconclusive, further investigation seems important not only due to how gestures help people learn and understand, but also that when we are prohibited from gesturing, our ability to communicate becomes less fluent (Rauscher et al., 1996) and therefore, finding the best way to include emojis in text could help people express themselves properly.

13.6 Conclusion

This study wanted to find placements for emojis into text which would provide universal benefits in reading comprehension due to the added benefit of visual information providing body behavior which is extremely crucial for proper communication and message transmission. Following a survey which indicated that participants thought the emoji mostly referred to the element they followed in the sentence, we tested how people processed a large number of moving emojis in this position in various sentences. We didn't find significant differences in the N400 effect between congruent and incongruent probes, and neither in reaction time. The main reason for this was probably that the participants thought that the emojis were ambiguous and chose to ignore them when processing the prime sentence. Furthermore, the sentence primes in which the emojis occurred were quite long which might have made the task too difficult. The results also indicate that there were differences between the participants who answered the survey on the internet and the subjects who participated in the EEG experiment in terms of knowledge of infrequent emojis. Another reason for the lack of conclusive results of the EEG experiment may be due to the fact that moving emojis were used in it while static emojis were shown in the online survey. Whether there are differences in the way people process static and moving emojis should be investigated in the future. Many steps need to be taken to create such a system which involves more pilot studies regarding people's interpretations of emojis, and using a smaller target of homogeneous subjects.

13.7 Appendix



Morten and Kenneth listened to music.

	Yes	No	Maybe
Was Morten happy?	○	○	○
Was Kenneth happy?	○	○	○
Was the music happy?	○	○	○

Trine and Signe listened to music

	Yes	No	Maybe
Was Trine happy?	○	○	○
Was Signe happy?	○	○	○
Was it happy music?	○	○	○

Figure 13.7: Examples of two questions from the survey.

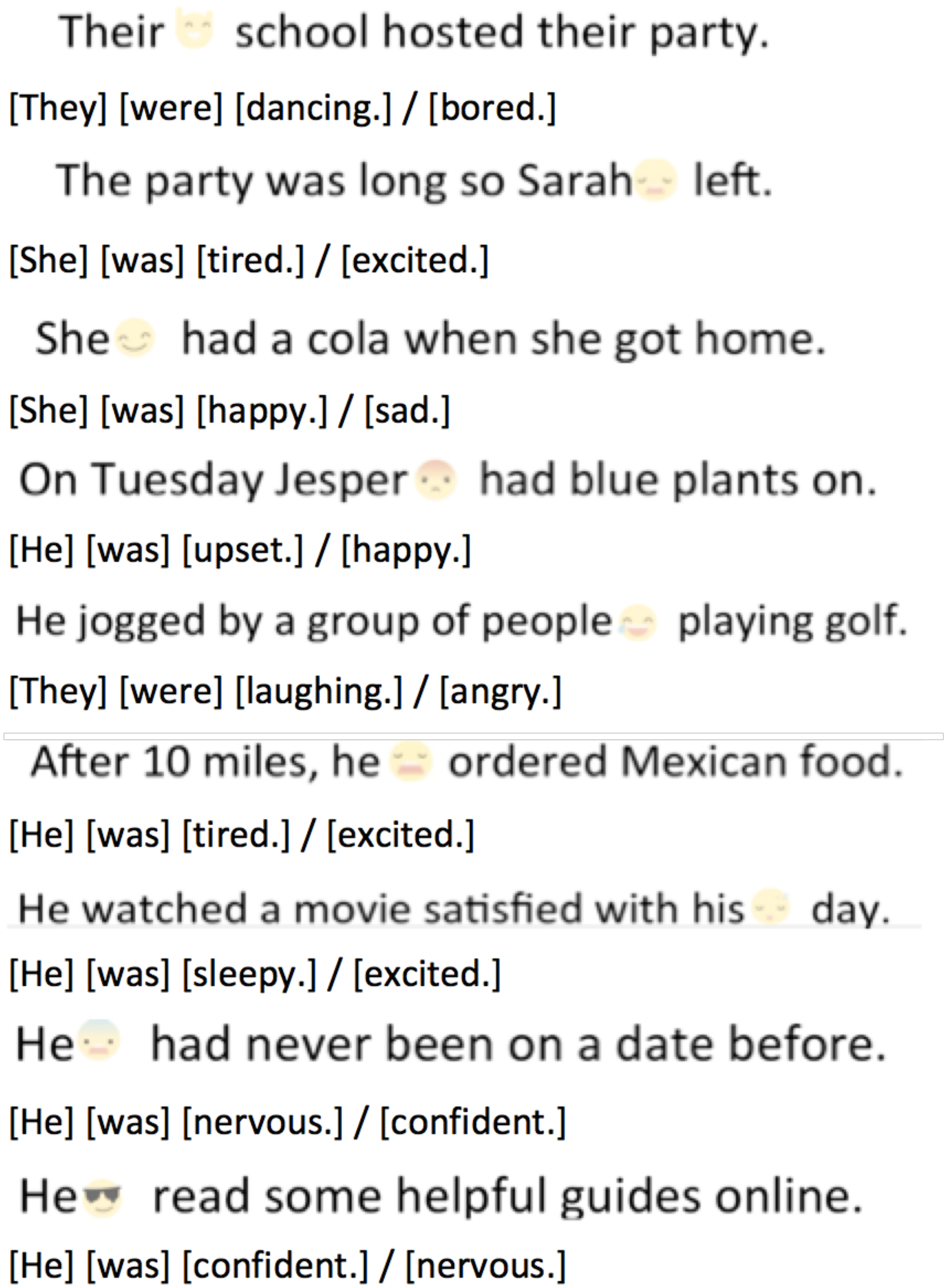


Figure 13.8: Examples of stimuli sentences with congruous and incongruous probe sequences.

14

Investigation of emoji type and placement in sentences with respect to the semantic priming effect

Thomas Ousterhout to be submitted to the CogInfoCom Conference 2017.

14.1 Abstract

The integration of emojis in everyday written messaging is becoming an increasingly popular and common phenomenon. Its popularity seems to be due to the fact that it can supplement written text by adding to it a visual feature that adds information similar to that provided via body behavior during face-to-face communication. These emojis, at the very least, seem to have the ability to reinforce written text, as well as provide additional information to the message. This study investigates if there is a difference in the semantic processing effect when different types of emojis, being hand gestures, pictures of faces, and moving faces, are placed in various positions in short text messages. The investigated positions are the beginning of the sentence, the end of the sentence, and following the element (a person, object or event) which the emoji refers to. The emojis present new information and their influence on the reader are measured through behavioral responses of accuracy and reaction time. The results showed that semantic priming was the same independently of the placement of the emoji. As expected, the reaction time was faster for facial emojis than for gestural emojis, while it was faster for

static emojis than for moving emojis suggesting that the latter distracted the readers from the main task.

14.2 Introduction

In face-to-face communication, people typically simultaneously utilize auditory and visual modalities. Auditory information typically only consists of speech, and visual information includes body behavior such as facial expressions, body posture, hand and arm gestures. These body behaviors are used primarily to disambiguate context with supplemental information to the dominantly used vocal information, to reinforce speech or to change its meaning (Goldin-Meadow, 1999; Kelly et al., 1999; Kendon, 2004; McNeill, 2005). The auditory modality typically provides most of the content in face-to-face communication and thus usually is classified as the dominant modality. However, the visual cues are very important and sometimes necessary to understand fully what the intended message is (Clark, 1996).

An integral part of our daily communication paradigm is hand gesturing. Hand gesture types while being part of a large continuum, can simultaneously be categorized into several groups. One type, for example, is called an emblem, which is a hand gesture that has a conventionalized meaning in a particular culture, such as the "thumbs up" gesture in western cultures, and requires no verbal supplement. Emblems can be useful in face-to-face communication because a single gesture alone can give a complicated and instantaneous message to the recipient (McNeill, 1992). Therefore, since there is a strong relationship between the gesture and its meaning, emblems can be considered unspoken words or phrases. Another gesture type is iconic and is used to represent an object or an action such as mimicking playing a piano. There are also deictic gestures, which comprise pointing hand gestures and batonic gestures which rhythmically accompany speech, emphasizing it. Finally, non-deictic indexical gestures, such as facial expressions, indicate the attitude of the conversational participant (Allwood et al., 2007).

In face-to-face communication, gestures clearly contribute to the semantic meaning of speech. Some aspects of semantic meaning can be measured with EEG by looking at so-called event-related potentials (ERPs), which are EEG amplitude deflections in the brain produced in response to certain events or stimuli. The N400 ERP is a negative deflecting component occurring 400 ms after the onset of an auditory or visual stimulus that has been studied for over thirty years to measure semantic processing (Duncan et al., 2009; Gunter and Bach, 2004; Kutas and Federmeier, 2000, 2011).

The N400 ERP is used to measure semantics because when there is an incongruous stimulus, in comparison to a congruous one, or one that is expected, there is a much larger negative deflection. With this ERP, the semantics of sentences, videos, or any other stimulus type can be measured to see if they are congruous or incongruous with the preceding context by looking at the ERP amplitude deflection. But not all electrode positions can measure this ERP since the responses to abstract words in semantic processing are typically found in centro-parietal sites, while concrete words, such as those referring to picturable objects, have a frontal distribution (Holcomb et al., 1999).

The way in which one processes and understands semantic information from probe stimuli concluding a particular context is highly dependent on one's ability to remember previous relevant semantically meaningful stimuli from various channels and/or modalities such as image, video, text and speech (Federmeier and Kutas, 2001; Lau et al., 2008; Petten and Luka, 2006).

A recent trend seems to be occurring in computer-mediated communication that consists in more and more people using semantically meaningful pictures (commonly called emojis) in their short text messages to supplement the written text with pictures that resemble body behavior in face-to-face communication. Evolutionary psychology explains that body behavior such as hand gestures, body posture, and facial expressions, have been around long before spoken language came about (Corballis, 2003; Hewes et al., 1973; Rizzolatti and Arbib, 1998; Tomasello, 2000). While semantically meaningful pictures attempt to replicate this body behavior in text messages, it is unclear if they are a sufficient substitute.

This study aims to measure whether there is an effect on probe comprehension in relation to the type of emoji, or its placement in a sentence prime as a step to determine possible uses of emojis to improve reading and interpretation of texts. The most common position of emojis is at the end of a sentence, but recently people have started placing them in other positions. In the present experiment, three different kinds of emoji are used in three specific locations. The three emoji types are of hand gestures, pictures of facial expressions, and videos of moving facial emojis. My preceding experiments indicated that there might be differences in the way they are processed. The placement of the emojis are before the sentence, directly after the word indicating a person, object or event, which the emoji modifies, and at the end of the sentence. These placements were chosen for the following reasons. The most common position of emojis in short messaging is the end of the sentence, and the emojis modify the whole sentence. Placing the emoji at the start of the sentence changes the prominence of the emoji, because it is topicalized. The 3rd placement, after the element the emojis refers to independently of the type of word (noun or verb), is chosen to compare the other positions with the emojis' placement inside the sentences used in Chapter 13. We wanted to determine whether the various placements influence reaction time. We have used the same position for all emojis in order to compare their types, but we are aware that the temporal relation of various types of hand gestures with speech as well as the relation between facial expressions and speech are different in face-to-face communication.

Our hypotheses with respect to the types of emojis are the following: a) semantic processing will be faster when facial emojis are used since they are more frequently used than hand gesture emojis; b) semantic processing will be different between moving emojis since the experiment in Chapter 13 seems to indicate this. With respect to the position of emojis in sentences, we wanted to systematically test whether the position of the emoji influences reaction time.

14.3 Method

14.3.1 Participants

Thirteen subjects participated in the experiment. They were native English or fluent English speaking adults. Their ages ranged from 24 to 62 years, 7 were males and 12 were right handed. All participants were university students or had a university education. They signed an informed consent form ensuring their understanding of the experiment to be conducted and their permission to use the data for research. All participants had normal or correct-to-normal vision with no reported psychiatric, neurological, or reading disorders that could disrupt this study's efficacy.

14.3.2 Stimuli and Procedure

Before running the experiment, the researcher reviewed all the emojis with the participants to ensure that they knew them. The stimuli consisted of short sentences primes each with one of the three types of emojis in one of the three locations. See Figure 14.1

Figure 14.1: Sentence example followed by [He][was][happy.] as congruous probe and [He][was][sad.] as incongruous probe.

for a sentence example and see the appendix to the chapter for some more examples. The emojis themselves consisted of 6 pictures of hand gestures, 6 pictures of facial emojis, and the same 6 emojis in video format playing on a replay-loop lasting 1-2 seconds. When each prime sentence was displayed, subjects had as much time as they needed to read the sentence and the experiment would continue with a button press. Then a sequence of three slides with a single word each appeared for 500 ms. The three words formed a phrase and the third word was directly congruent or incongruent to the emoji. The third word, which was the probe, usually was one syllable long and common to ensure that it could be read at equal speed regardless of English proficiency level. Also, the longer and slightly less frequent probes were matched equally for congruent and incongruent conditions so that they were balanced.

Participants had to read the prime sentence at their own speed and the experiment would continue with a button press on a keyboard. Once a button was pressed a fixation cross for a random duration between 750 and 1250 ms would appear. Afterwards the three words of the probe sentence would appear in the center of the screen, each for 500 ms, and on the final probe word, which was always the 3rd word, the participant was required to make a button press response for whether the word was congruent or incongruent with the prime sentence containing the emoji. This experiment consisted of 1 cycle of all 108 prime sentences and took about 20 minutes to complete. The experiment was conducted in a university office with potential for noise disturbances outside to support continued research with this device in more natural environments. After the experiment was concluded, participants were asked about their opinion of the experiment as a whole, the sentences' understandability, and their thoughts about the emojis themselves. Everyone claimed that the sentences made sense, nothing was confusing, and the emojis were clearly interpretable. Some of the participants said that the moving emoji were the most salient ones.

Table 14.1: Behavioral Results

	Emoji																	
Emoji Types	Gesture						Picture						Video					
Position	Start		Middle		End		Start		Middle		End		Start		Middle		End	
Congruency	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I
Accuracy	96.1%	92.8%	90.9%	95.7%	94.8%	92.9%	93.5%	90.2%	88.0%	100%	88.1%	98.5%	94.8%	94.1%	98.6%	97.0%	97.4%	98.6%
Reaction time (ms)	1117.9	1309.9	1107.6	1502.9	1157.8	1150.0	972.8	999.0	887.3	1029.3	927.9	970.8	960.9	1057.1	948.64	1179.4	939.0	1248.6
RT Significance	0.38		0.07		0.92		0.87		0.34		0.52		0.37		0.09		0.08	

14.4 Results

A 3 x 3 repeated measures ANOVA was performed comparing the three emoji types with the three positions. There was an effect for emoji type $F(2,130) = 3.952$, $p = 0.022$, but not for position $F(2,130) = 0.403$, $p = 0.67$. The grand total for the correct response times for hand gesture emojis (in all three positions combined) was 1128.02 ms, for picture facial emojis it was 930.82 ms, and for video facial emojis it was 951.4 ms. The behavioral results can bee seen in Table 14.1. The three types of emoji are divided into their appropriate groups for hand gestural emojis, pictures of facial emojis, and movies of facial emojis. The position is also listed with "Start" for being placed at the beginning of the sentence, "Middle" for being placed inside the sentence after the relevant word, and "End" for being placed at the end of the sentence. Congruency is also included with "C" for congruous stimuli and "I" for incongruous ones. 16 out of the 18 subgroups had an accuracy of over 90%. Their individual reaction time in milliseconds is listed, and the statistical significance is also listed with a Student's T-Test between the congruous and incongruous reaction times for each subgroup. None of the subgroups had statistically significant differences in reaction time to congruous and incongruous stimuli and no other significant effects were found.

14.5 Discussion

The results about the placement of the emojis confirm our impression from the preceding study (Chapter 13) that the position of emojis does not influence reaction time. With respect to the emojis type the results show that facial emojis were processed faster than gestural emojis. This could be due to the fact that facial emojis are more commonly used. The results could also have been affected by saliency reasons since facial expressions are the most immediately perceived body behavior (Rizzolatti and Fabbri-Destro, 2008), and thus are retained in short-term memory better.

The fact that the reaction time in cases of moving emojis was lower than in the case of picture emojis is surprising since some of the participants reported that they perceived the former as more salient than the latter. However, these results might be due to the fact that the moving emojis surprised the subjects distracting them from the task. The same effect was found in Chapter 11 where the picture of a man was processed with more difficulty than the picture of a hand gesture.

Regardless of the type of emoji or its placement, the accuracy of the responses was almost always over 90% meaning that participants were easily able to retain the information that the emojis provided. This shows that the extra visual information that the emojis provide to the text does produce a cognitive effect regardless of if the emoji is in the beginning, middle or end of the sentence. This indicates that even though users typically use emojis at the end of the sentence, when emojis are used in other positions, they provide the same kind of supplementation to the text. This is interesting since it demonstrates that emojis do not necessarily need to refer to the whole sentence as one could conclude from their most frequent use. This can also be observed in tweets where people use emojis with no clear pattern of placement.

This can explain the lack of clear results in the preceding chapter's survey with respect to an ideal position for facial emojis. In that survey people had many different opinions about what the emojis referred to when they were placed in different positions. The present experiment indicates more clearly that there might not be an ideal position for emojis when they add information to text.

The results help answer the question as to whether or not emojis influence the way readers perceive sentences. There seems to be a positive result in their usage with no rules on how to use them since they can be integrated freely. Writers can freely have their own style and be creative in generating text-based communication that has multiple visual forms.

There was no difference in the reaction time between the congruent and incongruent conditions. This could be due to the overly repetitive nature of the experiment. Participants could have learned what to expect even in incongruous conditions and therefore there was no longer processing time. Similarly, it is possible that subjects learned to focus on the emoji more than on the entire sentence. A future study could have a mixture of questions relating to the text and to the emojis in order to control this.

14.6 Conclusion

People have a tendency to use emojis in their short messages. These emojis provide supplemental information to the sentence aiding in behavioral responses. While there is no clear difference if the way participants process emojis with respect to their placement in the sentence, we found that the use of facial emojis results in faster reaction time than the use of gestural emojis and moving facial emojis. We also found that all emojis contributed to the semantic processing of the sentence. Emoji usage in general seem to be increasing with popularity, and a new kind of writing style is being created. It would be advantageous to study the effects of this new style on the reader and how to enhance its production and effectiveness in transmitting the intended message. It could also be interesting to determine whether users with different language backgrounds use the emojis differently.

The results of this experiment indicate that the use of emojis in sentences is complex and requires further investigations especially to determine how they can be used to help people with reading difficulties, or even as a way to support people who have difficulties in identifying the affective valence of sentences.

Since emojis can either reinforce words or replace them, it is valuable to investigate whether they can be used to help people who have reading difficulties or have problems in understanding the semantics of words that express emotions.

14.7 Appendix

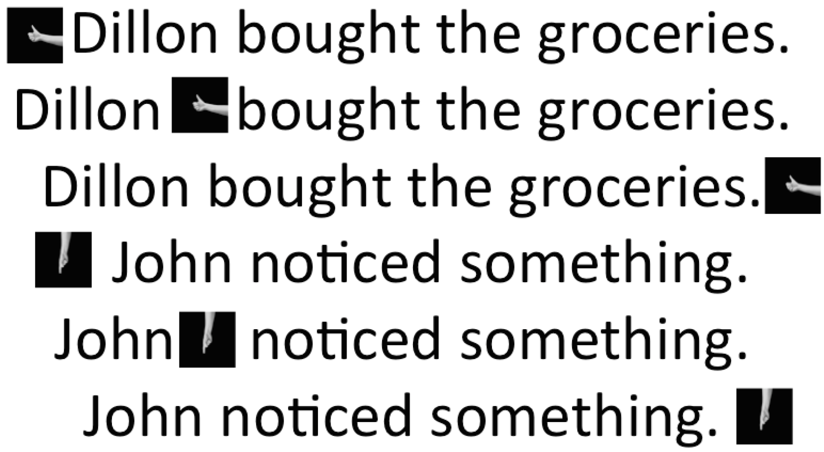


Figure 14.2: Sentence examples of gesture emojis in all three positions. The congruous/incongruous probes for the first three were [He][was][happy.] / [upset.], and for the second three were [On] [the] [floor] / [roof.]

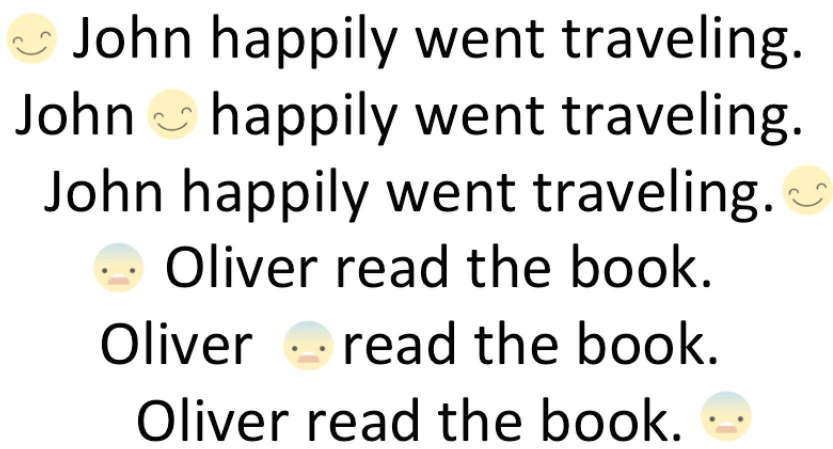


Figure 14.3: Sentence examples of moving facial emojis in all three positions. The congruous/incongruous probes for the first three were [He][was][happy.] / [sad.], and for the second three were [He] [was] [scared.] / [bored.]

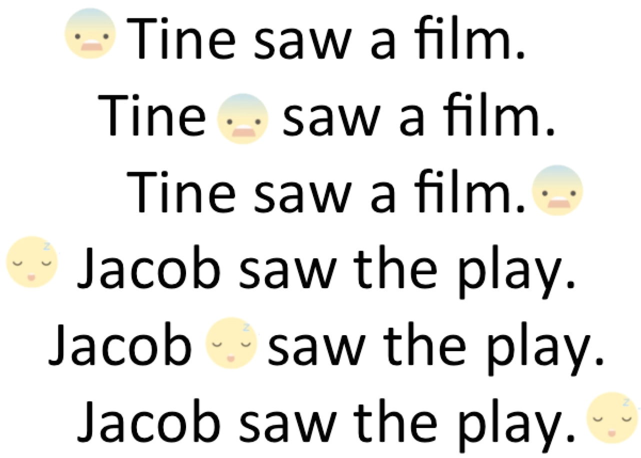


Figure 14.4: Sentence examples of pictural facial emojis in all three positions. The congruous/incongruous probes for the first three were [She][was][scared.] / [glad.], and for the second three were [He] [was] [bored] / [excited.]

15

Conclusions

This chapter presents the conclusions of the entire dissertation and discusses future research. The dissertation has investigated a number of cognitive and behavioral effects of incorporating pictures and moving pictures of hand gestures and facial expressions, all called emojis, into text. The purpose was to determine if these pictural stimuli complemented the written text in a way that facilitated reading ability and/or comprehension. First, since face-to-face communication is multimodal, but written computer-mediated messaging is not, people have started using emojis, and especially facial emojis, in order to add attitudes and other nonverbal signals to text compensating for the missing body behavior. It is unclear how emojis and text are processed. Secondly, since emojis are becoming increasingly popular, it is valuable to investigate whether text and pictures can be integrated in more ways than it is presently done with the purpose of improving reading and, possibly, making human-computer interaction and computer-mediated messaging more similar to the natural face-to-face communication. Finally, the investigation of how humans process simple emojis and texts can be valuable for the production of short messages and pictures in e.g. advertising and traffic signals.

In Chapter 2, types of gesture were discussed and theories about their functions in communication were presented. This was relevant because gesture type, function and use are connected and these relations must be understood when analyzing emojis as a replacement for different gestures. The visual modality of face-to-face communication was further discussed in Chapter 3. Here, emotion expression was the focus, since many of the emojis used in the thesis and in text messaging represent facial expressions and are connected to the emotions which these expressions can show. The topic covered in

Chapter 4 was mirroring theory research, which provided neurological information for the importance of how exactly we respond to each others' behaviors when performing actions, and also when looking at pictures of actions. This justifies on a neurological level why emojis can be considered as a replacement for some body actions. The information provided in Chapter 5 introduced the method of EEG data acquisition and analyses. Chapter 6 accounted for a number of studies that have used EEG to measure gesture processing which were the ground work for my research articles.

These articles provided a number of contributions to the research investigating the use of emojis in text and EEG studies. First, I demonstrated through a replication study of Gunter and Bach (2004) that the Emotiv headset, which is a relatively cheap commercial grade EEG acquisition device, can give results comparable to those of expensive high resolution medical grade equipment when measuring some ERP signals such as the N400. This is useful for a number of reasons. First, researchers like myself and students can perform EEG experiments and contribute to the scientific world through published scholarly articles without the need of a medical environment and expensive equipment. This also means that the rather exclusive field of EEG research can be easily accessed by virtually anyone allowing the volume of contribution to the field to increase dramatically. Furthermore, it can be assumed that the participants in the EEG experiments behave more naturally in an office than in an experimental laboratory and when they wear a simple EEG headset. I also showed that the electrodes that were best for detecting the N400 were the four in the back of the Emotiv since the Pz electrode, which is the electrode mostly used for N400 detection, is not available in it.

Second, I showed by measuring the N400 effect with the Emotiv that the semantic priming effect could be produced even though the prime and probe stimuli were two different channels within the same modality (text and pictures). Similarly to results found by Bajo (1988), this shows that the visual differences of the two channels do not produce too much cognitive load when they must be integrated, and that their semantic meanings can be processed congruently. Since integrated text and emojis are processed easily, I tested whether emojis can be easily processed when they occur in other places than in the most common sentence final position. Using the results from the previous experiment, it was shown that individuals perceive hand gesture emojis following or preceding the text to which they are related in a way that is parallel to how they process co-speech hand gestures. More specifically, it was found that emojis following the related text are perceived more naturally when they add information to the text than when they reinforce its content.

In the following experiment, I investigated the placement of text and pictures of gestures further, and I found that having the emoji to the left of the word which it was related to, and provided reinforcing information for, produced the fastest reaction time. This study also corroborates with preceding research regarding the temporal relationship between speech and hand gestures stating that co-speech gestures are produced slightly before the speech they are reinforcing (Bergmann et al., 2011).

The next contribution was testing the integration of emojis in full sentences, and the emojis represented not only emblematic hand gestures, but also simple happy/sad faces. In this experiment, the emojis were used instead of adjectives. Since an N400 effect was found, semantic priming was produced showing semantic integration of the emojis in the sentences. Also there were faster reaction times for congruent emojis placed in the sentences versus incongruent ones. This further demonstrates the efficacy of strategically incorporating emojis into text in order to add information otherwise not available to it. The experiment also shows the behavioral and physiological benefits of the reading level that are produced thereafter. Although the reading benefits could have been due to a bias in the probes, this is unlikely since the frequency/length of the probe words used were mostly only one syllable long and very common so that they could be read at equal speed regardless of English proficiency. Moreover, the longer and slightly less frequent probes were matched equally for congruent and incongruent conditions in order to balance out. Therefore, it seems unlikely that their could have been any biases to either congruent or incongruent conditions.

With the information thus far, I wanted to progress further by incorporating moving facial emojis into text. But since facial expressions are cognitively processed differently than hand gestures and their temporal relationship to speech can be different, a pilot study was first conducted to see where participants liked the placement of the facial emojis in the sentence examples the most. The survey results indicated that participants had a slight preference for facial emoji to be placed to the right of the element which they referred to. Thus 45 sentences were created using 14 different moving facial emojis placed to the right of the subject they described. The results turned out negative with no N400 semantic priming effect between congruous and incongruous probes and no benefit in reaction time or accuracy.

These results could have been due to a number of variables that were not controlled. Participants who answered the survey and people who participated to the last EEG experiment ranged greatly in country of origin and country of residence, preferences and experience of using emojis, language, age, gender, and education level. Furthermore, in the online survey picture emojis were used while in the experiment moving emojis were added to the text, and this might also have had an influence on the results. Finally the sentences in this experiment were longer than in preceding experiments and this added possible ambiguities in the interpretation of the emojis and the text.

In order to control inconclusive results in the preceding experiment, a final study was produced where emoji type and placement were the focus. This study used gestural emojis, static facial emojis, and moving facial emojis, in three different locations, before the sentence, right after the object being modified, and at the conclusion of the sentence. Since all emojis added information to the sentences, this position was found acceptable also for hand gestural emojis according to the preceding experiment (Chapter 10). While facial emojis had better response times than hand gestural ones, there was no effect of the emoji's placement since it aided in the transmission of the message equally in all locations. This could explain the difficulty of finding one ideal location in the previous experiment, since there seems not to be an ideal location for the emojis. Participants in the survey had different opinions and interpretations of what was ideal. When using different types of locations, people integrate emojis and text nicely. This means that authors of sentences with emojis can add information to the text with tremendous freedom.

I found that facial emojis were processed faster than hand gesture emojis either because they are more common or because pictures of facial expressions are processed more immediately than pictures of hand gestures. A little surprising was the fact that the reaction time was slower with moving emojis than with picture emojis since some participants said that they perceived the former as more salient than the latter. However, the results might be due to the fact that the moving emojis can have distracted the participants influencing negatively the reaction time.

Another experiment determined that a single hand preceding a word was processed much faster than the figure of the entire man, probably because the more complex picture required more cognitive load and distracted the participants from the task as was the case with moving emojis. I also found that if sentences were kept simple, participants had an easier time processing the emojis in them. This is not surprising since emojis inside long sentences can refer to more words than in short sentences.

The results of the last experiment indicating that subjects processed the semantics of combined emojis and text independently of their position in the sentence show that people have become very accepting of emojis in short sentences. While people usually use emojis in sentence final positions, they can quickly adapt to other placements. The inclusion of emojis as either a word substitute, or as a reinforcement of a word, is increasing. It will be interesting to see in the long term whether short messaging including emojis will influence the writing style.

The increasing popularity of emoji shows that there is a need and a desire to include them in text. Even though this dissertation presented various studies analyzing the perception of different types of emoji primes and probes combined with single words or sentences of different complexities, more analyses of their reception are needed. One limitation of the work presented in this thesis is the number of subjects which participated in each experiment. It was hard to find voluntaries and some individuals participated to several experiments. Therefore, more large scale experiments are needed and several aspects still need to be investigated, such as the use of other emoji types than facial expressions and hand gestures, their position, their relation to syntax, the effect of emojis on short and long memory and their impact on the reader in terms of style. Furthermore, every participant in all of the experiments in this dissertation had university level education. On one side this is an advantage since the population in the experiments was the same. On the other side, this is problematic since I didn't test how people with lower education levels process emojis in text.

Bibliography

Martha W Alibali. Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4):307–331, 2005. 10, 12

Martha W Alibali, Dana C Heath, and Heather J Myers. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44(2):169–188, 2001. 12

Truett Allison, Aina Puce, Dennis D Spencer, and Gregory McCarthy. Electrophysiological studies of human face perception. i: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral cortex*, 9(5):415–430, 1999. 59

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9(1):1–26, 1992. 7, 15, 16

Jens Allwood, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta, and Patrizia Paggio. The mumin coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3-4):273–287, 2007. 16, 94, 95, 148

Michael Andric, Ana Solodkin, Giovanni Buccino, Susan Goldin-Meadow, Giacomo Rizzolatti, and Steven L Small. Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, 51(8):1619–1629, 2013. 55

David F Armstrong, William C Stokoe, and Sherman E Wilcox. *Gesture and the nature of language*. Cambridge University Press, 1995. 131

James R Augustine. Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews*, 22(3):229–244, 1996. 26, 28

Nicholas A Badcock, Petroula Mousikou, Yatin Mahajan, Peter de Lissa, Johnson Thie, and Genevieve McArthur. Validation of the emotiv epoc® eeg gaming system for measuring research quality auditory erps. *PeerJ*, 1:e38, 2013. 133

Nicholas A Badcock, Kathryn A Preece, Bianca de Wit, Katharine Glenn, Nora Fieder, Johnson Thie, and Genevieve McArthur. Validation of the emotiv epoc eeg system for research quality auditory event-related potentials in children. *PeerJ*, 3:e907, 2015. 133

Ou Bai, Varun Rathi, Peter Lin, Dandan Huang, Harsha Battapady, Ding-Yu Fei, Logan Schneider, Elise Houdayer, Xuedong Chen, and Mark Hallett. Prediction of human voluntary movement before it occurs. *Clinical Neurophysiology*, 122(2):364–372, 2011. 74

Lyndsay M Baird and A Mike Burton. The bilateral advantage for famous faces: Interhemispheric communication or competition? *Neuropsychologia*, 46(5):1581–1587, 2008. 106, 107

Maria-Teresa Bajo. Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4):579, 1988. 83, 106, 158

Filippo Barbieri, Antimo Buonocore, Riccardo Dalla Volta, and Maurizio Gentilucci. How symbolic gestures and words interact with each other. *Brain and Language*, 110 (1):1–11, 2009. 119

Rafael Barea, Luciano Boquete, Manuel Mazo, and E López. Wheelchair guidance strategies using eog. *Journal of intelligent and robotic systems*, 34(3):279–299, 2002a. 74

Rafael Barea, Luciano Boquete, Manuel Mazo, and Elena López. System for assisted mobility using eye movements based on electrooculography. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(4):209–218, 2002b. 74

Rafael Barea, Luciano Boquete, Luis Miguel Bergasa, Elena López, and Manuel Mazo. Electro-oculographic guidance of a wheelchair using eye movements codification. *The International Journal of Robotics Research*, 22(7-8):641–652, 2003. 74

Sarah E Barrett and Michael D Rugg. Event-related potentials and the semantic matching of pictures. *Brain and cognition*, 14(2):201–212, 1990. 80, 82, 90, 119, 126

Sarah E Barrett, Michael D Rugg, and David I Perrett. Event-related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia*, 26(1):105–117, 1988. 119

SE Barrett and Michael D Rugg. Event-related potentials and the semantic matching of faces. *Neuropsychologia*, 27(7):913–922, 1989. 119, 126

Janet Bavelas, Jennifer Gerwing, Chantelle Sutton, and Danielle Prevost. Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58(2):495–520, 2008. 18

Geoffrey Beattie and Heather Shovelton. Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? an experimental investigation. *Semiotica*, 123(1-2):1–30, 1999. 13, 94

Geoffrey Beattie and Heather Shovelton. What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology*, 41(3):403–417, 2002. 10, 12

S Bentin. Visual word perception and semantic processing: an electrophysiological perspective. *Israel journal of medical sciences*, 23(1-2):138–144, 1986. 56, 80

Shlomo Bentin, Gregory McCarthy, and Charles C Wood. Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology*, 60(4):343–355, 1985. 56, 80, 119, 123, 126

Shlomo Bentin, Truett Allison, Aina Puce, Erik Perez, and Gregory McCarthy. Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, 8(6):551–565, 1996. 59

Kirsten Bergmann, Volkan Aksu, and Stefan Kopp. The relation of speech and gestures: temporal synchrony follows semantic synchrony. In *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction (GeSpIn 2011)*, 2011. 94, 95, 96, 97, 101, 102, 113, 159

Jos J Berkum, Peter Hagoort, Colin Brown, et al. Semantic integration in sentences and discourse: Evidence from the n400. *Cognitive Neuroscience, Journal of*, 11(6): 657–671, 1999. 50

Jos JA Van Berkum, Pienie Zwitserlood, Peter Hagoort, and Colin M Brown. When and how do listeners relate a sentence to the wider discourse? evidence from the n400 effect. *Cognitive brain research*, 17(3):701–718, 2003. 50

Paolo Bernardis and Maurizio Gentilucci. Speech and gesture share the same communication system. *Neuropsychologia*, 44(2):178–190, 2006. 119

Ferdinand Binkofski, Katrin Amunts, Klaus Martin Stephan, Stefan Posse, Thorsten Schormann, Hans-Joachim Freund, Karl Zilles, and Rüdiger J Seitz. Broca's region subserves imagery of motion: a combined cytoarchitectonic and fmri study. *Human brain mapping*, 11(4):273–285, 2000. 8

Eva Bonda, Michael Petrides, S Frey, and AC Evans. Frontal cortex involvement in organized sequences of hand movements: evidence from a positron emission tomography study. In *Soc Neurosci Abstr*, volume 20, 1994. 8

Arielle Borovsky, Jeffrey L Elman, and Marta Kutas. Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8(3):278–302, 2012. 119

Arielle Borovsky, Marta Kutas, and Jeffrey L Elman. Getting it right: word learning across the hemispheres. *Neuropsychologia*, 51(5):825–837, 2013. 119

Hidenori Boutani and Mieko Ohsuga. Applicability of the "emotiv eeg neuroheadset" as a user-friendly input interface. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 1346–1349. IEEE, 2013. 133

Marcel Brass, Ruth M Schmitt, Stephanie Spengler, and György Gergely. Investigating action understanding: inferential processes versus action simulation. *Current Biology*, 17(24):2117–2121, 2007. 25, 29

Sara C Broaders and Susan Goldin-Meadow. Truth is at hand how gesture adds information during investigative interviews. *Psychological Science*, 21(5):623–628, 2010. 10, 13

Sara C Broaders, Susan Wagner Cook, Zachary Mitchell, and Susan Goldin-Meadow. Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, 136(4):539, 2007. 139

Catherine P Browman and Louis Goldstein. Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, pages 341–376, 1990. 9

Colin Brown and Peter Hagoort. The processing nature of the n400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1):34–44, 1993. 119

Robbins Burling. Motivation, conventionalization, and arbitrariness in the origin of language. *The origins of language: What nonhuman primates can tell us*, pages 307–350, 1999. 118

Brian Butterworth and Geoffrey Beattie. Gesture and silence as indicators of planning in speech. In *Recent advances in the psychology of language*, pages 347–360. Springer, 1978. 8

Brian Butterworth and Uri Hadar. Gesture, speech, and computational stages: a reply to McNeill. *Psychological Review*, 96(1):168–174, 1989. 8

Andrew J Calder, Jill Keane, Facundo Manes, Nagui Antoun, and Andrew W Young. Impaired recognition and experience of disgust following brain injury. *Nature neuroscience*, 3(11):1077–1078, 2000. 27

Andrew Campbell, Tanzeem Choudhury, Shaohan Hu, Hong Lu, Matthew K Mukerjee, Mashfiqui Rabbi, and Rajeev DS Raizada. Neurophone: brain-mobile phone interface using a wireless eeg headset. In *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*, pages 3–8. ACM, 2010. 74, 83, 108

Laurie Carr, Marco Iacoboni, Marie-Charlotte Dubeau, John C Mazziotta, and Gian Luigi Lenzi. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences*, 100(9):5497–5502, 2003. 27

Thomas H Carr, Charley McCauley, Richard D Sperber, and CM Parmelee. Words, pictures, and priming: on semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 8(6):757, 1982. 126

Justine Cassell, David McNeill, and Karl-Erik McCullough. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & cognition*, 7(1):1–34, 1999. 5

Luigi Cattaneo, Maddalena Fabbri-Destro, Sonia Boria, Cinzia Pieraccini, Annalisa Monti, Giuseppe Cossu, and Giacomo Rizzolatti. Impairment of actions chains in autism and its possible role in intention understanding. *Proceedings of the National Academy of Sciences*, 104(45):17825–17830, 2007. 25, 29

Pritom Chowdhury and SS Kibria Shakim. *Optimizing cognitive efficiency of emotiv EPOC and controlling wheelchair through it*. PhD thesis, BRAC University, 2014. 74

Pulak Chowdhury, SS Kibria Shakim, Muhammad Rezaul Karim, and Md Khalilur Rhaman. Cognitive efficiency in robot control by emotiv epoc. In *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*, pages 1–6. IEEE, 2014. 74

Mingyuan Chu and Sotaro Kita. The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140(1):102, 2011. 139

R Breckinridge Church and Susan Goldin-Meadow. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23(1):43–71, 1986. 10

Ruth Breckinridge Church, Philip Garber, and Kathryn Rogalski. The role of gesture in memory and social communication. *Gesture*, 7(2):137–158, 2007. 13

Owen Churches, Mike Nicholls, Myra Thiessen, Mark Kohler, and Hannah Keage. Emoticons in mind: An event-related potential study. *Social neuroscience*, 9(2): 196–202, 2014. 59

Herbert H Clark. *Using language*, volume 1996. Cambridge university press Cambridge, 1996. 75, 82, 148

Akiba A Cohen and Randall P Harrison. Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, 28(2):276, 1973. 12

Olivier Collignon, Simon Girard, Frederic Gosselin, Sylvain Roy, Dave Saint-Amour, Maryse Lassonde, and Franco Lepore. Audio-visual integration of emotion expression. *Brain research*, 1242:126–135, 2008. 107

Olivier Collignon, S Girard, Frédéric Gosselin, D Saint-Amour, Franco Lepore, and Maryse Lassonde. Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia*, 48(1):220–225, 2010. 107

Stephan M Collishaw and Graham J Hole. Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29(8):893–909, 2000. 58

Montserrat Comesaña, Ana Paula Soares, Manuel Perea, Ana P Piñeiro, Isabel Fraga, and Ana Pinheiro. Erp correlates of masked affective priming with emoticons. *Computers in Human Behavior*, 29(3):588–595, 2013. 61, 132

Jeffrey Coney and Marjorie Abernethy. Picture-name priming in the cerebral hemispheres. *Brain and language*, 47(2):300–313, 1994. 126

Michael C Corballis. *From hand to mouth: The origins of language.* Princeton University Press, 2002. 131

Michael C Corballis. From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, 26(02):199–208, 2003. 118, 149

Antonio R Damasio. *Looking for Spinoza: Joy, sorrow, and the feeling brain.* Houghton Mifflin Harcourt, 2003. 27

Iwan de Kok and Dirk Heylen. The multilis corpus–dealing with individual differences in nonverbal listening behavior. In *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues*, pages 362–375. Springer, 2011. 7, 15

Stefan Debener, Falk Minow, Reiner Emkes, Katharina Gandras, and Maarten Vos. How about taking a low-cost, small, and wireless eeg for a walk? *Psychophysiology*, 49(11):1617–1621, 2012. 74, 83, 108

J Decety, D Peranif, M Jeannerod, V Bettinardif, B Tadary, and R Woods. Mapping motor representations with positron emission. *Nature*, 371:13, 1994. 8

Daantje Derks, Arjan ER Bos, and Jasper Von Grumbkow. Emoticons and social interaction on the internet: the importance of social context. *Computers in human behavior*, 23(1):842–849, 2007. 17, 18

Eli Dresner and Susan C Herring. Functions of the nonverbal in cmc: Emoticons and illocutionary force. *Communication theory*, 20(3):249–268, 2010. 19

James E Driskell and Paul H Radtke. The effect of gesture on speech production and comprehension. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3):445–454, 2003. 10, 12

Connie C Duncan, Robert J Barry, John F Connolly, Catherine Fischer, Patricia T Michie, Risto Näätänen, John Polich, Ivar Reinvang, and Cyma Van Petten. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, p300, and n400. *Clinical Neurophysiology*, 120(11):1883–1908, 2009. 75, 82, 97, 107, 149

Francis T Durso and Marcia K Johnson. Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, 5(5):449, 1979. 106

Matthieu Duvinage, Thierry Castermans, Thierry Dutoit, M Petieau, T Hoellinger, C De Saedeleer, K Seetharaman, and G Cheron. A p300-based quantitative comparison between the emotiv epoc headset and a medical eeg device. *Biomedical Engineering*, 765:2012–764, 2012. 74, 83, 108

Matthieu Duvinage, Thierry Castermans, Mathieu Petieau, Thomas Hoellinger, Guy Cheron, and Thierry Dutoit. Performance of the emotiv epoc headset for p300-based applications. *Biomed Eng Online*, 12:56, 2013. 74, 83, 108

EEGlab. EEGlab - Swartz Center for Computational Neuroscience. `https://sccn.ucsd.edu/eeglab/`, 2016. 136

Martin Eimer. Effects of face inversion on the structural encoding and recognition of faces: Evidence from event-related brain potentials. *Cognitive Brain Research*, 10(1): 145–158, 2000. 59

Hiran Ekanayake. P300 and emotiv epoc: Does emotiv epoc capture real eeg? *Web publication http://neurofeedback. visaduma. info/emotivresearch. htm*, 2010. 133

Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98, 1969. 16

Karen Emmorey and Shannon Casey. Gesture, thought, and spatial language. In *Spatial language*, pages 87–101. Springer, 2002. 10

Emotiv. EMOTIV - Brainwear® Wireless EEG Technology. `http://emotiv.com`, 08 2016. 65, 98, 136

Hong J Eoh, Min K Chung, and Seong-Han Kim. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics*, 35(4):307–320, 2005. 74

Gijsbert Erkens, Johan Theil, Gellof Kanselaar, Maaike Prangsma, and Jos Jaspers. 'w8 ff:-)'. chattalk in een coöperatieve leeromgeving. *Levende Talen Tijdschrift*, 3(2): 24–33, 2002. 19

ERPLab. ERPLAB Toolbox — ERPinfo. `http://erpinfo.org/erplab`, 2016. 136

A Esposito, A Natale, S Duncan, D McNeill, and F Quek. Speech and gestures pauses relationships: A hypothesis of synchronization. In *Proceedings of the V National Conference on Italian Psychology, AIP, Grafica80-Modugno, Bari, Italy*, pages 95–98, 2003. 8

Anna Esposito and Antonietta M Esposito. On speech and gestures synchrony. In *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pages 252–272. Springer, 2011. 8

Anna Esposito, Karl E McCullough, and Francis Quek. Disfluencies in gesture: Gestural correlates to filled and unfilled speech pauses. In *Proceedings of IEEE Workshop on Cues in Communication*, pages 1–6. Citeseer, 2001. 8, 96

Anna Esposito, Susan Duncan, and Francis Quek. Holds as gestural correlates to empty and filled speech pauses. In *Seventh International Conference on Spoken Language Processing*, 2002. 8, 96

Anna Esposito, Daniela Esposito, Mario Refice, Michelina Savino, and Stefanie Shattuck-Hufnagel. A preliminary investigation of the relationship between gestures and prosody in italian. *Fundamentals of verbal and nonverbal communication and the biometric issue*, 18:65, 2007. 8

Maddalena Fabbri-Destro and Giacomo Rizzolatti. Mirror neurons and mirror systems in monkeys and humans. *Physiology*, 23(3):171–179, 2008. 26

Maddalena Fabbri-Destro, Pietro Avanzini, Elisa De Stefani, Alessandro Innocenti, Cristina Campi, and Maurizio Gentilucci. Interaction between words and symbolic gestures as revealed by n400. *Brain topography*, pages 1–15, 2014. 57, 119

Maddalena Fabbri-Destro, Pietro Avanzini, Elisa De Stefani, Alessandro Innocenti, Cristina Campi, and Maurizio Gentilucci. Interaction between words and symbolic gestures as revealed by n400. *Brain topography*, 28(4):591–605, 2015. 126

Martha J Farah, Kevin D Wilson, Maxwell Drain, and James N Tanaka. What is" special" about face perception? *Psychological review*, 105(3):482, 1998. 58

Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988. 74

Kara D Federmeier and Marta Kutas. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(1):202, 2001. 56, 76, 80, 82, 90, 149

Kara D Federmeier, Edward W Wlotko, Esmeralda De Ochoa-Dewald, and Marta Kutas. Multiple effects of sentential constraint on word processing. *Brain research*, 1146: 75–84, 2007. 119

David Freedberg and Vittorio Gallese. Motion, emotion and empathy in esthetic experience. *Trends in cognitive sciences*, 11(5):197–203, 2007. 27, 29, 137

Arao Funase, Tohru Yagi, Yoslhiaki Kuno, and Yoshiki Uchikawa. Prediction of eye movements from eeg. In *Neural Information Processing, 1999. Proceedings. ICONIP'99. 6th International Conference on*, volume 3, pages 1127–1131. IEEE, 1999. 74

Susan R Fussell. *The verbal communication of emotions: interdisciplinary perspectives*. Psychology Press, 2002. 15, 17, 18

Giovanni Galfano, Veronica Mazza, Alessandro Angrilli, and Carlos Umiltà. Electrophysiological correlates of stimulus-driven multiplication facts retrieval. *Neuropsychologia*, 42(10):1370–1382, 2004. 97

Giovanni Galfano, Barbara Penolazzi, Ineke Vervaeck, Alessandro Angrilli, and Carlo Umiltà. Event-related brain potentials uncover activation dynamics in the lexicon of multiplication facts. *Cortex*, 45(10):1167–1177, 2009. 97

Helen L Gallagher and Christopher D Frith. Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia*, 42(13):1725–1736, 2004. 55

Vittorio Gallese, Christian Keysers, and Giacomo Rizzolatti. A unifying view of the basis of social cognition. *Trends in cognitive sciences*, 8(9):396–403, 2004. 27

Giorgio Ganis, Marta Kutas, and ML Sereno. The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Cognitive Neuroscience, Journal of*, 8(2):89–106, 1996. 89, 119

Markus Giray and Rolf Ulrich. Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance*, 19(6):1278, 1993. 106

Alec Go, Lei Huang, and Richa Bhayani. Twitter sentiment analysis. *Entropy*, 17, 2009. 20

Georg Goldenberg. Matching and imitation of hand and finger posturesin patients with damage in the left or right hemispheres. *Neuropsychologia*, 37(5):559–566, 1999. 24

Susan Goldin-Meadow. The role of gesture in communication and thinking. *Trends in cognitive sciences*, 3(11):419–429, 1999. 75, 82, 107, 132, 148

Susan Goldin-Meadow. *Hearing gesture: How our hands help us think*. Harvard University Press, 2005. 8

Susan Goldin-Meadow and Catherine Momeni Sandhofer. Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, 2 (1):67–74, 1999. 10, 13

Susan Goldin-Meadow, Debra Wein, and Cecilia Chang. Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction*, 9(3): 201–219, 1992. 10, 13

Susan Goldin-Meadow, Susan Wagner Cook, and Zachary A Mitchell. Gesturing gives children new ideas about math. *Psychological Science*, 20(3):267–272, 2009. 139

Alvin I Goldman and Chandra Sekhar Sripada. Simulationist models of face-based emotion recognition. *Cognition*, 94(3):193–213, 2005. 27

G Robert Grice, Lyn Canham, and Joseph M Boroughs. Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, 35(5):451–463, 1984. 106

Christoph Guger, Werner Harkam, Carin Hertnaes, and Gert Pfurtscheller. Prosthetic control by an eeg-based brain-computer interface (bci). In *Proc. aaate 5th european conference for the advancement of assistive technology*, pages 3–6. Citeseer, 1999. 74

Thomas C Gunter and Patric Bach. Communicating hands: Erps elicited by meaningful symbolic hand postures. *Neuroscience Letters*, 372(1):52–56, 2004. xii, 56, 57, 75, 76, 77, 78, 80, 82, 83, 84, 89, 90, 91, 94, 98, 107, 108, 118, 119, 120, 126, 127, 149, 158

Emmanuel Guzman, German Palafox, Marcia Grabowecky, and Satoru Suzuki. A visual redundant-signal effect strongly depends on attention even for probability summation. *Journal of Vision*, 8(6):135–135, 2008. 106

Boukje Habets, Sotaro Kita, Zeshu Shao, Asli Özyurek, and Peter Hagoort. The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, 23(8):1845–1854, 2011. 97

Uri Hadar and Brian Butterworth. Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, 115(1-2):147–172, 1997. 8

Peter Hagoort. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, 20:S18–S29, 2003a. 50

Peter Hagoort. Interplay between syntax and semantics during sentence comprehension: Erp effects of combining syntactic and semantic violations. *Cognitive Neuroscience, Journal of*, 15(6):883–899, 2003b. 50

Jeff P Hamm, Blake W Johnson, and Ian J Kirk. Comparison of the n300 and n400 erps to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, 113(8):1339–1350, 2002. 53, 83, 91

Randall P Harrison. Nonverbal communication. *Human Communication As a Field of Study: Selected Contemporary Views*, 113, 1989. 16

Donald Olding Hebb. *The organization of behavior: A neuropsychological approach.* John Wiley & Sons, 1949. 107

Maurice Hershenson. Reaction time as a measure of intersensory facilitation. *Journal of experimental psychology*, 63(3):289–293, 1962. 106

Gordon W Hewes, RJ Andrew, Louis Carini, Hackeny Choe, R Allen Gardner, A Kortlandt, Grover S Krantz, Glen McBride, Fernando Nottebohm, John Pfeiffer, et al. Primate communication and the gestural origin of language [and comments and reply]. *Current Anthropology*, pages 5–24, 1973. 118, 149

Gregory Hickok. *The myth of mirror neurons: The real neuroscience of communication and cognition*. WW Norton & Company, 2014. 23

David Hines. Effect of masked picture primes on semantic priming for easy-and difficult-to-name words. *The Journal of general psychology*, 120(2):149–164, 1993. 126

Philip J Holcomb and Warren B McPherso. Event-related brain potentials reflect semantic priming in an object decision task. *Brain and cognition*, 24(2):259–276, 1994. 119

Phillip J Holcomb. Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and language*, 35(1):66–85, 1988. 119, 123

Phillip J Holcomb and Jane E Anderson. Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, 8(4): 379–411, 1993. 119

Phillip J Holcomb, John Kounios, Jane E Anderson, and W Caroline West. Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3):721, 1999. 56, 76, 82, 149

Henning Holle and Thomas C Gunter. The role of iconic gestures in speech disambiguation: Erp evidence. *19*, 19(7):1175–1192, 2007. 54, 119, 133

Barry Horwitz, Katrin Amunts, Rajan Bhattacharyya, Debra Patkin, Keith Jeffries, Karl Zilles, and Allen R Braun. Activation of broca's area during the production of spoken and signed language: a combined cytoarchitectonic mapping and pet analysis. *Neuropsychologia*, 41(14):1868–1876, 2003. 9

Autumn B Hostetter. When do gestures communicate? a meta-analysis. *Psychological bulletin*, 137(2):297, 2011. 9, 10, 13

Fatima T Husain, Debra J Patkin, Jieun Kim, Allen R Braun, and Barry Horwitz. Dissociating neural correlates of meaningful emblems from meaningless gestures in deaf signers and hearing non-signers. *Brain research*, 1478:24–35, 2012. 57

Marco Iacoboni, Istvan Molnar-Szakacs, Vittorio Gallese, Giovanni Buccino, John C Mazziotta, and Giacomo Rizzolatti. Grasping the intentions of others with one's own mirror neuron system. *PLoS biology*, 3(3), 2005. 25, 29

Jerom F Janssen and Carl Vogel. Politics makes the swedish:-) and the italians:-. *Sentiment Analysis: Emotion, Metaphor, Ontology & Terminology*, pages 53–61, 2008. 20

Joris H Janssen, Wijnand A IJsselsteijn, and Joyce HDM Westerink. How affective technologies can influence intimate interactions and improve social connectedness. *International Journal of Human-Computer Studies*, 72(1):33–43, 2014. 19

Budi Thomas Jap, Sara Lal, Peter Fischer, and Evangelos Bekiaris. Using eeg spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications*, 36(2):2352–2359, 2009. 74

Marc Jeannerod. The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain sciences*, 17(02):187–202, 1994. 25, 130, 137

Jacob Jolij and Victor AF Lamme. Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10747–10751, 2005. 62

Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. Geometry-based muscle modeling for facial animation. In *Graphics Interface*, volume 2001, pages 37–46, 2001. 8

Aleksandra Kawala-Janik, Mariusz Pelc, and Michal Podpora. Method for eeg signals pattern recognition in embedded systems. *Elektronika ir Elektrotechnika*, 21(3):3–9, 2015. 133

Spencer D Kelly and R Breckinridge Church. A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child development*, 69(1):85–93, 1998. 10, 13

Spencer D Kelly and Leslie H Goldsmith. Gesture and right hemisphere involvement in evaluating lecture material. *Gesture*, 4(1):25–42, 2004. 9, 10, 13

Spencer D Kelly, Dale J Barr, R Breckinridge Church, and Katheryn Lynch. 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, 1999. 5, 75, 82, 107, 132, 148

Spencer D Kelly, Corinne Kravitz, and Michael Hopkins. Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, 89(1):253–260, 2004. 47, 119, 133

A Kendon. Gestures, their origins and distribution-morris, d, collett, p, marsh, p, oshaughnessy, m, 1981. 55

Adam Kendon. Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, 25(1980):207–227, 1980. 8

Adam Kendon. Current issues in the study of gesture. *The biological foundations of gestures: Motor and semiotic aspects*, 1:23–47, 1986. 8

Adam Kendon. How gestures can become like words. *Cross-cultural perspectives in nonverbal communication*, 13:1–141, 1988a. 6

Adam Kendon. *Sign Languages of Aboriginal Australia: Cultural, semiotic and communicative perspectives.* Cambridge University Press, 1988b. 8

Adam Kendon. Do gestures communicate? a review. *Research on language and social interaction*, 27(3):175–200, 1994. 10, 12, 13

Adam Kendon. *Gesture: Visible action as utterance.* Cambridge University Press, 2004. 8, 63, 75, 82, 94, 96, 97, 113, 132, 133, 148

Adam Kendon. Spacing and orientation in co-present interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony*, pages 1–15. Springer, 2010. 8

Simon Khalaf. Mobile use grows 115% in 2013, propelled by messaging apps. *Flurry Analytics, January*, 13, 2014. 19

Rami N Khushaba, Sarath Kodagoda, Sara Lal, and Gamini Dissanayake. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *Biomedical Engineering, IEEE Transactions on*, 58(1):121–131, 2011. 74

Sara Kiesler, Jane Siegel, and Timothy W McGuire. Social psychological aspects of computer-mediated communication. *American psychologist*, 39(10):1123, 1984. 16

James M Kilner and Chris D Frith. Action observation: inferring intentions without mirror neurons. *Current Biology*, 18(1):R32–R33, 2008. 25, 29

Michael Kipp. *Gesture generation by imitation: From human behavior to computer character animation.* Universal-Publishers, 2005. 7

Sotaro Kita. *Pointing: Where language, culture, and cognition meet.* Psychology Press, 2003. 6

Sotaro Kita and Asli Özyürek. 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, 2003. 8

Kristine M Knutson, Erin M McClellan, and Jordan Grafman. Observing social gestures: an fmri study. *Experimental brain research*, 188(2):187–198, 2008. 55

Evelyne Kohler, Christian Keysers, M Alessandra Umilta, Leonardo Fogassi, Vittorio Gallese, and Giacomo Rizzolatti. Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582):846–848, 2002. 131

John Kounios and Phillip J Holcomb. Structure and process in semantic memory: evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, 121(4):459, 1992. 119

Robert M Krauss. Why do we gesture when we speak? *Current Directions in Psychological Science*, pages 54–60, 1998. 5, 10

Robert M Krauss, Palmer Morrel-Samuels, and Christina Colasante. Do conversational hand gestures communicate? *Journal of personality and social psychology*, 61(5):743, 1991. 5

Robert M Krauss, Robert A Dushay, Yihsiu Chen, and Frances Rauscher. The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, 31(6):533–552, 1995. 9

Robert M Krauss, Yihsiu Chen, and Purnima Chawla. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in experimental social psychology*, 28:389–450, 1996. 9

Marta Kutas and Kara D Federmeier. Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, 4(12):463–470, 2000. 75, 82, 97, 107, 119, 149

Marta Kutas and Kara D Federmeier. Thirty years and counting: Finding meaning in the n400 component of the event related brain potential (erp). *Annual review of psychology*, 62:621, 2011. 75, 82, 89, 97, 107, 119, 123, 133, 149

Marta Kutas and Steven A Hillyard. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205, 1980. 50, 119, 133

Marta Kutas and Steven A Hillyard. Brain potentials during reading reflect word expectancy and semantic association. 1984. 119, 123, 126

Sarah Laszlo and Kara D Federmeier. The n400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, 48(2):176–186, 2011. 76, 82

Ellen F Lau, Colin Phillips, and David Poeppel. A cortical network for semantics:(de) constructing the n400. *Nature Reviews Neuroscience*, 9(12):920–933, 2008. 56, 76, 82, 119, 126, 149

Victoria Lee and Hugh Wagner. The effect of social presence on the facial and verbal expression of emotion and the interrelationships among emotion components. *Journal of Nonverbal Behavior*, 26(1):3–25, 2002. 16, 18

Thomas Leonard and Fred Cummins. The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10):1457–1471, 2011. 97, 113

By Chin-Teng Lin, Li-Wei Ko, Jin-Chern Chiou, Jeng-Ren Duann, Ruey-Song Huang, Sheng-Fu Liang, Tzai-Wen Chiu, and Tzyy-Ping Jung. Noninvasive neural prostheses using mobile and wireless eeg. *Proceedings of the IEEE*, 96(7):1167–1183, 2008. 74

Chin-Teng Lin, Che-Jui Chang, Bor-Shyh Lin, Shao-Hang Hung, Chih-Feng Chao, and I-Jan Wang. A real-time wireless brain–computer interface system for drowsiness detection. *Biomedical Circuits and Systems, IEEE Transactions on*, 4(4):214–222, 2010. 74

Magdalena Lis. Multimodal representation of entities: A corpus-based investigation of co-speech hand gesture. 2014. 94

Yue Liu, Xiao Jiang, Teng Cao, Feng Wan, Peng Un Mak, Pui-In Mak, and Mang I Vai. Implementation of ssvep based bci with emotiv epoc. In *Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2012 IEEE International Conference on*, pages 34–37. IEEE, 2012. 74, 83, 108

Daniel Loehr. Aspects of rhythm in gesture and speech. *Gesture*, 7(2):179–214, 2007. 94, 97, 113

Daniel P Loehr. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1):71–89, 2012. 94, 97, 113

Steven J Luck. *An introduction to the event-related potential technique*. MIT press, 2014. 31, 32, 38, 39

Peter F MacNeilage. The frame/content theory of evolution of speech production. *Behavioral and brain sciences*, 21(04):499–511, 1998. 131

Anna Mari Mäkelä, Ville Mäkinen, Minna Nikkilä, Risto J Ilmoniemi, and Hannu Tiitinen. Magnetoencephalographic (meg) localization of the auditory n400m: effects of stimulus duration. *NeuroReport*, 12(2):249–253, 2001. 97

Fridanna Maricchiolo, Augusto Gnisci, Marino Bonaiuto, and Gianluca Ficca. Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language and Cognitive Processes*, 24(2):239–266, 2009. 10, 13

Lee-Ellen Marvin. Spoof, spam, lurk, and lag: the aesthetics of text-based virtual realities. *Journal of Computer-Mediated Communication*, 1(2):0–0, 1995. 18

Matlab. MATLAB - MathWorks. `http://se.mathworks.com/products/matlab/`, 08 2016. 136

Daphne Maurer, Richard Le Grand, and Catherine J Mondloch. The many faces of configural processing. *Trends in cognitive sciences*, 6(6):255–260, 2002. 59

Louis Mayaud, Marco Congedo, Aurélien Van Laghenhove, D Orlikowski, M Figère, E Azabou, and F Cheliout-Heraut. A comparison of recording modalities of p300 event-related potentials (erp) for brain-computer interface (bci) paradigm. *Neurophysiologie Clinique/Clinical Neurophysiology*, 43(4):217–227, 2013. 133

Gregory McCarthy, Aina Puce, Aysenil Belger, and Truett Allison. Electrophysiological studies of human face perception. ii: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, 9(5):431–444, 1999. 59

David McNeil. Hand and mind, 1992. 5, 6

Nicole M McNeil, Martha W Alibali, and Julia L Evans. The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, 24(2):131–150, 2000. 10

David McNeill. *Hand and mind: What gestures reveal about thought.* University of Chicago Press, 1992. 5, 8, 75, 82, 94, 95, 107, 133, 148

David McNeill. *Gesture and thought.* University of Chicago Press, 2005. 8, 24, 75, 82, 94, 132, 148

David McNeill and Susan D Duncan. Growth points in thinking-for-speaking. 1998. 8

David McNeill and Laura L Pedelty. Right brain and gesture. *Language, gesture, and space*, pages 63–85, 1995. 9

David McNeill, Justine Cassell, and Karl-Erik McCullough. Communicative effects of speech-mismatched gestures. *Research on language and social interaction*, 27(3): 223–237, 1994. 10, 13

W Brian McPherson and Phillip J Holcomb. An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, 36(01):53–65, 1999. 53, 80, 82, 89, 90

George Herbert Mead. *Mind, self, and society: From the standpoint of a social behaviorist*, volume 1. University of Chicago press, 2009. 24

Ingo Gerrit Meister, Babak Boroojerdi, Henrik Foltys, Roland Sparing, Walter Huber, and Rudolf Töpper. Motor cortex hand area and speech: implications for the development of language. *Neuropsychologia*, 41(4):401–406, 2003. 28

Alissa Melinger and Willem JM Levelt. Gesture and the communicative intention of the speaker. *Gesture*, 4(2):119–141, 2004. 9, 12

David E Meyer and Roger W Schvaneveldt. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, 90(2):227, 1971. 118, 132

Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. "blissfully happy" or "ready to fight": Varying interpretations of emoji. *ICWSM'16*, 2016. 143

Jeff Miller. Divided attention: Evidence for coactivation with redundant signals. *Cognitive psychology*, 14(2):247–279, 1982. 106

Jeff Miller. Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, 40(5):331–343, 1986. 106

Jeff Miller. Contralateral and ipsilateral motor activation in visual simple reaction time: A test of the hemispheric coactivation model. *Experimental brain research*, 176(4): 539–558, 2007. 106

Bettina Mohr, Friedemann Pulvermüller, and Eran Zaidel. Lexical decision after left, right and bilateral presentation of function words, content words and non-words: Evidence for interhemispheric interaction. *Neuropsychologia*, 32(1):105–124, 1994a. 106, 107

Bettina Mohr, Friedemann Pulvermüller, Janice Rayman, and Eran Zaidel. Interhemispheric cooperation during lexical processing is mediated by the corpus callosum: evidence from the split-brain. *Neuroscience letters*, 181(1):17–21, 1994b. 107

Bettina Mohr, Friedemann Pulvermüller, Katrin Mittelstädt, and Janice Rayman. Multiple simultaneous stimulus presentation facilitates lexical processing. *Neuropsychologia*, 34(10):1003–1013, 1996. 106, 107

Bettina Mohr, Anna Landgrebe, and Stefan R Schweinberger. Interhemispheric cooperation for familiar but not unfamiliar face processing. *Neuropsychologia*, 40(11): 1841–1848, 2002. 106, 107

Kimberly J Montgomery and James V Haxby. Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20(10):1866–1877, 2008. 55

Valerie Morash, Ou Bai, Stephen Furlani, Peter Lin, and Mark Hallett. Classifying eeg signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. *Clinical neurophysiology*, 119(11):2570–2578, 2008. 74

J Toby Mordkoff and Steven Yantis. Dividing attention between color and shape: Evidence of coactivation. *Perception & Psychophysics*, 53(4):357–366, 1993. 106

J Toby Mordkoff, Jeff Miller, and Anne-Catherine Roch. Absence of coactivation in the motor component: Evidence from psychophysiological measures of target detection. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1):25, 1996. 106

Gernot R Müller-Putz and Gert Pfurtscheller. Control of an electrical prosthesis with an ssvep-based bci. *Biomedical Engineering, IEEE Transactions on*, 55(1):361–364, 2008. 74

Paul A Mullin and Howard E Egeth. Capacity limitations in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance*, 15(1):111, 1989. 107

Akinori Nakamura, Burkhard Maess, Thomas R Knösche, Thomas C Gunter, Patric Bach, and Angela D Friederici. Cooperation of different neuronal systems during hand sign recognition. *Neuroimage*, 23(1):25–34, 2004. 57

James H Neely. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general*, 106(3):226, 1977. 118, 132

Koen Nelissen, Giuseppe Luppino, Wim Vanduffel, Giacomo Rizzolatti, and Guy A Orban. Observing others: multiple action representation in the frontal lobe. *Science*, 310(5746):332–336, 2005. 26, 29, 131, 139

Anselmo Frizera Neto, Wanderley Cardoso Celeste, Vinicius Ruiz Martins, et al. Human-machine interface based on electro-biological signals for mobile vehicles. In *Industrial Electronics, 2006 IEEE International Symposium on*, volume 4, pages 2954–2959. IEEE, 2006. 74

Michael Niedeggen, Frank Rösler, and Kerstin Jost. Processing of incongruous mental calculation problems: Evidence for an arithmetic n400 effect. *Psychophysiology*, 36 (3):307–324, 1999. 97

Arti Nigam, J Hoffman, and R Simons. N400 to semantically anomalous pictures and words. *Cognitive Neuroscience, Journal of*, 4(1):15–22, 1992. 91

Nobuyuki Nishitani and Riitta Hari. Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences*, 97(2):913–918, 2000. 8, 24

Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006. 74

Ela I Olivares, Jaime Iglesias, and Marıa Antonieta Bobes. Searching for face-specific long latency erps: a topographic study of effects associated with mismatching features. *Cognitive Brain Research*, 7(3):343–356, 1999. 119

Parmy Olson. Teenagers say goodbye to facebook and hello to messenger apps. *The Guardian*, 10, 2013. 19

Lee Osterhout, Judith McLaughlin, and Michael Bersick. Event-related brain potentials and human language. *Trends in Cognitive Sciences*, 1(6):203–209, 1997. 50

Thomas Ousterhout. Cross-form facilitation effects from simultaneous gesture/word combinations with erp analysis. In *Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on*, pages 493–497. IEEE, 2015a. 133

Thomas Ousterhout. N400 congruency effects from emblematic gesture probes following sentence primes. In *Intelligent Engineering Systems (INES), 2015 IEEE 19th International Conference on*, pages 411–415. IEEE, 2015b. 119, 126, 127, 133

Thomas Ousterhout and Mads Dyrholm. Cortically coupled computer vision with emotiv headset using distractor variables. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on*, pages 245–250. IEEE, 2013. 74, 83, 108, 133

Thomas Ousterhout and Costanza Navarretta. Reaction time for two types of semantically related gesture and sentence pairs. In *Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on*, pages 499–503. IEEE, 2015. 106

Aslı Özyürek, Roel M Willems, Shinichi Kita, and Peter Hagoort. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Cognitive Neuroscience, Journal of*, 19(4):605–616, 2007. 50, 97, 119, 133

Allan Paivio. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3):255, 1991. 11, 13

Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 1320–1326, 2010. 20

Paradigm. Paradigm - Stimulus Presentation Without Limitations. `http://www.paradigmexperiments.com`, 08 2016. 98

Charles Sanders Peirce. *Collected papers of charles sanders peirce*, volume 5. Harvard University Press, 1974. 95

Giuseppe Di Pellegrino, Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese, and Giacomo Rizzolatti. Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1):176–180, 1992. 24

CYMA PETTEN. Words and sentences: Event-related brain potential measures. *Psychophysiology*, 32(6):511–525, 1995. 76, 82

Cyma Van Petten and Barbara J Luka. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and language*, 97(3):279–293, 2006. 76, 82, 149

Cyma Van Petten and Heather Rheinfelder. Conceptual relationships between spoken words and environmental sounds- event-related brain potential measures. *Neuropsychologia*, 33(4):485–508, 1995. 119

Raedy M Ping and Susan Goldin-Meadow. Hands in the air: using ungrounded iconic gestures to teach children conservation of quantity. *Developmental psychology*, 44(5): 1277, 2008. 13

Isabella Poggi and Emanuela Magno Caldognetto. A score for the analysis of gestures in multimodal communication. In *Proceedings of the Workshop on the Integration of Gesture and Language in Speech, Applied Science and Engineering Laboratories*, pages 235–244, 1996. 96

Tom Postmes, Russell Spears, Khaled Sakhel, and Daphne De Groot. Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, 27(10):1243–1254, 2001. 18

Peter Praamstra and Dick F Stegeman. Phonological effects on the auditory n400 event-related brain potential. *Cognitive Brain Research*, 1(2):73–86, 1993. 97

Marc E Pratarelli. Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain and Cognition*, 24(1):137–157, 1994. 126

Alice Mado Proverbio, Veronica Gabaro, Andrea Orlandi, and Alberto Zani. Semantic brain areas are involved in gesture comprehension: An electrical neuroimaging study. *Brain and language*, 147:30–40, 2015. 57, 119

Frances H Rauscher, Robert M Krauss, and Yihsiu Chen. Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7 (4):226–231, 1996. 143

Jonathon Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop*, pages 43–48. Association for Computational Linguistics, 2005. 20

Brice Rebsamen, Etienne Burdet, Cuntai Guan, Haihong Zhang, Chee Leong Teo, Qiang Zeng, Marcelo Ang, and Christian Laugier. A brain-controlled wheelchair based on p300 and path guidance. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pages 1101–1106. IEEE, 2006. 74

Brice Rebsamen, Chee Leong Teo, Qiang Zeng, Marcelo H Ang Jr, Etienne Burdet, Cuntai Guan, Haihong Zhang, and Christian Laugier. Controlling a wheelchair indoors using thought. *Intelligent Systems, IEEE*, 22(2):18–24, 2007. 74

Landra L Rezabek and John J Cochenour. Visual cues in computer-mediated communication: Supplementing text with emoticons. *Journal of Visual Literacy*, 18(2), 1998. 17

Ronald E Rice and Gail Love. Electronic emotion socioemotional content in a computer-mediated communication network. *Communication research*, 14(1):85–108, 1987. 17

Giacomo Rizzolatti and Michael A Arbib. Language within our grasp. *Trends in neurosciences*, 21(5):188–194, 1998. 118, 131, 149

Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27:169–192, 2004. 24, 28, 130, 137

Giacomo Rizzolatti and Maddalena Fabbri-Destro. The mirror system and its role in social cognition. *Current opinion in neurobiology*, 18(2):179–184, 2008. 130, 137, 152

Giacomo Rizzolatti, Luciano Fadiga, Vittorio Gallese, and Leonardo Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141, 1996. 24, 25, 130, 137

Helios De Rosario, Jose S Solaz, N Rodrıguez, and Luis M Bergasa. Controlled inducement and measurement of drowsiness in a driving simulator. *IET intelligent transport systems*, 4(4):280–288, 2010. 74

Bruno Rossion and Isabel Gauthier. How does the brain process upright and inverted faces? *Behavioral and cognitive neuroscience reviews*, 1(1):63–75, 2002. 58

Bruno Rossion, Isabel Gauthier, M J Tarr, P Despland, Raymond Bruyer, S Linotte, and Marc Crommelinck. The n170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11(1):69–72, 2000. 59

Jean-P Royet, Jane Plailly, Chantal Delon-Martin, David A Kareken, and Christoph Segebarth. fmri of emotional responses to odors:: influence of hedonic valence and judgment, handedness, and gender. *Neuroimage*, 20(2):713–728, 2003. 26, 28

Michael D Rugg. The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22(6):642–647, 1985. 119, 123

Michael D Rugg and Margaret E Nagy. Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory & Cognition*, 15(6):473–481, 1987. 126

Jan De Ruiter. The production of gesture and speech. *Language and gesture*, 2000. 8

Derek R Rutter. *Communicating by telephone.* Pergamon Press, 1987. 16

DR Rutter and Geoffrey M Stephenson. The role of visual communication in social interaction. *Current Anthropology*, 1979. 16

Noam Sagiv and Shlomo Bentin. Structural encoding of human and schematic faces: holistic and part-based processes. *Cognitive Neuroscience, Journal of*, 13(7):937–951, 2001. 59

Dean F Salisbury. Semantic memory and verbal working memory correlates of n400 to subordinate homographs. *Brain and cognition*, 55(2):396–399, 2004. 97

Silvia Savazzi and Carlo A Marzi. Speeding up reaction time with invisible stimuli. *Current Biology*, 12(5):403–407, 2002. 107

Haline E Schendan and Marta Kutas. Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, 40(7):931–945, 2002. 53

Haline E Schendan and Marta Kutas. Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, 19(5):734–749, 2007. 53

Haline E Schendan and Stephen M Maher. Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage*, 44(4):1423–1438, 2009. 53

Anne Schienle, R Stark, B Walter, C Blecker, U Ott, P Kirsch, G Sammer, and D Vaitl. The insula is not specifically involved in disgust processing: an fmri study. *Neuroreport*, 13(16):2023–2026, 2002. 26, 28

Wolfgang Schwarz and Anja Ischebeck. Coactivation and statistical facilitation in the detection of lines. *PERCEPTION-LONDON-*, 23:157–157, 1994. 107

Stefan R Schweinberger, Lyndsay M Baird, Margarethe Blümler, Jürgen M Kaufmann, and Bettina Mohr. Interhemispheric cooperation for face recognition but not for affective facial expressions. *Neuropsychologia*, 41(4):407–414, 2003. 107

M Seyal, B Mull, N Bhullar, T Ahmad, and B Gage. Anticipation and execution of a simple reading task enhance corticospinal excitability. *Clinical Neurophysiology*, 110 (3):424–429, 1999. 28

Stefanie Shattuck-Hufnagel, Yelena Yasinnik, Nanette Veilleux, and Margaret Renwick. A method for studying the time alignment of gestures and prosody in american english:hits' and pitch accents in academic-lecture-style speech. *NATO SECURITY THROUGH SCIENCE SERIES E HUMAN AND SOCIETAL DYNAMICS*, 18:34, 2007. 8

John Short, Ederyn Williams, and Bruce Christie. The social psychology of telecommunications. 1976. 16

Jane Siegel, Vitaly Dubrovsky, Sara Kiesler, and Timothy W McGuire. Group processes in computer-mediated communication. *Organizational behavior and human decision processes*, 37(2):157–187, 1986. 16

Tania Singer, Ben Seymour, John O'Doherty, Holger Kaube, Raymond J Dolan, and Chris D Frith. Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661):1157–1162, 2004. 26, 28, 131, 139

Tatiana Sitnikova, Phillip J Holcomb, Kristi A Kiyonaga, and Gina R Kuperberg. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of cognitive neuroscience*, 20(11):2037–2057, 2008. 97

Russell Spears and Martin Lea. Panacea or panopticon? the hidden power in computer-mediated communication. *Communication Research*, 21(4):427–459, 1994. 18

Lee Sproull and Sara Kiesler. Reducing social context cues: Electronic mail in organizational communication. *Management science*, 32(11):1492–1512, 1986. 16

SPSS. Ibm - statistical analysis software package - spss statistics. `http://www.ibm.com/analytics/us/en/technology/spss/`, 2016. 69, 136

Jürgen Streeck. Gesture as communication i: Its coordination with gaze and speech. *Communications Monographs*, 60(4):275–299, 1993. 9

Kirill Stytsenko, Evaldas Jablonskis, and Cosima Prahm. Evaluation of consumer eeg device emotiv epoc. In *MEi: CogSci Conference 2011, Ljubljana*, 2011. 74, 83, 108

Ayano Sueyoshi and Debra M Hardison. The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4):661–699, 2005. 9, 10, 13

Ted Supalla. The classifier system in american sign language. *Noun classes and categorization*, pages 181–214, 1986. 82

Marco Tamietto, Luca Latini Corazzini, Beatrice de Gelder, and Giuliano Geminiani. Functional asymmetry and interhemispheric cooperation in the perception of emotions from facial expressions. *Experimental brain research*, 171(3):389–404, 2006. 107

Marco Tamietto, Mauro Adenzato, Giuliano Geminiani, and Beatrice de Gelder. Fast recognition of social emotions takes the whole brain: interhemispheric cooperation in the absence of cerebral asymmetry. *Neuropsychologia*, 45(4):836–843, 2007. 107

Philip A Thompsen and Davis A Foulger. Effects of pictographs and quoting on flaming in electronic mail. *Computers in Human Behavior*, 12(2):225–243, 1996. 17

Laura A Thompson and Dominic W Massaro. Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, 42(1):144–168, 1986. 8

Monique Timmers, Agneta H Fischer, and Antony SR Manstead. Gender differences in motives for regulating emotions. *Personality and Social Psychology Bulletin*, 24(9): 974–985, 1998. 15

H Tokimura, T Asakura, Y Tokimura, A Oliviero, and JC Rothwell. Speech-induced changes in corticospinal excitability. *Annals of neurology*, 40(4):628–634, 1996. 28

Michael Tomasello. Primate cognition: introduction to the issue. *Cognitive Science*, 24 (3):351–361, 2000. 118, 149

Chad C Tossell, Philip Kortum, Clayton Shepard, Laura H Barg-Walkow, Ahmad Rahmati, and Lin Zhong. A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior*, 28(2):659–663, 2012. 19

Devashree Tripathy and Jagdish Lal Raheja. Design and implementation of brain computer interface based robot motion control. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, pages 289–296. Springer, 2015. 74

Maria Alessandra Umilta, Evelyne Kohler, Vittorio Gallese, Leonardo Fogassi, Luciano Fadiga, Christian Keysers, and Giacomo Rizzolatti. I know what you are doing: A neurophysiological study. *Neuron*, 31(1):155–165, 2001. 131

Laura Valenzeno, Martha W Alibali, and Roberta Klatzky. Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, 28 (2):187–204, 2003. 13

Mary Vanderwart. Priming by pictures in lexical decision. *Journal of Verbal Learning and Verbal Behavior*, 23(1):67–83, 1984. 126

Maria G Veldhuizen, Timothy G Shepard, Miao-Fen Wang, and Lawrence E Marks. Coactivation of gustatory and olfactory signals in flavor perception. *Chemical senses*, 35(2):121–133, 2010. 107

Mirta Villarreal, Esteban A Fridman, Alejandra Amengual, German Falasco, Eliana Roldan Gerscovich, Erlinda R Ulloa, and Ramon C Leiguarda. The neural substrate of gesture recognition. *Neuropsychologia*, 46(9):2371–2382, 2008. 55

Carl Vogel and Jerom F Janssen. Emoticonsciousness. In *Multimodal Signals: Cognitive and Algorithmic Issues*, pages 271–287. Springer, 2009. 20

Maarten De Vos, Katharina Gandras, and Stefan Debener. Towards a truly mobile auditory brain–computer interface: Exploring the p300 to take away. *International Journal of Psychophysiology*, 91(1):46–53, 2014a. 74, 83, 108

Maarten De Vos, Markus Kroesen, Reiner Emkes, and Stefan Debener. P300 speller bci with a mobile eeg system: comparison to a traditional amplifier. *Journal of neural engineering*, 11(3):036008, 2014b. 74, 83, 108

Hugh Wagner and Victoria Lee. 10. facial behavior alone and in the presence of others. *The social context of nonverbal behavior*, page 262, 1999. 15

JB Walter. Interpersonal effects in computer-mediated interaction: A meta-analysis of social and anti-social communication. *Communication Research*, 21:460–487, 1994. 16

Joseph B Walther. Interpersonal effects in computer-mediated interaction a relational perspective. *Communication research*, 19(1):52–90, 1992. 16

Joseph B Walther and Kyle P D'Addario. The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review*, 19 (3):324–347, 2001. 17, 18

Shaojung Sharon Wang. More than words? the effect of line character sticker use on intimacy in the mobile communication environment. *Social Science Computer Review*, page 0894439315590209, 2015. 19

W Caroline West and Phillip J Holcomb. Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13(3):363–375, 2002. 53, 80, 83, 90

Bruno Wicker, Christian Keysers, Jane Plailly, Jean-Pierre Royet, Vittorio Gallese, and Giacomo Rizzolatti. Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40(3):655–664, 2003. 26, 28, 131, 139

Roel M Willems and Peter Hagoort. Neural evidence for the interplay between language, gesture, and action: A review. *Brain and language*, 101(3):278–289, 2007. 57

Ying Choon Wu and Seana Coulson. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6):654–667, 2005. 51, 52, 53, 119, 126, 127, 133

Ying Choon Wu and Seana Coulson. How iconic gestures enhance communication: An erp study. *Brain and language*, 101(3):234–245, 2007. 52, 119, 133

Ying Choon Wu and Seana Coulson. Are depictive gestures like pictures? commonalities and differences in semantic processing. *Brain and language*, 119(3):184–195, 2011. 53

Ying Choon Wu and Seana Coulson. Co-speech iconic gestures and visuo-spatial working memory. *Acta psychologica*, 153:39–50, 2014. 11, 12, 13

Jiang Xu, Patrick J Gannon, Karen Emmorey, Jason F Smith, and Allen R Braun. Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences*, 106(49):20664–20669, 2009. 55

Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. Emotion classification using web blog corpora. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, pages 275–278. IEEE, 2007. 20

Yelena Yasinnik, Margaret Renwick, and Stefanie Shattuck-Hufnagel. The timing of speech-accompanying gestures with respect to prosody. In *Proceedings of the International Conference: From Sound to Sense*, volume 50, pages 10–13, 2004. 8

Robert K Yin. Looking at upside-down faces. *Journal of experimental psychology*, 81 (1):141, 1969. 58

Masahide Yuasa, Keiichi Saito, and Naoki Mukawa. Emoticons convey emotions without cognition of faces: an fmri study. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1565–1570. ACM, 2006. 60, 61

Masahide Yuasa, Keiichi Saito, and Naoki Mukawa. Brain activity associated with graphic emoticons. the effect of abstract faces in communication over a computer network. *Electrical Engineering in Japan*, 177(3):36–45, 2011. 61

Chunlin Zhao, Chongxun Zheng, Min Zhao, Yaling Tu, and Jianping Liu. Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Systems with Applications*, 38(3): 1859–1865, 2011. 74

Jichang Zhao, Li Dong, Junjie Wu, and Ke Xu. Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1528–1531. ACM, 2012. 20

Catharina Zich, Maarten De Vos, Cornelia Kranczioch, and Stefan Debener. Wireless eeg with individualized channel layout enables efficient motor imagery training. *Clinical Neurophysiology*, 2014. 83, 108